Two-dimensional frustrated J_1-J_2 model studied with neural network quantum states

Kenny Choo [1], Titus Neupert,[1] and Giuseppe Carleo[2]

[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*

The use of artificial neural networks to represent quantum wave functions has recently attracted interest as a way to solve complex many-body problems. The potential of these variational parametrizations has been supported by analytical and numerical evidence in controlled benchmarks. While approaching the end of the early research phase in this field, it becomes increasingly important to show how neural-network states perform for models and physical problems that constitute a clear open challenge for other many-body computational methods. In this paper, we start addressing this aspect, concentrating on a presently unsolved model describing two-dimensional frustrated magnets. Using a fully convolutional neural network model as a variational ansätz, we study the frustrated spin-1/2 J_1-J_2 Heisenberg model on the square lattice. We demonstrate that the resulting predictions for both ground-state energies and properties are competitive with, and often improve upon, existing state-of-the-art methods. In a relatively small region in the parameter space, corresponding to the maximally frustrated regime, our ansätz exhibits comparatively good but not the best performance. The gap between the complexity of the models adopted here and those routinely adopted in deep-learning applications is, however, still substantial, such that further improvements in future generations of neural-network quantum states are likely to be expected.

I. INTRODUCTION

With ever-improving computational resources, techniques, and data sets, machine learning has in recent years proven itself to be an extremely versatile tool for solving tasks previously thought impossible for a computer in the near future [1]. Deep learning, a branch of machine learning based on the use of deep artificial neural networks (ANNs) has played a very important role in these developments and it is currently largely believed that it will be an important computational method for years to come. In the field of condensed-matter physics, several machine-learning applications have been put forward in recent years. For example, they have been used to successfully classify phases of matter [2–5], to perform quantum state tomography [6–8], to simulate quantum computers [9], to classify experimental data [10–12], to infer phase transitions [13], and much more.

In the realm of computational quantum physics, the restricted Boltzmann machine (RBM), a type of ANN, was proposed as a variational ansätz [14] for many-body quantum systems. Since then, there has been a burst of research investigating the viability of such an ansätz. It has been shown that unlike other variational ansätze [15], wave functions based on RBMs can potentially capture long-ranged and volume-law scaling entanglement [16]. Its representation properties have also been extensively characterized, and by now it is known that RBM and related states can efficiently describe the ground states of many physical Hamiltonians [17–20]. Motivated by these theoretical successes, the community has pushed ahead and explored a wide variety of ANNs such as feedforward neural networks [21–24], deep Boltzmann machines [25,26], and a variety of other neural-network in-

spired ansätze [27,28]. These approaches have undergone an extensive phase of benchmarks, and have been compared to existing exact results, both in one and two dimensions, typically showing very good accuracy for ground- and excited-state properties. While this phase of benchmark has overall been important to assess the potential of this approach, the method is yet to be fully deployed on manifestly open problems, for which the application of ANN states could prove beneficial to resolve inconclusive results from other methods.

Here, we consider the case of the antiferromagnetic spin-1/2 J_1-J_2 model on the square lattice, a prototypical frustrated magnetic system for which no exact solution is known. Despite active research on the model for the past few decades, one of the chiefly open questions is whether a spin-liquid phase exists around the point of maximum frustration. Numerous computational methods [29–46] have been used to study this problem, often finding conflicting conclusions. The active and long-lasting interest in the J_1-J_2 model makes it an especially useful and nontrivial benchmark for ANN techniques, since the best of the currently available many-body techniques have been used to study it.

In this paper, we demonstrate that ANNs are a viable and competitive variational ansätz to study frustrated spin models in two dimensions. Our wave function is parametrized as a type of feedforward neural network known as a convolutional neural network (CNN), routinely used in top applications of deep learning in industry. We concentrate on the two-dimensional J_1-J_2 model on the square lattice, providing results for both the ground-state energy and spin correlation functions. We benchmark our results on the 6×6 cluster with exact diagonalization (ED) results [47] and on the 10×10 cluster with density matrix renormalization group (DMRG)

calculations [41] as well as traditional variational Monte Carlo (VMC) based on Gutzwiller-projected mean field Fermionic wave functions [40]. We show that for several points in the phase diagram, the CNN variational energies we obtain improve upon those obtained by the other techniques. We also find that there is a small window of frustration ratios for which our method, while being competitive with the state of the art, is not yet delivering cutting-edge results. Finally, we discuss the origin of these limitations and some strategy to further improve ANN-based methods in future works.

II. MODEL

The spin-1/2 J_1-J_2 Heisenberg model is defined by the Hamiltonian

$$\hat{H} = J_1 \sum_{\langle ij \rangle} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_j, \tag{1}$$

where $\hat{\boldsymbol{S}}_i = (\hat{S}_i^x, \hat{S}_i^y, \hat{S}_i^z)$ here represents the spin operators at site i of a square lattice with periodic boundary conditions. The symbols $\langle \cdots \rangle$ and $\langle\langle \cdots \rangle\rangle$ indicate pairs of nearest- and next-nearest-neighbor sites, respectively. We are interested in the case where both the nearest and next-nearest neighbor interactions are antiferromagnetic, i.e., $J_1, J_2 \geqslant 0$, so the magnetic interactions are frustrated. For simplicity, we fix $J_1 = 1$ throughout this paper. In addition, since total magnetization is a conserved quantity in this model and it is expected that the ground state is in the zero magnetization sector, we shall restrict our analysis to this sector, denoting by $\boldsymbol{\sigma}$ spin configurations belonging to it.

When $J_2 = 0$, the system is unfrustrated and there is a well-established Neel long-range order [48,49]. In the opposite limit $J_2 \gg J_1$, the system is also magnetically ordered with pitch vector $\boldsymbol{Q} = (\pi, 0)$ or $(0, \pi)$. However, in the intermediate regime, where $J_2/J_1 \approx 0.5$, the system is highly frustrated. There have been numerous conflicting proposals for the candidate ground state, such as the plaquette valence-bond state [29–32], the columnar valence-bond state [33–35], or a gapless spin liquid [36,37,39–42,46], but the correct answer is still unknown.

While we make no claim about the nature of the physics in the frustrated regime $J_2/J_1 \approx 0.5$, we hope to present a new and potentially viable ansätz for variational methods. We shall compare our variational energies with other methods such as DMRG and traditional VMC based on a projected fermionic ansätz and show that there are regimes in parameter space, where neural network states have the lowest variational energies.

III. NEURAL NETWORK QUANTUM STATES

Neural network quantum states (NQSs) were first proposed in Ref. [14], where a RBM was used as a variational ansätz for the Heisenberg model on a square lattice, corresponding to $J_2 = 0$ in Eq. (1). Since neural networks are essentially functions with a large number of parameters, the idea behind NQSs is to interpret the output of a network as the complex amplitudes of a wave function $\Psi(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ is a vector representing a spin configuration of the system. In this paper, we shall use a variational wave function that is expressed as

a deep feedforward CNN, which has been shown to be able to support volume law entanglement more efficiently than shallow networks such as the RBM [50].

A feedforward neural network has the most basic structure of a variational function composed of a series of transformations called layers. It takes an input vector $\boldsymbol{v}^{(0)}$ and successively applies a sequence of layers to map it to the output, i.e., $\boldsymbol{v}^{(0)} \rightarrow \boldsymbol{v}^{(1)} \rightarrow \cdots \rightarrow \boldsymbol{v}^{(L)}$ for a L layer network. A generic layer implements an affine transformation followed by a nonlinear transformation, which is usually taken to be a coefficient wise operation,

$$\boldsymbol{v}_i^{(n)} = g\left(\sum_j W_{ij} v_j^{(n-1)} + b_i \right), \tag{2}$$

where W_{ij} are the elements of a weight matrix, $\boldsymbol{b}_i$ are the elements of a bias vector, and g is some nonlinear function. Since wave functions are in general complexvalued, we use complex-valued weights and biases. The nonlinear function is then a function over the complex numbers, i.e., $g : \mathbb{C} \rightarrow \mathbb{C}$. For the simulations done in this paper, g is either the rectified linear unit (ReLU) generalized to complex numbers [51,52],

$$g_{\text{ReLU}}(z) = \begin{cases} z & \text{if } \frac{3\pi}{4} > \arg z \geqslant -\frac{\pi}{4} \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

or the logarithm of the hyperbolic cosine (LnCosh) $g_{\text{lncosh}}(z) = \log[\cosh(z)]$. Although complex ReLU as defined above is not continuous, in practice we do not notice any convergence issues. On the other hand, we experienced some optimization problems when using the LnCosh activation in the deeper layers of the network, possibly due to divergences in the LnCosh function. Hence, for the simulations done in this paper, we use LnCosh activation only in the first layer and the complex ReLU in all subsequent layers.

In a CNN, the layers have a spatially local structure and are called convolutional layers. A convolutional layer is also an affine transformation as in Eq. (2), but with certain constraints being placed on the weights and biases: The transformation is separated into several independent "channels" and each channel is characterized by a "filter" matrix. A convolutional layer indexed by n with C_{n-1} input channels and C_n output channels will perform the transformation

$$\boldsymbol{v}_{m,j}^{(n)} = g\left(\sum_{l=1}^{C_{n-1}} \sum_{k=1}^{\mathcal{K}^{(n)}} K_{m,k}^{(n)} \boldsymbol{v}_{l,a_{jk}}^{(n-1)} + b_m^{(n)} \right), \tag{4}$$

where m, l are channel indices, $K_{m,k}^{(n)}$ are filter parameters in channel m, and $b_m^{(n)}$ is bias in channel m. The index a_{jk} indicates the position of the input image to be acted on by the k parameter of the filter so as to contribute to the j position of the output vector. The complete structure of our CNN and shape of the convolutional filters are shown in Fig. 1. In the final layer, all outputs of all channels are summed over, constituting a so-called average pooling layer. The full network has a total of 3838 complex-valued parameters, independent of the system size that we will apply it to. Notice that, contrary to a conventional CNN, we do not use pooling layers between each convolutional layer. The rationale is that it has been shown in Ref. [50] that the maximum entanglement entropy in a CNN with pooling layers obeys an area law as opposed
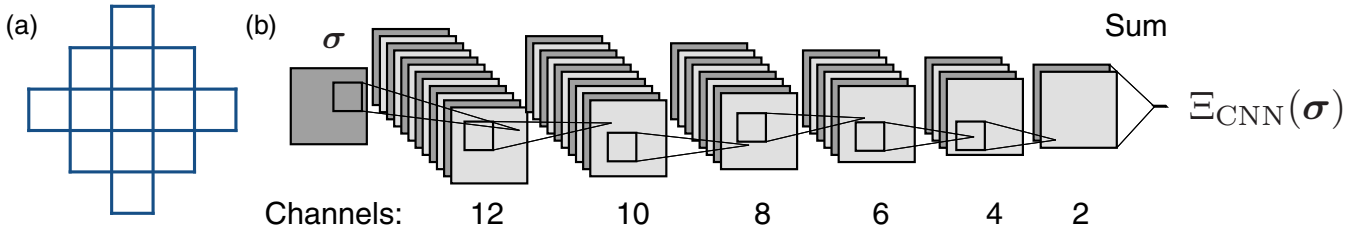
FIG. 1. Network architecture. (a) Shape of the filter which we apply across the square lattice. (b) Full architecture of the convolutional neural network used in this work. There are six convolutional layers followed by an output layer which simply sums the values of the prenultimate layer. In the first layer, we use the logarithm of the hyperbolic cosine as the activation function $g_{\text{lncosh}}(z) = \log[\cosh(z)]$, while in all other layers the activation function is given by the complex generalization of the ReLU [51,52]. The total number of complex-valued parameters in the network is 3838.

to the volume law scaling entanglement entropy in a CNN without pooling. Since the ground states of the J_1-J_2 model are potentially critical,with entanglement entropies exceeding an area law, it is important that the expressibility of the variational ansätz not be limited as such.

The complete CNN thus represents an explicitly translationally invariant function with zero momentum defined on the configurations of the Hilbert space, i.e., $\Xi_{\text{CNN}} : \{\boldsymbol{\sigma}\} \to \mathbb{C}$, where $\boldsymbol{\sigma}$ refers to a computational basis of the Hilbert space. The variational wave function represented by the CNN is then given by

$$\Psi_{\text{CNN}}(\boldsymbol{\sigma}) = \exp[\Xi_{\text{CNN}}(\boldsymbol{\sigma})]. \quad (5)$$

One of the advantages using such a fully convolutional structure resides in an intrinsically more efficient learning procedure, as opposed to fully connected layers. For example, the set of kernels optimized for a smaller system size provides a good starting point for the optimization of a larger system size such that relatively few iterations are needed for convergence. We use the same network structure, with the same number of variational parameters, for all system sizes studied.

A. Sign structure of the ground state

It is known that in the extremal limits ($J_1 = 0$ or $J_2 = 0$), the ground-state wave function obeys a simple sign rule. In those limits, the ground-state wave function takes the form

$$|\text{GS}\rangle = \sum_{\boldsymbol{\sigma}} (-1)^{M_{\mathcal{A}}(\boldsymbol{\sigma})} \Psi(\boldsymbol{\sigma}) |\boldsymbol{\sigma}\rangle, \quad (6)$$

where $M_{\mathcal{A}}(\boldsymbol{\sigma})$ is the total number of up spins on a subset $\mathcal{A}$ of the sites and $\Psi(\boldsymbol{\sigma}) \geqslant 0$. When $J_2 = 0$, the subset $\mathcal{A}$ is given by one of the two bipartite components of the square lattice, leading to the so-called Marshall-Peierls sign rule [53]. On the other hand, when $J_1 = 0$, $\mathcal{A}$ can be chosen either to be every other row or every other column of spins on the square lattice.

These sign conventions can be exactly expressed by a suitable choice of the variational parameters and can then be in principle learned during the variational optimization. However, we find in general that it's more convenient to initialize our ansätz in one of the two sign conventions, and then optimize the resulting state. Our full variational ansätz then takes the form

$$|\Psi\rangle = \sum_{\boldsymbol{\sigma}} (-1)^{M_{\mathcal{A}}(\boldsymbol{\sigma})} \Psi_{\text{CNN}}(\boldsymbol{\sigma}) |\boldsymbol{\sigma}\rangle = \sum_{\boldsymbol{\sigma}} \Psi_{\text{CNN}}^{\mathcal{A}}(\boldsymbol{\sigma}) |\boldsymbol{\sigma}\rangle, \quad (7)$$

where $\Psi_{\text{CNN}}(\boldsymbol{\sigma})$ is given by Eq. (5). Since $\Psi_{\text{CNN}}(\boldsymbol{\sigma})$ is complex valued, the sign structure can in principle be changed by the network. However, the choice of the subset $\mathcal{A}$ does unavoidably present a bias in the variational ansätz and we unfortunately find that optimization is extremely challenging if an appropriate sign structure is not imposed. In this paper, we perform an optimization with both of the limiting sign structures and pick the one which gives the better variational energy.

B. Enforcing $\mathcal{C}_4$ Symmetry

In addition, as the model in Eq. (1) is defined on a square lattice, we can expect that the ground state of the model transforms within an irreducible representation of the symmetry. However, while the CNN we use is explicitly translationally invariant, it is not explicitly $\mathcal{C}_4$ symmetric, i.e., it does not need to be an irreducible representation of the $\mathcal{C}_4$ group of fourfold rotations. As the $\mathcal{C}_4$ group is Abelian, the irreducible representations are one-dimensional.

To ensure that the variational wave function is $\mathcal{C}_4$ symmetric, we symmetrize the wave function Ψ_{CNN}. This can be done generically as follows: Denoting $\hat{c}_4$ to be the generator of the $\mathcal{C}_4$ group, the symmetrized wave function

$$\tilde{\Psi}_{\text{CNN}}(\boldsymbol{\sigma}) = \sum_{r=0}^{3} \omega^r \Psi_{\text{CNN}}^{\mathcal{A}}(\hat{c}_4^r \boldsymbol{\sigma}), \quad (8)$$

transforms within an irreducible representation with character ω. Here, $\Psi_{\text{CNN}}^{\mathcal{A}}$ is defined through Eq. (7). This symmetrization ensures that correlation functions have the correct spatially symmetry, a circumstance especially important in the striped order phase at large J_2, where the *a priori* sign structure we start from is not rotationally invariant.

IV. VARIATIONAL MONTE CARLO OPTIMISATION

To optimize the parameters of the variational ansätz Eq. (8), we use the method known as stochastic reconfiguration (SR) [54], which can can be seen as an imaginary time evolution.

Consider a variational wave function $\Psi(\{\alpha_k^0\}) \in \mathbb{C}^{2^n}$ which depends on a set of variational parameters $\{\alpha_k^0\}_{k=1,\ldots,p}$. If we have a small variation in the parameters $\alpha_k = \alpha_k^0 + \delta\alpha_k$, then the corresponding wave function can be written as

$$\Psi(\{\alpha_k\}) = \Psi(\{\alpha_k^0\}) + \sum_{k=1}^{p} \delta\alpha_k \mathcal{O}_k \Psi(\{\alpha_k^0\}), \quad (9)$$

where $\mathcal{O}_k = \frac{\partial}{\partial \alpha_k} \log [\Psi(\{\alpha_k^0\})]$ are the logarithmic derivatives.

The SR scheme is then essentially an imaginary time evolution, which is given to first order by

$$\Psi'_{\text{exact}} = (1 - \epsilon \hat{H})\Psi. \tag{10}$$

The optimal coefficients $\{\delta \alpha_k^0\}_{k=1,\dots,p}$ that minimize the distance to the new wave function Ψ' with respect to the Fubini-Study metric,

$$\gamma(\phi, \Psi) = \arccos \sqrt{\frac{\langle \Psi | \phi \rangle \langle \phi | \Psi \rangle}{\langle \Psi | \Psi \rangle \langle \phi | \phi \rangle}}, \tag{11}$$

are then given by the solution to the linear equation:

$$\sum_{k'} [\langle \mathcal{O}_k^\dagger \mathcal{O}_{k'} \rangle - \langle \mathcal{O}_k^\dagger \rangle \langle \mathcal{O}_{k'} \rangle] \delta \alpha_{k'} = -\epsilon [\langle \mathcal{O}_k^\dagger \hat{H} \rangle - \langle \mathcal{O}_k^\dagger \rangle \langle \hat{H} \rangle]. \tag{12}$$

We update the parameters as $\alpha_k = \alpha_k^0 + \delta \alpha_k$ and repeat the procedure until convergence is achieved.

Since each SR iteration requires solving the linear system, the computational complexity of each step is $\mathcal{O}(N_w^2)$ (using the iterative conjugate gradients algorithm to solve the linear system), as compared to $\mathcal{O}(N_w)$ for the stochastic gradient descent (SGD) method, where N_w is the number of variational parameters. However, the SR method is known to perform better than SGD for variational optimization of small to midsized networks.

The expectation values $\langle \cdots \rangle$ can be estimated using Monte Carlo sampling. For instance, the energy can be estimated as

$$
\begin{aligned}
\langle \hat{H} \rangle &= \frac{\sum_{\sigma', \sigma} \Psi^*(\sigma') \langle \sigma | \hat{H} | \sigma' \rangle \Psi(\sigma)}{\sum_\sigma |\Psi(\sigma)|^2} \\
&= \sum_\sigma \left(\sum_{\sigma'} \langle \sigma | \hat{H} | \sigma' \rangle \frac{\Psi(\sigma')}{\Psi(\sigma)} \right) \frac{|\Psi(\sigma)|^2}{\sum_{\sigma'} |\Psi(\sigma')|^2} \\
&\approx \left\langle \sum_{\sigma'} \langle \sigma | \hat{H} | \sigma' \rangle \frac{\Psi(\sigma')}{\Psi(\sigma)} \right\rangle_M,
\end{aligned}
\tag{13}
$$

where $\langle \cdots \rangle_M$ denotes an average over a sample of configuration $\{\sigma\}$ drawn from the probability distribution given by $|\Psi(\sigma)|^2$. The sample is easily obtained by the Metropolis algorithm [55]. This average can be evaluated efficiently when the matrix $\langle \sigma | \hat{H} | \sigma' \rangle$ is sparse, which is the case for the Hamiltonian in Eq. (1) when the basis corresponds to a tensor product of local spin degrees of freedom. Finally, the linear system Eq. (12) is known to be highly ill conditioned, especially for a network with numerous parameters. To alleviate this problem, we regularize the matrix by adding a multiple of the identity. The details regarding the choice of the optimization hyperparameters are provided in the Appendix.

V. RESULTS AND DISCUSSION

Using the variational ansätz presented in Eq. (8) together with the SR method described above, we now discuss the results obtained on the 6×6 as well as the 10×10 square lattice with periodic boundary conditions.

A. Comparison with ED

In the case of the 6×6 square lattice with periodic boundary conditions, the system is amenable to ED. In Fig. 2(a), we compare our variational energies with that of the Lanczos ED results found in Ref. [47]. We show also the relative error defined by $|\frac{E - E_{\text{exact}}}{E_{\text{exact}}}|$. While the relative error is of order 10^{-3} or lower for most of the parameter space, it is immediately clear that the simulation is the most challenging close to the point of maximum frustration $J_2 \approx 0.5$. In our simulations, the largest relative error occurred at $J_2 = 0.55$, which roughly corresponds to the point where our prior sign structure is maximally violated. This suggests that although the phases of the wave functions can be changed by the network, the prior sign structure presents a strong bias for the simulation.

To study the properties of the NQS beyond the energy expectation value alone, we also measure the spin-spin correlation structure factors defined by

$$S^2(\vec{q}) = \frac{1}{N(N+2)} \sum_{i,j} \langle \hat{S}_i \cdot \hat{S}_j \rangle e^{i\vec{q} \cdot (\vec{r}_i - \vec{r}_j)}, \tag{14}$$

where $\vec{q}$ is the pitch vector and N is the total number of spins in the system. For $J_2 \lesssim 0.5$, the system is Neel ordered with pitch vector $\vec{q} = (\pi, \pi)$, while for $J_2 \gtrsim 0.5$, the system has a stripe order with $\vec{q} = (\pi, 0)$ or $(0, \pi)$. We plot these two structure factors in Fig. 2(b) in comparison with the exact values. The result is accurate for most of the parameter range apart from the transition region where frustration is maximal.

Finally, we measure the total spin $\langle \hat{S}^2 \rangle$ of our variational wave function as shown in Fig. 2(c). The model has SU(2) symmetry and the ground state is known to be in the singlet sector where the total spin is zero. It is apparent from Fig. 2 that the spike in the total spin coincides with the spike in relative error. As we have an even number of spin-1/2 degrees of freedom, the eigenvalues of the $\hat{S}^2$ are of the form $s(s+1)$ with $s = 0, 1, 2, \dots, N/2$ such that the next lowest eigenvalue is 2. Since the expectation value of the total spin of our variational ansätz is much less than 2 we can be certain that our wave function has a good overlap with the singlet sector.

B. Benchmarking with state-of-the-art methods

We now proceed to a system size which cannot be reached by ED calculations. The purpose is to benchmark our ansätz with the more established ones such as matrix product states and traditional VMC wave functions. In Fig. 3(a), we compare our variational energies on the 10×10 cluster with the DMRG results in Ref. [41] [where 8192 SU(2) states or equivalently 32 000 U(1) states were kept] as well as the Gutzwiller-projected mean-field fermionic ansätz, which is closely related to that used in Ref. [40]. We see that the CNN ansätz has competitive energies in all the range of J_2/J_1. With the notable exception of the reported point at $J_2/J_1 = 0.55$, NQS energies are very close or better than those reported in the literature. This point is around the region where the prior sign structures are most violated, and indicates a residual, unoptimized sign structure as a likely source of systematic error. Nevertheless, it is encouraging that by using a relatively simple ansätz, without much prior input information about the physics of the system, one is able to achieve variational
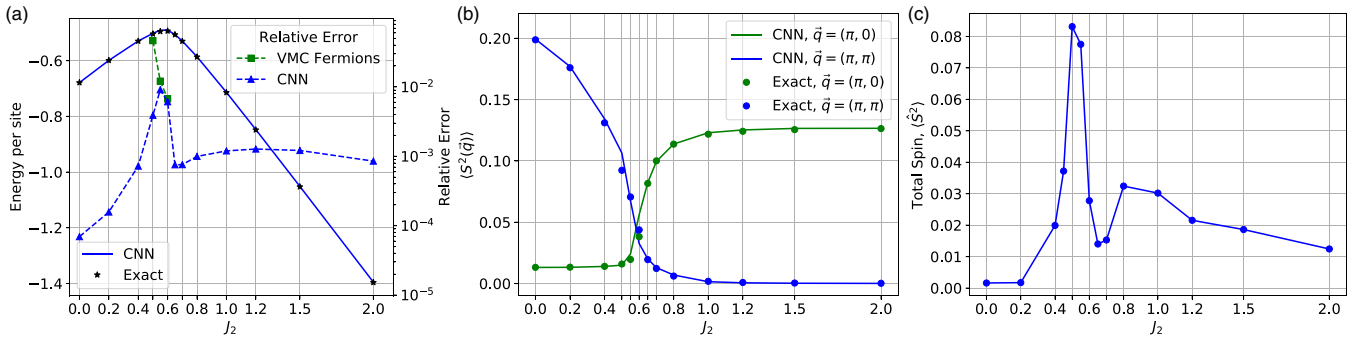
FIG. 2. Simulation results on the 6×6 square lattice with periodic boundary conditions. (a) Energy comparison with ED results from Ref. [47]. The CNN energies are indicated by the blue line while the black stars gives the exact values. The relative errors are plotted with respect to the right axis. The blue dashed line shows the relative error for our CNN ansätz and the green dashed line corresponds the the Gutzwiller-projected mean-field fermionic variational ansätz from Ref. [40]. (b) Spin-spin structure factor as defined in Eq. (14). (c) The total spin $\langle \hat{\mathbf{S}}^2 \rangle$, an extensive quantity. In the exact case, this value should be zero since the ground state is in the singlet sector.

energies competitive with, and in most of the parameter space better than, other state-of-the-art methods. In the Appendix, we provide a table containing all the variational energies obtained. A further possible origin for the nonoptimal performance at $J_2/J_1 = 0.55$ can also be traced in symmetry violations in our CNN ansätz. Most notably, whereas the fermionic ansätz as well as the DMRG calculations preserve SU(2) symmetry explicitly, the CNN does not. This can be seen in Fig. 3(c), where the total spin of our variational form is substantially different from zero, albeit still being less than 2 (the next lowest eigenvalue of the total spin operator $\hat{\mathbf{S}}^2$) and hence ensuring a decent overlap with the singlet sector. Despite the mixing with other spin sectors, the features of the Neel and striped magnetic orders can still be seen in the spin-spin correlations in Fig. 3(b).

Additionally, we show some results using a RBM with approximately the same number of parameters as the CNN. We can see that the CNN indeed has better performance, especially in the frustrated region, further corroborating the point that deep networks can capture entanglement more efficiently than shallower ones [50].

Before concluding the section, we would like to mention also a previous work Ref. [24], which used a slightly different

CNN as a variational ansätz to study the same problem. On the 10×10 cluster with $J_2 = 0.5$, a variational energy per site of -0.4736 was obtained as compared to our result -0.4952. This differs significantly from other state-of-the-art methods as can be seen from the scale on the energy axis of Fig. 3(a). There are some qualitative differences between their ansätz and ours: (1) Reference [24] used real parameters while ours are complex-valued. (2) The nonlinearity in Ref. [24] is introduced via max-pooling operations while we use nonlinear activation functions. In addition, we symmetrize our ansätz to have $\mathcal{C}_4$ rotation symmetry and also provided an initial prior sign structure. (3) The optimization technique used in Ref. [24] is different from the SR method we have employed.

VI. CONCLUSION

The results obtained in this paper provide tangible evidence that NQSs are a competitive variational ansätz to study challenging open problems such as frustrated magnets. While we have provided here direct numerical evidence for their suitability to study the stability of the spin-liquid phase in the J_1-J_2 model, some open aspects have not been yet addressed in this paper, and will be the focus of future research.
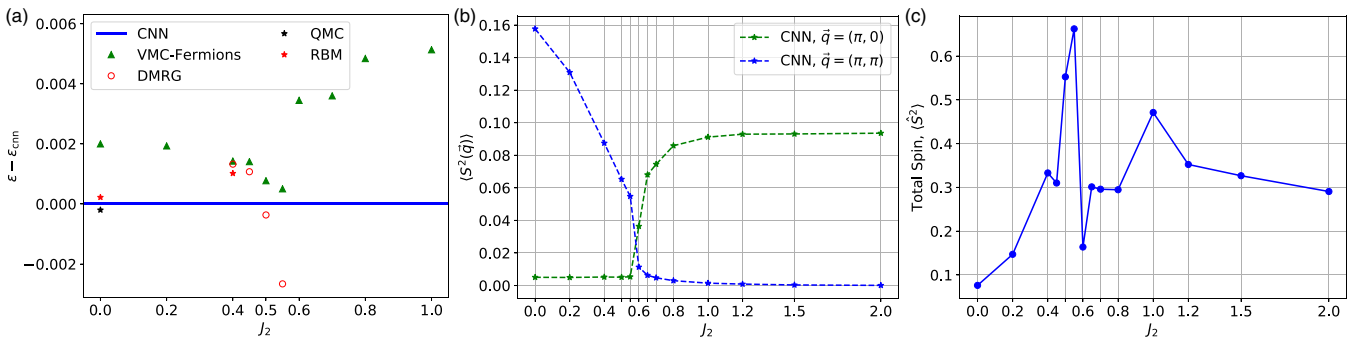

FIG. 3. Simulation results on the 10×10 square lattice with periodic boundary conditions. (a) Energy comparison with other state-of-the-art methods. We plot the energy difference per site with respect to the energy obtained from our CNN variational ansätz. The red open circles show the values for the DMRG results of Ref. [41] and the green triangles correspond to the Gutzwiller-projected mean-field fermionic variational ansätz. The black star shows the exact result from Green's function quantum Monte Carlo for the sign-problem free point ($J_2 = 0$). The red stars show the results of a restricted Boltzmann machine with 3600 hidden units (3636 parameters). (b) Spin-spin structure factor as defined in Eq. 14. (c) The total spin $\langle \hat{\mathbf{S}}^2 \rangle$. In the exact case, this value should be zero since the ground state is in the singlet sector.

TABLE I. Comparison between Gutzwiller-projected mean-field fermionic wave functions VMC [40] (on the 10×10 case the energies were provided by F. Ferrari and F. Becca), DMRG with 8192 SU(2) states, or equivalently 32 000 U(1) states [41] and the CNN used in this paper. The exact energies on the 6×6 case were take from Ref. [47].

6×6	$J_2 = 0.0$	$J_2 = 0.2$	$J_2 = 0.4$	$J_2 = 0.45$	$J_2 = 0.5$	$J_2 = 0.55$	$J_2 = 0.6$	$J_2 = 0.8$	$J_2 = 1.0$
Exact	−0.678 872	−0.599 046	−0.529 745	–	−0.503 810	−0.495 178	−0.493 239	−0.586 487	−0.714 360
VMC	–	–	−0.527 15(1)	−0.513 64(1)	−0.501 17(1)	−0.489 92(1)	–	–	–
DMRG	–	–	−0.529 744	−0.515 655	−0.503 805	−0.495 167	–	–	–
CNN	−0.678 82(1)	−0.598 95(1)	−0.529 36(1)	−0.514 52(1)	−0.501 85(1)	−0.490 67(2)	−0.490 23(1)	−0.585 90(1)	−0.713 51(1)

10×10	$J_2 = 0.0$	$J_2 = 0.2$	$J_2 = 0.4$	$J_2 = 0.45$	$J_2 = 0.5$	$J_2 = 0.55$	$J_2 = 0.6$	$J_2 = 0.8$	$J_2 = 1.0$
VMC	−0.669 35(1)	−0.590 82(1)	−0.522 29(1)	−0.507 64(1)	−0.494 39(1)	−0.482 27(1)	−0.472 59(1)	−0.568 99(1)	−0.691 23(1)
DMRG	–	–	−0.522 391	−0.507 976	−0.495 530	−0.485 434	–	–	–
CNN	−0.671 35(1)	−0.592 75(1)	−0.523 71(1)	−0.509 05(1)	−0.495 16(1)	−0.482 77(1)	−0.476 04(1)	−0.573 83(1)	−0.696 36(1)

First, since the system is gapless, finite size effects are large such that accurate extrapolations to the thermodynamic limit are necessary. A more computationally demanding simulation campaign would be required to provide a firm finite-size extrapolation of the magnetic correlations presented here.

Second, the networks we have used here are comparably much smaller, in terms of depth and number of trainable parameters, than state-of-the-art models used in modern deep-learning applications. Our networks contain at least three orders of magnitude less parameters than what is found in typical deep CNNs routinely used for image recognition, and the margin for future improvements therefore seems quite substantial. To bridge this gap, more expressive models could be adopted, for example, along the lines of the recently introduced auto-regressive models [56] for quantum states. Finally, it is rather clear from our simulations that the performance of our ansätz is also linked to the sign structure of the many-body state. This implies that "learning" the correct sign structure within the parametrization we adopted presents a challenge for the SR optimization we use. In this context, it would be interesting to see how modern machine-learning techniques such as reinforcement learning can help tackle this problem.

ACKNOWLEDGMENTS

APPENDIX A: VARIATIONAL ENERGIES

In the table below, we show the exact values (including error bars) for the variational energies obtained in this paper.

APPENDIX B: OPTIMIZATION DETAILS

The optimization of a variational ansätz $\Psi(\{\alpha_k\}) \in \mathbb{C}^{2^n}$ depending on parameters $\{\alpha_k\}$ is achieved using the SR scheme, which involves solving the linear equation

$$\sum_{k'} [\langle \mathcal{O}_k^\dagger \mathcal{O}_{k'} \rangle - \langle \mathcal{O}_k^\dagger \rangle \langle \mathcal{O}_{k'} \rangle + \lambda \delta_{kk'}] \delta \alpha_{k'}$$
$$= -\epsilon [\langle \mathcal{O}_k^\dagger \hat{H} \rangle - \langle \mathcal{O}_k^\dagger \rangle \langle \hat{H} \rangle], \tag{B1}$$

where $\mathcal{O}_k = \frac{\partial}{\partial \alpha_k} \log [\Psi(\{\alpha_k^0\})]$ are the logarithmic derivatives, ϵ is the the step size, and λ is the regularization parameter. For the simulations done in this paper, we take $\epsilon = 0.01$ and $\lambda = 0.01$. The expectation values $\langle \cdots \rangle$ are estimated with Markov chain Monte Carlo sampling (using local spin exchange moves to conserve the total magnetization) with a sample size of 5000.

The CNN architecture used in this work is shown in Fig. 1. There are six convolutional layers which are initialized from 0 mean random normal distributions with standard deviations 0.1, 0.03, 0.04, 0.05, 0.06, 0.07, respectively, starting from the first layer.

[1] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, Nature **550**, 354 (2017).

[2] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).

[3] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Nat. Phys. **13**, 435 (2017).

[4] F. Schindler, N. Regnault, and T. Neupert, Phys. Rev. B **95**, 245134 (2017).

[5] J. Venderley, V. Khemani, and E.-A. Kim, Phys. Rev. Lett. **120**, 257204 (2018).

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

[7] A. Rocchetto, E. Grant, S. Strelchuk, G. Carleo, and S. Severini, npj Quantum Inform. **4**, 28 (2018).

[8] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Nat. Machine Intell. **1**, 155 (2019).

[9] B. Jónsson, B. Bauer, and G. Carleo, arXiv:1808.05232.

[10] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. S. Davis, E. Khatami, and E.-A. Kim, Nature **570**, 484 (2019).

[11] B. S. Rem, N. Käming, M. Tarnowski, L. Asteria, N. Fläschner, C. Becker, K. Sengstock, and C. Weitenberg, Nat. Phys. **15**, 917 (2019).

[12] A. Bohrdt, C. S. Chiu, G. Ji, M. Xu, D. Greif, M. Greiner, E. Demler, F. Grusdt, and M. Knap, Nature Phys. **15**, 921 (2019).

[13] R. A. Vargas-Hernández, J. Sous, M. Berciu, and R. V. Krems, Phys. Rev. Lett. **121**, 255702 (2018).

[14] G. Carleo and M. Troyer, Science **355**, 602 (2017).

[15] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[16] D.-L. Deng, X. Li, and S. Das Sarma, Phys. Rev. X **7**, 021021 (2017).

[17] X. Gao and L.-M. Duan, Nat. Commun. **8**, 662 (2017).

[18] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Phys. Rev. B **96**, 205152 (2017).

[19] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, Phys. Rev. B **97**, 085104 (2018).

[20] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Phys. Rev. X **8**, 011006 (2018).

[21] H. Saito, J. Phys. Soc. Jpn. **86**, 093001 (2017).

[22] H. Saito and M. Kato, J. Phys. Soc. Jpn. **87**, 014001 (2018).

[23] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Phys. Rev. Lett. **121**, 167204 (2018).

[24] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, Phys. Rev. B **98**, 104426 (2018).

[25] G. Carleo, Y. Nomura, and M. Imada, Nat. Commun. **9**, 5322 (2018).

[26] L. Pastori, R. Kaubruegger, and J. C. Budich, Phys. Rev. B **99**, 165123 (2019).

[27] D. Luo and B. K. Clark, Phys. Rev. Lett. **122**, 226401 (2019).

[28] D. Kochkov and B. K. Clark, arXiv:1811.12423.

[29] M. E. Zhitomirsky and K. Ueda, Phys. Rev. B **54**, 9007 (1996).

[30] L. Capriotti and S. Sorella, Phys. Rev. Lett. **84**, 3173 (2000).

[31] M. Mambrini, A. Läuchli, D. Poilblanc, and F. Mila, Phys. Rev. B **74**, 144422 (2006).

[32] J.-F. Yu and Y.-J. Kao, Phys. Rev. B **85**, 094407 (2012).

[33] S. Sachdev and R. N. Bhatt, Phys. Rev. B **41**, 9323 (1990).

[34] A. V. Chubukov and T. Jolicoeur, Phys. Rev. B **44**, 12050 (1991).

[35] R. R. P. Singh, Z. Weihong, C. J. Hamer, and J. Oitmaa, Phys. Rev. B **60**, 7278 (1999).

[36] L. Capriotti, F. Becca, A. Parola, and S. Sorella, Phys. Rev. Lett. **87**, 097201 (2001).

[37] G.-M. Zhang, H. Hu, and L. Yu, Phys. Rev. Lett. **91**, 067201 (2003).

[38] H.-C. Jiang, H. Yao, and L. Balents, Phys. Rev. B **86**, 024424 (2012).

[39] L. Wang, D. Poilblanc, Z.-C. Gu, X.-G. Wen, and F. Verstraete, Phys. Rev. Lett. **111**, 037202 (2013).

[40] W.-J. Hu, F. Becca, A. Parola, and S. Sorella, Phys. Rev. B **88**, 060402(R) (2013).

[41] S.-S. Gong, W. Zhu, D. N. Sheng, O. I. Motrunich, and M. P. A. Fisher, Phys. Rev. Lett. **113**, 027201 (2014).

[42] J. Richter, R. Zinke, and D. J. J. Farnell, Eur. Phys. J. B **88**, 2 (2015).

[43] S. Morita, R. Kaneko, and M. Imada, J. Phys. Soc. Jpn. **84**, 024720 (2015).

[44] D. Poilblanc and M. Mambrini, Phys. Rev. B **96**, 014414 (2017).

[45] R. Haghshenas and D. N. Sheng, Phys. Rev. B **97**, 174408 (2018).

[46] W.-Y. Liu, S. Dong, C. Wang, Y. Han, H. An, G.-C. Guo, and L. He, Phys. Rev. B **98**, 241109(R) (2018).

[47] H. Schulz, T. Ziman, and D. Poilblanc, J. Phys. I (France) **6**, 675 (1996).

[48] A. W. Sandvik, Phys. Rev. B **56**, 11678 (1997).

[49] M. Calandra Buonaura and S. Sorella, Phys. Rev. B **57**, 11446 (1998).

[50] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Phys. Rev. Lett. **122**, 065301 (2019).

[51] N. Guberman, arXiv:1602.09046.

[52] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, in *6th International Conference on Learning Representations (ICLR 2018),Vancouver, BC, Canada* (OpenReview.net, 2018).

[53] W. Marshall, Proc. R. Soc. London A **232**, 48 (1955).

[54] S. Sorella, M. Casula, and D. Rocca, J. Chem. Phys. **127**, 014105 (2007).

[55] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[56] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, arXiv:1902.04057.

[57] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, and A. Wietek, SoftwareX **10**, 100311 (2019).