

Neural-Network Quantum States, String-Bond States, and Chiral Topological States

Ivan Glasser,¹ Nicola Pancotti,¹ Moritz August,² Ivan D. Rodriguez,¹ and J. Ignacio Cirac¹

¹Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, D-85748 Garching, Germany

²Department of Informatics, Technical University of Munich, Boltzmannstraße 3, D-85748 Garching, Germany



(Received 17 October 2017; revised manuscript received 8 December 2017; published 11 January 2018)

Neural-network quantum states have recently been introduced as an Ansatz for describing the wave function of quantum many-body systems. We show that there are strong connections between neural-network quantum states in the form of restricted Boltzmann machines and some classes of tensor-network states in arbitrary dimensions. In particular, we demonstrate that short-range restricted Boltzmann machines are entangled plaquette states, while fully connected restricted Boltzmann machines are string-bond states with a nonlocal geometry and low bond dimension. These results shed light on the underlying architecture of restricted Boltzmann machines and their efficiency at representing many-body quantum states. String-bond states also provide a generic way of enhancing the power of neural-network quantum states and a natural generalization to systems with larger local Hilbert space. We compare the advantages and drawbacks of these different classes of states and present a method to combine them together. This allows us to benefit from both the entanglement structure of tensor networks and the efficiency of neural-network quantum states into a single Ansatz capable of targeting the wave function of strongly correlated systems. While it remains a challenge to describe states with chiral topological order using traditional tensor networks, we show that, because of their nonlocal geometry, neural-network quantum states and their string-bond-state extension can describe a lattice fractional quantum Hall state exactly. In addition, we provide numerical evidence that neural-network quantum states can approximate a chiral spin liquid with better accuracy than entangled plaquette states and local string-bond states. Our results demonstrate the efficiency of neural networks to describe complex quantum wave functions and pave the way towards the use of string-bond states as a tool in more traditional machine-learning applications.

DOI: [10.1103/PhysRevX.8.011006](https://doi.org/10.1103/PhysRevX.8.011006)

Subject Areas: Computational Physics,
Condensed Matter Physics,
Quantum Physics

I. INTRODUCTION

Recognizing complex patterns is a central problem that pervades all fields of science. The increased computational power of modern computers has allowed the application of advanced methods to the extraction of such patterns from humongous amounts of data, and we are witnessing an ever-increasing effort to find novel applications in numerous disciplines. This led to a line of research now called quantum machine learning [1], which is divided into two main branches. The first tries to develop quantum algorithms capable of learning, i.e., to exploit speed-ups from quantum computers to make machines learn faster and better. The second branch, which we consider in this work,

uses classical machine-learning algorithms to extract insightful information about quantum systems.

The versatility of machine learning has allowed scientists to employ it in a number of problems, which span from quantum control [2–4] and error correcting codes [5] to tomography [6]. In the last few years, we have also been experiencing interesting developments for some central problems in condensed matter, such as quantum phase classification or recognition [7–10], improvement of dynamical mean field theory [11], enhancement of quantum Monte Carlo methods [12,13], or approximations of thermodynamic observables in statistical systems [14].

An idea that received a lot of attention from the scientific community consists in using neural networks as variational wave functions to approximate ground states of many-body quantum systems [15]. These networks are trained or optimized by the standard variational Monte Carlo (VMC) method, and while a few different neural-network architectures have been tested [15–17], the most promising results so far have been achieved with Boltzmann machines [18]. In particular, state-of-the-art numerical results have

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

been obtained on popular models with restricted Boltzmann machines (RBM), and recent effort has demonstrated the power of deep Boltzmann machines to represent ground states of many-body Hamiltonians with polynomial-size gap and quantum states generated by any polynomial size quantum circuits [19,20].

Other seemingly unrelated classes of states that are widely used in condensed-matter physics are tensor-network states. In 1D, matrix product states (MPS) can approximate ground states of physical Hamiltonians efficiently [21,22], and their structure has led to both analytical insights over the entanglement properties of physical systems and efficient variational algorithms for approximating them [23–25]. The natural extension of MPS to larger-dimensional systems are projected entangled pair states (PEPS) [26]; however, their exact contraction is $\#P$ hard [27], and algorithms for optimizing them need to rely on approximations. Another approach to define higher-dimensional tensor networks consists in first dividing the lattice into overlapping clusters of spins. The wave function of the spins in each cluster is then described by a simple tensor network. The global wave function is finally taken to be the product of these tensor networks, which introduces correlations among the different clusters. This construction for local clusters parametrized by a full tensor gives rise to entangled plaquette states (EPS) [28–30], while taking one-dimensional clusters of spins each described by a MPS leads to a string-bond states (SBS) Ansatz [31,32]. These states can be variationally optimized using the VMC method [31,33] and have been applied to 2D and 3D systems.

All these variational wave functions have been successful in describing strongly correlated, quantum many-body systems, including topologically ordered states. The toric code [34] is a prototypical example, which can be written exactly as a PEPS [35], an EPS [30], a SBS [31], or a short-range RBM [36]. This shows that, in some cases, tensor-network and neural-network quantum states can be related. Indeed, it was recently shown that local tensor networks can be represented efficiently by deep Boltzmann machines [19,20,37]. However, not every topological state can be easily represented by local tensor networks. A class of states for which this is challenging are chiral topological states breaking time-reversal symmetry. Such states were first realized in the context of the fractional quantum Hall (FQH) effect [38], and significant progress has since been made towards the construction of lattice models displaying the same physics, either in Hamiltonians realizing fractional Chern insulators [39–44] or in quantum antiferromagnets on several lattices [45–48]. One approach to describe the wave function of these antiferromagnets is to use parton-constructed wave functions [49–52]. It has also been suggested to construct chiral lattice wave functions from the FQH continuum wave functions, the paradigmatic example being the Kalmeyer-Laughlin wave function [53]. Efforts to construct chiral topological states

with PEPS have been undertaken recently [54–58], but the resulting states are critical. In the noninteracting case, it has moreover been proven that the local parent Hamiltonian of a chiral fermionic Gaussian PEPS has to be gapless [55].

In this work, we show that there is a strong relation between restricted Boltzmann machines and tensor-network states in arbitrary dimensions. We demonstrate that short-range RBM are a special subclass of EPS, while fully connected RBM are a subclass of SBS with a flexible nonlocal geometry and low bond dimension. This relation provides additional insights over the geometric structure of RBM and their efficiency. We discuss the advantages and drawbacks of RBM and SBS and provide a way to combine them together. This generalization in the form of nonlocal string-bond states takes advantage of both the entanglement structure of tensor networks and the efficiency of RBM. It allows for the description of states with larger local Hilbert space and has a flexible geometry. Moreover, it can be combined with more traditional Ansatz wave functions that serve as an initial approximation of the ground state.

We then apply these methods to the challenging problem of approximating chiral topological states. We prove that any Jastrow wave function, and thus the Kalmeyer-Laughlin wave function, can be written exactly as a RBM. Moreover, we show that a remarkable accuracy can be achieved numerically with much less parameters than is required for an exact construction. We numerically evaluate the power of EPS, SBS, and RBM to approximate the ground state of a chiral spin liquid for which the Laughlin state is already a good approximation [45], and we find that RBM and nonlocal SBS are able to achieve lower energy than the Laughlin wave function. By combining these classes of states with the Laughlin wave function, we are able to reach even lower energies and to characterize the properties of the ground state of the model.

The paper is organized as follows: In Sec. II, we introduce the variational Monte Carlo method and show how it can be used to optimize both tensor-network and neural-network states. In Sec. III, the mapping between RBM, EPS, and SBS is derived, and its geometric implications are discussed. Finally, we apply these techniques to the approximation of chiral topological states in Sec. IV.

II. VARIATIONAL MONTE CARLO WITH TENSOR-NETWORK AND NEURAL-NETWORK STATES

A. Variational Monte Carlo method

Given a general Hamiltonian H , one of the main challenges of quantum many-body physics is to find its ground state $|\psi_0\rangle$ satisfying the Schrödinger equation $H|\psi_0\rangle = E_0|\psi_0\rangle$. This eigenvalue problem can be mapped to an optimization problem through the variational principle, stating that the energy of any quantum state is higher than the energy of the ground state. A general pure quantum state on a lattice with N spins can be expressed in the basis

spanned by $|s_1, \dots, s_N\rangle$, where s_i are the projections of the spins on the z axis, as

$$|\psi\rangle = \sum_{s_1, \dots, s_N} \psi(s_1, \dots, s_N) |s_1, \dots, s_N\rangle. \quad (1)$$

Finding the ground state amounts to finding the exponentially many parameters $\psi(s_1, \dots, s_N)$ minimizing the energy, which can only be done exactly for small sizes. Instead of searching for the ground state in the full Hilbert space, one may restrict the search to an Ansatz class specified by a particular form for the function $\psi_w(s_1, \dots, s_N)$ depending on polynomially many variational parameters w . The VMC method [59,60] provides a general algorithm for optimizing the energy of such a wave function. One can compute the energy by expressing it as

$$E_w = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} = \sum_{\mathbf{s}} p(\mathbf{s}) E_{\text{loc}}(\mathbf{s}), \quad (2)$$

where $\mathbf{s} = s_1, \dots, s_N$ is a spin configuration, $p(\mathbf{s}) = [|\psi_w(\mathbf{s})|^2 / (\sum_{\mathbf{s}} |\psi_w(\mathbf{s})|^2)]$ is a classical probability distribution, and the local energy $E_{\text{loc}}(\mathbf{s}) = \sum_{s'} \langle \mathbf{s} | H | \mathbf{s}' \rangle \{ \psi_w(\mathbf{s}') / [\psi_w(\mathbf{s})] \}$ can be evaluated efficiently for Hamiltonians involving few-body interactions. The energy is therefore an expectation value with respect to a probability distribution p that can be evaluated using Markov chain Monte Carlo sampling techniques such as the Metropolis-Hastings algorithm [61,62]. The second ingredient required to minimize the energy with respect to the parameters w is the gradient of the energy, which can be expressed in a similar form since

$$\frac{\partial E_w}{\partial w_i} = 2 \sum_{\mathbf{s}} p(\mathbf{s}) \Delta_{w_i}(\mathbf{s})^* (E_{\text{loc}}(\mathbf{s}) - E_w), \quad (3)$$

where we have defined $\Delta_{w_i}(\mathbf{s}) = \{1/[\psi_w(\mathbf{s})]\} \{[\partial \psi_w(\mathbf{s}) / \partial w_i]\}$ as the log-derivative of the wave function with respect to some parameter w_i . This is also an expectation value with respect to the same probability distribution p and can therefore be sampled at the same time, which allows for the use of gradient-based optimization methods. At each iteration, the energy and its gradient are computed with Monte Carlo, the parameters w are updated by small steps in the direction of the negative energy derivative ($w_i \leftarrow w_i - \alpha [\partial E_w / (\partial w_i)]$), and the process is repeated until convergence of the energy. The VMC method, in its simplest form, only requires the efficient computation of $\{[\psi_w(\mathbf{s}') / \psi_w(\mathbf{s})]\}$ for two spin configurations s and s' , as well as the log-derivative of the wave function $\Delta_w(\mathbf{s})$. More efficient optimization methods can be used, such as conjugate-gradient descent, stochastic reconfiguration [63,64], the Newton method [65], or the linear method [66–68].

At this point, one has to choose a special form for the wave function ψ_w . One of the traditional variational wave

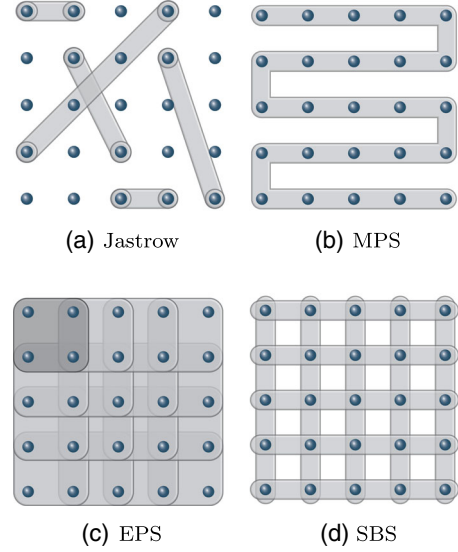


FIG. 1. Geometry of Ansatz wave functions: (a) Jastrow wave functions include correlations within all pairs of spins. (b) MPS in 2D cover the lattice with one snake. (c) EPS include all spin correlations within each plaquette (2×2 on the figure) and mediate correlations between distant spins through overlapping plaquettes. (d) SBS cover the lattice with many 1D strings on which the interactions within spins are captured by a MPS.

functions for a many-body quantum system is a Jastrow wave function [59,69], which consists, in its most general form, of a product of wave functions for all pairs of spins:

$$\psi_w(\mathbf{s}) = \prod_{i < j} f_{ij}(s_i, s_j), \quad (4)$$

where each f_{ij} is fully specified by its four values $f_{ij}(s_i, s_j)$, $s_i, s_j \in \{-1, 1\}$. Such an Ansatz does not presuppose a particular local geometry of the many-body quantum state: In general, this Ansatz can be nonlocal because of the correlations between all pairs of spins [Fig. 1(a)]. A local structure can be introduced by choosing a form for f_{ij} that decays with the distance between positions i and j .

B. Variational Monte Carlo method with tensor networks

In condensed-matter physics, important assets to simplify the problem are the geometric structure and locality of physical Hamiltonians. In 1D, it has been proven that ground states of gapped local Hamiltonians have an entanglement entropy of a subsystem that grows only like the boundary of the subsystem [21]. States satisfying such an area law can be efficiently approximated by MPS [22]. Matrix product states are one-dimensional tensor-network states whose wave function for a spin configuration reads

$$\psi_w(\mathbf{s}) = \text{Tr} \left(\prod_{j=1}^N A_j^{s_j} \right). \quad (5)$$

For each spin and lattice site, the matrix $A_i^{s_i}$ of dimension $D \times D$, where D is called the bond dimension, contains the variational parameters. Matrix product states can be efficiently optimized using the density matrix renormalization group (DMRG) [70], but the previously described VMC method can also be applied [31,33] by observing that the ratio of two configurations is straightforward to compute and that the log-derivative with respect to some matrix $A_k^{s'_k}$ is given by

$$\Delta_{A_k^{s'_k}}(\mathbf{s}) = \frac{\delta_{s_k, s'_k} (A_{k+1}^{s_k} \dots A_N^{s_N} A_1^{s_1} A_{k-1}^{s_{k-1}})^\top}{\text{Tr}(A_1^{s_1} \dots A_N^{s_N})}. \quad (6)$$

In some cases, this method is less likely to be trapped in a local minimum than DMRG since all coefficients can be updated at once. In addition, the cost only scales as $O(D^3)$ in the bond dimension for periodic boundary conditions.

In higher dimensions, matrix product states can be defined by mapping the system to a line [Fig. 1(b)]. The problem of this construction is evident from Fig. 1(b). Spins that sit close to each other might be separated by a long distance on the line; the Ansatz thus fails to reproduce the local structure of the state, which leads to an exponential scaling of the computing resources needed with the system size [71]. The natural extensions of MPS to 2D systems are projected entangled pair states (PEPS) [26], for which the wave function can be written as a contraction of local tensors on the 2D lattice. While PEPS have been successful in describing strongly correlated quantum many-body systems, their exact contraction is $\#P$ hard [27], and their optimization cannot rely on the standard VMC method without approximations. In the following, we instead consider other classes of tensor-network states in more than one dimension for which the exact computation of the wave function is efficient, which allows for the direct use of the VMC method.

One approach consists in cutting a lattice in P small clusters of n_p spins, or plaquettes, and constructing the wave function exactly on each plaquette. The wave function of the full quantum system is then taken to be the product of the wave functions in each plaquette, in a mean-field fashion. Choosing overlapping plaquettes allows one to go beyond the mean field and include correlations between different plaquettes [Fig. 1(c)]. The wave function of such an EPS (also called a correlated product state) is written as [28–30]

$$\psi_w(\mathbf{s}) = \prod_{p=1}^P C_p^{s_p}, \quad (7)$$

where a coefficient $C_p^{s_p}$ is assigned to each of the 2^{n_p} (for spin-1/2 particles) configurations $\mathbf{s}_p = s_{a_1}, \dots, s_{a_{n_p}}$ of the spins on the plaquette p . Each C_p can be seen as the most general function on the Hilbert space corresponding to the spins in plaquette p . The accuracy can be improved by

enlarging the size of the plaquettes, and the Ansatz is exact once the size of the plaquettes reaches the size of the lattice (which can only be achieved on small lattices). Moreover, once the spin configuration \mathbf{s}_p is fixed, the log-derivative of the wave function with respect to the variational parameters is simply

$$\Delta_{C_p^{s_p}}(\mathbf{s}) = \frac{1}{C_p^{s_p}}, \quad (8)$$

which is efficient to compute.

EPS are limited to small plaquettes since, for each plaquette, the number of coefficients scales exponentially with the size of the plaquette. However, one can generalize this Ansatz by describing the state of clusters of spins by a MPS, avoiding the exponentially many coefficients needed. The lattice is now cut in overlapping 1D strings, which can mediate correlations on longer distances compared to local plaquettes [Fig. 1(d)]. The resulting Ansatz is a SBS [31] defined by a set of strings $i \in S$ (each string i is an ordered subset of the set of spins) and a MPS for each string:

$$\psi_w(\mathbf{s}) = \prod_i \text{Tr} \left(\prod_{j \in i} A_{i,j}^{s_j} \right). \quad (9)$$

The descriptive power of this Ansatz is highly dependent on the choice of strings: For example, by using small strings covering small plaquettes and a large bond dimension, it includes EPS, whereas a single long string in a snake pattern includes MPS in 2D. In 3D, it has been used by choosing strings parallel to the axes of the lattice [32]. Since the form of the wave function is a product of MPS, the log-derivative with respect to some elements present in one of the MPS is simply the log-derivative for the corresponding MPS [Eq. (6)]. The VMC procedures for optimizing SBS and MPS thus have the same cost. In addition, the ratio of two configurations that differ only by a few spins can be computed by considering only the strings including these spins, which speeds up the computation considerably. Let us note that a SBS can be mapped analytically to a MPS but that the resulting MPS would have a bond dimension exponential in the number of strings.

C. Variational Monte Carlo method with neural networks

Recently, it was realized that the VMC method can be viewed as a form of learning, which motivated the use of another class of seemingly unrelated states for describing the ground state of many-body quantum states: Neural-network quantum states [15] are quantum states for which the wave function has the structure of an artificial neural network. While a few different networks have been investigated [6,15–17], the most promising results so far have been obtained with Boltzmann machines [18]. Boltzmann

machines are types of generative stochastic artificial neural networks that can learn a distribution over the set of their inputs. In quantum many-body physics, the inputs are spin configurations, and the wave function is interpreted as a (complex) probability distribution that the networks try to approximate. Boltzmann machines consist of two sets of binary units (classical spins): the visible units v_i , $i \in \{1, \dots, N\}$, corresponding to the configurations of the original spins in a chosen basis, and hidden units h_j , $j \in \{1, \dots, M\}$, which introduce correlations between the visible units. The whole system interacts through an Ising interaction, which defines a joint probability distribution over the visible and hidden units as the Boltzmann weight of this Hamiltonian:

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{\mathcal{H}(\mathbf{v}, \mathbf{h})}, \quad (10)$$

where the Hamiltonian \mathcal{H} is defined as

$$\begin{aligned} \mathcal{H} = & \sum_j a_j v_j + \sum_i b_i h_i + \sum_{i < j} c_{ij} v_i v_j \\ & + \sum_{i,j} w_{ij} h_i v_j + \sum_{i < j} d_{ij} h_i h_j, \end{aligned}$$

and Z is the partition function. The marginal probability of a visible configuration is then given by summing over all possible hidden configurations:

$$P(\mathbf{v}) = \sum_{\mathbf{h}} \frac{1}{Z} e^{\mathcal{H}(\mathbf{v}, \mathbf{h})}, \quad (11)$$

and we take this quantity as the Ansatz for the wave function: $\psi_w(\mathbf{s}) = P(\mathbf{s})$. The variational parameters are the complex parameters of the Ising Hamiltonian. In the case where there are interactions between the hidden units [Fig. 2(a)], the Boltzmann machine is called a deep Boltzmann machine. It has been shown that deep

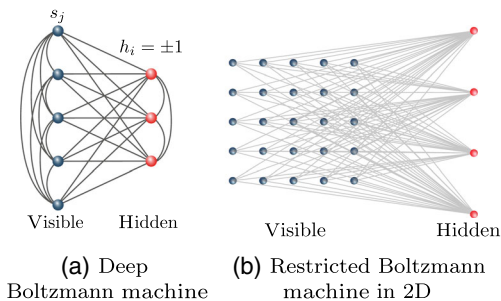


FIG. 2. (a) Boltzmann machines approximate a probability distribution by the Boltzmann weights of an Ising Hamiltonian on a graph including visible units (corresponding to the spins s_j) and hidden units h_i , which are summed over. (b) Restricted Boltzmann machines (here in 2D) only include interactions between the visible and the hidden units.

Boltzmann machines can efficiently represent ground states of many-body Hamiltonians with polynomial-size gaps, local tensor-network states, and quantum states generated by any polynomial-size quantum circuits [19,20,37]. On the other hand, computing the wave function $\psi_w(\mathbf{s})$ of such a deep Boltzmann machine in the general case is intractable because of the exponential sum over the hidden variables, so the VMC method cannot be applied to deep Boltzmann machines without approximations. We therefore turn to the investigation of restricted Boltzmann machines (RBM), which only include interactions between the visible and hidden units (as well as the one-body interaction terms that correspond to biases). In this case, the sum over the hidden units can be performed analytically, and the resulting wave function can be written as (here we take the hidden units to have values ± 1):

$$\psi_w(\mathbf{s}) = e^{\sum_j a_j s_j} \prod_i \cosh \left(b_i + \sum_j w_{ij} s_j \right). \quad (12)$$

RBM can represent many quantum states of interest, such as the toric code [36], any graph state, cluster states, and coherent thermal states [19]; however, the possibility of efficiently computing $\psi_w(\mathbf{s})$ prevents it from approximating all PEPS and ground states of local Hamiltonians [19]. On the other hand, since computing $\psi_w(\mathbf{s})$ and its derivative is very efficient, RBM can be optimized numerically via the VMC method.

III. RELATIONSHIP BETWEEN TENSOR-NETWORK AND NEURAL-NETWORK STATES

While the machine-learning perspective that leads to the application of Boltzmann machines to quantum many-body systems seems quite different from the information-theoretic approach to the structure of tensor-network states, we see that they are in fact intimately related. It was recently shown that, while fully connected RBM can exhibit volume-law entanglement, contrary to local tensor networks, all short-range RBM satisfy an area law [72]. Moreover, short-range and sufficiently sparse RBM can be written as a MPS [37], but doing so for a fully connected RBM would require an exponential scaling of the bond dimension with the size of the system. In this section, we show that there is a tighter connection between RBM and the previously introduced tensor networks in arbitrary dimensions.

A. Jastrow wave functions, RBM, and the Majumdar-Gosh model

Before turning to tensor networks, let us first consider the simple case of the Jastrow wave function [Eq. (4)]. Boltzmann machines that include only interactions between the visible units lead to a wave function

$$\psi_w(\mathbf{s}) = \prod_k e^{a_k s_k} \prod_{i < j} e^{c_{ij} s_i s_j}, \quad (13)$$

which has the form of a product between functions of pairs of spins and is thus a Jastrow wave function. More generally, semirestricted Boltzmann machines, including interactions between visible units as well as between hidden and visible units, are a product of a RBM and a Jastrow factor.

Nevertheless, one may ask whether a RBM alone is enough to describe a Jastrow factor. We first rewrite the RBM as

$$\psi_w(\mathbf{s}) = \prod_j A_j^{s_j} \prod_i \left(B_i \prod_j W_{ij}^{s_j} + \frac{1}{B_i \prod_j W_{ij}^{s_j}} \right), \quad (14)$$

where we have redefined the parameters with uppercase letters as the exponential of the original parameters, thus removing the exponentials in the hyperbolic cosine. This form will be convenient for the numerical simulations presented later. Since Jastrow wave functions are a product of functions of all pairs of spins, let us show that a RBM with one hidden unit can represent any function of two spins. It then follows that a RBM with $M = N(N-1)/2$ hidden units, each representing a function of one pair of spins, can represent a Jastrow wave function with polynomial resources. We thus have to solve for a system of four nonlinear equations with $s_1, s_2 \in \{-1, 1\}$ and f the most general function of two spins: $\psi_w(s_1, s_2) = f(s_1, s_2)$. This system is solved in Appendix A, which provides an analytical solution for the parameters of the RBM to represent the Jastrow wave function exactly, or to arbitrary precision if $f(s_1, s_2) = 0$ for some spins.

As an application, we use this result to write the ground state of the Majumdar-Gosh model [73] exactly as a RBM. The Majumdar-Ghosh model is defined by the following spin-1/2 Hamiltonian:

$$H = J \sum_{i=1}^{N-1} \mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{J}{2} \sum_{i=1}^{N-2} \mathbf{S}_i \cdot \mathbf{S}_{i+2}. \quad (15)$$

The ground-state wave function is a product of singlets formed by neighboring pairs of spins:

$$|\psi\rangle \propto \prod_{n=1}^{N/2} (|\uparrow_{2n-1}\rangle |\downarrow_{2n}\rangle - |\downarrow_{2n-1}\rangle |\uparrow_{2n}\rangle). \quad (16)$$

This wave function can also be expanded in the computational basis as

$$\psi(s_1, \dots, s_N) \propto \prod_{n=1}^{N/2} (-1)^{(s_{2n-1}+3)/2} \delta_{s_{2n-1} \neq s_{2n}}, \quad (17)$$

$$\propto \prod_{n=1}^{N/2} f(s_{2n-1}, s_{2n}). \quad (18)$$

Using the previous result, each function of two spins f can be written as a RBM using one hidden unit, which leads to a RBM representation of the ground states with $M = N/2$ hidden units. We also find numerically on small systems that a RBM using less than $M = N/2$ has higher energy than the ground state, which suggests that $M = N/2$ could be optimal.

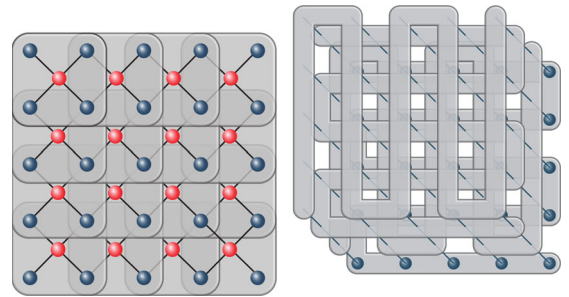
B. Short-range RBM are EPS

Let us now turn to the specific case of RBM with short-range connections (sRBM). This encompasses all quantum states that have previously been written exactly as a RBM, such as the toric code or the 1D symmetry-protected topological cluster state [36]. Such states have weight connections between visible hidden units that are local. Each hidden unit is connected to a local region with, at most, d neighboring spins. If we divide the lattice into M subsets p_i , $i \in \{1, \dots, M\}$, the wave function can be rewritten as (we omit here the biases a_j , which are local one-body terms)

$$\psi_w(\mathbf{s}) = \prod_{i=1}^M \cosh \left(b_i + \sum_{j \in p_i} w_{ij} s_j \right) \quad (19)$$

$$= \prod_{i=1}^M C_i^{s_i}, \quad (20)$$

where \mathbf{s}_i is the spin configuration in the subset p_i . This is the form [Eq. (7)] of an EPS [Fig. 3(a)]. For translational-invariant systems, the short-range RBM becomes a convolutional RBM, which corresponds to a translational-invariant EPS. The main difference between a short-range RBM and an EPS is that the RBM considers a very specific function among all possible functions of the spins inside a plaquette; hence, EPS are more general than short-range RBM. This also directly implies that the



(a) Local RBM as an EPS (b) RBM as a nonlocal SBS

FIG. 3. (a) A locally connected RBM is an EPS where each plaquette encodes the local connections to a hidden unit. (b) Once expressed as a SBS, a fully connected RBM can be represented by many strings on top of each other. Enlarging the RBM by using noncommuting matrices to nonlocal SBS induces a geometry in each string.

entanglement of short-range RBM follows an area law. The main advantage of short-range RBM over EPS is that, because of the exponential scaling of EPS with the size of the plaquettes, larger plaquettes can be used in short-range RBM than in EPS. Since, in practice, for finite systems it is possible to work directly with fully connected RBM, we argue that EPS or fully connected RBM should be preferred to short-range RBM for numerical purposes.

C. Fully connected RBM are SBS

Fully connected RBM, on the other hand, do not always satisfy an area law [72] and hence cannot always be approximated by local tensor networks. Nevertheless, one can express the RBM wave function as (here, we also omit the bias a_j)

$$\psi_w(\mathbf{s}) = \prod_i \cosh \left(b_i + \sum_j w_{ij} s_j \right) \quad (21)$$

$$\propto \prod_i \left(e^{b_i + \sum_j w_{ij} s_j} + e^{-b_i - \sum_j w_{ij} s_j} \right) \quad (22)$$

$$\propto \prod_i \text{Tr} \begin{pmatrix} e^{b_i + \sum_j w_{ij} s_j} & 0 \\ 0 & e^{-b_i - \sum_j w_{ij} s_j} \end{pmatrix} \quad (23)$$

$$\propto \prod_i \text{Tr} \left(\prod_{j \in i} A_{i,j}^{s_j} \right), \quad (24)$$

where

$$A_{i,j}^{s_j} = \begin{pmatrix} e^{b_i/N + w_{ij} s_j} & 0 \\ 0 & e^{-b_i/N - w_{ij} s_j} \end{pmatrix} \quad (25)$$

are diagonal matrices of bond dimension 2. This shows that RBM are string-bond states, as the wave function can be written as a product of MPS over strings, where each hidden unit corresponds to one string. The only difference between the SBS as depicted in Fig. 1(d) and the RBM is the geometry of the strings. In a fully connected RBM, each string goes over the full lattice, while SBS have traditionally been used with smaller strings and with, at most, a few strings overlapping at each lattice site.

D. Generalizing RBM to nonlocal SBS

In the SBS language, RBM consists of many strings overlapping on the full lattice. The matrices in each string in the RBM are diagonal and hence commute, so they can be moved in the string up to a reordering of the spins. This means that each string does not have a fixed geometry and can adapt to stronger correlations in different parts of the lattice, even over long distances. This motivates us to generalize RBM to SBS with diagonal matrices in which each string covers the full lattice [Fig. 3(b)]. In the following, we denote these states as nonlocal dSBS.

This amounts to relaxing the constraints on the RBM parameters to the most general diagonal matrix and enlarging the bond dimension of the matrices. For example, taking the matrices

$$A_{i,j}^{s_j} = \begin{pmatrix} a_{i,j}^{s_j} & 0 & 0 \\ 0 & b_{i,j}^{s_j} & 0 \\ 0 & 0 & c_{i,j}^{s_j} \end{pmatrix}, \quad (26)$$

with different parameters $a_{i,j}^{s_j}$ for each string, lattice site, and spin direction, leads to the wave function (here, $D = 3$)

$$\psi_w(\mathbf{s}) = \prod_i \left(\prod_j a_{i,j}^{s_j} + \prod_j b_{i,j}^{s_j} + \prod_j c_{i,j}^{s_j} \right). \quad (27)$$

Note that even for 2×2 matrices, the nonlocal dSBS is more general than a RBM since the coefficients in each of the two matrices corresponding to one spin are independent from each other, which is not the case in the RBM.

Generalizing such a wave function to larger spins than spin-1/2 is straightforward since the spin s_i is just indexing the parameters. This provides a way of defining a natural generalization of RBM that can handle systems with larger physical dimension. For instance, this can be applied to spin-1 systems, while a naive construction for a RBM with spin-1 visible and hidden units leads to additional constraints, as well as to approximate bosonic systems by truncating the local Hilbert space of the bosons.

A further way to extend this class of states is to include noncommuting matrices. This fixes the geometry of each string by defining an order and also enables us to represent more complicated interactions. In the following, we refer to SBS in such a geometry as nonlocal SBS. The advantage of this approach is that it can capture more complex correlations within each string while introducing additional geometric information about the problem at hand. However, it comes at a greater numerical cost than nonlocal dSBS or RBM because of the additional number of parameters. In practice, one can use an already-optimized RBM or dSBS as a way of initializing a nonlocal SBS.

In some cases, the SBS representation is more compact than the RBM/dSBS representation. Let us consider again the ground state of the Majumdar-Gosh Hamiltonian, which we previously wrote as a RBM with $M = N/2$ hidden units. The ground state of the Majumdar-Gosh Hamiltonian can also be written as a simple MPS with bond dimension 3 and periodic boundary conditions, with matrices [24]

$$A_n^{s_n=-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 0 & 0 \end{pmatrix}, \quad A_n^{s_n=1} = \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad (28)$$

or for open boundary conditions with

$$A_{2n}^{s=-1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad A_{2n}^{s=1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (29)$$

$$A_{2n-1}^{s=-1} = (1 \ 0), \quad A_{2n-1}^{s=1} = (0 \ 1). \quad (30)$$

Since this state is a MPS, it is also a SBS with one string. The RBM representation of the same state requires $N/2$ strings. In practice, the number of nonzero coefficients is comparable since, in both cases, the representation is sparse; however, for numerical purposes, a fully connected RBM needs of the order $O(N^2)$ parameters before finding the exact ground state, while a MPS or SBS with one string will need $O(N)$ parameters for both open and periodic boundary conditions.

Another example is the AKLT model [74] defined by the following spin-1 Hamiltonian in periodic boundary conditions:

$$H = \sum_{i=1}^N \left[\frac{1}{2} \mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{6} (\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2 + \frac{1}{3} \right]. \quad (31)$$

Its ground state has a simple form as a MPS of bond dimension 2. It can also be written as an exact RBM by mapping the system to a spin-1/2 chain, but the number of hidden units needed for an exact representation scales as $O(N^2)$ in the system size [75]. We have numerically optimized the spin-1 extension of a RBM with the form Eq. (27) (see Appendix B for the details of the numerical optimization) and found that, already for small sizes of the chain, a much higher number of parameters is required to approach the ground-state energy as compared to a SBS with noncommuting matrices, which is exact with one string of bond dimension 2 (Fig. 4). We also show in Sec. IV that, in some other cases, the RBM needs less parameters than a SBS to obtain a similar energy. This demonstrates that both RBM and SBS have advantages and that their efficiency depends on the particular model that is investigated. It remains an open question whether there exist MPS or SBS that can provably not be efficiently approximated by a RBM (for which the RBM would need exponentially many parameters).

To be able to use both the advantages of RBM (efficient to compute, few parameters) and of SBS (complex representation, geometric interpretation), one can use the flexibility of SBS by including some strings that have a full MPS over the whole lattice, some strings that include only local connections and that will ensure that the locality of the system is preserved, and some strings that have the form of a RBM and that can easily capture large entanglement and long-range correlations. In many cases of interest, an initial approximation of the ground state can be obtained, either by optimizing simpler wave functions or by first applying DMRG to optimize a MPS. This initial approximation can

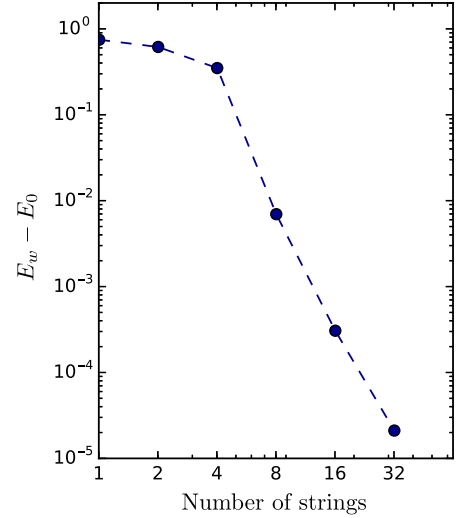


FIG. 4. Energy difference with the exact ground-state energy of a spin-1 extension of a RBM [Eq. (27)] with $D = 2$ and different number of strings for the AKLT model on a spin-1 chain with eight spins. A nonlocal SBS with noncommuting matrices and one string is exact within numerical accuracy.

then be used in conjunction with the previous Ansatz classes by multiplying an Ansatz wave function by the initial approximation. For the resulting wave function,

$$\psi_w(\mathbf{s}) = \psi_w^{\text{init}}(\mathbf{s}) \psi_w^{\text{SBS}}(\mathbf{s}), \quad (32)$$

the ratio of the wave function on two configurations, as well as the log-derivatives, depends only on the respective ratio and log-derivatives of each separate wave function, making the application of the VMC method straightforward. This procedure has the advantage of reducing the number of parameters necessary for obtaining a good approximation to the ground state and making the optimization procedure more stable since the initial state is not a completely random state. Such a procedure provides a generic way to enhance the power of more specific Ansatz wave functions tailored to particular problems, as we demonstrate in the next section. A similar technique has been used to construct tensor-product projected states with tensor networks in Ref. [76], and more generally, it can be used to project the wave function of an initial reference state in a Fock space and is thus also suitable to describe fermionic systems.

IV. APPLICATION TO CHIRAL TOPOLOGICAL STATES

In this section, we turn to a practical application on a challenging problem for traditional tensor-network methods, namely, the approximation of a state with chiral topological order. While chiral topological PEPS have been constructed, the resulting states are critical. Moreover, the local parent Hamiltonian of a chiral fermionic Gaussian PEPS has to be gapless [55]. In the

following, we investigate if this obstruction carries on to the tensor-network and neural-network states that we have introduced previously.

A. RBM can describe a Laughlin state exactly

Let us consider a lattice version of the Laughlin wave function at filling factor $1/2$ defined for a spin- $1/2$ system as

$$\psi_{\text{Laughlin}}(\mathbf{s}) = \delta_{\mathbf{s}} \prod_k \chi_k^{s_k} \prod_{i < j} (z_i - z_j)^{\frac{1}{2} s_i s_j}, \quad (33)$$

where $\delta_{\mathbf{s}}$ fixes the total spin to 0, the z_i are the complex coordinates of the positions of the lattice sites, and the phase factor is defined as $\chi_k^{s_k} = e^{i\pi(k-1)(s_k+1)/2}$, ensuring that the state is a singlet. This wave function is equivalent to the Kalmeyer-Laughlin wave function in the thermodynamic limit and has been shown to describe a lattice state sharing the topological properties of the continuum Laughlin states on several lattices [77–79]. In addition, it can be written as a correlator from conformal fields, which has enabled the exact derivation of parent Hamiltonians for this state on any finite lattice [80].

The Laughlin wave function has the structure of a Jastrow wave function, and we have shown in Sec. III A that any Jastrow wave function can be written as a RBM with $M = N(N-1)/2$ hidden units. It follows that RBM and nonlocal SBS can represent a gapped chiral topological state exactly. This is in sharp contrast to local tensor-network states for which there is no exact description known for a (noncritical) chiral topological state. This difference is due to the nonlocal connections in the RBM and Jastrow wave function, which allow them to easily describe a Laughlin state. We note that a chiral p -wave superconductor is another example of a gapped chiral topological state that has been recently written as a (fermionic) quasilocal Boltzmann machine [20].

However, the previous construction is not satisfactory in the sense that the RBM requires a number of hidden units scaling as $O(N^2)$, which is too high for numerical purposes on lattices that are not extremely small. We thus turn to the approximate representation of the Laughlin wave function using a RBM.

B. Numerical approximation of a Laughlin state

The lattice Laughlin wave function we consider has an exact parent Hamiltonian on a finite lattice [80] defined as

$$H_{\text{parent}} = \frac{2}{3} \sum_{i \neq j} |w_{ij}|^2 \mathbf{S}_i \cdot \mathbf{S}_j + \frac{2}{3} \sum_{i \neq j \neq k} \bar{w}_{ij} w_{ik} \mathbf{S}_j \cdot \mathbf{S}_k - \frac{2i}{3} \sum_{i \neq j \neq k} \bar{w}_{ij} w_{ik} \mathbf{S}_i \cdot (\mathbf{S}_j \times \mathbf{S}_k), \quad (34)$$

where $w_{ij} = [(z_i + z_j)/(z_i - z_j)]$ and $\mathbf{S}_j = (S_j^x, S_j^y, S_j^z)$ is the spin operator at site j . We specialize to the square lattice

TABLE I. Energy per site difference with the ground-state energy and overlap with the Laughlin state of different Ansatz wave functions optimized with respect to the Hamiltonian H_{parent} on a 6×6 square lattice with open boundary conditions. Note that sRBM have M' hidden units connected to all spins in each plaquette of size 3×3 , while RBM have M hidden units connected to all spins of the lattice.

Ansatz	$(E_w - E_0)/N$	$ \langle \psi_w \psi_{\text{Laughlin}} \rangle $
EPS 2×2	4.3×10^{-2}	46.10%
EPS 3×3	2.2×10^{-2}	75.79%
sRBM $M' = 1$	8.3×10^{-2}	0.01%
sRBM $M' = 2$	3.1×10^{-2}	46.32%
sRBM $M' = 4$	2.5×10^{-2}	59.07%
RBM $M = N$	5.8×10^{-4}	99.7%
RBM $M = 2N$	1.1×10^{-5}	99.99%

with open boundary conditions and minimize the energy of different wave functions with respect to this Hamiltonian by applying the VMC method presented in Sec. II B with a stochastic reconfiguration optimization, which is equivalent to the natural gradient descent [63,81,82] (details of the numerical optimization can be found in Appendix B). Results are presented in Table I.

We find that EPS with plaquettes of size up to 3×3 have an energy difference with the Laughlin state of the order 10^{-2} , which is better than a short-range RBM (denoted sRBM) on 3×3 plaquettes and up to $M' = 4$ hidden units per plaquette, while the energy of a fully connected RBM with $M = 2N$ hidden units is within 10^{-5} of the energy of the ground state. The resulting RBM uses much less hidden units than would be required for it to be exact, yet it reaches an overlap of 99.99% with the Laughlin wave function. This result shows that the fully connected structure of the RBM is an advantage to describe this state and that EPS can be used instead of short-range RBM. Moreover, we have found that EPS are easier to optimize numerically than a short-range RBM: They are more stable since each coefficient is considered separately, no exponentials or products that lead to unstable behavior are present, and the derivatives have a very simple form [Eq. (8)].

C. Numerical approximation of a chiral spin liquid

The previous results indicate that RBM might be useful for approximating chiral topological states numerically, but they are limited to relatively small sizes because of the nonlocal nature of the parent Hamiltonian, which includes interactions between all triplets of spins on the lattice. In Ref. [45], a local Hamiltonian stabilizing a state in the same class as the Laughlin state was obtained by restricting H_{parent} to local terms and setting the long-range interactions to zero. This leads to the Hamiltonian

$$H_l = J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j + J_\chi \sum_{(i,j,k)_\mathcal{O}} \mathbf{S}_i \cdot (\mathbf{S}_j \times \mathbf{S}_k), \quad (35)$$

where $\langle i, j \rangle$ indicates indices of nearest neighbors on the lattice and $\langle i, j, k \rangle_{\cup}$ indicates indices of all triangles of neighboring spins, with vertices labeled in the counter-clockwise direction. We focus on the case $J = 1, J_{\chi} = 1$ for which the ground state of H_I has above 98% overlap with the Laughlin wave function [Eq. (33)] on a 4×4 lattice. We minimize the energy of different classes of states on 4×4 and 10×10 square lattices with open boundary conditions. For optimizing wave functions with tens of thousands of parameters, we use a batch version of stochastic reconfiguration, which optimizes a random subset of the parameters at each iteration (see Appendix B). We consider several Ansatz wave functions, including EPS with plaquettes of size $2 \times 2, 3 \times 2, 4 \times 2$, and 3×3 ; local SBS covering the lattice with horizontal, vertical, and diagonal strings and increasing bond dimension; RBM with an increasing number of hidden units; nonlocal SBS with diagonal matrices (denoted dSBS) or with noncommuting matrices of bond dimension 2, and different numbers of strings covering the full lattice. We observe that, while the optimization of EPS and SBS is particularly stable, the optimization of RBM can lead to numerical instabilities that are resolved by writing the RBM in the form presented in Eq. (14). Since we use the same optimization procedure for all Ansatz wave functions and since the required time (and memory) to perform the optimization is mainly a function of the number of parameters and of the accuracy, we can compare the Ansatz classes by comparing how many parameters are needed to obtain a similar energy.

We first focus [Fig. 5(a)] on a 4×4 lattice for which the exact ground state can be obtained using exact diagonalization. Local SBS have an energy higher than the Laughlin state, and the energy is saturated with increasing bond dimension, which means that the pattern of horizontal, vertical, and diagonal strings is not enough to capture all correlations in the ground state. While a large 4×4 plaquette would make EPS exact on this small lattice, this would require 2^{16} parameters. The energy of the Laughlin state is already reached for 3×3 plaquettes. RBM with a number of hidden units larger than N and nonlocal SBS with a corresponding number of strings have lower energy than the Laughlin state or the Jastrow wave function. When the number of strings grows, the energy decreases even further. On a larger 10×10 lattice [Fig. 5(b)], the exact ground-state energy is unknown, but we can compare the energy of the different Ansatz wave functions and observe similar results. Only the Jastrow wave function, nonlocal SBS, and RBM have an energy comparable to the Laughlin state. Notice that nonlocal SBS have a constant factor more parameters than a RBM for the same number of strings. On the one side, this allows SBS to achieve better energy than RBM with the same number of strings. On the other side, this comes with the drawback that we can only optimize fewer strings, and on the large lattice, we are numerically limited to nonlocal dSBS with up to N strings. We can conclude that RBM are particularly efficient in this

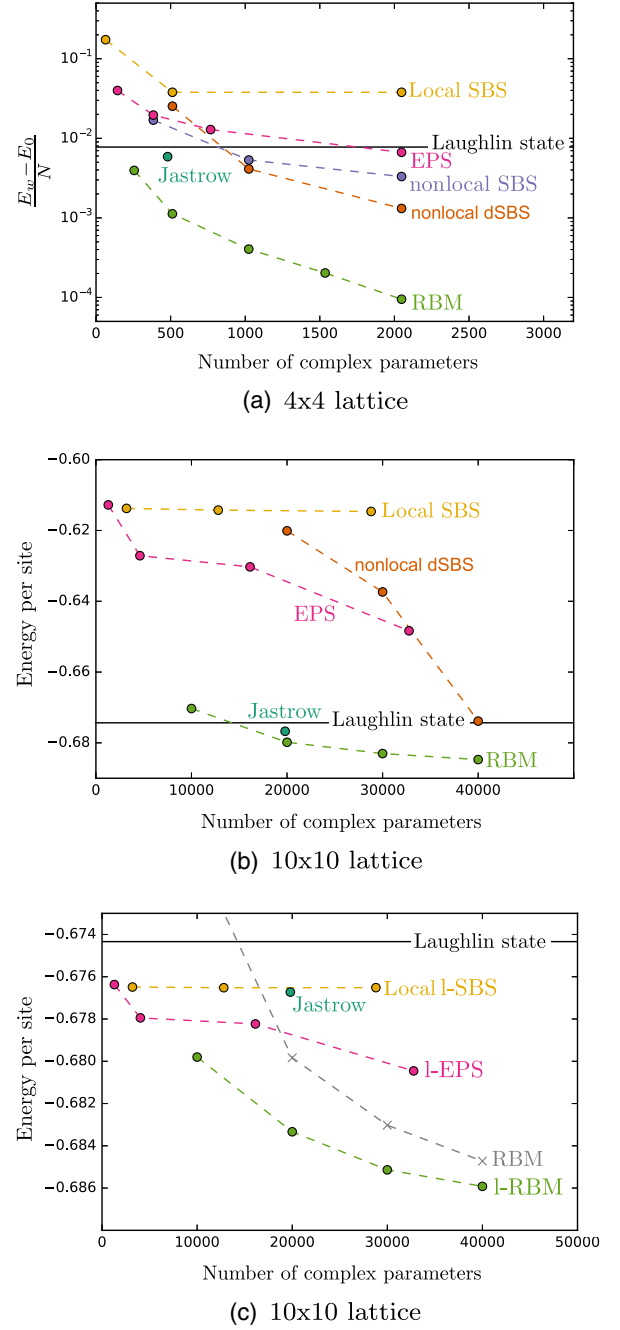


FIG. 5. Energy of H_I per site for different optimized Ansatz wave functions on a square lattice. The number of parameters (N_p) is modified by increasing the bond dimension D (local SBS, $N_p \propto D^2$), the size of the plaquettes (EPS, $N_p \propto M_p 2^P$, where M_p is the number of plaquettes and P is the number of spins in one plaquette), the number of strings M_S (nonlocal SBS and dSBS, $N_p \propto M_S$), or the number of hidden units M_h (RBM, $N_p \propto M_h$). (a) The 4×4 lattice for which the energy difference with the exact ground-state energy is plotted. (b) The 10×10 lattice for which the exact ground-state energy is unknown and the reference energy of the Laughlin state is indicated as a black line. (c) Optimization of wave functions that have been multiplied by the Laughlin wave function on a 10×10 lattice. The original RBM results are indicated for reference as grey crosses.

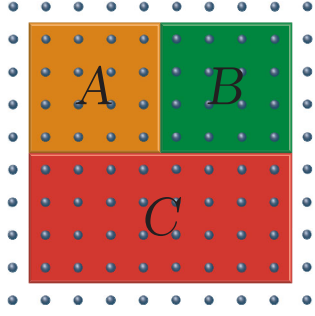


FIG. 6. Partition of the lattice used to compute the topological entanglement entropy.

example since they require significantly less parameters than SBS for attaining the same energy. This has to be contrasted with the previous examples of the Majumdar-Gosh and AKLT models where the opposite was true. Therefore, each class of states has advantages and drawbacks depending on the model we are looking at. We note, in addition, that a nonlocal SBS can be initialized with the results of a previous optimization with a RBM, which could provide a way of minimizing the difficulties of optimizing a large number of parameters.

As we have previously noticed, we can also use an initial approximation of the ground state in combination with the previous Ansatz classes. In the case of the Hamiltonian H_l , the analytical Laughlin wave function can be used as our initial approximation in Eq. (32). We denote l-EPs (resp. l-SBS, l-RBM) a wave function that consists of a product of the Laughlin wave function and an EPs (resp. SBS, RBM) and minimize the energy of the resulting states. This allows us to obtain lower energies for each Ansatz class [Fig. 5(c)]. Once the wave functions are optimized, their properties can be computed using Monte Carlo sampling. To check that the ground state is indeed in the same class as the Laughlin state, we compute the topological entropy of some of the optimized states by dividing the lattice into four regions (Fig. 6) and computing the Renyi entropy $S_A^{(2)} = -\ln \text{Tr} \rho_A^2$ of each subregion using the Metropolis-Hastings Monte Carlo algorithm with two independent spin chains [83,84]. The topological entanglement entropy is then defined as [85,86]

$$S_{\text{topo}} = S_A^{(2)} + S_B^{(2)} + S_C^{(2)} - S_{AB}^{(2)} - S_{AC}^{(2)} - S_{BC}^{(2)} + S_{ABC}^{(2)}, \quad (36)$$

TABLE II. Topological entanglement entropy (TEE) of the analytical Laughlin state and optimized l-EPs, RBM, and l-RBM.

Ansatz	TEE
Laughlin	-0.339(3)
l-EPs 3×3	-0.36(1)
RBM $M = 4N$	-0.34(1)
l-RBM $M = 4N$	-0.34(1)

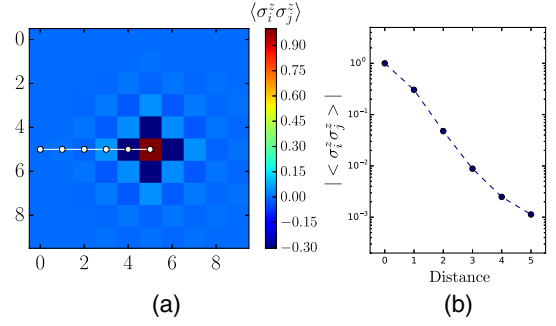


FIG. 7. (a) The spin-spin correlation function between one lattice site (in red) and all other spins on the lattice measured on the optimized l-RBM with lowest energy reveals the antiferromagnetic behavior of the correlations. (b) Decay of the correlations with the distance across the direction indicated in (a) as a white solid line. The error bars are within dot size, and finite-size effects can already be seen for the last point.

and it is expected to be equal to $-\ln 2 \approx -0.347$ for the Laughlin state [87]. The results we obtain are presented in Table II and provide additional evidence that the ground state of H_l has the same topological properties as the Laughlin state. The Hamiltonian H_l was recently investigated on an infinite lattice using infinite PEPS [88], and further evidence was provided that the ground state is chiral. The PEPS results suggest the presence of long-range algebraically decaying correlations that may be a feature of the model or a restriction of PEPS to study chiral systems. The correlations on short distances agree with the correlations that we can compute on our finite system [Fig. 7(a)], but our method does not allow us to make claims about the long-distance behavior of the correlation function. We also observe that fully connected RBM cannot be defined directly in the thermodynamic limit without a truncation of the distance of the interaction between visible and hidden units, thus transforming the RBM into a short-range RBM (albeit of larger range than an EPs). In Ref. [72], it was observed that the entanglement entropy of some specific short-range RBM can be computed analytically from the weights of the RBM. The method we use here works in the general case and also for a fully connected RBM, but it requires Monte Carlo sampling of the wave function. The optimized RBM weights encode all of the information about the wave function; thus, it would be interesting to understand more precisely which quantities can be extracted directly from them. Whether direct information about the phase of the system can be obtained in this way without requiring Monte Carlo sampling remains an interesting open problem for future work.

V. CONCLUSION

We have shown that there is a strong connection between neural-network quantum states in the form of Boltzmann machines and some tensor-network states that can be optimized using the variational Monte Carlo method:

While short-range restricted Boltzmann machines are a subclass of entangled plaquette states, fully connected restricted Boltzmann machines are a subclass of string-bond states. However, these string-bond states are different from traditional string-bond states because of their nonlocal structure, which connects every spin on the lattice to every string. This enabled us to generalize restricted Boltzmann machines by introducing nonlocal (diagonal or noncommuting) string-bond states, which can be defined for larger local Hilbert space and with additional geometric flexibility. We compared the power of these different classes of states and showed that, while there are cases where string-bond states require less parameters than fully connected restricted Boltzmann machines to describe the ground state of a many-body Hamiltonian, there are also cases where the additional parameters in each string make string-bond states less efficient to optimize numerically. We applied these methods to the challenging problem of describing states with chiral topological order, which is hard for traditional tensor networks. We showed that every Jastrow wave function, and thus a Laughlin wave function, can be written as an exact restricted Boltzmann machine. In addition, we gave numerical evidence that a restricted Boltzmann machine with a much smaller number of hidden units can still give a good approximation to the Laughlin state. Finally, we turned to the approximation of the ground state of a chiral spin liquid and showed that restricted Boltzmann machines achieve a lower energy than the Laughlin state and the same topological entanglement entropy. We argued that combining different classes of states allows us to take advantage of the initial knowledge of the model and of the particularities of each class. This was demonstrated by combining a Jastrow wave function to tensor networks and restricted Boltzmann machines, which allowed us to get lower energies than the initial states and to characterize the ground state.

Our work sheds some light on the representative power of restricted Boltzmann machines and establishes a bridge between their optimization and the optimization of tensor-network states. On the one hand, the methods developed in this work can be used to target the ground state of other Hamiltonians, and it would be interesting to know whether similar results can be achieved, for example, for non-Abelian chiral spin liquids [89,90] or generalized to fermionic systems of electrons in the continuum displaying the fractional quantum Hall effect. On the other hand, we also showed that some tools used in machine learning can be rephrased in tensor-network language, thus providing additional physical insights about the systems they describe. Matrix product states have already been used as a tool for supervised learning [91,92], and our work opens up the possibility of using not only restricted Boltzmann machines but also string-bond states to represent a probability distribution over some data while encoding additional information about its geometric structure.

ACKNOWLEDGMENTS

We would like to thank Martin Ganahl, Xun Gao, and Giuseppe Carleo for discussions. This work was supported by the ERC grant QUENOCOBA, ERC-2016-ADG (Grant No. 742102). N.P. and M.A. acknowledge financial support from the Exploring Quantum Matter program.

Note added.—Recently, related independent work came to our attention. Nomura *et al.* [93] combine RBM with pair-product wave functions and apply them to the Heisenberg and Hubbard models. Clark [94] constructs a mapping between RBM and EPS/correlator product states. Kaubruegger *et al.* [95] give further analytical and numerical evidence supporting the application of RBM to chiral topological states such as the Laughlin state.

APPENDIX A: JASTROW WAVE FUNCTIONS ARE RESTRICTED BOLTZMANN MACHINES

Let us show that a RBM with one hidden unit can represent any function f of two spins. It then follows that a RBM with $M = N(N - 1)/2$ hidden units, each representing a function of one pair of spins, can represent a Jastrow wave function. We parametrize f by its four values on two spins $s_1, s_2 \in \{-1, 1\}$ and solve for a system of four nonlinear equations:

$$F_{11} = A_1 A_2 \left(W_1 W_2 + \frac{1}{W_1 W_2} \right), \quad (\text{A1})$$

$$F_{-1-1} = \frac{1}{A_1 A_2} \left(W_1 W_2 + \frac{1}{W_1 W_2} \right), \quad (\text{A2})$$

$$F_{1-1} = \frac{A_1}{A_2} \left(\frac{W_1}{W_2} + \frac{W_2}{W_1} \right), \quad (\text{A3})$$

$$F_{-11} = \frac{A_2}{A_1} \left(\frac{W_2}{W_1} + \frac{W_1}{W_2} \right), \quad (\text{A4})$$

where we have set $B_1 = B_2 = 1$. The RBM is well defined when all parameters are nonzero, and we change variables by defining $X = W_1 W_2$, $Y = (W_1/W_2)$, $A = A_1 A_2$, $B = (A_1/A_2)$, obtaining a new set of equations:

$$F_{-1-1} A^2 = F_{11}, \quad (\text{A5})$$

$$F_{-11} B^2 = F_{1-1}, \quad (\text{A6})$$

$$X^2 - \frac{1}{A} X + 1 = 0, \quad (\text{A7})$$

$$Y^2 - \frac{1}{B} Y + 1 = 0. \quad (\text{A8})$$

We first suppose that the values $F_{s_i s_j}$ are nonzero. These quadratic equations all have nonzero analytical solutions

in the complex plane, which we denote A_0 , B_0 , X_0 , and Y_0 . The original parameters are then the solutions of

$$W_1^2 = X_0 Y_0, \quad (\text{A9})$$

$$W_2^2 = X_0 / Y_0, \quad (\text{A10})$$

$$A_1^2 = A_0 B_0, \quad (\text{A11})$$

$$A_2^2 = A_0 / B_0, \quad (\text{A12})$$

which is again a set of quadratic equations with nonzero analytical solutions. If $F_{11} = F_{-1-1} = 0$ (resp. $F_{1-1} = F_{-11} = 0$), the exact solution is given directly by $A_0 = 1$, $X_0 = i$ (resp. $B_0 = 1$, $Y = i$). In the remaining cases where some $F_{s_i s_j}$ are zeros, the equations do not always have an exact solution, but the function can still be approximated to arbitrary precision. This case corresponds to strong restrictions on the part of the Hilbert space, which is used to write the wave function, and these constraints can also be imposed on the states directly by adding a delta function to the wave function, which is equal to 1 only when the constraints on the spins are satisfied. Having a Markov chain Monte Carlo sampling that does not visit these states then allows for a more efficient sampling.

APPENDIX B: OPTIMIZATION PROCEDURE

The goal is to minimize the energy $E_{\mathbf{w}}$ depending on some vector of parameters \mathbf{w} . We define \mathbf{f} to be the energy gradient vector at \mathbf{w} . Expanding the energy to first order around \mathbf{w} leads to the steepest gradient descent, where the variational parameters are updated at each iteration according to $\mathbf{w}' = \mathbf{w} + \boldsymbol{\gamma}$, with a change of parameters given by $\boldsymbol{\gamma} = -\alpha \mathbf{f}$. Here, α is a small step size. Expanding the energy to second order instead would result in the Newton method with a change of parameters given by

$$\boldsymbol{\gamma} = -\alpha \mathbf{H}^{-1} \mathbf{f}, \quad (\text{B1})$$

where \mathbf{H} is the Hessian of the energy. Small changes of the variational parameters may, however, lead to big changes in the wave function, especially in the case of compact nonlocal representations like RBM in which each parameter affects each part of the wave function. Taking into account the metric of changes of the wave function leads to the stochastic reconfiguration [63] method, which is equivalent to the natural gradient descent [82] and replaces the Hessian in Eq. (B1) by the covariance matrix of the derivatives of the wave function, avoiding the computation of the second-order derivatives of the energy.

The stochastic reconfiguration method can also be viewed as an approximate imaginary-time evolution in the tangent space of the wave function. Consider the normalized wave function $|\bar{\psi}_0\rangle$ and its derivatives

$$|\bar{\psi}_0\rangle = \frac{|\psi_0\rangle}{\sqrt{\langle\psi_0|\psi_0\rangle}}, \quad (\text{B2})$$

$$|\bar{\psi}_i\rangle = \frac{|\psi_i\rangle}{\sqrt{\langle\psi_0|\psi_0\rangle}} - \frac{\langle\psi_0|\psi_i\rangle}{\langle\psi_0|\psi_0\rangle} \frac{|\psi_0\rangle}{\sqrt{\langle\psi_0|\psi_0\rangle}}, \quad (\text{B3})$$

defining a nonorthogonal basis set Ω . Expanding the wave function to linear order around some parameters \mathbf{w} leads to

$$|\bar{\psi}(\mathbf{w} + \boldsymbol{\gamma})\rangle = \sum_{i=0}^{N_w} \gamma_i |\bar{\psi}_i\rangle. \quad (\text{B4})$$

To minimize the energy, one can apply the imaginary-time evolution operator $e^{-\alpha H}$, which, expanded to first order for small α , is $1 - \alpha H$. The change of coefficients $\boldsymbol{\gamma}$ is found by applying this operator to $|\bar{\psi}(\mathbf{w} + \boldsymbol{\gamma})\rangle$ and projecting in the set Ω , which leads to the equation

$$-\alpha \langle\bar{\psi}_i|H|\bar{\psi}_0\rangle = \sum_{j=1}^M \langle\bar{\psi}_i|\bar{\psi}_j\rangle \gamma_j, \quad (\text{B5})$$

which can be rewritten as

$$\boldsymbol{\gamma} = -\alpha \mathbf{S}^{-1} \mathbf{f}, \quad (\text{B6})$$

where $S_{ij} = \langle\bar{\psi}_i|\bar{\psi}_j\rangle$ and $f_i = \langle\bar{\psi}_i|H|\bar{\psi}_0\rangle$. If we expand these expressions as expectation values over the probability distribution $p(\mathbf{s}) = \{[|\psi_{\mathbf{w}}(\mathbf{s})|^2] / \sum_{\mathbf{s}} |\psi_{\mathbf{w}}(\mathbf{s})|^2\}$, we obtain

$$f_i = \langle\Delta_i^* E_{\text{loc}}\rangle - \langle\Delta_i^*\rangle \langle E_{\text{loc}}\rangle, \quad (\text{B7})$$

$$S_{ij} = \langle\Delta_i^* \Delta_j\rangle - \langle\Delta_i^*\rangle \langle\Delta_j\rangle, \quad (\text{B8})$$

where the local energy is defined as $E_{\text{loc}}(\mathbf{s}) = \sum_{\mathbf{s}'} \langle\mathbf{s}|H|\mathbf{s}'\rangle \{\psi_{\mathbf{w}}(\mathbf{s}') / \psi_{\mathbf{w}}(\mathbf{s})\}$ and the log-derivative of the wave function as $\Delta_{\mathbf{w}}(\mathbf{s}) = \{1 / [\psi_{\mathbf{w}}(\mathbf{s})]\} \{[\partial \psi_{\mathbf{w}}(\mathbf{s}) / \partial \mathbf{w}]\}$. Finally, the complete algorithm is as follows:

- (1) Using a Metropolis-Hastings algorithm, generate samples of the probability p and compute stochastic estimates for the expectation values $\langle\Delta_j\rangle$, $\langle E_{\text{loc}}\rangle$, $\langle\Delta_i^* E_{\text{loc}}\rangle$, $\langle\Delta_i^* \Delta_j\rangle$.
- (2) Construct the vector \mathbf{f} and matrix \mathbf{S} .
- (3) Update the parameters according to $\mathbf{w} \leftarrow \mathbf{w} - \alpha \mathbf{S}^{-1} \mathbf{f}$.
- (4) Repeat the full procedure until convergence of the energy.

In practice, we repeat the full procedure 1000 to 20 000 times until the energy is converged. To optimize a large number of parameters, we randomly select a subset of the parameters of size up to 10 000 at each iteration of the algorithm and update only these parameters. This reduces the computational cost associated with the operations dealing with \mathbf{f} and \mathbf{S} . Moreover, we can avoid forming the full matrix \mathbf{S} by instead solving Eq. (B6) with a

conjugate-gradient solver [81]. Numerical stability can be achieved by adding a small constant ϵ to the diagonal elements of the matrix \mathbf{S} , rotating the direction of change towards the steepest descent direction. We find that a step size α of the order $1/\sqrt{i}$, where i is the iteration step, works well in conjunction with a large stabilization at the beginning, while a fixed step size can also be chosen in conjunction with a small stabilization of the order $10^{-4} - 10^{-8}$ by performing several optimizations. At the later stages of the optimization, the step size is lowered to ensure that the energy is converged. Further improvements are achieved by projecting the wave functions in a subset of fixed total spin when it is conserved by the Hamiltonian we consider [96]. The spin-flip symmetry can be enforced in a RBM by choosing the bias $b_i = 0$.

-
- [1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, *Quantum Machine Learning*, *Nature (London)* **549**, 195 (2017).
- [2] E. Zahedinejad, J. Ghosh, and B. C. Sanders, *Designing High-Fidelity Single-Shot Three-Qubit Gates: A Machine-Learning Approach*, *Phys. Rev. Applied* **6**, 054005 (2016).
- [3] M. August and X. Ni, *Using Recurrent Neural Networks to Optimize Dynamical Decoupling for Quantum Memory*, *Phys. Rev. A* **95**, 012335 (2017).
- [4] L. Banchi, N. Pancotti, and S. Bose, *Quantum Gate Learning in Qubit Networks: Toffoli Gate without Time-Dependent Control*, *npj Quantum Inf.* **2**, 16019 (2016).
- [5] G. Torlai and R. G. Melko, *Neural Decoder for Topological Codes*, *Phys. Rev. Lett.* **119**, 030501 (2017).
- [6] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Many-Body Quantum State Tomography with Neural Networks*, [arXiv:1703.05334](https://arxiv.org/abs/1703.05334).
- [7] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Learning Phase Transitions by Confusion*, *Nat. Phys.* **13**, 435 (2017).
- [8] J. Carrasquilla and R. G. Melko, *Machine Learning Phases of Matter*, *Nat. Phys.* **13**, 431 (2017).
- [9] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, *Machine Learning Quantum Phases of Matter Beyond the Fermion Sign Problem*, *Sci. Rep.* **7**, 8823 (2017).
- [10] L. Wang, *Discovering Phase Transitions with Unsupervised Learning*, *Phys. Rev. B* **94**, 195105 (2016).
- [11] L.-F. Arsenault, A. Lopez-Bezanilla, O. A. von Lilienfeld, and A. J. Millis, *Machine Learning for Many-Body Physics: The Case of the Anderson Impurity Model*, *Phys. Rev. B* **90**, 155136 (2014).
- [12] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, *Self-Learning Monte Carlo Method*, *Phys. Rev. B* **95**, 041101 (2017).
- [13] L. Huang and L. Wang, *Accelerated Monte Carlo Simulations with Restricted Boltzmann Machines*, *Phys. Rev. B* **95**, 035105 (2017).
- [14] G. Torlai and R. G. Melko, *Learning Thermodynamics with Boltzmann Machines*, *Phys. Rev. B* **94**, 165134 (2016).
- [15] G. Carleo and M. Troyer, *Solving the Quantum Many-Body Problem with Artificial Neural Networks*, *Science* **355**, 602 (2017).
- [16] Z. Cai, *Approximating Quantum Many-Body Wave-Functions Using Artificial Neural Networks*, [arXiv:1704.05148](https://arxiv.org/abs/1704.05148).
- [17] H. Saito, *Solving the Bose-Hubbard Model with Machine Learning*, *J. Phys. Soc. Jpn.* **86**, 093001 (2017).
- [18] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, *A Learning Algorithm for Boltzmann Machines*, *Cogn. Sci.* **9**, 147 (1985).
- [19] X. Gao and L.-M. Duan, *Efficient Representation of Quantum Many-Body States with Deep Neural Networks*, *Nat. Commun.* **8**, 662 (2017).
- [20] Y. Huang and J. E. Moore, *Neural Network Representation of Tensor Network and Chiral States*, [arXiv:1701.06246](https://arxiv.org/abs/1701.06246).
- [21] M. B. Hastings, *An Area Law for One-Dimensional Quantum Systems*, *J. Stat. Mech.* (2007) P08024.
- [22] F. Verstraete and J. I. Cirac, *Matrix Product States Represent Ground States Faithfully*, *Phys. Rev. B* **73**, 094423 (2006).
- [23] F. Verstraete, V. Murg, and J. Cirac, *Matrix Product States, Projected Entangled Pair States, and Variational Renormalization Group Methods for Quantum Spin Systems*, *Adv. Phys.* **57**, 143 (2008).
- [24] U. Schollwöck, *The Density-Matrix Renormalization Group in the Age of Matrix Product States*, *Ann. Phys. (Amsterdam)* **326**, 96 (2011).
- [25] S. R. White, *Density-Matrix Algorithms for Quantum Renormalization Groups*, *Phys. Rev. B* **48**, 10345 (1993).
- [26] F. Verstraete and J. I. Cirac, *Renormalization Algorithms for Quantum-Many Body Systems in Two and Higher Dimensions*, [arXiv:0407066](https://arxiv.org/abs/0407066).
- [27] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Computational Complexity of Projected Entangled Pair States*, *Phys. Rev. Lett.* **98**, 140506 (2007).
- [28] A. Gendiar and T. Nishino, *Latent Heat Calculation of the Three-Dimensional $q = 3, 4$, and 5 Potts Models by the Tensor Product Variational Approach*, *Phys. Rev. E* **65**, 046702 (2002).
- [29] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, *Ground-State Properties of Quantum Many-Body Systems: Entangled-Plaquette States and Variational Monte Carlo*, *New J. Phys.* **11**, 083026 (2009).
- [30] H. J. Changlani, J. M. Kinder, C. J. Umrigar, and G. K.-L. Chan, *Approximating Strongly Correlated Wave Functions with Correlator Product States*, *Phys. Rev. B* **80**, 245116 (2009).
- [31] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Simulation of Quantum Many-Body Systems with Strings of Operators and Monte Carlo Tensor Contractions*, *Phys. Rev. Lett.* **100**, 040501 (2008).
- [32] A. Sfondrini, J. Cerrillo, N. Schuch, and J. I. Cirac, *Simulating Two- and Three-Dimensional Frustrated Quantum Systems with String-Bond States*, *Phys. Rev. B* **81**, 214426 (2010).
- [33] A. W. Sandvik and G. Vidal, *Variational Quantum Monte Carlo Simulations with Tensor-Network States*, *Phys. Rev. Lett.* **99**, 220602 (2007).
- [34] A. Kitaev, *Fault-Tolerant Quantum Computation by Anyons*, *Ann. Phys. (Amsterdam)* **303**, 2 (2003).
- [35] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, *Criticality, the Area Law, and the Computational Power of*

- Projected Entangled Pair States*, *Phys. Rev. Lett.* **96**, 220601 (2006).
- [36] D.-L. Deng, X. Li, and S. Das Sarma, *Exact Machine Learning Topological States*, *Phys. Rev. B* **96**, 195145 (2017).
- [37] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *On the Equivalence of Restricted Boltzmann Machines and Tensor Network States*, arXiv:1701.04831.
- [38] D. C. Tsui, H. L. Stormer, and A. C. Gossard, *Two-Dimensional Magnetotransport in the Extreme Quantum Limit*, *Phys. Rev. Lett.* **48**, 1559 (1982).
- [39] M. Levin and A. Stern, *Fractional Topological Insulators*, *Phys. Rev. Lett.* **103**, 196803 (2009).
- [40] D. Sheng, Z.-C. Gu, K. Sun, and L. Sheng, *Fractional Quantum Hall Effect in the Absence of Landau Levels*, *Nat. Commun.* **2**, 389 (2011).
- [41] T. Neupert, L. Santos, C. Chamon, and C. Mudry, *Fractional Quantum Hall States at Zero Magnetic Field*, *Phys. Rev. Lett.* **106**, 236804 (2011).
- [42] Y.-F. Wang, Z.-C. Gu, C.-D. Gong, and D. N. Sheng, *Fractional Quantum Hall Effect of Hard-Core Bosons in Topological Flat Bands*, *Phys. Rev. Lett.* **107**, 146803 (2011).
- [43] K. Sun, Z. Gu, H. Katsura, and S. Das Sarma, *Nearly Flatbands with Nontrivial Topology*, *Phys. Rev. Lett.* **106**, 236803 (2011).
- [44] N. Regnault and B. A. Bernevig, *Fractional Chern Insulator*, *Phys. Rev. X* **1**, 021014 (2011).
- [45] A. E. B. Nielsen, G. Sierra, and J. I. Cirac, *Local Models of Fractional Quantum Hall States in Lattices and Physical Implementation*, *Nat. Commun.* **4**, 2864 (2013).
- [46] B. Bauer, L. Cincio, B. P. Keller, M. Dolfi, G. Vidal, S. Trebst, and A. W. W. Ludwig, *Chiral Spin Liquid and Emergent Anyons in a Kagome Lattice Mott Insulator*, *Nat. Commun.* **5**, 5137 (2014).
- [47] Y.-C. He, D. N. Sheng, and Y. Chen, *Chiral Spin Liquid in a Frustrated Anisotropic Kagome Heisenberg Model*, *Phys. Rev. Lett.* **112**, 137202 (2014).
- [48] S.-S. Gong, W. Zhu, and D. N. Sheng, *Emergent Chiral Spin Liquid: Fractional Quantum Hall Effect in a Kagome Heisenberg Model*, *Sci. Rep.* **4**, 6317 (2014).
- [49] G. Baskaran, Z. Zou, and P. Anderson, *The Resonating Valence Bond State and High- T_c Superconductivity—A Mean Field Theory*, *Solid State Commun.* **63**, 973 (1987).
- [50] I. Affleck, Z. Zou, T. Hsu, and P. W. Anderson, *SU(2) Gauge Symmetry of the Large- U Limit of the Hubbard Model*, *Phys. Rev. B* **38**, 745 (1988).
- [51] X.-G. Wen, *Projective Construction of Non-Abelian Quantum Hall Liquids*, *Phys. Rev. B* **60**, 8827 (1999).
- [52] W.-J. Hu, W. Zhu, Y. Zhang, S. Gong, F. Becca, and D. N. Sheng, *Variational Monte Carlo Study of a Chiral Spin Liquid in the Extended Heisenberg Model on the Kagome Lattice*, *Phys. Rev. B* **91**, 041124 (2015).
- [53] V. Kalmeyer and R. B. Laughlin, *Equivalence of the Resonating-Valence-Bond and Fractional Quantum Hall States*, *Phys. Rev. Lett.* **59**, 2095 (1987).
- [54] T. B. Wahl, H.-H. Tu, N. Schuch, and J. I. Cirac, *Projected Entangled-Pair States Can Describe Chiral Topological States*, *Phys. Rev. Lett.* **111**, 236805 (2013).
- [55] J. Dubail and N. Read, *Tensor Network Trial States for Chiral Topological Phases in Two Dimensions and a No-Go Theorem in Any Dimension*, *Phys. Rev. B* **92**, 205307 (2015).
- [56] S. Yang, T. B. Wahl, H.-H. Tu, N. Schuch, and J. I. Cirac, *Chiral Projected Entangled-Pair State with Topological Order*, *Phys. Rev. Lett.* **114**, 106803 (2015).
- [57] D. Poilblanc, J. I. Cirac, and N. Schuch, *Chiral Topological Spin Liquids with Projected Entangled Pair States*, *Phys. Rev. B* **91**, 224431 (2015).
- [58] S. Yang, T. B. Wahl, H.-H. Tu, N. Schuch, and J. I. Cirac, *Chiral Projected Entangled-Pair State with Topological Order*, *Phys. Rev. Lett.* **114**, 106803 (2015).
- [59] W. L. McMillan, *Ground State of Liquid He⁴*, *Phys. Rev.* **138**, A442 (1965).
- [60] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Quantum Monte Carlo Simulations of Solids*, *Rev. Mod. Phys.* **73**, 33 (2001).
- [61] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines*, *J. Chem. Phys.* **21**, 1087 (1953).
- [62] W. K. Hastings, *Monte Carlo Sampling Methods Using Markov Chains and Their Applications*, *Biometrika* **57**, 97 (1970).
- [63] S. Sorella, *Generalized Lanczos Algorithm for Variational Quantum Monte Carlo*, *Phys. Rev. B* **64**, 024512 (2001).
- [64] S. Sorella, *Wave Function Optimization in the Variational Monte Carlo Method*, *Phys. Rev. B* **71**, 241103 (2005).
- [65] C. J. Umrigar and C. Filippi, *Energy and Variance Optimization of Many-Body Wave Functions*, *Phys. Rev. Lett.* **94**, 150201 (2005).
- [66] M. P. Nightingale and V. Melik-Alaverdian, *Optimization of Ground- and Excited-State Wave Functions and van der Waals Clusters*, *Phys. Rev. Lett.* **87**, 043401 (2001).
- [67] J. Toulouse and C. J. Umrigar, *Optimization of Quantum Monte Carlo Wave Functions by Energy Minimization*, *J. Chem. Phys.* **126**, 084102 (2007).
- [68] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, *Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions*, *Phys. Rev. Lett.* **98**, 110201 (2007).
- [69] R. Jastrow, *Many-Body Problem with Strong Forces*, *Phys. Rev.* **98**, 1479 (1955).
- [70] S. R. White, *Density Matrix Formulation for Quantum Renormalization Groups*, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [71] S. Liang and H. Pang, *Approximate Diagonalization Using the Density Matrix Renormalization-Group Method: A Two-Dimensional-Systems Perspective*, *Phys. Rev. B* **49**, 9214 (1994).
- [72] D.-L. Deng, X. Li, and S. Das Sarma, *Quantum Entanglement in Neural Network States*, *Phys. Rev. X* **7**, 021021 (2017).
- [73] C. K. Majumdar and D. K. Ghosh, *On Next Nearest Neighbor Interaction in Linear Chain. I*, *J. Math. Phys. (N.Y.)* **10**, 1388 (1969).
- [74] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, *Rigorous Results on Valence-Bond Ground States in Antiferromagnets*, *Phys. Rev. Lett.* **59**, 799 (1987).
- [75] X. Gao (private communication).

- [76] O. Sikora, H.-W. Chang, C.-P. Chou, F. Pollmann, and Y.-J. Kao, *Variational Monte Carlo Simulations Using Tensor-Product Projected States*, *Phys. Rev. B* **91**, 165113 (2015).
- [77] A. E. B. Nielsen, J. I. Cirac, and G. Sierra, *Laughlin Spin-Liquid States on Lattices Obtained from Conformal Field Theory*, *Phys. Rev. Lett.* **108**, 257206 (2012).
- [78] H.-H. Tu, A. E. B. Nielsen, J. I. Cirac, and G. Sierra, *Lattice Laughlin States of Bosons and Fermions at Filling Fractions $1/q$* , *New J. Phys.* **16**, 033025 (2014).
- [79] I. Glasser, J. I. Cirac, G. Sierra, and A. E. B. Nielsen, *Lattice Effects on Laughlin Wave Functions and Parent Hamiltonians*, *Phys. Rev. B* **94**, 245104 (2016).
- [80] A. E. B. Nielsen, J. I. Cirac, and G. Sierra, *Quantum Spin Hamiltonians for the $SU(2)_k$ WZW Model*, *J. Stat. Mech.* (2011) P11014.
- [81] E. Neuscamman, C. J. Umrigar, and G. K.-L. Chan, *Optimizing Large Parameter Sets in Variational Quantum Monte Carlo*, *Phys. Rev. B* **85**, 045103 (2012).
- [82] S. Amari, *Natural Gradient Works Efficiently in Learning*, *Neural Comput.* **10**, 251 (1998).
- [83] M. B. Hastings, I. González, A. B. Kallin, and R. G. Melko, *Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations*, *Phys. Rev. Lett.* **104**, 157201 (2010).
- [84] J. Wildeboer and N. E. Bonesteel, *Spin Correlations and Topological Entanglement Entropy in a Non-Abelian Spin-One Spin Liquid*, *Phys. Rev. B* **94**, 045125 (2016).
- [85] A. Kitaev and J. Preskill, *Topological Entanglement Entropy*, *Phys. Rev. Lett.* **96**, 110404 (2006).
- [86] M. Levin and X.-G. Wen, *Detecting Topological Order in a Ground State Wave Function*, *Phys. Rev. Lett.* **96**, 110405 (2006).
- [87] O. S. Zozulya, M. Haque, K. Schoutens, and E. H. Rezayi, *Bipartite Entanglement Entropy in Fractional Quantum Hall States*, *Phys. Rev. B* **76**, 125310 (2007).
- [88] D. Poilblanc, *Investigation of the Chiral Antiferromagnetic Heisenberg Model Using PEPs*, *Phys. Rev. B* **96**, 121118 (2017).
- [89] M. Greiter and R. Thomale, *Non-Abelian Statistics in a Quantum Antiferromagnet*, *Phys. Rev. Lett.* **102**, 207203 (2009).
- [90] I. Glasser, J. I. Cirac, G. Sierra, and A. E. B. Nielsen, *Exact Parent Hamiltonians of Bosonic and Fermionic Moore-Read States on Lattices and Local Models*, *New J. Phys.* **17**, 082001 (2015).
- [91] A. Novikov, M. Trofimov, and I. Oseledets, *Exponential Machines*, [arXiv:1605.03795](https://arxiv.org/abs/1605.03795).
- [92] E. M. Stoudenmire and D. J. Schwab, *Supervised Learning with Quantum-Inspired Tensor Networks*, *Adv. Neural Inf. Process. Syst.* **29**, 4799 (2016).
- [93] Y. Nomura, A. Darmawan, Y. Yamaji, and M. Imada, *Restricted-Boltzmann-Machine Learning for Solving Strongly Correlated Quantum Systems*, *Phys. Rev. B* **96**, 205152 (2017).
- [94] S. R. Clark, *Unifying Neural-Network Quantum States and Correlator Product States via Tensor Networks*, [arXiv:1710.03545](https://arxiv.org/abs/1710.03545).
- [95] R. Kaubruegger, L. Pastori, and J. C. Budich, *Chiral Topological Phases from Artificial Neural Networks*, [arXiv:1710.04713](https://arxiv.org/abs/1710.04713).
- [96] D. Tahara and M. Imada, *Variational Monte Carlo Method Combined with Quantum-Number Projection and Multi-Variable Optimization*, *J. Phys. Soc. Jpn.* **77**, 114701 (2008).