

**Chiral topological phases from artificial neural networks**Raphael Kaubruegger,<sup>1,2</sup> Lorenzo Pastori,<sup>1,3</sup> and Jan Carl Budich<sup>1,3</sup><sup>1</sup>*Department of Physics, University of Gothenburg, SE 412 96 Gothenburg, Sweden*<sup>2</sup>*Institute for Theoretical Physics, University of Innsbruck, A-6020 Innsbruck, Austria*<sup>3</sup>*Institute of Theoretical Physics, Technische Universität Dresden, 01062 Dresden, Germany*

(Received 20 October 2017; revised manuscript received 27 April 2018; published 18 May 2018)

Motivated by recent progress in applying techniques from the field of artificial neural networks (ANNs) to quantum many-body physics, we investigate to what extent the flexibility of ANNs can be used to efficiently study systems that host chiral topological phases such as fractional quantum Hall (FQH) phases. With benchmark examples, we demonstrate that training ANNs of restricted Boltzmann machine type in the framework of variational Monte Carlo can numerically solve FQH problems to good approximation. Furthermore, we show by explicit construction how  $n$ -body correlations can be kept at an exact level with ANN wave functions exhibiting polynomial scaling with power  $n$  in system size. Using this construction, we analytically represent the paradigmatic Laughlin wave function as an ANN state.

DOI: [10.1103/PhysRevB.97.195136](https://doi.org/10.1103/PhysRevB.97.195136)**I. INTRODUCTION**

The quest for methods to solve, at least approximately, the quantum many-body problem has been a major focus of research in physics for many years. The paramount issue in this context is the exponential complexity of the wave function, which severely limits the system sizes tractable with exact diagonalization. An important challenge for the study of larger quantum many-body systems is to efficiently parametrize the physically relevant states. Along these lines, Carleo and Troyer [1] recently demonstrated the potential of artificial neural networks (ANNs) as an ansatz for variational wave functions. There, the synaptic coupling strengths between the physical (visible) and auxiliary (hidden) spin variables (neurons) of the ANN play the role of the variational parameters, and the quantum state is obtained by tracing out the auxiliary variables.

The purpose of this work is to harness the flexibility of ANNs to study chiral topological phases (CTPs) in two spatial dimensions (2D), such as fractional quantum Hall states [2–4] and chiral spin liquids [5–7] in the framework of variational Monte Carlo (VMC). Furthermore, we analytically demonstrate how CTP model wave functions can be exactly represented with ANNs *at polynomial cost*. This is of particular relevance as these exotic phases so far have fairly successfully eluded efficient numerical methods: Quantum Monte Carlo approaches to finding CTP ground states are generically stymied by the negative-sign problem, and fundamental limitations regarding the exact representability of such complex many-body states with tensor networks have been proven [8,9]. However, despite these difficulties, it is fair to say that impressive progress has been made in the computational treatment of CTPs, e.g., using matrix product states at the expense of exponential scaling of resources in only one of the spatial directions [10–12], and tree-tensor network methods [13–15]. Another promising direction is to resort to tensor network states of mixed-state density matrices, the effective temperature of which decreases with increasing resources [16].

Furthermore, Monte Carlo techniques have been successfully applied using, e.g., the fixed-phase method [17,18], and to sample various observables from CTP model wave functions (see, e.g., [19] and references therein).

Below, we study 2D lattice systems hosting CTPs within the ANN architecture of restricted Boltzmann machines (RBMs; see Fig. 1 for an illustration) [1]. Using VMC techniques to train the network, we investigate the efficiency of this method in finding the ground state of chiral spin liquid and lattice fractional quantum Hall models such as the Kapit-Mueller model [20,21]. As a benchmark for small systems, we compare our VMC results to exact diagonalization. Remarkably, we find that systems whose size exceeds the scope of exact diagonalization can be solved with the ANN approach by increasing the number of variational parameters *polynomially* with system size [22]. Besides this numerical study, we construct a modified RBM architecture, coined cluster neural network quantum states (CNQSs; see Fig. 2 for an illustration), to capture CTP model states. While many tensor network methods rely on the truncation of entanglement in real space, the CNQS ansatz is based on limiting the number of particles that are directly correlated in the wave function as a means to contain its complexity. For example, the Laughlin state as a paradigmatic representative of CTPs is characterized by the constraint of simultaneously maximizing the relative angular momentum between any pair of particles. Such two-body constraints of Jastrow form are exactly captured by a CNQS with quadratic scaling (see Fig. 2), as we show analytically. Three-body constraints which appear in non-Abelian phases such as the Moore-Read state [23] require a CNQS ansatz with a cubic effort in system size.

This paper is structured as follows. In Sec. II, we discuss how variational wave functions are obtained from the RBM architecture. Thereafter, in Sec. III we apply this RBM variational ansatz to numerically study chiral topological phases and introduce the CNQS architecture in Sec. IV to obtain analytical insights into how CTP model states can be exactly described

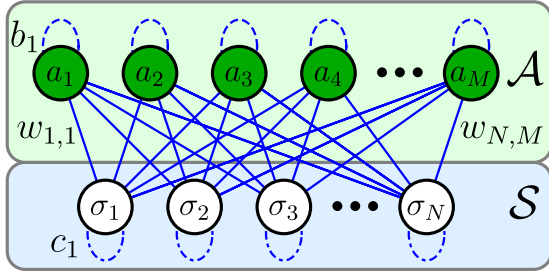


FIG. 1. Graphical representation of the restricted Boltzmann machine (RBM) network for the variational wave function [see Eq. (1)]. The physical spins (in the blue shaded area  $\mathcal{S}$ ) are denoted by  $\sigma_j, j = 1, \dots, N$ , and the auxiliary variables (in the green shaded area  $\mathcal{A}$ ) are denoted by green dots  $a_j, j = 1, \dots, M$ , with  $M = \alpha N$ . The coupling strengths on the links between  $i$  and  $j$  (solid lines) are labeled  $w_{ij}$ , while the local fields are denoted by  $b_i$  (dashed lines) for the auxiliary variables and by  $c_j$  (dash-dotted lines) for the physical spins.

with ANNs. Finally, a concluding discussion is presented in Sec. V. Technical details about the numerical methods we use in this work are provided in the Appendix.

## II. RESTRICTED BOLTZMANN MACHINE STATES

The general ANN framework considered here is that of an RBM consisting of a set of  $N$  physical spins  $\{\sigma_1, \dots, \sigma_N\} = \mathcal{S}$  coupled to a set  $\mathcal{A}$  of  $M$  classical Ising spins called the auxiliary (hidden) variables via a set  $\mathcal{W}$  of complex parameters [1]. The network energy of the RBM is then defined as  $\mathcal{E}_{\text{nw}}(\mathcal{S}, \mathcal{W}, \mathcal{A}) = \sum_j \sigma_j c_j + \sum_i (\sum_j w_{ij} \sigma_j + b_i) a_i$ , where  $w_{ij} \in \mathcal{W}$  are the couplings between the auxiliary and the physical spins, while  $b_i, c_j \in \mathcal{W}$  play the role of a complex local field for the auxiliary variables  $a_i = \pm 1$  and the physical spins  $\sigma_j = 0, 1$ , respectively. The network energy  $\mathcal{E}_{\text{nw}}$  does not have the meaning of a physical energy but specifies the connectivity of the RBM via the functional form of a Boltzmann weight. The defining constraint of an RBM is that there are no direct couplings within  $\mathcal{A}$  which allow

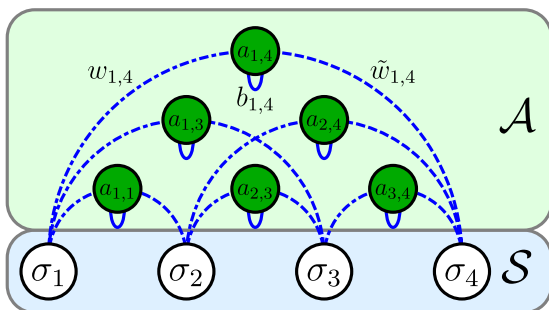


FIG. 2. Sketch of the cluster neural network quantum state (CNQS) architecture with cluster size  $n = 2$  and  $m = 1$  [see Eq. (3)]. The physical spins (in the blue shaded area  $\mathcal{S}$ ) are denoted by  $\sigma_i$ . The auxiliary variables (in the green shaded area  $\mathcal{A}$ ) are denoted by  $a_{ij}$ . The coupling strengths on the links between  $i \neq j$  (dash-dotted lines) are labeled  $w_{ij}$  and  $\tilde{w}_{ij}$ , while the local fields (solid lines) at  $a_{ij}$  are labeled  $b_{ij}$ . For  $n > 2$  (not shown),  $m$  auxiliary variables  $a_{i_1, \dots, i_n}^v, v = 1, \dots, m$  are associated with every cluster of  $n$  distinct sites labeled  $i_1, \dots, i_n$ .

us to analytically trace out the auxiliary variables, yielding the explicit form of the variational wave function at fixed couplings  $\mathcal{W}$ :

$$\begin{aligned} \psi_{\mathcal{W}}(\mathcal{S}) &= \sum_{\{a_i\}} e^{-\mathcal{E}_{\text{nw}}(\mathcal{S}, \mathcal{W}, \mathcal{A})} \\ &= e^{-\sum_j c_j \sigma_j} \prod_i 2 \cosh \left( \sum_j w_{ij} \sigma_j + b_i \right). \end{aligned} \quad (1)$$

Choosing a constant density  $\alpha$  of auxiliary variables per physical spin, i.e.,  $M = \alpha N$ , the number of variational parameters scales as  $\alpha N^2$ .

## III. CHIRAL TOPOLOGICAL PHASES FROM RBM STATES

We now demonstrate how the RBM variational wave function [see Eq. (1)] approach can be used to solve systems hosting CTPs. Concretely, we study the lattice model introduced by Kapit and Mueller [21] on a 2D square lattice. Considering the limit of hard-core bosons, the model Hamiltonian can be readily cast into the spin-1/2 form,

$$H = \sum_{jk} J_{j,k} S_j^+ S_k^-, \quad (2)$$

where the spin operators  $S_j^\pm = S_j^x \pm i S_j^y$  at site  $j = (x_j, y_j)$  replace the bosonic creation and annihilation operators  $\hat{a}_j^\dagger$  and  $\hat{a}_j$ , respectively. Introducing the complex notation  $z_j = x_j + iy_j$  for the 2D lattice indices  $j$ , the complex coupling matrix elements  $J_{j,k}$  take the form [21]  $J_{j,k} = W(z) e^{i\pi\phi (y_k - y_j)(x_k + x_j)}$ , with  $z = z_k - z_j = x + iy$ , and the exponentially decaying prefactor  $W(z)$  reads  $W(z) = (-1)^{x+y+xy} \exp\{-\frac{\pi(1-\phi)}{2} |z|^2\}$ , where  $\phi$  is the magnetic flux per plaquette. The single-particle states of the Kapit-Mueller Hamiltonian constitute a lattice version of the lowest Landau level in the continuum, and the appearance of fractional quantum Hall states as its many-body ground states has been proven in several studies [21,24,25]. In our present numerical study, we consider a quarter filling of the lattice with hard-core bosons (i.e.,  $N/4$  spin-up sites in the spin language) at flux  $\phi = 1/2$ . At these parameters, a bosonic  $\nu = \frac{1}{2}$  Laughlin phase and the corresponding chiral spin liquid phase in the spin language are the ground states of this model. Specifically, we consider the Hamiltonian of Eq. (2) in a cylinder geometry, with periodic boundary conditions in the  $y$  direction. As chiral edge states appear in this geometry, reaching variational energies close to the actual ground-state energy implies that these edge states are also well captured by the RBM wave function (1).

We initialize the RBM with a set of random parameters  $\mathcal{W}$ , generally using  $\alpha = 4$ , and search for the ground state of the Hamiltonian (2) by minimizing the energy expectation value of the RBM state (1) using the stochastic reconfiguration (SR) method to update the RBM wave function [1,26–29]. In Table I, we compare the results we obtain from exact diagonalization (ED) to those from the RBM ansatz for various system sizes. For system size  $L_x \times L_y = 8 \times 8$ , the Hilbert space dimension after taking into account particle number conservation and translation symmetry is  $6.1 \times 10^{13}$  and thus beyond the scope of direct study with ED. However, for such larger systems we

TABLE I. Comparison between the ground-state energy of the Kapit-Mueller Hamiltonian [see Eq. (2)] in cylinder geometry obtained with exact diagonalization (ED) and variational Monte Carlo (VMC) for different system sizes  $L_x \times L_y$ . In the ED column, the values marked with an asterisk are interpolated from shorter cylinders with the same circumference. The last column shows the relative deviation of the VMC result from the ED, defined as  $\Delta E_{\text{rel}} = (E_{\text{ED}} - E_{\text{VMC}})/E_{\text{ED}}$ .

Size	ED	VMC	$\Delta E_{\text{rel}}$
$4 \times 4$	-3.8776	-3.8769(3)	$1.7 \times 10^{-4}$
$6 \times 4$	-5.8773	-5.8767(3)	$1.0 \times 10^{-4}$
$8 \times 4$	-7.8773	-7.8764(3)	$1.1 \times 10^{-4}$
$4 \times 6$	-5.7125	-5.7019(8)	$1.9 \times 10^{-3}$
$6 \times 6$	-8.712*	-8.7010(8)	$1.3 \times 10^{-3}$
$4 \times 8$	-7.6632	-7.658(1)	$6.7 \times 10^{-4}$
$8 \times 8$	-15.663*	-15.652(2)	$6.9 \times 10^{-4}$

interpolate the expected ground-state energy by noticing that the deviation of the ground-state energy from  $-N/4$  up to small fluctuations depends on only the circumference of the cylinder (see values marked with an asterisk in Table I). With our VMC calculations, we reach down to the ground-state energy up to a relative deviation  $\Delta E_{\text{rel}}$  on the order of  $10^{-4}$  to  $10^{-3}$ , where the difference to the exact energy is found to be least at the smallest circumference  $L_y = 4$ , owing to the smaller influence of the metallic edge effects at longer aspect ratios.

In Fig. 3, we show an example of the variational energy of the RBM wave function towards the exact ground-state energy (red horizontal line) as a function of the number of SR iterations for a cylinder of size  $L_x = 6, L_y = 4$ .

#### IV. CLUSTER NEURAL NETWORK QUANTUM STATES AND CHIRAL TOPOLOGICAL PHASES

To gain analytical insight into how ANN states can exactly represent model wave functions for CTPs, we now construct a modified RBM architecture coined cluster neural network states. To this end, we associate a fixed number  $m$  of aux-

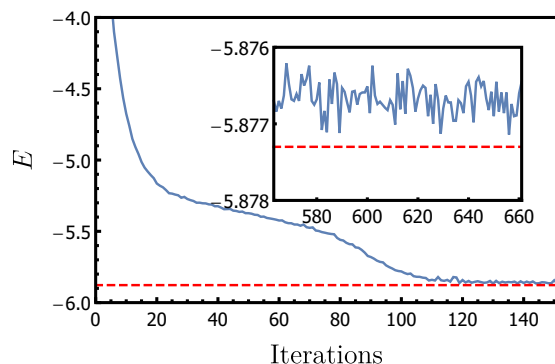


FIG. 3. Energy expectation value  $E = \langle \psi_{\mathcal{W}} | H | \psi_{\mathcal{W}} \rangle / \langle \psi_{\mathcal{W}} | \psi_{\mathcal{W}} \rangle$  as a function of the stochastic reconfiguration (SR) iterations for system size  $6 \times 4$ . For this system size, the expectation values are calculated using a sample of 10 000 configurations drawn with a standard Metropolis algorithm. The inset shows the final SR iterations, and the horizontal red line marks the exact ground-state energy from ED.

iliary variables to every subset of  $n$  physical spins, termed a cluster of size  $n$ . We illustrate our construction for the case  $n = 2, m = 1$ , where an auxiliary variable  $a_{ij} \in \{-1, 1\}$  is associated with every bond between two distinct physical sites (spins)  $i \neq j$  (see Fig. 2) to which it is coupled by the complex weights  $w_{ij}, \tilde{w}_{ij} \in \mathcal{W}$ . The network energy of this RBM is then defined as  $\mathcal{E}_{\text{nw}}(\mathcal{S}, \mathcal{W}, \mathcal{A}) = \sum_{i < j} (w_{ij} \sigma_i + \tilde{w}_{ij} \sigma_j + b_{ij}) a_{ij}$ , where  $b_{ij} \in \mathcal{W}$  are the complex local fields for the auxiliary variables. The explicit form of the variational wave function at fixed couplings  $\mathcal{W}$  then reads

$$\begin{aligned} \psi_{\mathcal{W}}(\mathcal{S}) &= \sum_{\{a_{ij}\}} e^{-\mathcal{E}_{\text{nw}}(\mathcal{S}, \mathcal{W}, \mathcal{A})} \\ &= \prod_{i < j} 2 \cosh(w_{ij} \sigma_i + \tilde{w}_{ij} \sigma_j + b_{ij}). \end{aligned} \quad (3)$$

The generalization of this CNQS to larger  $n, m$  is straightforward with the number of couplings in  $\mathcal{W}$  as well as of the auxiliary variables in  $\mathcal{A}$  scaling as  $mN^n$ . The generalization of the product structure of  $\psi_{\mathcal{W}}$  in Eq. (3) then contains  $m$  factors for each cluster labeled by  $n$  indices  $i_1 < i_2 \dots < i_n$ , capturing  $n$ -body correlations.

*Chiral topological phases from CNQSs.* As a concrete example, we now demonstrate how the above CNQS construction can be used to exactly represent chiral topological states. As a paradigmatic example, we explicitly parametrize a chiral spin liquid ground state of a spin-1/2 system, or, equivalently, the  $\nu = 1/2$  bosonic Laughlin state in the language of hard-core bosons. The desired state  $|\psi_L\rangle$  in the complex position representation  $z_j = x_j + iy_j$  is written as

$$\psi_L(z_1, \dots, z_p) = \prod_{i < j=1}^p (z_i - z_j)^2 e^{-\frac{|z_i|^2 + |z_j|^2}{p-1}}, \quad (4)$$

where  $p$  is the number of particles. In our spin-1/2 representation, where we choose  $\sigma_i \in \{0, 1\}$ , the positions of the up spins, i.e., sites with  $\sigma_i = 1$ , are simply identified with the positions  $z_i$  of hard-core bosons. In order to represent  $|\psi_L\rangle$  as a CNQS, it is helpful to rewrite Eq. (4) as

$$\psi_L(\mathcal{S}) = \prod_{i < j=1}^N \left[ 1 + \left( (z_i - z_j)^2 e^{-\frac{|z_i|^2 + |z_j|^2}{p-1}} - 1 \right) \sigma_i \sigma_j \right], \quad (5)$$

where in the CNQS language only pairs  $i, j$  (two-clusters) with both sites occupied ( $\sigma_i = \sigma_j = 1$ ) contribute a non-trivial factor to the wave function. Equation (5) is of the general Jastrow form  $\prod_{i < j=1}^N (c_{ij} \sigma_i + d_{ij} \sigma_j + e_{ij} \sigma_i \sigma_j + f_{ij})$  with arbitrary complex coefficients  $c, d, e, f$ . Simple parameter counting shows that any such state can be *exactly* represented as a CNQS with  $n = m = 2$ . This already tells us that the exact Laughlin wave function  $\psi_L$  is part of the variational space for  $n = m = 2$ .

Going beyond this general argument, we analytically find that even with  $m = 1$  and  $w_{ij} = \tilde{w}_{ij} = -2b_{ij}$ , i.e., with a single complex parameter per  $ij$  pair, the  $ij$  factor of the CNQS wave function (3) can be decomposed as

$$\begin{aligned} &\cosh(w_{ij} \sigma_i + \tilde{w}_{ij} \sigma_j + b_{ij}) \\ &= \cosh(b_{ij}) \left( 1 + \left[ \frac{\cosh(3b_{ij})}{\cosh(b_{ij})} - 1 \right] \sigma_i \sigma_j \right). \end{aligned} \quad (6)$$

Comparing Eq. (6) to Eq. (5), we find that any Laughlin wave function up to a global prefactor can be exactly represented with analytically determined parameters  $b_{ij}$ .

## V. CONCLUDING DISCUSSION

Using ANN constructions for variational quantum many-body wave functions has already led to several promising insights, including the parametrization of states with volume law entanglement [30], the approximate representation of  $p + ip$  superconductors [31], the exact representation of topological stabilizer states [32], a numerical study of the 2D Hubbard model [33,34], and that on the relation between ANN states and conventional tensor networks [35].

Here, we have shown that RBM states can be efficiently used as an ansatz to describe chiral topological phases both at the numerical level and at an exact analytical level. With small-scale numerical benchmark studies not imposing any symmetry constraints except particle number conservation, we could already significantly exceed the system sizes amenable to direct study with exact diagonalization. However, due to the expected polynomial cost of our RBM simulations [22], even larger systems sizes should be tractable. This may be of particular importance for gapless topological phases exhibiting severe finite-size effects [36]. Moreover, as generally shown in Ref. [1], the ANN approach is capable of describing unitary time evolution. This may open up the possibility to study dynamical aspects such as nonequilibrium response functions and quantum transport properties of CTPs, where comparably large system sizes are required to clearly observe topologically quantized features and where capturing quantum correlations beyond area-law entanglement is important.

The fact that certain CTP model states can naturally be parameterized with polynomial cost within the ANN approach is generally promising, as their exact parametrization with the most well known tensor network methods such as matrix product states requires exponential cost in at least one spatial direction [37]. However, it remains an open question whether the fundamental limitation [9] to the representability of non-trivial CTPs with tensor network states using finite resources in the thermodynamic limit can be overcome with ANN states. An important challenge and interesting direction of future research hence is to devise ANN architectures that are flexible enough to parametrize even in the thermodynamic limit CTPs and other strongly correlated topological phases with no known exact tensor network representative.

*Note added.* Recently, two related papers appeared [38,39]. Clark constructs a mapping [38] between RBM states and correlator product states, with relevance for CTP states such as Laughlin wave functions. Glasser *et al.* [39] establish a correspondence between string-bond network states and RBM states, also presenting VMC data for the  $\nu = 1/2$  Laughlin phase but studying a different model Hamiltonian [40] from the one in this paper.

## ACKNOWLEDGMENTS

We acknowledge discussions with E. Bergholtz, G. Carleo, M. Heyl, M. Hohenadler, N. Cooper, C. Repellin, and A. Sterdyniak. The numerical calculations were performed on

resources at the Chalmers Centre for Computational Science and Engineering (C3SE) provided by the Swedish National Infrastructure for Computing (SNIC). We acknowledge financial support from the German Research Foundation (DFG) through the Collaborative Research Centre SFB 1143.

## APPENDIX

### 1. Stochastic reconfiguration

In the following we provide a brief description of the stochastic reconfiguration (SR) method [26,27,41]. The problem the SR method addresses is the minimization of the energy expectation value within the subspace of the variational wave functions. In order to carry out this minimization procedure we interpret the variational state as effectively depending on  $2N_w$  real parameters, which are the real and imaginary parts of the  $N_w$  complex weights. We denote with the real vector  $\mathbf{w}$  a certain configuration of real and imaginary parts of the weights and with  $|\hat{\psi}_{\mathbf{w}}\rangle = \frac{|\psi_{\mathbf{w}}\rangle}{\sqrt{\langle\psi_{\mathbf{w}}|\psi_{\mathbf{w}}\rangle}}$  the normalized variational state for this set of values. We adopt the following convention:  $w_j$  for  $j = 2\ell - 1$  is the real part of the  $\ell$ th complex weight, and for  $j = 2\ell$  it is the imaginary part of the  $\ell$ th complex weight, where  $\ell = 1, \dots, N_w$ . In the VMC algorithm, after the samples from the probability distribution  $\langle\hat{\psi}_{\mathbf{w}}|\hat{\psi}_{\mathbf{w}}\rangle$  have been generated and the energy expectation value  $E_{\mathbf{w}} = \langle\hat{\psi}_{\mathbf{w}}|H|\hat{\psi}_{\mathbf{w}}\rangle$  has been calculated, an updating step  $d\mathbf{w}$  in parameter space is made such that  $E_{\mathbf{w}+d\mathbf{w}}$  is reduced. The SR method ensures an optimal direction of  $d\mathbf{w}$  by effectively implementing an imaginary-time evolution projected onto the variational manifold [1]. In the following, we discuss the practical implementation of this method to first order in the imaginary-time step  $d\tau$ , as used in our present simulations.

Let us introduce the local tangent space  $\mathcal{T}_{\mathbf{w}}$  at point  $\mathbf{w}$  to the manifold of variational states parametrized by the weights  $w_k$  ( $k = 1, \dots, 2N_w$ ).  $\mathcal{T}_{\mathbf{w}}$  is spanned by the nonorthogonal basis states:

$$|j_{\mathbf{w}}\rangle = |\partial_{w_j}\hat{\psi}_{\mathbf{w}}\rangle - |\hat{\psi}_{\mathbf{w}}\rangle\langle\hat{\psi}_{\mathbf{w}}|\partial_{w_j}\hat{\psi}_{\mathbf{w}}\rangle. \quad (\text{A1})$$

Notice that  $\langle\hat{\psi}_{\mathbf{w}}|j_{\mathbf{w}}\rangle = 0 \forall j = 1, \dots, 2N_w$ . We again point out that the derivatives  $\partial_{w_j}$  are derivatives with respect to the real parts for odd  $j = 2\ell - 1$  and with respect to the imaginary parts of the complex weight  $\ell$  for even  $j = 2\ell$ . We denote with  $(S_{\mathbf{w}})_{j,k} = \langle j_{\mathbf{w}}|k_{\mathbf{w}}\rangle$  the components of the local metric tensor at  $\mathbf{w}$ , also referred to as the covariance matrix, which take the form

$$(S_{\mathbf{w}})_{j,k} = \langle\partial_{w_j}\hat{\psi}_{\mathbf{w}}|\partial_{w_k}\hat{\psi}_{\mathbf{w}}\rangle - \langle\partial_{w_j}\hat{\psi}_{\mathbf{w}}|\hat{\psi}_{\mathbf{w}}\rangle\langle\hat{\psi}_{\mathbf{w}}|\partial_{w_k}\hat{\psi}_{\mathbf{w}}\rangle. \quad (\text{A2})$$

With  $\tau$  being the imaginary time and assuming that the wave function depends on  $\tau$  through the variational parameters  $w_k(\tau)$ , the imaginary-time evolution is governed by the equation

$$|\hat{\psi}_{\mathbf{w}(\tau+d\tau)}\rangle = e^{-d\tau H}|\hat{\psi}_{\mathbf{w}(\tau)}\rangle. \quad (\text{A3})$$

Expanding the left-hand side of the above equation to first order in  $d\tau$ , we obtain

$$\begin{aligned} |\hat{\psi}_{\mathbf{w}(\tau+d\tau)}\rangle &\simeq |\hat{\psi}_{\mathbf{w}(\tau)}\rangle + d\tau \sum_{k=1}^{2N_w} \dot{w}_k(\tau) \left[ |\partial_{w_k}\hat{\psi}_{\mathbf{w}(\tau)}\rangle \right. \\ &\quad \left. - |\hat{\psi}_{\mathbf{w}(\tau)}\rangle\langle\hat{\psi}_{\mathbf{w}(\tau)}|\partial_{w_k}\hat{\psi}_{\mathbf{w}(\tau)}\rangle \right], \end{aligned}$$

where the second term in the sum subtracts the variation of the state parallel to  $|\hat{\psi}_{w(\tau)}\rangle$  (to keep the norm fixed) and  $\dot{w}_k$  denotes the imaginary-time derivative of  $w_k$ . The right-hand side expanded to first order reads

$$e^{-d\tau H}|\hat{\psi}_{w(\tau)}\rangle \simeq |\hat{\psi}_{w(\tau)}\rangle - d\tau H|\hat{\psi}_{w(\tau)}\rangle. \quad (\text{A4})$$

Equating the two terms and multiplying from the left by  $\langle j_{w(\tau)}|$  (i.e., projecting the imaginary-time evolution onto the tangent space  $\mathcal{T}_{w(\tau)}$ ), we obtain (we drop the  $\tau$  dependence now for simplicity)

$$\begin{aligned} \sum_{k=1}^{2N_w} \dot{w}_k [\langle \partial_{w_j} \hat{\psi}_w | \partial_{w_k} \hat{\psi}_w \rangle - \langle \partial_{w_j} \hat{\psi}_w | \hat{\psi}_w \rangle \langle \hat{\psi}_w | \partial_{w_k} \hat{\psi}_w \rangle] \\ = -\langle \partial_{w_j} \hat{\psi}_w | H | \hat{\psi}_w \rangle + \langle \partial_{w_j} \hat{\psi}_w | \hat{\psi}_w \rangle \langle \hat{\psi}_w | H | \hat{\psi}_w \rangle, \end{aligned}$$

which can be rewritten in vector notation as

$$S_w \frac{d\mathbf{w}}{d\tau} = -\mathbf{F}_w, \quad (\text{A5})$$

where  $S_w$  is the  $2N_w \times 2N_w$  metric tensor [see Eq. (A2)] and  $\mathbf{F}_w$  is the force vector, whose components are given by

$$F_j(\mathbf{w}) = \langle \partial_{w_j} \hat{\psi}_w | H | \hat{\psi}_w \rangle - \langle \partial_{w_j} \hat{\psi}_w | \hat{\psi}_w \rangle \langle \hat{\psi}_w | H | \hat{\psi}_w \rangle. \quad (\text{A6})$$

Introducing the imaginary-time step size  $\gamma$ , we then have

$$d\mathbf{w} = -\gamma S_w^{-1} \mathbf{F}_w. \quad (\text{A7})$$

At each imaginary-time step the covariance matrix and the force-vector elements are calculated from the samples of  $\langle \hat{\psi}_w | \hat{\psi}_w \rangle$  by computing the local variational derivative estimators [1,26,27,41]

$$O_k(\mathcal{S}) = \frac{\partial}{\partial w_k} \ln(\langle \mathcal{S} | \psi_w \rangle) \quad (\text{A8})$$

at spin configuration  $\mathcal{S}$  and using

$$S_{k,k'}(\mathbf{w}) = \langle O_k^* O_{k'} \rangle - \langle O_k^* \rangle \langle O_{k'} \rangle, \quad (\text{A9})$$

$$F_k(\mathbf{w}) = \langle O_k^* E_{\text{loc}} \rangle - \langle E_{\text{loc}} \rangle \langle O_k^* \rangle, \quad (\text{A10})$$

with  $E_{\text{loc}}(\mathcal{S}) = \frac{\langle \mathcal{S} | H | \psi_w \rangle}{\langle \mathcal{S} | \psi_w \rangle}$  and the angle brackets denoting the Monte Carlo average over the samples.

The step  $d\mathbf{w}$  calculated in Eq. (A7) is a *complex* vector with  $2N_w$  components which correspond to the variations of the real and imaginary parts of the  $N_w$  complex weights. Denoting with  $u_\ell$  the  $\ell$  th complex weight, the SR update  $du_\ell = du_\ell^R + i du_\ell^I$  is calculated from  $d\mathbf{w}$  as

$$\begin{aligned} du_\ell^R &= \text{Re}(dw_{2\ell-1}) - \text{Im}(dw_{2\ell}), \\ du_\ell^I &= \text{Re}(dw_{2\ell}) + \text{Im}(dw_{2\ell-1}). \end{aligned} \quad (\text{A11})$$

## 2. Efficient calculation of the step in parameter space

Rather than explicitly evaluating the  $S$ -matrix inverse for calculating the step in parameter space from Eq. (A7), it is numerically more efficient to solve the linear system

$$S_w d\mathbf{w} = -\gamma \mathbf{F}_w \quad (\text{A12})$$

for  $d\mathbf{w}$ . We adopt the MINRES-QLP algorithm [28], which is an iterative linear solver based on the Lanczos method. Lanczos tridiagonalization requires the calculation of the Krylov space,

which involves matrix-vector multiplications of the form  $S_w \mathbf{v}$ , where  $\mathbf{v}$  is a generic vector with  $2N_w$  entries. Since the explicit calculation of the  $S$  matrix has a computational cost of  $O(N_s N_w^2)$ , with  $N_s$  being the number of samples, we exploit the product structure of the covariance matrix to avoid its explicit calculation [1]. At every SR iteration, for each sample  $\mathcal{S}_n$  we store the local variational derivative estimator  $O_j(\mathcal{S}_n)$  defined in Eq. (A8) in the  $N_s \times 2N_w$  matrix  $\mathbb{O}$ , with elements  $(\mathbb{O})_{nj} = O_j(\mathcal{S}_n)$ . After  $\mathbb{O}$  has been evaluated, we compute

$$\underline{u} = \mathbb{O} \mathbf{v}, \quad (\text{A13})$$

where  $\underline{u}$  is a  $N_s$ -component vector with elements

$$(\underline{u})_n = \sum_{j=1}^{2N_w} \frac{(\partial_{w_j} \psi_w)(\mathcal{S}_n)}{\psi_w(\mathcal{S}_n)} (\mathbf{v})_j. \quad (\text{A14})$$

Then the evaluation of  $\mathbf{v}' = \frac{1}{N_s} \mathbb{O}^\dagger \underline{u}$  leads to

$$\begin{aligned} (\mathbf{v}')_k &= \sum_{j=1}^{2N_w} \frac{1}{N_s} \sum_{n=1}^{N_s} O_j^*(\mathcal{S}_n) O_j(\mathcal{S}_n) (\mathbf{v})_j \\ &= \sum_{j=1}^{2N_w} \langle O_k^* O_j \rangle (\mathbf{v})_j, \end{aligned}$$

and it is sufficient to shift these components by

$$(\mathbf{v}')_k \rightarrow (\mathbf{v}')_k - \langle O_k^* \rangle \sum_{j=1}^{2N_w} \langle O_j \rangle (\mathbf{v})_j$$

to retrieve  $\mathbf{v}' = S_w \mathbf{v}$  at an overall computational cost of  $O(N_s N_w)$ .

## 3. Metric rescaling of step length

In our numerical simulations we used a time-dependent imaginary-time step  $\gamma$  and adopted a local metric rescaling (LMR) technique for the optimization of its length [42], as we explain below. At each imaginary-time step, the length of the step in parameter space is rescaled according to the local metric in order to keep the effective step length in the variational manifold constant despite the nonorthogonal frame. Let us consider a generic real function  $f_w$  which depends on the variational parameters  $\mathbf{w}$  through the state  $|\hat{\psi}_w\rangle$ , i.e., a real function on the variational manifold embedded in the Hilbert space. This function could be the energy expectation value, the squared modulus of the overlap of  $|\hat{\psi}_w\rangle$  with a given state, or the distance between  $A|\hat{\psi}_w\rangle$  and  $B|\hat{\psi}_w\rangle$ , with  $A$  and  $B$  being some operators. Our problem is to find the optimal variation  $d\mathbf{w}$  of the variational parameters in the context of minimizing  $f_w$ . To this end we Taylor expand to first order

$$f_{w+\epsilon d\mathbf{w}} \simeq f_w + \epsilon d\mathbf{w} \cdot \nabla_{\mathbf{w}} f_w, \quad (\text{A15})$$

where  $\epsilon$  is a free small parameter chosen to be small enough that the above first-order approximation is justified. We want to find  $d\mathbf{w}$  such that  $f_{w+\epsilon d\mathbf{w}}$  is minimal, under the constraint of a fixed step length on the variational manifold, as measured by  $ds_{w+d\mathbf{w}}$  for a variation  $\mathbf{w} \rightarrow \mathbf{w} + d\mathbf{w}$ . Explicitly, we get

$$ds_{w+d\mathbf{w}} = \sqrt{\sum_{i,j} dw_i^* (S_w)_{i,j} dw_j} = 1, \quad (\text{A16})$$

where  $(S_w)_{i,j}$  are the components of the metric (or covariance matrix) defined before [see Eq. (A2)]. In the following we drop the subscripts in  $ds_{w+d}$  and  $S_w$  for simplicity. This constrained optimization problem amounts to the minimization of the function

$$\mathcal{L}(d\mathbf{w}, \lambda) = d\mathbf{w} \cdot \nabla_{\mathbf{w}} f_w + \lambda(ds^2 - 1), \quad (\text{A17})$$

yielding the system

$$\begin{aligned} \nabla_{d\mathbf{w}} \mathcal{L} &= \nabla_{\mathbf{w}} f_w + 2\lambda S d\mathbf{w} = 0, \\ \partial_{\lambda} \mathcal{L} &= ds^2 - 1 = 0. \end{aligned}$$

From the first equation we have

$$d\mathbf{w} = -\frac{1}{2\lambda} S^{-1} \nabla_{\mathbf{w}} f_w \equiv \frac{1}{2\lambda} \delta\mathbf{w}, \quad (\text{A18})$$

where we have introduced the *bare* step in parameter space  $\delta\mathbf{w} = -S^{-1} \nabla_{\mathbf{w}} f_w$ . Plugging this result into the second equation, we obtain

$$\lambda = \pm \frac{1}{2} \sqrt{\delta\mathbf{w}^\dagger S \delta\mathbf{w}}, \quad (\text{A19})$$

which we call the LMR factor. Since we want the variation of the function  $f_w$  to be negative, we pick the positive root for  $\lambda$  and finally arrive at

$$d\mathbf{w} = -\frac{S^{-1} \nabla_{\mathbf{w}} f}{\sqrt{(S^{-1} \nabla_{\mathbf{w}} f)^\dagger S (S^{-1} \nabla_{\mathbf{w}} f)}}. \quad (\text{A20})$$

At each SR iteration the *bare* step in parameter space  $\delta\mathbf{w}$  is calculated, and its length is rescaled with the LMR factor  $\lambda$  to make the effective step length in the variational manifold constant. The rescaled step  $d\mathbf{w}$  of Eq. (A20) is then multiplied by the free parameter  $\epsilon$  in order for the first-order expansion of Eq. (A15) to be valid. In the simulations we have used the

SR method; thus we substitute  $\nabla_{\mathbf{w}} f_w \rightarrow \mathbf{F}_w$ , where  $\mathbf{F}_w$  is the force vector defined in Eq. (A6). The *bare* step then becomes  $\delta\mathbf{w} = -S^{-1} \mathbf{F}_w$ , and the effective update of the weights is  $\epsilon d\mathbf{w} = \frac{\epsilon \delta\mathbf{w}}{\sqrt{\delta\mathbf{w}^\dagger S \delta\mathbf{w}}} = \gamma \delta\mathbf{w}$ . The  $\epsilon$  parameter is generally chosen to be time dependent. We started with an initial value  $\epsilon = 0.1$ , and close to the end we reduced it by a factor of 10 for a more accurate minimum search.

#### 4. Regularization of the metric tensor

Finally, we discuss some common issues related to the inversion of the  $S$  matrix. The matrix elements of  $S$  are calculated as Monte Carlo averages, and they are subject to statistical fluctuations. These may lead to very small, not even positive eigenvalues of  $S$ , which could amplify the fluctuations in the force vector when  $S^{-1} \mathbf{F}_w$  is calculated, leading to numerical instabilities of the SR scheme. One can adopt different regularization schemes to avoid those numerical instabilities [1,27]. One scheme amounts to adding a term proportional to the identity matrix, shifting all the diagonal elements by the same amount

$$S^{\text{reg.}} = S + \lambda_{\text{reg.}} I, \quad (\text{A21})$$

and the other one is a rescaling of the diagonal elements

$$(S^{\text{reg.}})_{k,k} = (1 + \lambda_{\text{reg.}}) (S)_{k,k}. \quad (\text{A22})$$

We found it useful to adopt the identity regularization of Eq. (A21) for the first ( $\sim 500$ ) iterations and then to switch to the second scheme [Eq. (A22)] towards the end of the simulation. With this choice we found better stability and smoother convergence towards the ground state, probably due to the fact that the diagonal regularization does not modify the ratio between the eigenvalues of the  $S$  matrix.

- 
- [1] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).  
[2] H. L. Stormer, A. Chang, D. C. Tsui, J. C. M. Hwang, A. C. Gossard, and W. Wiegmann, *Phys. Rev. Lett.* **50**, 1953 (1983).  
[3] R. B. Laughlin, *Phys. Rev. Lett.* **50**, 1395 (1983).  
[4] R. Prange and S. Girvin, *The Quantum Hall Effect* (Springer, Berlin, 1990).  
[5] W. Anderson, *Mater. Res. Bull.* **8**, 153 (1973).  
[6] V. Kalmeyer and R. B. Laughlin, *Phys. Rev. Lett.* **59**, 2095 (1987).  
[7] X.-G. Wen, *Quantum Field Theory of Many-Body Systems* (Oxford University Press, Oxford, 2007).  
[8] J. Dubail and N. Read, *Phys. Rev. B* **92**, 205307 (2015).  
[9] N. Read, *Phys. Rev. B* **95**, 115309 (2017).  
[10] S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).  
[11] I. P. McCulloch, [arXiv:0804.2509](https://arxiv.org/abs/0804.2509).  
[12] M. P. Zaletel, R. S. K. Mong, C. Karrasch, J. E. Moore, and F. Pollmann, *Phys. Rev. B* **91**, 165112 (2015).  
[13] L. Tagliacozzo, G. Evenbly, and G. Vidal, *Phys. Rev. B* **80**, 235127 (2009).  
[14] V. Murg, F. Verstraete, Ö. Legeza, and R. M. Noack, *Phys. Rev. B* **82**, 205105 (2010).  
[15] M. Gerster, M. Rizzi, P. Silvi, M. Dalmonte, and S. Montangero, *Phys. Rev. B* **96**, 195123 (2017).  
[16] B. Béri and N. R. Cooper, *Phys. Rev. Lett.* **106**, 156401 (2011).  
[17] G. Ortiz, D. M. Ceperley, and R. M. Martin, *Phys. Rev. Lett.* **71**, 2777 (1993).  
[18] V. Melik-Alaverdian, N. E. Bonesteel, and G. Ortiz, *Phys. Rev. Lett.* **79**, 5286 (1997).  
[19] J. Wang, S. D. Geraedts, E. H. Rezayi, and F. D. M. Haldane, [arXiv:1710.09729](https://arxiv.org/abs/1710.09729).  
[20] D. F. Schroeter, E. Kapit, R. Thomale, and M. Greiter, *Phys. Rev. Lett.* **99**, 097202 (2007).  
[21] E. Kapit and E. Mueller, *Phys. Rev. Lett.* **105**, 215303 (2010).  
[22] We note that the convergence and thus the overall polynomial cost of the VMC calculations can be challenged by complex energy landscapes with quasistable minima in the variational space (see Ref. [43] for a recent discussion in the context of neural networks).  
[23] G. Moore and N. Read, *Nucl. Phys. B* **360**, 362 (1991).  
[24] Z. Liu, E. J. Bergholtz, and E. Kapit, *Phys. Rev. B* **88**, 205101 (2013).  
[25] E. J. Bergholtz and Z. Liu, *Int. J. Mod. Phys. B* **27**, 1330017 (2013).  
[26] S. Sorella, *Phys. Rev. B* **64**, 024512 (2001).  
[27] S. Sorella, M. Casula, and D. Rocca, *J. Chem. Phys.* **127**, 014105 (2007).

- [28] S. C. T. Choi and M. A. Saunders, *ACM Trans. Math. Software* **40**, 16 (2014).
- [29] See the Appendix for technical details regarding the SR method.
- [30] D.-L. Deng, X. Li, and S. D. Sarma, *Phys. Rev. X* **7**, 021021 (2017).
- [31] Y. Huang and J. E. Moore, [arXiv:1701.06246](https://arxiv.org/abs/1701.06246).
- [32] D.-L. Deng, X. Li, and S. D. Sarma, *Phys. Rev. B* **96**, 195145 (2017).
- [33] H. Saito and M. Kato, *J. Phys. Soc. Jpn.* **87**, 014001 (2018).
- [34] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, *Phys. Rev. B* **96**, 205152 (2017).
- [35] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *Phys. Rev. B* **97**, 085104 (2018).
- [36] J. Shao, E.-A. Kim, F. D. M. Haldane, and E. H. Rezayi, *Phys. Rev. Lett.* **114**, 206402 (2015).
- [37] M. P. Zaletel and R. S. K. Mong, *Phys. Rev. B* **86**, 245305 (2012).
- [38] S. R. Clark, *J. Phys. A: Math. Theor.* **51**, 135301 (2018).
- [39] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, *Phys. Rev. X* **8**, 011006 (2018).
- [40] A. E. B. Nielsen, J. I. Cirac, and G. Sierra, *J. Stat. Mech.* (2011) P11014.
- [41] M. Casula, C. Attaccalite, and S. Sorella, *J. Chem. Phys.* **121**, 7110 (2004).
- [42] S.-I. Amari, *Neural Comput.* **10**, 251 (1998).
- [43] C. D. Freeman and J. Bruna, [arXiv:1611.01540](https://arxiv.org/abs/1611.01540).