Symmetries and Many-Body Excitations with Neural-Network Quantum States

Kenny Choo,[1] Giuseppe Carleo,[2] Nicolas Regnault,[3] and Titus Neupert[1]
[1]Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[2]Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, New York, New York 10010, USA
[3]Laboratoire Pierre Aigrain, Ecole normale supérieure, PSL University, Sorbonne Université, Université Paris Diderot, Sorbonne Paris Cité, CNRS, 24 rue Lhomond, 75005 Paris France

(Received 16 July 2018; published 19 October 2018)

Artificial neural networks have been recently introduced as a general ansatz to represent many-body wave functions. In conjunction with variational Monte Carlo calculations, this ansatz has been applied to find Hamiltonian ground states and their energies. Here, we provide extensions of this method to study excited states, a central task in several many-body quantum calculations. First, we give a prescription that allows us to target eigenstates of a (nonlocal) symmetry of the Hamiltonian. Second, we give an algorithm to compute low-lying excited states without symmetries. We demonstrate our approach with both restricted Boltzmann machines and feed-forward neural networks. Results are shown for the one-dimensional spin-1/2 Heisenberg model, and for the one-dimensional Bose-Hubbard model. When comparing to exact results, we obtain good agreement for a large range of excited-states energies. Interestingly, we find that deep networks typically outperform shallow architectures for high-energy states.

DOI: 10.1103/PhysRevLett.121.167204

Introduction.—Artificial neural networks (ANN) for machine learning (ML) are quickly becoming an indispensable tool in fundamental sciences. In the context of statistical physics, for instance, machine learning techniques have been used successfully for classifying phases of matter and phase transitions [1–6], speeding up Monte Carlo simulations [7,8], molecular modeling [9,10] and more. These applications are close in spirit to classical ML tasks, in that the networks are trained using labeled data to learn a certain target function known only on a finite number of data points. In the context of many-body quantum physics, a representation of the many-body wave function based on ANN has been proposed in Ref. [11]. ANN representations can be used in unsupervised applications of ANN and ML, where no labeled data are given *a priori*. Applications in this sense include the simulation of ground states [11–17], and the reconstruction of quantum states from experimental measurements [18,19].

The key difficulty in many-body problems is the exponential growth of the wave-function complexity with the number of particles. This can be circumvented in interesting physical applications using either stochastic sampling approaches or compact representations of the many-body states. Popular techniques belonging to the two categories are, respectively, quantum Monte Carlo methods [20,21], and tensor-network approaches [22,23]. Known limitations of these approaches are, however, the sign problem [24] for quantum Monte Carlo calculations, and the entanglement problem for tensor networks. As a result, interesting many-body problems are currently inaccessible by state of the art techniques, including key strongly interacting fermionic problems in two dimensions, out-of-equilibrium dynamics, and excited states. The learning scheme proposed in Ref. [11] leverages the ability of ANN to compactly represent highly dimensional functions, and thus belongs to the second category of variational wave function approaches. A distinct feature of this approach is its ability to capture longer range correlations and entanglement structures [15,25] leading to highly accurate representations of many-body states [11,13,14,18,26,27].

Previous works [11–16] focused on obtaining ground states with ANN variational quantum states. However, for the method to become a comprehensive tool for quantum many-body calculations, it is crucial to have controlled access to low-lying excited states. This is needed to answer questions such as: Is the ground state gapped or gapless? What is the ground state degeneracy? What are the structure and the dispersion of low-lying excitations? In this Letter, we use ANN variational quantum states to compute excited states and target states with fixed quantum numbers (e.g., momentum). We achieve this in two ways, first by taking advantage of Abelian spatial symmetries such as translational symmetry and second by orthogonalizing the wave function with respect to the ground state. We demonstrate our approach with both restricted Boltzmann machine (RBM) states and three-layer feed-forward neural networks (FFNN) as variational wave functions. We test the methods on the one-dimensional spin-1/2 Heisenberg model and on the one-dimensional Bose-Hubbard model at filling one. When comparing to available exact results, we obtain relative errors between 10^{-5}–10^{-3} on the variational energies.

 © 2018 American Physical Society

Restricted Boltzmann machine.—For concreteness, consider a system of L spin-1/2 degrees of freedom denoted by $\sigma_j = \pm 1$, $j = 1, \ldots, L$. RBMs were proposed in Ref. [11] as a variational ansatz for many-body wave functions of such systems. The RBM wave function is given by

$$\Psi(\boldsymbol{\sigma}) = \sum_{\boldsymbol{h}} e^{\sum_j a_j \sigma_j + \sum_i b_i h_i + \sum_{ij} h_i W_{ij} \sigma_j}, \tag{1}$$

where the sum runs over $\boldsymbol{h} = (h_1, h_2, \ldots, h_M)$ with the spin variables $h_i \in \{-1, 1\}$ for $i = 1, \ldots, M$. The physical spins $\boldsymbol{\sigma}$ are called a visible layer, and $\boldsymbol{h}$ is interpreted as a second—hidden—layer of auxiliary spins. The visible and hidden layers can be thought of as classical spins subjected to an interaction energy,

$$E(\boldsymbol{\sigma}, \boldsymbol{h}) = -\sum_j a_j \sigma_j - \sum_i b_i h_i - \sum_{ij} h_i W_{ij} \sigma_j, \tag{2}$$

where a_j and b_i are known as the visible and hidden bias, respectively, analogous to a local magnetic field, and W_{ij} are the weights corresponding to interactions between visible and hidden nodes. Here, to apply the formalism to general wave functions, complex-valued weights and biases are used [11]. In this case, Eq. (2) does not have an analog in classical statistical physics.

Tracing over the hidden variables $\boldsymbol{h}$, Eq. (1) reduces to

$$\log \Psi(\boldsymbol{\sigma}) = \sum_j a_j \sigma_j + \sum_i \log \left[\cosh \left(b_i + \sum_j W_{ij} \sigma_j \right) \right] \tag{3}$$

up to some additive constant corresponding to an overall factor of the wave function.

Feed-forward neural network.—The second type of network that we consider is a FFNN. The input to the network is a configuration $\boldsymbol{\sigma}$, indexing the many-body basis states. It could be a binary vector for a spin half system or a vector of integers for spinless bosons.

We construct an ℓ-layer FFNN as follows. Let $\boldsymbol{v}_n$ be the M_n-component vector output from layer n and define $\boldsymbol{v}_0 = \boldsymbol{\sigma}$, where M_n is the number of neurons in layer n. At each layer, we apply an affine map followed by elementwise nonlinear function f (the so-called activation function) $\boldsymbol{v}_n \to \boldsymbol{v}_{n+1} = f(\boldsymbol{W}_n \boldsymbol{v}_n + \boldsymbol{b}_n)$, where $\boldsymbol{W}_n$ is a matrix of size $M_{n+1} \times M_n$ known as the weight matrix and $\boldsymbol{b}_n$ is a vector called the bias. The activation function f can be chosen freely. Since we would like the ansatz to be generic, we again have to allow for complex-valued parameters. Inspired by the effectiveness of the RBM, we choose $f(x) = \log[\cosh(x)]$ from here on.

The final layer consists of one neuron so the output $\boldsymbol{v}_\ell$ is a one-dimensional vector which corresponds to the value $\boldsymbol{v}_\ell = \log[\Psi(\boldsymbol{\sigma})]$. For the case of a single hidden layer followed by the final output layer, the ansatz can be written as

$$\log \Psi(\boldsymbol{\sigma}) = \boldsymbol{W}_1 [f(\boldsymbol{W}_0 \boldsymbol{\sigma} + \boldsymbol{b}_0)], \tag{4}$$

which reduces to the RBM in Eq. (3) without visible bias, if $\boldsymbol{W}_1 = (1, 1, 1, \ldots)$. We do observe in our tests that the single hidden layer FFNN has a similar performance to the RBM. Therefore, to go beyond the RBM we focus on a FFNN with three layers: two hidden layers followed by an output layer. The ansatz becomes

$$\log \Psi(\boldsymbol{\sigma}) = \boldsymbol{W}_2 (f\{\boldsymbol{W}_1 [f(\boldsymbol{W}_0 \boldsymbol{\sigma} + \boldsymbol{b}_0)] + \boldsymbol{b}_1\}). \tag{5}$$

See the Supplemental Material [28] for information regarding the optimization of the networks.

Abelian symmetries.—We now explain how to use a network to represent an eigenstate with a certain symmetry. Let $\{\hat{T}_1, \ldots, \hat{T}_\nu\}$ be the generators of a finite Abelian symmetry group G of order ν, where the elements $g \in G$ act on the configurations of the system as $g\boldsymbol{\sigma} = \boldsymbol{\sigma}'$. Since G is Abelian, its irreducible representations are purely one dimensional. A wave function belongs to an irreducible representation with character $\{\omega_1, \ldots, \omega_\nu\}$ corresponding to the ν generators if

$$\hat{T}_i |\Psi\rangle = \omega_i |\Psi\rangle \Rightarrow \Psi(\hat{T}_i \boldsymbol{\sigma}) = \omega_i \Psi(\boldsymbol{\sigma}). \tag{6}$$

In order to obtain the eigenstate corresponding to this irreducible representation, we want the output of the network to obey Eq. (6). Since the network represents the logarithm of the wave function, this means $\log \Psi(\hat{T}_i \boldsymbol{\sigma}) = \log \omega_i + \log \Psi(\boldsymbol{\sigma})$. Because of the highly nonlinear form of the wave function representation, it is not straightforward to adjust the weights of the network such that this condition is strictly satisfied. Instead, we solve the problem of obtaining a neural-network representation with a specific eigenvalue ω_i as follows: Let $\log \tilde{\Psi}(\boldsymbol{\sigma})$ represent the value obtained from the network as given by Eqs. (3) or (5). Next, consider the equivalence classes of configurations related by the symmetry group G, i.e., $[\boldsymbol{\sigma}] = \{g\boldsymbol{\sigma} \colon \forall\ g \in G\}$. For each equivalence class, we pick a canonical configuration $\boldsymbol{\sigma}_{\text{canonical}}$. We then define the amplitude of a configuration $\boldsymbol{\sigma}$ to be

$$\log \Psi(\boldsymbol{\sigma}) = \sum_{i=1}^{\nu} r_{i,\boldsymbol{\sigma}} \log \omega_i + \log [\tilde{\Psi}(\boldsymbol{\sigma}_{\text{canonical}})], \tag{7}$$

where the integers $r_{i,\boldsymbol{\sigma}}$ are the number of times the generator $\hat{T}_i$ needs to be applied to map the canonical configuration back to $\boldsymbol{\sigma}$. They are implicitly defined though the equation $\boldsymbol{\sigma} = \prod_{i=1}^{\nu} \hat{T}_i^{r_{i,\boldsymbol{\sigma}}} \boldsymbol{\sigma}_{\text{canonical}}$. Such a procedure guarantees that the condition in Eq. (6) is satisfied. Then, instead of evaluating $\log \Psi(\boldsymbol{\sigma})$ directly for generic $\boldsymbol{\sigma}$, we evaluate $\log \tilde{\Psi}(\boldsymbol{\sigma}_{\text{canonical}})$ and obtain $\log \Psi(\boldsymbol{\sigma})$ from Eq. (7). Minimizing the energy with this expression for $\log \Psi(\boldsymbol{\sigma})$ in turn gives the lowest eigenstate in the selected symmetry sector.

Let us illustrate this procedure for translational symmetry in one dimension. In this case, there is only a single generator $\hat{T}$. For a state $|\Psi\rangle$ with momentum k, the amplitude of a configuration $\boldsymbol{\sigma}$ is given by

$$\log \Psi(\boldsymbol{\sigma}) = \mathrm{i} r_{\boldsymbol{\sigma}} k + \log\left[\tilde{\Psi}(\boldsymbol{\sigma}_{\mathrm{canonical}})\right], \tag{8}$$

where $\boldsymbol{\sigma} = \hat{T}^{r_\sigma}\boldsymbol{\sigma}_{\mathrm{canonical}}$ [33].

Excited states without symmetry.—Many interesting physical problems possess (nearly) degenerate ground states that are not distinguished by good quantum numbers, for instance topologically ordered systems or spin glasses. In this case, the following procedure can be applied. The task is as follows: Given an ANN wave function which represents the ground state of a Hamiltonian, say $\Phi_0(\boldsymbol{\sigma})$, we would like to find the wave function Ψ with the lowest energy but orthogonal to Φ_0. To that end, we define

$$\Psi = \Phi_1 - \lambda \Phi_0, \tag{9}$$

where λ is a complex scalar and Φ_1 corresponds to a different ANN variational wave function with its own set of parameters. To enforce orthogonality between Ψ and Φ_0, i.e., $\langle\Phi_0|\Psi\rangle = 0$ we set $\lambda = (\langle\Phi_0|\Phi_1\rangle/\langle\Phi_0|\Phi_0\rangle)$, which can be computed in standard Monte Carlo fashion

$$\lambda = \sum_{\boldsymbol{\sigma}} \left(\frac{\Phi_1(\boldsymbol{\sigma})}{\Phi_0(\boldsymbol{\sigma})}\right) \frac{|\Phi_0(\boldsymbol{\sigma})|^2}{\sum_{\boldsymbol{\sigma}'}|\Phi_0(\boldsymbol{\sigma}')|^2} \approx \left\langle \frac{\Phi_1(\boldsymbol{\sigma})}{\Phi_0(\boldsymbol{\sigma})} \right\rangle_{N_s}, \tag{10}$$

where the average is carried over N_s samples generated from the distribution $|\Phi_0(\boldsymbol{\sigma})|^2$ through Monte Carlo sampling.

The optimization scheme then proceeds in two steps. (i) Sample the ground state wave function $|\Phi_0(\boldsymbol{\sigma})|^2$ to compute λ as in Eq. (10). (ii) Perform the imaginary time evolution with stochastic reconfiguration [29] on the full wave function $\Psi = \Phi_1 - \lambda\Phi_0$ using the updated λ.

Notice that in this scheme it is crucial to recompute λ at each step of the energy minimization, due to the intrinsic statistical noise of the optimization algorithm used here [see Eq. (4) of the Supplemental Material [28]]. As a result of the stochastic process, the state Ψ cannot be guaranteed to be exactly orthogonal to the ground state, but only approximately. In order to quantify the accuracy of the result, we can monitor the normalized overlap $(\langle\Phi_0|\Psi\rangle\langle\Psi|\Phi_0\rangle)/(\langle\Phi_0|\Phi_0\rangle\langle\Psi|\Psi\rangle)$, which can also be computed as a Monte Carlo average.

Results.—To test the methods introduced above, we study two one-dimensional models: the spin-1/2 antiferromagnetic Heisenberg chain and the nonintegrable [34] Bose-Hubbard chain. The former is defined by the Hamiltonian

$$\hat{H} = 4\sum_{i=1}^{L} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_{i+1}, \tag{11}$$

where $\hat{\boldsymbol{S}}_i$ are the spin-1/2 operators on site i and we choose periodic boundary conditions. The momentum-resolved spectrum of this model can be obtained using the Bethe ansatz [35].

As a first benchmark, we computed the momentum spectrum of the model with $L = 36$ sites using both the RBM and the three-layer FFNN and compared them to the results from exact diagonalization (ED). We set the hidden unit density defined by $\alpha_n = M_n/L$ to be $\alpha_1 = 3$ for the RBM and $\alpha_1 = 2$ ($\alpha_2 = 0.5$) for the first (second) layer of the FFNN. The ANN results, compared to those obtained from ED, are shown in Fig. 1(a). One can observe that the relative error $\epsilon = |(E - E_{\mathrm{exact}})/E_{\mathrm{ground}}|$ is much larger for

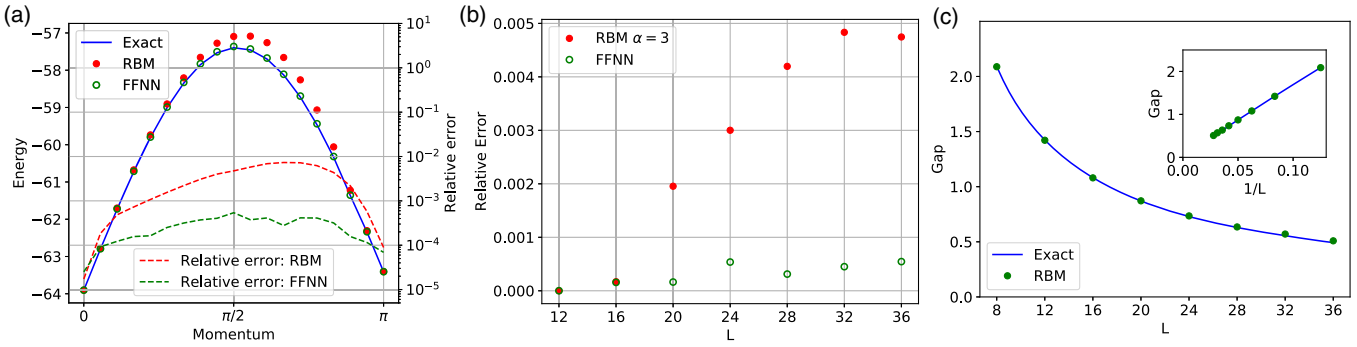

FIG. 1. (a) Momentum-resolved spectrum of the one-dimensional Heisenberg model with $L = 36$ spins. The blue line shows the exact values from ED, the green circles represent the energy obtained from a three-layer FFNN with hidden unit density $\alpha_1 = 2$ ($\alpha_2 = 0.5$) in the first (second) hidden layer (corresponding to 3996 free parameters) and red dots show the energy from an RBM with hidden unit density $\alpha_1 = 3$. (b) Relative error ϵ as a function of system size, for the $k = \pi/2$ state. For the RBM, we fix the hidden unit density $\alpha_1 = 3$, whereas for the FFNN we use a density of $\alpha_1 = 2$ ($\alpha_2 = 0.5$) in the first (second) hidden layer. For the $k = 0$ sector the relative error is $\sim 10^{-5}$. (c) Energy gap from the ground state to the first excited state of one-dimensional spin-1/2 Heisenberg model. The blue line shows the exact values computed using ED, the green circles represents the energy gap obtained from an RBM with hidden unit density $\alpha = 2$. FFNN results are identical to the RBM ones and are thus not shown here. The inset shows that the gap is inversely proportional to system size. The relative error of the excited states is less than 3×10^{-4} for all cases.

higher energy states, i.e., for momenta away from 0 or π. Moreover, the relative error for the RBM is higher than that of the three-layer FFNN, possibly suggesting that either the RBM ansatz is less efficient at representing those excited states or that the optimization of the network is caught in a local minimum. We checked that increasing the number of hidden units systematically improves the accuracy of the network.

In Fig. 1(b), we show the scaling of the relative error with system size for the two different network architectures, which shows that the three-layer FFNN systematically performs better than an RBM with a comparable number of parameters. Whereas the relative error remains roughly constant with system size for the FFNN, the RBM error instead seems to increase linearly. Once again, this circumstance does not strictly imply that RBM machines are less expressive than FFNNs, since optimization is also an extremely crucial ingredient to be considered.

Next, using the two-step method described above, we obtained the energy gap from the ground state to the first excited state as a function of system size L. This way, we do not use any information about the translation symmetry. Exact values were computed using ED. The results are shown in Fig. 1(c). Here, the hidden unit density of Φ_1 [see Eq. (9)] was fixed at $\alpha_1 = 2$ (except the $L = 40$ computation where we used $\alpha_1 = 4$), while the ground state Φ_0 was obtained using $\alpha_1 = 4$. This choice of hidden unit densities gives us a relative error below 3×10^{-5} for the ground states and below 2×10^{-4} for the excited states. It is necessary to compute the ground state accurately, since the error necessarily propagates to the excited state wave function due to the relation $\Psi = \Phi_1 - \lambda\Phi_0$. We verified that the overlap with the ground state is below 1% for a sample size of about 2000.

We now turn to the Bose-Hubbard model in one dimension with periodic boundary conditions,

$$\hat{H} = -t\sum_{i=1}^{L}(\hat{c}_i^\dagger \hat{c}_{i+1} + \text{H.c.}) + \frac{U}{2}\sum_{i=1}^{L}\hat{n}_i(\hat{n}_i - 1), \quad (12)$$

where $\hat{c}_i^\dagger$ and $\hat{c}_i$ are the boson creation and annihilation operators on site i, respectively, and $\hat{n} = \hat{c}_i^\dagger \hat{c}_i$ represents the local density at site i. For this problem, we experienced significant difficulty in lowering the relative error in both the two-layer FFNN and the RBM even with a large number of hidden units, suggesting that either optimization is difficult or that the expressiveness of the ansatz is limited. A three-layer FFNN, on the other hand, converged significantly better.

We set $U = 1$ and consider two system sizes. First the case of $N = 10$ bosons on a one-dimensional periodic lattice with $L = 10$ sites, for which exact results are easily obtained. We used a three-layer FFNN with hidden unit density $\alpha_1 = 4$ ($\alpha_2 = 1$) in the first (second) hidden layer (860 free parameters), and a RBM with hidden unit density $\alpha_1 = 8$ (890 free parameters). The relative error on the FFNN was lower than 5×10^{-4} for all momenta, whereas for the RBM one can see the error is increasing for larger momenta. The results are shown in Fig. 2(b).

Next, we show in Fig. 2(b) the results for $N = 40$ bosons in $L = 40$ sites. Here, the full (within the fixed particle number sector) Hilbert space dimension ($\sim 5 \times 10^{22}$) is too large to obtain results using ED. We could only infer the eigenenergies of the lowest few momentum sectors by matching with the lowest few eigenstates computed with MPS, since it is not straightforward to include momentum resolution in MPS. Although MPS [36,37] can in principle

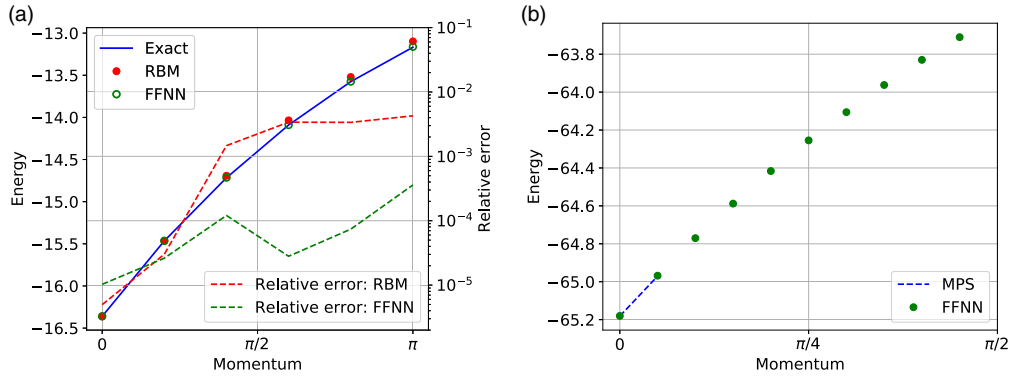

FIG. 2. Momentum-resolved spectrum of weakly interacting $U = 1$ bosons on a one-dimensional periodic lattice. (a) $N = 10$ bosons in $L = 10$ sites. The blue line shows the analytically calculated value and the green circles indicate the value obtained from a three-layer FFNN with hidden unit density $\alpha_1 = 4$ ($\alpha_2 = 1$) in the first (second) hidden layer (860 free parameters). The red circles show the value from an RBM with hidden unit density $\alpha_1 = 8$ (890 free parameters). The dashed lines indicate the relative error. (b) $N = 40$ bosons in $L = 40$ sites. The dashed blue line shows values inferred from MPS calculations. The green circles indicate the values obtained from a three-layer feed-forward neural network with hidden unit density $\alpha_1 = 2$ ($\alpha_2 = 1$) in the first (second) hidden layer (6560 free parameters) except for the last point $k = 18\pi/40$ where we used $\alpha_1 = 2$ ($\alpha_2 = 1$) in the first (second) hidden layer (8180 free parameters). We show only the first 10 momenta.

be used to determine momentum spectra, it is challenging to efficiently use this method to higher dimensions and bosons.

Conclusions and outlook.—We showed that artificial neural networks can be used as a variational Monte Carlo ansatz for obtaining excited states. In particular, we showed two ways to achieve this: first, by using Abelian spatial symmetries such as translational symmetry and, second, by using a superposition of two neural networks such that the combined network represents a state orthogonal to the ground state. While the methods presented here were demonstrated using simple networks (RBM and FFNN), they can in principle be used with any network architecture. Note that the models we employed do not have a sign problem overall. However, solving them in a sector of constant nonzero momentum generally introduces a sign problem: The wave function necessarily acquires complex amplitudes. Future work on more challenging models may require the use of more powerful networks, such as the convolutional arithmetic circuits or recurrent neural networks, which were shown to be highly efficient in representing entangled states [38]. Our strategy can be generalized to cases beyond spatial symmetries, e.g., permutational symmetry in fermionic systems would be a natural extension for future studies. Our results will be useful for the study generic many-body quantum systems which are hard to access with other methods. For instance, all gapped chiral states in two dimensions are hard to represent by tensor networks, but can be captured by RBM states [39–41]. In addition, any three-dimensional strongly correlated system, including topologically ordered spin liquids, is a natural candidate for future studies. This opens a new class of systems for numerical studies, which are beyond the reach of ED and density matrix renormalisation group.

We would like to thank Frank Schindler for insightful ideas regarding translational symmetries. K. C. was supported by the European Unions Horizon 2020 research and innovation program (ERC-StG-Neupert-757867-PARATOP). K. C. thanks the Flatiron Institute, a division of the Simons foundation, for hospitality. The ED computations were carried out with the DIAGHAM library. The MPS computations were done using the ALPS package [42]. The ANN computations were based on NETKET [43].

[1] E. van Nieuwenburg, Y. Liu, and S. Huber, Nat. Phys. **13**, 435 (2017).
[2] J. Carrasquilla and R. Melko, Nat. Phys. **13**, 431 (2017).
[3] F. Schindler, N. Regnault, and T. Neupert, Phys. Rev. B **95**, 245134 (2017).
[4] N. Yoshioka, Y. Akagi, and H. Katsura, Phys. Rev. B **97**, 205110 (2018).
[5] R. Kaubruegger, L. Pastori, and J. C. Budich, Phys. Rev. B **97**, 195136 (2018).
[6] J. Venderley, V. Khemani, and E.-A. Kim, Phys. Rev. Lett. **120**, 257204 (2018).
[7] H. Shen, J. Liu, and L. Fu, Phys. Rev. B **97**, 205140 (2018).
[8] L. Huang and L. Wang, Phys. Rev. B **95**, 035105 (2017).
[9] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Phys. Rev. Lett. **108**, 058301 (2012).
[10] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, Phys. Chem. Chem. Phys. **18**, 13754 (2016).
[11] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[12] H. Saito, J. Phys. Soc. Jpn. **86**, 093001 (2017).
[13] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Phys. Rev. B **96**, 205152 (2017).
[14] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Phys. Rev. X **8**, 011006 (2018).
[15] D.-L. Deng, X. Li, and S. Das Sarma, Phys. Rev. X **7**, 021021 (2017).
[16] H. Saito and M. Kato, J. Phys. Soc. Jpn. **87**, 014001 (2018).
[17] G. Carleo, Y. Nomura, and M. Imada, arXiv:1802.09558.
[18] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. **14**, 447 (2018).
[19] A. Rocchetto, E. Grant, S. Strelchuk, G. Carleo, and S. Severini, npj Quantum Inf. **4**, 28 (2018).
[20] D. Ceperley and B. Alder, Science **231**, 555 (1986).
[21] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, New York, 2017).
[22] S. White, Phys. Rev. Lett. **69**, 2863 (1992).
[23] F. Verstraete, F. Murg, and J. Cirac, Adv. Phys. **57**, 143 (2008).
[24] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).
[25] X. Gao and L.-M. Duan, Nat. Commun. **8**, 662 (2017).
[26] R. Kaubruegger, L. Pastori, and J. C. Budich, Phys. Rev. B **97**, 195136 (2018).
[27] Z. Cai and J. Liu, Phys. Rev. B **97**, 035116 (2018).
[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.121.167204 for details and information regarding the optimization [29] and sampling [30–32] of the networks.
[29] S. Sorella, M. Casula, and D. Rocca, J. Chem. Phys. **127**, 014105 (2007).
[30] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[31] R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **57**, 2607 (1986).
[32] H. G. Katzgraber, S. Trebst, D. A. Huse, and M. Troyer, J. Stat. Mech. (2006) P03018.
[33] We choose the canonical configuration to be the lexicographically smallest one, e.g., $\boldsymbol{\sigma} = (1,0,1,1,0,0) \rightarrow (0,0,1,0,1,1) = \hat{T}^2\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\text{canonical}}$.
[34] A. R. Kolovsky and A. Buchleitner, Europhys. Lett. **68**, 632 (2004).
[35] A. Klumper, M. T. Batchelor, and P. A. Pearce, J. Phys. A **24**, 3111 (1991).
[36] J. Haegeman, T. J. Osborne, and F. Verstraete, Phys. Rev. B **88**, 075133 (2013).

[37] V. Zauner-Strauber, L. Vanderstraeten, J. Haegeman, I.-P. McCulloch, and F. Verstraete, Phys. Rev. B **97**, 235155 (2018).
[38] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, arXiv: 1803.09780.
[39] S. R. Clark, J. Phys. A **51**, 135301 (2018).
[40] R. Kaubruegger, L. Pastori, and J. C. Budich, Phys. Rev. B **97**, 195136 (2018).
[41] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Phys. Rev. X **8**, 011006 (2018).
[42] M. Dolfi, B. Bauer, S. Keller, A. Kosenkov, T. Ewart, A. Kantian, T. Giamarchi, and M. Troyer, Comput. Phys. Commun. **185**, 3430 (2014).
[43] G. Carleo *et al.*, NetKet (2018), https://www.netket.org.