

This is a copy of the published version, or version of record, available on the publisher's website. This version does not track changes, errata, or withdrawals on the publisher's site.

High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks

Christopher Roth, Attila Szabó, and Allan H. MacDonald

Published version information

Citation: C Roth, A Szabó and AH MacDonald. High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks. Phys Rev B 108, no. 5 (2023): 054410

DOI: [10.1103/PhysRevB.108.054410](https://doi.org/10.1103/PhysRevB.108.054410)

This version is made available in accordance with publisher policies. Please cite only the published version using the reference above. This is the citation assigned by the publisher at the time of issuing the APV. Please check the publisher's website for any updates.

This item was retrieved from **ePubs**, the Open Access archive of the Science and Technology Facilities Council, UK. Please contact epublications@stfc.ac.uk or go to <http://epubs.stfc.ac.uk/> for further information and policies.

High-accuracy variational Monte Carlo for frustrated magnets with deep neural networksChristopher Roth,¹ Attila Szabó^{2,3} , and Allan H. MacDonald¹¹*Physics Department, University of Texas at Austin, Austin 78712-1710, Texas*²*Rudolf Peierls Centre for Theoretical Physics, University of Oxford, Oxford OX1 3PU, United Kingdom*³*ISIS Facility, Rutherford Appleton Laboratory, Harwell Campus, Didcot OX11 0QX, United Kingdom*

(Received 24 May 2023; accepted 18 July 2023; published 8 August 2023)

We show that neural quantum states based on very deep (4–16-layered) neural networks can outperform state-of-the-art variational approaches on highly frustrated quantum magnets, including quantum-spin-liquid candidates. We focus on group convolutional neural networks that allow us to efficiently impose space-group symmetries on our ansätze. We achieve state-of-the-art ground-state energies for the $J_1 - J_2$ Heisenberg models on the square and triangular lattices, in both ordered and spin-liquid phases, and discuss ways to access low-lying excited states in nontrivial symmetry sectors. We also compute spin and dimer correlation functions for the quantum paramagnetic phase on the triangular lattice, which do not indicate either conventional or valence-bond ordering.

DOI: [10.1103/PhysRevB.108.054410](https://doi.org/10.1103/PhysRevB.108.054410)**I. INTRODUCTION**

Frustrated magnetism has long been an exciting playground for discovering new physics. Perhaps the most interesting outcome of this activity has been the concept of a *quantum spin liquid*, a phase of matter without any long-range order even in the ground state [1,2], characterized by long-range entanglement, fractionalized excitations, and topological order. Establishing spin-liquid physics in specific microscopic models, however, remains a great theoretical challenge, apart from a few exactly solvable systems and perturbative constructions [3–5]: While spin-liquid phases have been proposed in Heisenberg and Hubbard models on a number of lattices [6–20], their presence and properties remain a matter of active debate. To a large extent, this is due to the difficulty of reliable numerical simulations of frustrated magnets: unbiased quantum Monte Carlo is plagued by the sign problem [21], while the most successful tensor-network approach, density-matrix renormalization group (DMRG) is limited by entanglement constraints of matrix-product states [22] to narrow cylinders in dimensions greater than one [23], where competing spin-liquid and ordered ground states may not be clearly resolved.

An interesting approach to overcome this difficulty is employing neural networks to parametrize variational wave functions [24]. Neural networks are universal function approximators [25], thus they can represent highly entangled quantum states [26], which makes them a promising tool to study quantum spin liquids. Furthermore, variational computations using such neural quantum states (NQS) benefit from the greater computational power of graphics processing units (GPUs) and powerful machine-learning optimization libraries [27]. NQS ansätze based on a variety of neural-network architectures have been proposed [28–33], however, they generally fall short of the reliability and accuracy required for state-of-the-art results on frustrated problems. A remarkable exception is the ansatz developed in Refs. [7,34], which,

however, uses neural networks merely to improve the accuracy of Gutzwiller-projected fermionic wave functions, an extremely successful ansatz in its own right.

By contrast, we demonstrate here that a “pure” NQS ansatz using very deep networks can achieve state-of-the-art variational energies. In particular, we use group-convolutional neural networks (GCNNs) [33,35], which allow us to impose the full space-group symmetry of the problem on the wave functions. We find two key design principles of a successful architecture: First, the network should not separate the amplitudes and phases of the network, as learning the latter in a frustrated system is beyond the capacity of even deep neural networks [36,37]. Second, imposing locality by using short-ranged convolutional filters in the GCNNs both makes using deeper networks computationally feasible and simplifies the learning landscape by structuring the representation of long-range correlation in the networks; the latter is reflected in faster convergence compared to full-width convolutional kernels. Our ansätze either match or surpass the best variational energies in the literature [7,10] in the quantum paramagnetic regimes of the square- and triangular-lattice $J_1 - J_2$ Heisenberg model on clusters as large as 16×16 . Notably, we are able to achieve this accuracy at a more modest computational cost than these works; compared to the immense resources used by some NQS approaches in the literature [38], this highlights the importance of carefully designed ansätze. We also present numerical experiments on finding low-lying excited states using GCNNs.

II. GROUP-CONVOLUTIONAL NEURAL NETWORKS

Lattice Hamiltonians are invariant under a large group of spatial symmetries, governed by the geometry of the lattice and anisotropies of the interactions: Wigner’s theorem [39] ensures that all eigenstates of such a Hamiltonian transform under an irreducible representation (irrep) of the same symmetry group. Imposing space-group symmetries explicitly in

a variational ansatz reduces the Hilbert space available for the variational algorithm, making it more reliable and efficient [40]. Furthermore, symmetry-broken phases are identified by symmetry quantum numbers of the lowest-lying excited states, known as the tower of states [41,42], thus access to the lowest-lying states in a number of symmetry sectors allow identifying distinct phases and transition points with a high accuracy [7].

Space-group symmetries can be imposed on variational wave functions by explicit projection. Consider an ansatz $\psi_0(\vec{\sigma}; \vec{\theta})$ with parameters $\vec{\theta}$ that maps a computational basis state $\vec{\sigma}$ onto its amplitude in the variational wave function. Then, given a group G of symmetries \hat{g} that maps each basis state $\vec{\sigma}$ onto another, $\hat{g}\vec{\sigma}$, we can construct an ansatz transforming under a given irrep of G using the projection formula [43]

$$|\psi\rangle = \frac{d_\chi}{|G|} \sum_{g \in G} \chi_g^* \hat{g} |\psi_0\rangle, \quad (1a)$$

$$\psi(\vec{\sigma}) = \frac{d_\chi}{|G|} \sum_{g \in G} \chi_g^* \psi_0(\hat{g}^{-1}\vec{\sigma}), \quad (1b)$$

where χ_g and d_χ are the characters and dimension of the irrep and $|G|$ is the number of symmetry elements.

Projection approaches based on Eq. (1) have been used with a variety of ansätze [7,28,44]. Usually, however, one has to evaluate ψ_0 for all symmetry-related basis states $\hat{g}\vec{\sigma}$: this can be prohibitively expensive, requiring setting all or most $\psi_0(\hat{g}\vec{\sigma})$ equal by construction, which in turn restricts which irreps of the symmetry group can be probed with the ansatz. Instead, one would prefer to generate all $\psi_g(\sigma) \equiv \psi_0(\hat{g}^{-1}\vec{\sigma})$ in a single evaluation of the ansatz. Such an ansatz would be *equivariant* under the symmetry group G in the sense that acting on its input by a symmetry element would cause the output to be acted on by the same symmetry in a nontrivial way (namely, by the left canonical action):

$$\psi_g(\hat{h}^{-1}\vec{\sigma}) = \psi_0(\hat{g}^{-1}\hat{h}^{-1}\vec{\sigma}) = \psi_{hg}(\vec{\sigma}). \quad (2)$$

Note that $(\hat{h}^{-1}\vec{\sigma})_{\vec{r}} = \sigma_{\hat{h}\vec{r}}$, so such an equivariant ansatz maps relabelling lattice sites by \hat{h} to left-multiplication of symmetry elements by h .

Convolutional neural networks (CNNs) are a famous example of such an equivariant function. Let us consider a (hypercubic) lattice of size $L_1 \times L_2 \times \dots \times L_d$ in periodic boundary conditions and its translation group, $\mathbb{Z}_{L_1} \times \dots \times \mathbb{Z}_{L_d}$, which has the same structure as the lattice itself. Now, the convolutional mapping

$$\psi(\vec{r}) = \sum_{\vec{r}'} K(\vec{r} - \vec{r}') \sigma(\vec{r}'), \quad (3)$$

with an arbitrary kernel K , is manifestly equivariant:

$$\begin{aligned} \sum_{\vec{r}'} K(\vec{r} - \vec{r}') \sigma(\vec{r}' + \vec{h}) &= \sum_{\vec{r}'} K(\vec{r} + \vec{h} - \vec{r}') \sigma(\vec{r}') \\ &= \psi(\vec{r} + \vec{h}), \end{aligned}$$

where vector addition winds around the periodic boundary conditions. This is equivalent to Eq. (2), as a translation \hat{h} acts by adding \vec{h} to both input and output coordinates. One can

use similar arguments to show that several iterations of the convolution (3), arbitrary “on-site” functions $y(\vec{r}) = f[x(\vec{r})]$, and thus arbitrarily deep neural networks built from alternating these building blocks, are all equivariant. The projection formula (1) also has a natural interpretation. The irreps of the translation group are all one-dimensional and labeled by the phases $e^{i\phi_1}, \dots, e^{i\phi_d}$ acquired by the wave function upon translation along each lattice vector: Eq. (1) thus becomes

$$\psi(\vec{\sigma}) = \frac{1}{L_1 \dots L_d} \sum_{\vec{r}} e^{-i\vec{\phi} \cdot \vec{r}} \psi_{\vec{r}}(\vec{\sigma}),$$

a Fourier transform that extracts a crystal momentum eigenstate.

GCNNs [33,35] generalize this idea to arbitrary (non-Abelian) groups. Since the group is no longer isomorphic to the input lattice, we will require two types of linear layer: first, an embedding layer

$$y_g = \sum_{\vec{r}} K(\hat{g}^{-1}\vec{r}) \sigma(\vec{r}) \quad (4a)$$

generates equivariant feature maps, indexed with group elements, from the input; this is followed by any number of group convolutional [45] layers

$$z_g = (W \circ y)_h \equiv \sum_{h \in G} W(h^{-1}g) y_h \quad (4b)$$

and “on-site” nonlinearities. One can show [27] that such a network satisfies Eq. (2).

A naïve alternative to GCNNs for space groups would be applying all point-group operations (rotations, reflections, etc.) on $\vec{\sigma}$ and feeding each of these configurations into a conventional CNN [28]. This too satisfies Eq. (2) and can be projected onto any irrep using Eq. (1): In fact, it is equivalent to a GCNN with the same number of features in each layer, in which every kernel W is restricted to pure translations rather than the whole space group. Thus, GCNNs of the same size can have more variational parameters, and in every layer, each output is determined by more inputs from the previous layers: both of these features allow the network to be more expressive at the same memory footprint. Indeed, GCNNs have outperformed symmetrized CNNs in both image-recognition [35] and variational Monte Carlo [33] tasks.

A. Details of the GCNN ansatz

In this paper, we will consider spins on a lattice and represent their wave functions in the σ^z basis. The largest symmetry group for which Eq. (1b) is applicable in this basis is the direct product of the space group of the lattice and the \mathbb{Z}_2 spin-parity group generated by $P = \prod_i \sigma_i^x$. [46]. Irreps of this group are specified by the eigenvalue of $P = \pm 1$ and an irrep of the space group. The latter are generally specified by a set of crystal momenta related by point-group operations (known as a *star*) and an irrep of the subgroup of the point group that leaves these momenta invariant (known as the *little group*) [27,47,48]. We shall denote space-group irreps using a representative wave vector of the star and the Mulliken symbol [49] of the little-group irrep: for example, the trivial irrep of the square-lattice space group is written as $\Gamma.A_1$ (cf. Fig. 1).

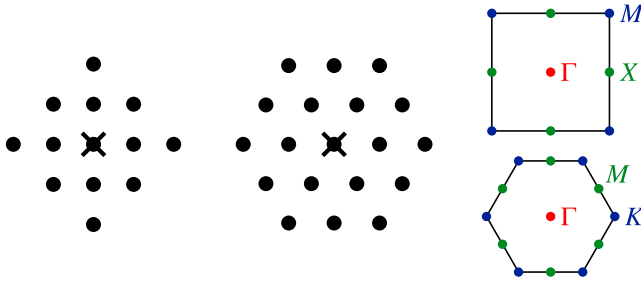


FIG. 1. Arrangement of nonzero kernel entries for the square (left) and triangular (center) lattices. The cross indicates the origin of the lattice and/or pure point-group symmetries. Right: illustration of high-symmetry points in the Brillouin zones of the square (top) and triangular (bottom) lattices.

Parity eigenvalues will be included through a \pm index, e.g., $\Gamma.A_1^+$.

All GCNN ansätze used in this work have a fixed number of feature maps in every layer, connected by real-valued kernels K, W . In most cases, we restrict the embedding kernels K to the local clusters shown in Fig. 1. We also impose locality on the convolutional kernels W : these are only nonzero for $g = tp$, where p is a point-group symmetry that leaves the origin in place, and t is a translation by a lattice vector indicated in Fig. 1.

All but the last layer is followed by the SELU activation function [50], which allows us to train very deep (up to 16 layers in this work) networks without severe instabilities. In the output layer, we combine pairs of feature maps into complex-valued features, exponentiate them, and project the result on the desired irrep:

$$\tilde{h}_{n,g} = h_{n,g} + ih_{n+F/2,g}, \quad (5a)$$

$$\psi(\vec{\sigma}) = \sum_{g \in G} \chi_g^* \sum_{n=1}^{F/2} \exp(\tilde{h}_{n,g}), \quad (5b)$$

where F is the number of (real-valued) feature maps. Including exponentiation in Eq. (5) is important to represent the wide dynamical range of wave-function amplitudes: for example, the amplitude of highly ferromagnetically correlated basis states is exponentially suppressed in the ground state of antiferromagnets [51], which neural networks do not represent efficiently without any structural bias [52]. Complex exponentiation, however, separates amplitudes and phases, which obstructs the learning of the accurate sign structure, a formidably complex object for a highly frustrated magnet [29,36,53]: constructing ψ as a sum of many terms alleviates this problem, as cancellation between the terms allows learning destructive interferences more easily [54].

B. GCNNs with residual layers

While the GCNN wave function described above performs quite well on the square lattice, we found that it struggles to converge to state-of-the-art energies on the more challenging triangular-lattice models. In principle, this could be remedied by making the network deeper; however, training these very deep networks becomes unreliable and unstable. In

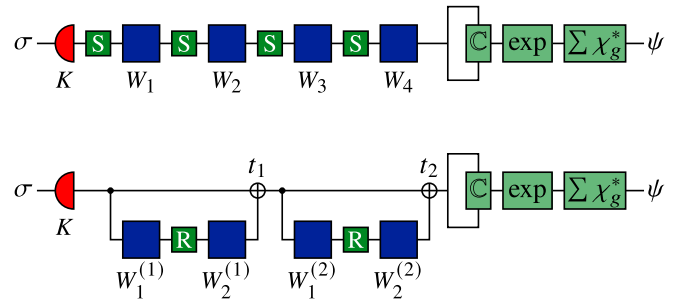


FIG. 2. Structure of the plain (top) and residual (bottom) GCNNs used in this work. Red and blue boxes stand for the embedding layer (4a) and the group convolutions (4b), respectively. Dark green boxes indicate SELU (S) or RELU (R) activation functions. Light green boxes represent the output layer (5).

machine-learning applications, this problem is remedied using residual layers [55]: Instead of the full mapping from one of hidden representation to the next, these learn the difference between the two, which effectively controls the importance of each layer and gives a controlled starting point for training (namely, where all these differences are zero). In practice, each layer learns to add a small but nonzero contribution, which helps the network avoid numerical instabilities without sacrificing the expressive power of deep neural networks. In particular, since the Jacobian of each layer, $\partial \mathbf{h}^{(i+1)} / \partial \mathbf{h}^{(i)}$, remains close to the identity matrix, residual networks do not suffer from vanishing or exploding gradients, a common and severe problem for training deep neural networks.

For the larger triangular-lattice simulations, therefore, we used residual GCNNs (ResGCNNs) of the structure shown in Fig. 2. Each hidden layer is of the functional form

$$\mathbf{h}^{(i+1)} = \mathbf{h}^{(i)} + t_i W_2^{(i)} \circ \text{relu} [W_1^{(i)} \circ \mathbf{h}^{(i)}], \quad (6)$$

that is, it applies a two-layer GCNN (with a RELU activation function) on the input and adds the result to the values in the previous hidden layer. The overall weight t_i is a learned parameter that gives a simple way for the training to control the relative importance of layers. We initialize $t = 0$ for all layers, which corresponds to the identity operation: Over the course of training, t_i move away from zero gradually, making the learning trajectory more controlled. The embedding layer Eq. (4a) and the construction Eq. (5) of ψ from the last hidden layer remains unchanged.

C. Training protocol

We optimize our variational wave function using the stochastic reconfiguration (SR) method [56], in which the parameter updates $\delta\theta$ are found by solving the equation

$$S \delta\vec{\theta} = -\eta \frac{\partial E}{\partial \vec{\theta}}, \quad (7)$$

where E is the variational energy $\langle \psi | H | \psi \rangle / \langle \psi | \psi \rangle$, and S is the quantum geometric tensor [57]. Equation (7) is ill-conditioned, requiring regularization of the S matrix [58]. We found that the standard approach of adding a constant to diagonal entries leads to poor convergence, while the scale-invariant regularizer of Ref. [58] is itself numerically unstable. Adding both types of shift, on the

other hand, stabilizes calculations and allows reliable convergence. We thus use $S_{ii} \mapsto S_{ii} + \varepsilon_1 S_{ii} + \varepsilon_2$ and set $\eta = 0.01$, $\varepsilon_1 = 10^{-2}$, $\varepsilon_2 = 10^{-3}$ in all simulations. We have not fine-tuned these hyperparameters and expect a wide range of η and $\varepsilon_2 \lesssim \varepsilon_1$ to yield similar results.

The above scheme works reliably relatively near the ground-state energy but it is often unstable at the start of training. We found that this is greatly improved by biasing the training towards states with similar amplitudes for all basis states in this stage. A physically motivated approach is minimizing a “free energy” $F = E - TS$, where the “entropy” is defined in the computational basis as [59,60]

$$\mathcal{S} = - \sum_{\vec{\sigma}} \frac{|\psi(\sigma)|^2}{\sum_{\vec{\tau}} |\psi(\tau)|^2} \log \frac{|\psi(\sigma)|^2}{\sum_{\vec{\tau}} |\psi(\tau)|^2}; \quad (8)$$

the denominators are necessary as our wave functions are generally not normalized. Even though we cannot calculate \mathcal{S} directly because of this, its gradients can be estimated via Monte Carlo sampling, see Appendix A. The effective temperature T is gradually lowered to zero as the training proceeds: we used $T_n = \exp(-n/50)/2$ in the n th step.

For the square lattice $J_1 - J_2$ model, the Marshall sign rule associated with the Néel order at $J_2 = 0$ [61] is a good first approximation of the true sign structure even in the quantum paramagnetic phases (cf. Appendix B), and applying it explicitly to NQS ansätze was found to improve their training stability [28]. However, doing so biases the training towards states that follow the sign rule exactly [29], which, while very low in energy, cannot be the true ground state. Therefore, we allowed the network to learn the sign structure without any explicit sign rule bias.

To find low-energy variational states in nontrivial symmetry sectors, we used networks of the same architecture as in the ground-state simulations and initialised them with the converged parameters of the latter. This amounts to using the ground state to generate an initial guess for the excited state by changing the irrep characters χ in Eq. (1). We found that this transfer learning process starts at variational energies close to that of the ground state and converge quickly (100–300 SR steps with 4096 Monte Carlo samples) to a stable variational energy; we used 1000 steps to allow for full annealing of the energy measurement. By contrast, networks trained from scratch either became unstable or levelled off at extremely high energies.

All simulations were carried out on a single NVIDIA A100 GPU. On both the square and triangular lattices, we found that the training does not get stuck at local minima, but reliably converges to the essentially the same variational energy for every random initialization. We verified this by running simulations on smaller systems three times, which yielded essentially identical variational energies; therefore, we ran the more time-consuming simulations only once.

III. SQUARE LATTICE $J_1 - J_2$ MODEL

We first apply our approach to the square-lattice Heisenberg model with first- and second-neighbor couplings,

$$H = J_1 \sum_{\langle ij \rangle} \vec{s}_i \cdot \vec{s}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{s}_i \cdot \vec{s}_j; \quad (9)$$

TABLE I. Optimized ground-state energies (in units of J_1/spin) and estimates of the total spin ($\langle S^2 \rangle$) for the $J_1 - J_2$ square-lattice Heisenberg model, compared to the best known variational energies for these systems [7].

J_2/J_1	System size	Our work	RBM+PP [7]	$\langle S^2 \rangle$
0.5	10×10	-0.497437(7)	-0.497629(1)	0.038(4)
	12×12	-0.496769(9)	-0.496791(4)	0.098(8)
	16×16	-0.496509(6)	-0.496213(3)	0.156(10)
0.55	10×10	-0.486772(11)	–	0.077(7)
	12×12	-0.486068(10)	-0.485735(7)	0.167(10)
	16×16	-0.485583(8)	-0.485208(4)	0.307(13)

we will set $J_1 = 1$. The model orders both at small and large values of J_2 , showing a Néel and a stripy pattern, respectively. Near the classically maximally frustrated point, $J_2/J_1 = 0.5$, both of these orders disappear: the current consensus is that this region is split between a \mathbb{Z}_2 Dirac spin liquid and a valence-bond solid (VBS) [7,8,62]. We benchmark our method at $J_2/J_1 = 0.5$ and 0.55, thought to lie inside the spin-liquid and VBS phases, respectively.

A. Ground-state energies

We estimated the ground state energy for the square-lattice $J_1 - J_2$ model for linear cluster sizes $L = 10, 12, 16$. We used GCNNs with L layers, each consisting of six group-indexed feature maps. For most simulations, we performed 1000 SR steps using 1024 Monte Carlo samples, followed by 500 steps with 4096 samples. The first stage allows approaching the ground state at relatively low computational cost, while training with more samples helps the network learn representations of the wave function that generalize better to portions of the Hilbert space that are not sampled [36].

Our best variational energies are summarized in Table I. In all experiments, the converged variational energies come very close to the best ones currently available in the literature [7], generated by VMC on an ansatz combining many-variable Gutzwiller-projected spinon-mean-field wave functions [44] and restricted Boltzmann machines (RBMs). We achieve the most marked improvements in energy (comparable to the Hamiltonian gap) for the largest, 16×16 lattices: While for small systems, a (small) linear combination of projected Slater determinants is an excellent approximation of spin-liquid and VBS ground states, this approximation deteriorates for larger lattices, beyond the expressivity of the RBM correction factors employed by Ref. [7]. By contrast, the deep networks used by us are not subject to such *a priori* limitations. Furthermore, the computational cost of our simulations are significantly smaller: for the 16×16 lattice in particular, we require approximately 300 GPU hours, compared to the approximately 60 000 CPU hours of Ref. [7].

To further verify the quality of our converged wave functions, we computed their total spin ($\langle S^2 \rangle$), expected to be exactly zero for an antiferromagnetic ground state. While we cannot project the wave function on the $S = 0$ sector explicitly, we find $\langle S^2 \rangle < 0.3$ even for the 16×16 lattice, a substantial improvement over $\langle S^2 \rangle \approx 0.6$ for $L = 10$ reported earlier for

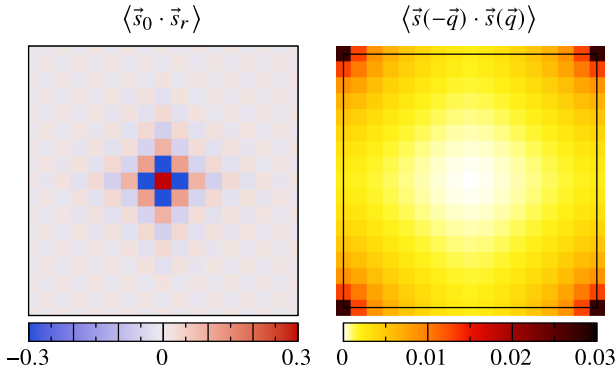


FIG. 3. Spin correlation function for the square lattice $J_1 - J_2$ model on a 16×16 lattice at $J_2 = 0.5$ in real space (left) and in the Brillouin zone (right). The true value of $\langle \vec{s}_0 \cdot \vec{s}_0 \rangle = 3/4$ is clipped for visibility.

various shallower NQS architectures [28,29]. The measured value of $\langle S^2 \rangle$ is consistently higher for larger system sizes as well as for $J_2 = 0.55$, reflecting the increasing difficulty of obtaining accurate wave functions. The spin structure factors $\langle \vec{s}(-\vec{q}) \cdot \vec{s}(\vec{q}) \rangle$ and $\langle s^z(-\vec{q})s^z(\vec{q}) \rangle$ showed similarly good agreement for all wave vectors \vec{q} , showing that our wave functions do not strongly break $SU(2)$ spin rotation symmetry, a common issue for earlier ansätze [63]. Furthermore, we demonstrate in Appendix B that our method overcomes the “sign problem” of neural quantum states [29,36], which has hampered the performance of such ansätze as recurrent neural networks for frustrated magnets [30].

B. Correlation functions

We computed the real-space spin correlation function $\langle \vec{s}_0 \cdot \vec{s}_r \rangle$ and its Fourier transform,

$$S(\vec{q}) = \langle \vec{s}(\vec{q}) \cdot \vec{s}(-\vec{q}) \rangle = \frac{1}{N} \sum_r \langle \vec{s}_0 \cdot \vec{s}_r \rangle e^{-i\vec{q} \cdot \vec{r}}, \quad (10)$$

and plotted them for the 16×16 cluster at $J_2 = 0.5$ in Fig. 3. As expected for a spin-liquid state, the pronounced antiferromagnetic correlations between nearby sites decay rapidly. Likewise, there are no pronounced Bragg peaks in reciprocal space, only a diffuse pattern with a smooth maximum at the M point. The picture for $J_2 = 0.55$ is very similar, with a somewhat faster decay of the real-space correlator and a less pronounced maximum at the M point.

To check for any residual spin ordering, we plotted the Néel order parameter $S(\pi, \pi)$ as a function of system size in Fig. 4. In an ordered phase, this is expected to scale as [32]

$$S(L) \simeq S_0 + \alpha/L^2, \quad (11)$$

where the extrapolated order parameter S_0 and the slope α are fitting parameters. The data do not fit this relationship very well (compared to the statistical error bars), and the error of the extrapolated order parameter is far larger than that of each data point. By contrast, we expect

$$S(L) \propto L^{-z} \quad (12)$$

in a disordered phase; Ref. [7] estimated $z \approx 1.5$ for the Néel order parameter. Fitting Eq. (12) to our data points returns

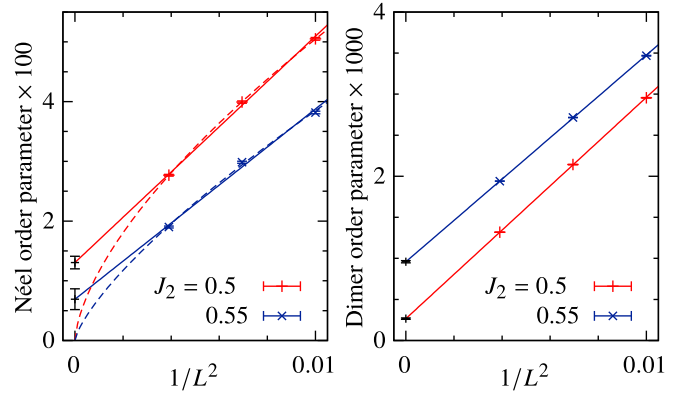


FIG. 4. Néel order parameter $S(\pi, \pi)$ (left) and columnar dimer order parameter D (right) as a function of system size for the square-lattice $J_1 - J_2$ model. Solid and dashed lines show scaling fits consistent with ordered (11) and disordered (12) ground states, respectively.

similar values for z , and yields a much better overall fit, indicating the lack of spin ordering at both $J_2 = 0.5$ and 0.55 .

We also computed the connected dimer correlation function

$$D(ij, kl) = \langle (\vec{s}_i \cdot \vec{s}_j)(\vec{s}_k \cdot \vec{s}_l) \rangle - \langle \vec{s}_i \cdot \vec{s}_j \rangle \langle \vec{s}_k \cdot \vec{s}_l \rangle. \quad (13)$$

As D can be rewritten as the covariance of local estimators of bond operators acting only on sites i, j and k, l , respectively (Appendix B), we could compute it for all pairs of nearest-neighbour sites at modest computational cost, allowing for very accurate measurements. These correlators are plotted for $J_2 = 0.5$ in Fig. 5: they show a short-range columnar pattern, consistent with Ref. [64], but decaying extremely quickly with the separation of dimers.

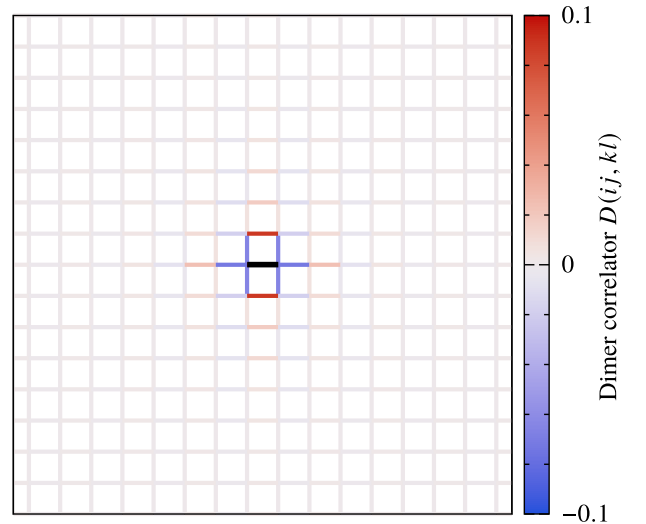


FIG. 5. Real-space dimer correlation function (13) for the square-lattice $J_1 - J_2$ model on a 16×16 lattice at $J_2 = 0.5$ in real space. The reference bond ij is shown in black; the true value of $D(ij, ij) = 0.2466(2)$ is clipped for visibility. The picture for $J_2 = 0.55$ is very similar; the dimer order parameter is too small to be directly visible.

TABLE II. Energy gaps (in units of J_1) to the optimized variational energies in several symmetry sectors for the square-lattice $J_1 - J_2$ Heisenberg model, compared to the best known figures in the literature [7]. Most of our gaps are marginally (by up to 10%) larger, owing to the lower ground-state energies we attain.

J_2/J_1	System size	Our work		RBM+PP [7]	
		MA_1^-	XB_2^+	M^-	X^+
0.5	10×10	0.2284(10)	0.2755(11)	–	–
	12×12	0.1820(17)	0.2460(16)	0.1659(9)	0.2282(4)
	16×16	0.1050(17)	0.1867(21)	0.0944(5)	0.1788(5)
0.55	10×10	0.2838(14)	0.2160(15)	–	–
	12×12	0.2170(20)	0.1522(19)	0.2202(4)	0.1371(4)
	16×16	0.2004(27)	0.1477(27)	0.1620(8)	0.0805(8)

Even though we expect dimer ordering at $J_2 = 0.55$, the plot of $D(ij, kl)$ is almost identical. To probe ordering more carefully, we define the dimer order parameter

$$\mathcal{D} = \frac{1}{2N^2} \sum_{\mu} \sum_{a,b \in \mu} D(a,b) e^{-i\vec{q}_{\mu} \cdot (\vec{r}_b - \vec{r}_a)}, \quad (14)$$

where $\mu = x, y$ are the possible orientations of nearest-neighbor bonds, a, b are bonds of the same orientation, $\vec{q}_x = (\pi, 0)$ and $\vec{q}_y = (0, \pi)$ are the columnar ordering wave vectors, and $\vec{r}_{a,b}$ are a reference point on each bond (e.g., their midpoints). \mathcal{D} is plotted in Fig. 4; it is quite small for every system we probed, consistent with the rapid decay of D in real space. For both $J_2 = 0.5$ and 0.55 , \mathcal{D} fits the scaling form (11). The extrapolated order parameter is very close to zero for $J_2 = 0.5$; it may well be eliminated by fitting to Eq. (12) with $z \approx 2$, consistent with the findings of Ref. [7]. By contrast, the extrapolated \mathcal{D} for $J_2 = 0.55$ is similar in magnitude to the finite-size measurements, which do not reasonably fit a power law, indicating dimer order.

C. Excited states

We further benchmarked our method by obtaining variational energies in the symmetry sectors corresponding to the lowest-lying excitations of the two ordered phases bordering the Dirac spin-liquid phase. In the Néel state, these come from a tower of states corresponding to the gapless magnon mode at the M point [41]: the lowest entry of this tower is a triplet transforming according to the MA_1 irrep of the space group. In the columnar valence-bond-solid proposed by Ref. [64], breaking translational symmetry yields four degenerate singlet ground states, including a pair transforming according to the XB_2 irrep, which become the lowest-lying excited states in a finite-size system.

We trained networks of the same structure as used for the ground-state estimates using the transfer learning procedure outlined in Sec. II C. Our best estimates for the gaps between the excited and ground-state energies in these two symmetry sectors are given in Table II. These gaps all match the corresponding results of Ref. [7] within 10%, suggesting that the necessarily more complex sign structure of the excited states poses little additional difficulty for the GCNN.

IV. TRIANGULAR-LATTICE $J_1 - J_2$ MODEL

The phase diagram of the triangular-lattice Heisenberg antiferromagnet is similar to the square-lattice case: the small and large J_2 limits show three-sublattice 120° and stripy orders, respectively, with a quantum paramagnetic phase between the two orders near the point of maximal frustration, $J_2 = J_1/8$ [10–12]. However, the triangular lattice is not bipartite: the resulting geometrical frustration heavily suppresses the 120° order parameter even at $J_2 = 0$ and gives rise to dynamics not captured accurately by linear spin-wave theory [65,66]. For the same reason, reliable numerical simulation is much more challenging than in the square-lattice case, and indeed, the extent and nature of the paramagnetic phase remains uncertain, with viable proposals of both spin-liquid and VBS ground states [12].

We benchmarked our GCNN wave functions at the nearest-neighbor point $J_2 = 0$ and at the point of maximal frustration, $J_2 = J_1/8$, by computing ground-state energies and correlation functions, as well as excited state gaps. For the 36-site simulations, as well as the 108-site simulation at $J_2 = 0$, the plain GCNN introduced in Sec. II A was able to find accurate ground and excited states. We used 4096 Monte Carlo samples and GCNNs with four (eight) layers of six feature maps each for the 36- (108-) site simulations. We trained the 36-site models for 12 h (~ 1500 steps) and the 108-site model for four days (~ 2000 steps).

For the more challenging $J_2 = J_1/8$ model, we found that even deep GCNNs struggled to resolve the ground states of the 108- and 144-site clusters with sufficient accuracy, and their training became unstable with increasing depth. To remedy this, we used the residual GCNN architecture shown in Sec. II B, with 12 residual blocks of six feature maps. Instead of the standard SR step (7), we trained these very deep networks using a modified algorithm, dubbed LayerSR, discussed in detail in Appendix C. Since this algorithm only requires constructing only a small portion of the quantum geometric tensor at a time, we were able to use more Monte Carlo samples, which considerably improved the converged wave functions. In particular, we used 4096 samples until the energy plateaued, which took four days (~ 600 steps), and then 16 384 samples for an additional eight days (~ 300 steps). After training, we further improved the variational energies by applying a Lanczos step as described in Appendix D.

A. Ground-state energies

Our variational energies are summarized in Table III, together with the best variational benchmarks in the literature [10,67,68]. On the 36-site cluster, GCNNs achieve energies extremely close to exact-diagonalization results for both the nearest-neighbour and maximally frustrated models within a few GPU hours. On the 108-site cluster too, GCNNs achieve state-of-the-art variational accuracy. GCNNs outperform other state-of-the-art neural-network benchmarks [68] by a wide margin, especially at the highly frustrated point $J_2 = J_1/8$. Our results also surpass those from Gutzwiller-projected fermionic mean-field ansätze with additional variational Lanczos steps [10]. Upon including one Lanczos step, we far outperform all other variational

TABLE III. Optimized ground-state energies (in units of J_1/spin) and estimates of the total spin $\langle S^2 \rangle$ for the $J_1 - J_2$ triangular-lattice Heisenberg model, compared to exact and variational benchmarks in the literature. The “exact/interpolated” energies are obtained from the thermodynamic limit estimated in Ref. [10] and the exact 36-site energy [67], assuming that finite-size effects scale as $1/L^3$ [68].

J_2/J_1	N	GCNN	GCNN+LS	Exact/Interpolated [10,67]	Graph NN [68]	Gutzwiller+LS [10]	$\langle S^2 \rangle$
0	36	-0.560313(3)	–	-0.5603734	–	–	0.0022(5)
	108	-0.55315(3)	–	–	-0.5519(4)	–	0.131(4)
1/8	36	-0.515386(9)	–	-0.515564	-0.5131(8)	-0.512503(3)	0.0067(8)
	108	-0.51175(7)	-0.51268(9)	-0.51297	-0.5069(8)	–	0.208(7)
	144	-0.51101(6)	-0.51218(9)	-0.51275	–	-0.510558(5)	–

approaches and reach relative errors of about 10^{-3} compared to ground-state energy estimates extrapolated from exact diagonalisation. We find that we need at least this level of precision in order to resolve excited-state gaps.

B. Correlation functions

We computed the reciprocal-space spin structure factors $S(q)$ for the converged 108-site GCNN wave functions and plotted them in Fig. 6. At $J_2 = 0$, we see strong Bragg peaks at the K points, consistent with 120° order; their relatively low intensity and the strong diffuse component is compatible with a small ordered moment suppressed by frustration [67,69]. At the point of maximal frustration, we find no distinct Bragg peak at either the K points or the M points expected for the large- J_2 stripy phase. Instead, we see a broad continuum around the edges of the Brillouin zone, consistent with an intermediate quantum paramagnetic phase [10–12].

To gain further insight into this intermediate phase, we computed the connected dimer-dimer correlators $D(ij, kl)$ for all relative positions of the nearest-neighbor pairs ij and kl . These are plotted in Fig. 7: dimer correlations appear to decay to very small values over a correlation length of about three unit cells. While simulations on larger system sizes would be needed to exclude the possibility of a small residual VBS order parameter, these results are most compatible with a spin-liquid state.

C. Excited states

In addition to the ground state, we also estimated the energy of low-lying excited states in nontrivial symmetry sectors using the transfer learning procedure described in Sec. II C. The tower of states for the 120° order is generated by gapless magnon modes at the K points: the lowest-lying entries are triplets transforming under the $K.A_1$ and $\Gamma.B_1$ irreps of the space group [42]. For the stripy phase at large J_2 , the lowest

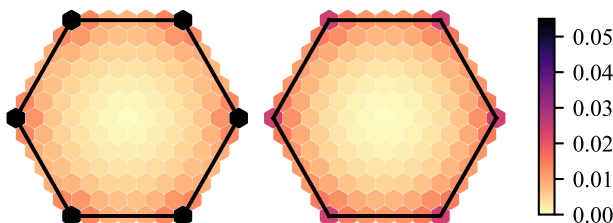


FIG. 6. Spin structure factor in the Brillouin zone for the 108-site triangular cluster at $J_2 = 0$ (left) and $J_2 = J_1/8$ (right).

lying triplet excitation transforms as $M.A_1$; in addition, the spontaneous breaking of point-group symmetry gives rise to a $\Gamma.E_2$ singlet. The proposed intermediate Dirac spin liquid phase has a plethora of gapless excitations that can be captured either as pairs of gapless spinons or monopoles of the $U(1)$ gauge field [70,71]; all states mentioned above lie in this manifold.

Our numerical results are summarized in Table IV. On 36-site clusters, we recover the exact gaps [12,42] to a good approximation. For $J_2 = 0$, the gap estimate of the $K.A_1^-$ state, corresponding to 120° order, is seen to decrease, while the others remain approximately the same, as expected for the ordered phase.

In the paramagnetic phase, we find that when we supplement our wave functions with a Lanczos step, the gaps for all three irreps decrease going from 36 to 108 sites, indicating that they may be gapless modes. However, they do not appear to scale as $N^{-1/2}$, as would be expected for a Dirac spin liquid [72]. This may be partially related to differences between the 36- and 108-site geometries, but we believe it is mostly due to the increased difficulty to obtain accurate variational energies on larger clusters. Indeed, we find that all of the gaps of all three irreps decrease when a Lanczos step is applied,

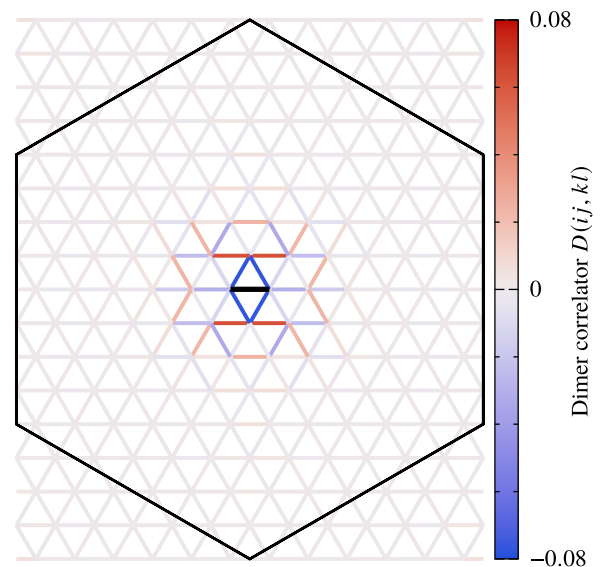


FIG. 7. Real-space dimer correlation function (13) for the triangular-lattice $J_1 - J_2$ model on a 144-site lattice at $J_2 = 1/8$ in real space. The reference bond ij is shown in black; the true value of $D(ij, ij) = 0.2445(8)$ is clipped for visibility.

TABLE IV. Energy gaps (in units of J_1) to the optimized variational energies in several symmetry sectors for the triangular-lattice $J_1 - J_2$ Heisenberg model. For $J_2 = J_1/8$, we list the gap both before and after applying a Lanczos step.

J_2/J_1	System size	KA_1^-	MA_1^-	$\Gamma.E_2^+$
0	36	0.3499(5)	0.9064(7)	0.9040(13)
	108	0.198(6)	0.780(7)	1.039(11)
1/8	36	0.4865(12)	0.5906(16)	0.2206(7)
	108	0.396(12)	0.591(13)	0.242(11)
	108 + LS	0.350(15)	0.414(16)	0.176(20)

indicating that the variational excited states are less accurate than the ground state.

V. CONCLUSION

In summary, we demonstrated the power of deep neural networks to represent highly entangled many-body wave functions. In particular, we used group convolutional neural networks (GCNNs) to study the $J_1 - J_2$ Heisenberg model on the square and triangular lattices, both of which are thought to host a quantum spin-liquid phase. We demonstrated that GCNNs achieve competitive variational energies in the quantum paramagnetic phases of the square lattice with a relatively modest computational effort. For the more difficult triangular lattice model, we showed that GCNNs with residual layers could improve upon previous state-of-the-art results by almost an order of magnitude, especially when supplemented with a Lanczos step. These results highlight deep neural networks, and GCNNs in particular, as promising ansätze to study challenging quantum many-body problems.

While we were able to consistently achieve excellent ground-state variational energies, doing so for excited states proved significantly harder. Projecting on nontrivial irreps of the space group often leads to an unstable training trajectory, as well as excessive variational gap estimates. In future work, we plan to understand the origin of this issue and propose ways to remedy it in order to gain reliable access to the low-lying spectrum, a key diagnostic for establishing phase diagrams.

Note added. While revising this manuscript, we became aware of Ref. [73], which systematically studies NQS design choices, including the role of the sign structure for frustrated systems. Their numerical results support the arguments leading to our choice of architecture in Sec. II A.

ACKNOWLEDGMENTS

We thank M. Bukov and A. Wietek for helpful discussions. We are especially grateful to Y. Nomura for sharing detailed numerical results from Ref. [7] with us. Simulations were performed using the NetKet [27] library. All heat maps use perceptually uniform color maps developed in Ref. [74]. Computing resources were provided by STFC Scientific Computing Department's SCARF cluster and STFC Cloud service, as well as the Texas Advanced Computing

Center (TACC) at the University of Texas at Austin. A.S. gratefully acknowledges the ISIS Neutron and Muon Source and the Oxford–ShanghaiTech collaboration for support of the Keeley–Rutherford fellowship at Wadham College, Oxford. C.R. acknowledges support from the University of Michigan Grant No. 2635103712 and Cornell University Grant No. 2635160012. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC-BY) license to any author accepted manuscript version arising.

APPENDIX A: GRADIENTS OF THE COMPUTATIONAL-BASIS ENTROPY

Consider a wave function ansatz $\psi(\vec{\sigma}; \theta)$ with real parameters θ . For brevity, we introduce $p(\vec{\sigma}) \equiv |\psi(\vec{\sigma}; \theta)|^2$, $P = \sum_{\vec{\sigma}} p(\vec{\sigma})$, and $q(\vec{\sigma}) = p(\vec{\sigma})/P$. Now, the derivative of Eq. (8) with respect to θ is

$$\begin{aligned}
 \partial_{\theta} \mathcal{S} &= - \sum_{\vec{\sigma}} \left[\partial_{\theta} q(\vec{\sigma}) \log q(\vec{\sigma}) + q(\vec{\sigma}) \frac{\partial_{\theta} q(\vec{\sigma})}{q(\vec{\sigma})} \right] \\
 &= - \sum_{\vec{\sigma}} \partial_{\theta} q(\vec{\sigma}) \log q(\vec{\sigma}) - \partial_{\theta} \left[\sum_{\vec{\sigma}} q(\vec{\sigma}) \right] \\
 &= - \sum_{\vec{\sigma}} \partial_{\theta} q(\vec{\sigma}) \log p(\vec{\sigma}) + \partial_{\theta} \left[\sum_{\vec{\sigma}} q(\vec{\sigma}) \right] \log P \\
 &= - \sum_{\vec{\sigma}} \left[\frac{\partial_{\theta} p(\vec{\sigma})}{P} - \frac{p(\vec{\sigma})}{P} \frac{\partial_{\theta} P}{P} \right] \log p(\vec{\sigma}) \\
 &= - \langle \log p \partial_{\theta} (\log p) \rangle + \langle \log p \rangle \langle \partial_{\theta} (\log p) \rangle, \quad (\text{A1})
 \end{aligned}$$

where the final expectation values are taken with respect to the Born distribution $q(\vec{\sigma})$. We repeatedly use the fact that $\sum_{\vec{\sigma}} q(\vec{\sigma}) = 1$, so its derivative vanishes. Eq. (A1) gives a Monte Carlo estimator for $\partial_{\theta} \mathcal{S}$ that can be used for unnormalized ansätze. Furthermore, as $\log p = 2 \operatorname{Re} \log \psi$, we can incorporate Eq. (A1) in the usual Monte Carlo estimator of the energy gradient [58] at minimal computational overhead by adding $T \log p$ to the local energy $\langle \vec{\sigma} | H | \psi \rangle / \langle \vec{\sigma} | \psi \rangle$.

APPENDIX B: COMPUTING SPIN AND DIMER CORRELATION FUNCTIONS

The expectation value of any Hermitian operator \hat{O} with respect to a variational wave function can be evaluated as the Monte Carlo average of the local estimator

$$O_{\text{loc}}(\vec{\sigma}) = \frac{\langle \vec{\sigma} | \hat{O} | \psi \rangle}{\langle \vec{\sigma} | \psi \rangle} \quad (\text{B1})$$

with respect to the Born distribution $q(\vec{\sigma})$ introduced above [58]. Furthermore, the variance of the local estimators converges to the quantum mechanical variance $\operatorname{var} \hat{O} = \langle \hat{O}^2 \rangle - \langle \hat{O} \rangle^2$. We used this directly for $\hat{O} = \vec{s}_i \cdot \vec{s}_j$ to estimate real-space spin correlators. To limit the effect of not fully translationally invariant sampling on the results, we computed $\langle \vec{s}_i \cdot \vec{s}_j \rangle$ for all pairs of spins, at the expense of using fewer independent Monte Carlo samples.

To ensure that our wave functions do not break spin-rotation symmetry, we also computed $\langle s_i^z s_j^z \rangle$ for the square-lattice ground state, expected to be one third of the

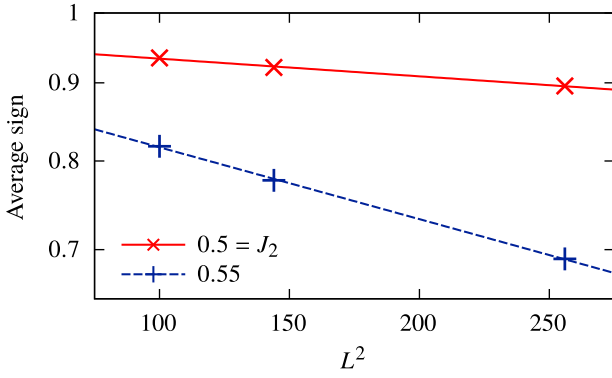


FIG. 8. Average sign of the best converged wave function after Marshall transformation. For both values of J_2 , they are very closely exponential in system size (fitted lines). The error bars are smaller than the symbols.

corresponding $\langle \vec{s}_i \cdot \vec{s}_j \rangle$. Since these operators are diagonal in the computational basis, they can be computed as the covariance of the appropriate σ^z in the sampled bit strings $\vec{\sigma}$. Not having to compute local estimators using Eq. (B1) allows us to use significantly more samples at the same computational cost. Their variance can be obtained by noting that $\langle s_i^z s_j^z \rangle^2$ is a constant (in the spin-operator normalization, $1/16$).

Making use of the identity

$$\begin{aligned} \langle \hat{A} \hat{B} \rangle &= \frac{1}{\langle \psi | \psi \rangle} \sum_{\vec{\sigma}} \langle \psi | \hat{A} | \vec{\sigma} \rangle \langle \vec{\sigma} | \hat{B} | \psi \rangle \\ &= \sum_{\vec{\sigma}} q(\vec{\sigma}) A_{\text{loc}}^*(\vec{\sigma}) B_{\text{loc}}(\vec{\sigma}), \end{aligned} \quad (\text{B2})$$

dimer correlators (13) were computed as the covariance of the local estimators of $\vec{s}_i \cdot \vec{s}_j$ and $\vec{s}_k \cdot \vec{s}_l$. The advantage of this method over computing $\langle (\vec{s}_i \cdot \vec{s}_j)(\vec{s}_k \cdot \vec{s}_l) \rangle$ directly using Eq. (B1) is that computing the local estimators for each nearest-neighbor bond operator $O(N_{\text{sample}}N)$ yields all $O(N^2)$ dimer correlators, allowing us to use substantially more independent Monte Carlo samples. (Computing the covariance takes negligible time compared to the wave-function evaluations needed for sampling and computing the local estimators.) However, this method does not allow us to directly obtain the variance of $D(ij, kl)$; instead, we estimated its error from the variance of translationally equivalent estimates.

On the square lattice, the sign structure of the $J_2 = 0$ ground state is governed by the Marshall sign rule [61]; the frustrated $J_1 - J_2$ model, by contrast, has a nontrivial sign structure even after applying the Marshall transformation. Monte Carlo estimates of the Marshall-adjusted average sign for our best converged wave functions are plotted in Fig. 8: they decay exponentially in the number of sites, as expected for a sign-problematic Hamiltonian [21]. The decay is much faster at $J_2 = 0.55$, consistent with the higher degree of frustration.

For the square-lattice wave functions, we used 2^{11} Monte Carlo samples to compute $\langle \vec{s}_i \cdot \vec{s}_j \rangle$ and the same 2^{18} samples to obtain $\langle s_i^z s_j^z \rangle$, $D(ij, kl)$, and the average sign. For the

36- (108-) site triangular lattice, we used 2^{17} (2^{14}) samples for both spin and dimer correlators.

APPENDIX C: LAYERSR ALGORITHM

To train the residual GCNN ansätze, we use an alternative to the standard stochastic reconfiguration (7). As shown in Ref. [75], if the number of samples is less than the number of neural-network parameters, Eq. (7) is equivalent to

$$\overline{O} \delta \vec{\theta} = -\eta \overline{E}_{\text{loc}}, \quad (\text{C1})$$

where E_{loc} is the local estimator (B1) of the Hamiltonian, $O = [\partial \log(\psi_i) / \partial \theta_j]_{ij}$ is the Jacobian of $\log \psi$ with respect to the parameters θ , and the overline stands for subtracting the mean across samples:

$$\overline{O}_{ij} = O_{ij} - \langle O_j \rangle; \quad \overline{E}_{i,\text{loc}} = E_{i,\text{loc}} - \langle E_{\text{loc}} \rangle. \quad (\text{C2})$$

If the number of samples is less than the number of neural-network parameters, Eq. (C1) is also underdetermined, so we are free to constrain the parameter updates further.

For LayerSR, we require that the updates $\delta \vec{\theta}_{\text{layer}}$ of the parameters of each layer in the network satisfy

$$\overline{O}_{\text{layer}} \delta \vec{\theta}_{\text{layer}} = -\frac{\eta \overline{E}_{\text{loc}}}{n_{\text{layer}}}, \quad (\text{C3})$$

where n_{layer} is the number of layers. Summing Eq. (C3) for all layers recovers Eq. (C1), so as long as there are fewer samples than parameters in each layer [i.e., Eq. (C3) is underdetermined too], the solutions of the former give a valid solution of the latter too. Equation (C3) can be recast in the standard SR form (7), which we solved using the regularization described in the main text.

Computationally, LayerSR improves on the implementations of standard SR available in NetKet [27], which either recompute the Jacobian O on the fly many times (and are thus impractically slow for a deep network) or store it in full (causing severe memory limitations on GPUs). LayerSR only requires storing the Jacobian of a single layer O^{layer} in memory, which results in a memory cost reduction proportional to the number of layers at the modest cost of performing the backpropagation needed to obtain O once for each layer. This allows us to use more samples on the single GPU, leading to significantly better converged energies. We also found that LayerSR tends to improve the convergence of the GCNN with residual layers during the early stages of training.

APPENDIX D: IMPLEMENTATION OF THE LANCZOS STEP

We implement one Lanczos step for the optimized GCNN wave functions $|\psi_0\rangle$; that is, we find the wave function with minimum variational energy in the two-dimensional Krylov space spanned by $|\psi_0\rangle$ and the orthogonal component of $H|\psi_0\rangle$,

$$|\psi_1\rangle = \frac{H - \langle H \rangle}{\sigma} |\psi_0\rangle, \quad (\text{D1})$$

which we normalized with the variance $\sigma^2 = \langle H^2 \rangle - \langle H \rangle^2$. We thus have to diagonalize the 2×2 matrix

$$\begin{aligned} H_L &= \begin{pmatrix} \langle \psi_0 | H | \psi_0 \rangle & \langle \psi_0 | H | \psi_1 \rangle \\ \langle \psi_1 | H | \psi_0 \rangle & \langle \psi_1 | H | \psi_1 \rangle \end{pmatrix} \\ &= \langle H \rangle + \begin{pmatrix} \langle \psi_0 | H - \langle H \rangle | \psi_0 \rangle & \langle \psi_0 | H - \langle H \rangle | \psi_1 \rangle \\ \langle \psi_1 | H - \langle H \rangle | \psi_0 \rangle & \langle \psi_1 | H - \langle H \rangle | \psi_1 \rangle \end{pmatrix} \\ &= \langle H \rangle + \sigma \begin{pmatrix} 0 & 1 \\ 1 & \mu_3 \end{pmatrix}, \end{aligned} \quad (D2)$$

where $\mu_3 = (\langle (H - \langle H \rangle)^3 \rangle) / \sigma^3$ is the standardized third moment of the Hamiltonian acting on $|\psi_0\rangle$. The lower eigenvalue of (D2) is

$$E_L = \langle H \rangle + \sigma \left(\frac{\mu_3}{2} - \sqrt{\frac{\mu_3^2}{4} + 1} \right), \quad (D3)$$

which is the first-order Lanczos estimate of the ground-state energy. We compute σ and μ_3 by finding the local estimators (B1) of the operators $H - \langle H \rangle$ and $(H - \langle H \rangle)^2$, labeled ε_{loc} and ν_{loc} , respectively, for 2^{14} Monte Carlo samples, from which we estimate

$$\sigma^2 = \langle \nu_{\text{loc}} \rangle; \quad \mu_3 = \frac{\text{cov}(\varepsilon_{\text{loc}}, \nu_{\text{loc}})}{\sigma^3}. \quad (D4)$$

This procedure is slightly different from Refs. [76,77], which use local estimators of H and H^2 (without the mean subtraction) to compute estimates of $\langle H^n \rangle$. This introduces an ambiguity in computing $\langle H^2 \rangle$ from the local estimators of either H or H^2 : We found that the combination of the two implied by subtracting the mean from our estimators significantly improves statistical accuracy.

-
- [1] L. Balents, Spin liquids in frustrated magnets, *Nature (London)* **464**, 199 (2010).
- [2] L. Savary and L. Balents, Quantum spin liquids: A review, *Rep. Prog. Phys.* **80**, 016502 (2017).
- [3] A. Y. Kitaev, Fault-tolerant quantum computation by anyons, *Ann. Phys.* **303**, 2 (2003).
- [4] A. Kitaev, Anyons in an exactly solved model and beyond, *Ann. Phys.* **321**, 2 (2006).
- [5] M. Hermele, M. P. A. Fisher, and L. Balents, Pyrochlore photons: The U(1) spin liquid in a $S = 1/2$ three-dimensional frustrated magnet, *Phys. Rev. B* **69**, 064404 (2004).
- [6] W.-J. Hu, F. Becca, A. Parola, and S. Sorella, Direct evidence for a gapless Z2 spin liquid by frustrating Néel antiferromagnetism, *Phys. Rev. B* **88**, 060402(R) (2013).
- [7] Y. Nomura and M. Imada, Dirac-Type Nodal Spin Liquid Revealed by Refined Quantum Many-Body Solver Using Neural-Network Wave Function, Correlation Ratio, and Level Spectroscopy, *Phys. Rev. X* **11**, 031034 (2021).
- [8] W.-Y. Liu, S.-S. Gong, Y.-B. Li, D. Poilblanc, W.-Q. Chen, and Z.-C. Gu, Gapless quantum spin liquid and global phase diagram of the spin-1/2 $J_1 - J_2$ square antiferromagnetic Heisenberg model, *Sci. Bull.* **67**, 1034 (2022).
- [9] W.-Y. Liu, J. Hasik, S.-S. Gong, D. Poilblanc, W.-Q. Chen, and Z.-C. Gu, Emergence of Gapless Quantum Spin Liquid from Deconfined Quantum Critical Point, *Phys. Rev. X* **12**, 031039 (2022).
- [10] Y. Iqbal, W.-J. Hu, R. Thomale, D. Poilblanc, and F. Becca, Spin liquid nature in the Heisenberg $J_1 - J_2$ triangular antiferromagnet, *Phys. Rev. B* **93**, 144411 (2016).
- [11] S. Hu, W. Zhu, S. Eggert, and Y.-C. He, Dirac Spin Liquid on the Spin-1/2 Triangular Heisenberg Antiferromagnet, *Phys. Rev. Lett.* **123**, 207203 (2019).
- [12] A. Wietek and A. M. Läuchli, Chiral spin liquid and quantum criticality in extended $s = 1/2$ heisenberg models on the triangular lattice, *Phys. Rev. B* **95**, 035141 (2017).
- [13] S. Yan, D. A. Huse, and S. R. White, Spin-liquid ground state of the $S = 1/2$ kagome Heisenberg antiferromagnet, *Science* **332**, 1173 (2011).
- [14] S. Depenbrock, I. P. McCulloch, and U. Schollwöck, Nature of the Spin-Liquid Ground State of the $S = 1/2$ Heisenberg Model on the Kagome Lattice, *Phys. Rev. Lett.* **109**, 067201 (2012).
- [15] Y. Iqbal, F. Becca, S. Sorella, and D. Poilblanc, Gapless spin-liquid phase in the kagome spin-1/2 Heisenberg antiferromagnet, *Phys. Rev. B* **87**, 060405(R) (2013).
- [16] Y.-C. He, M. P. Zaletel, M. Oshikawa, and F. Pollmann, Signatures of Dirac Cones in a DMRG Study of the Kagome Heisenberg Model, *Phys. Rev. X* **7**, 031020 (2017).
- [17] A. M. Läuchli, J. Sudan, and R. Moessner, The $S = 1/2$ kagome Heisenberg antiferromagnet revisited, *Phys. Rev. B* **100**, 155142 (2019).
- [18] D. Kiese, F. Ferrari, N. Astrakhantsev, N. Niggemann, P. Ghosh, T. Müller, R. Thomale, T. Neupert, J. Reuther, M. J. P. Gingras, S. Trebst, and Y. Iqbal, Pinch-points to half-moons and up in the stars: The kagome skymap, *Phys. Rev. Res.* **5**, L012025 (2023).
- [19] A. Szasz, J. Motruk, M. P. Zaletel, and J. E. Moore, Chiral Spin Liquid Phase of the Triangular Lattice Hubbard Model: A Density Matrix Renormalization Group Study, *Phys. Rev. X* **10**, 021042 (2020).
- [20] F. Mezzacapo, Ground-state phase diagram of the quantum $J_1 - J_2$ model on the square lattice, *Phys. Rev. B* **86**, 045115 (2012).
- [21] M. Troyer and U. J. Wiese, Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations, *Phys. Rev. Lett.* **94**, 170201 (2005).
- [22] R. Orús, A practical introduction to tensor networks: Matrix product states and projected entangled pair states, *Ann. Phys.* **349**, 117 (2014).
- [23] E. M. Stoudenmire and S. R. White, Studying two-dimensional systems with the density matrix renormalization group, *Annu. Rev. Condens. Matter Phys.* **3**, 111 (2012).
- [24] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [25] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359 (1989).
- [26] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Neural-Network Quantum States, String-Bond States, and Chiral Topological States, *Phys. Rev. X* **8**, 011006 (2018).

- [27] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev, and G. Carleo, NetKet 3: Machine learning toolbox for many-body quantum systems, *SciPost Phys. Codebases* **7** (2022).
- [28] K. Choo, T. Neupert, and G. Carleo, Study of the two-dimensional frustrated $J_1 - J_2$ model with neural network quantum states, *Phys. Rev. B* **100**, 125124 (2019).
- [29] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, *Phys. Rev. Res.* **2**, 033075 (2020).
- [30] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wavefunctions, *Phys. Rev. Res.* **2**, 023358 (2020).
- [31] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, Deep Autoregressive Models for the Efficient Variational Simulation of Many-Body Quantum Systems, *Phys. Rev. Lett.* **124**, 020503 (2020).
- [32] N. Astrakhantsev, T. Westerhout, A. Tiwari, K. Choo, A. Chen, M. H. Fischer, G. Carleo, and T. Neupert, Broken-Symmetry Ground States of the Heisenberg Model on the Pyrochlore Lattice, *Phys. Rev. X* **11**, 041021 (2021).
- [33] C. Roth and A. H. MacDonald, Group Convolutional Neural Networks Improve Quantum State Accuracy, [arXiv:2104.05085](https://arxiv.org/abs/2104.05085) (2021).
- [34] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted boltzmann machine learning for solving strongly correlated quantum systems, *Phys. Rev. B* **96**, 205152 (2017).
- [35] T. S. Cohen and M. Welling, Group Equivariant Convolutional Networks, [arXiv:1602.07576](https://arxiv.org/abs/1602.07576) (2016).
- [36] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, Generalization properties of neural network approximations to frustrated magnet ground states, *Nat. Commun.* **11**, 1593 (2020).
- [37] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, *SciPost Phys.* **10**, 147 (2021).
- [38] M. Li, J. Chen, Q. Xiao, F. Wang, Q. Jiang, X. Zhao, R. Lin, H. An, X. Liang, and L. He, Bridging the gap between deep learning and frustrated quantum spin system for extreme-scale simulations on new generation of sunway supercomputer, *IEEE Trans. Parallel Distrib. Syst.* **33**, 2846 (2022).
- [39] E. Wigner, *Group Theory: And its Application to the Quantum Mechanics of Atomic Spectra* (Elsevier, Amsterdam, 2012), Vol. 5.
- [40] S. d'Ascoli, L. Sagun, J. Bruna, and G. Biroli, Finding the needle in the haystack with convolutions: On the benefits of architectural bias, [arXiv:1906.06766](https://arxiv.org/abs/1906.06766) (2019).
- [41] P. W. Anderson, An approximate quantum theory of the antiferromagnetic ground state, *Phys. Rev.* **86**, 694 (1952).
- [42] A. Wietek, M. Schuler, and A. M. Läuchli, Studying continuous symmetry breaking using energy level spectroscopy, [arXiv:1704.08622](https://arxiv.org/abs/1704.08622) (2017).
- [43] V. Heine, *Group Theory in Quantum Mechanics*, International Series in Natural Philosophy, Vol. 91 (Pergamon, Oxford, 1960).
- [44] T. Misawa, S. Morita, K. Yoshimi, M. Kawamura, Y. Motoyama, K. Ido, T. Ohgoe, M. Imada, and T. Kato, mVMC—Open-source software for many-variable variational Monte Carlo method, *Comput. Phys. Commun.* **235**, 447 (2019).
- [45] Our convention differs from that of Ref. [35], which in fact implements group correlation rather than convolution. The two conventions are equivalent (the indexing of the kernels differs by taking the inverse of each element).
- [46] The full $SU(2)$ spin-rotation symmetry of Heisenberg models can only be imposed for a few ansätze that show $SU(2)$ invariance explicitly [44,78,79]. We note that for $M_z = 0$, states whose total spin quantum number S is even transform as $P = +1$, while odd- S states have $P = -1$.
- [47] C. J. Bradley and A. P. Cracknell, *The Mathematical Theory of Symmetry in Solids: Representation Theory for Point Groups and Space Groups* (Clarendon Press, London, 1972).
- [48] *Space-Group Symmetry*, International Tables for Crystallography, 2nd ed., Vol. A, edited by M. I. Aroyo (Wiley, New York, 2016).
- [49] R. S. Mulliken, Electronic structures of polyatomic molecules and valence. IV. Electronic states, quantum theory of the double bond, *Phys. Rev.* **43**, 279 (1933).
- [50] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, Self-normalizing neural networks, in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Red Hook, New York, 2017).
- [51] D. A. Huse and V. Elser, Simple Variational Wave Functions for Two-Dimensional Heisenberg Spin- $\frac{1}{2}$ Antiferromagnets, *Phys. Rev. Lett.* **60**, 2531 (1988).
- [52] A. Trask, F. Hill, S. Reed, J. Rae, C. Dyer, and P. Blunsom, Neural arithmetic logic units, [arXiv:1808.00508](https://arxiv.org/abs/1808.00508) (2018).
- [53] G. Valle-Pérez, C. Q. Camargo, and A. A. Louis, Deep learning generalizes because the parameter-function map is biased towards simple functions, [arXiv:1805.08522](https://arxiv.org/abs/1805.08522) (2018).
- [54] Y. Nomura, Helping restricted boltzmann machines with quantum-state representation by restoring symmetry, *J. Phys.: Condens. Matter* **33**, 174003 (2021).
- [55] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, New York, 2016), pp. 770–778.
- [56] S. Sorella, Green Function Monte Carlo with Stochastic Reconfiguration, *Phys. Rev. Lett.* **80**, 4558 (1998).
- [57] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum Natural Gradient, *Quantum* **4**, 269 (2020).
- [58] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).
- [59] C. Roth, Iterative retraining of quantum spin models using recurrent neural networks, [arXiv:2003.06228](https://arxiv.org/abs/2003.06228) (2020).
- [60] M. Hibat-Allah, R. G. Melko, and J. Carrasquilla, Supplementing recurrent neural network wave functions with symmetry and annealing to improve accuracy, [arXiv:2207.14314](https://arxiv.org/abs/2207.14314) (2022).
- [61] W. Marshall and R. E. Peierls, Antiferromagnetism, *Proc. R. Soc. London A* **232**, 48 (1955).
- [62] H. Shackleton and S. Sachdev, Anisotropic deconfined criticality in Dirac spin liquids, *J. High Energy Phys.* **07** (2022) 141.
- [63] K. Zhang, S. Lederer, K. Choo, T. Neupert, G. Carleo, and E.-A. Kim, Hamiltonian reconstruction as metric for variational studies, *SciPost Phys.* **13**, 063 (2022).
- [64] F. Ferrari and F. Becca, Gapless spin liquid and valence-bond solid in the $J_1 - J_2$ Heisenberg model on the square lattice:

- Insights from singlet and triplet excitations, *Phys. Rev. B* **102**, 014417 (2020).
- [65] E. A. Ghioldi, M. G. Gonzalez, S.-S. Zhang, Y. Kamiya, L. O. Manuel, A. E. Trumper, and C. D. Batista, Dynamical structure factor of the triangular antiferromagnet: Schwinger boson theory beyond mean field, *Phys. Rev. B* **98**, 184403 (2018).
- [66] D. Macdougall, S. Williams, D. Prabhakaran, R. I. Bewley, D. J. Voneshen, and R. Coldea, Avoided quasiparticle decay and enhanced excitation continuum in the spin- $\frac{1}{2}$ near-Heisenberg triangular antiferromagnet $\text{Ba}_3\text{CoSb}_2\text{O}_9$, *Phys. Rev. B* **102**, 064421 (2020).
- [67] B. Bernu, P. Lecheminant, C. Lhuillier, and L. Pierre, Exact spectra, spin susceptibilities, and order parameter of the quantum Heisenberg antiferromagnet on the triangular lattice, *Phys. Rev. B* **50**, 10048 (1994).
- [68] D. Kochkov, T. Pfaff, A. Sanchez-Gonzalez, P. Battaglia, and B. K. Clark, Learning ground states of quantum Hamiltonians with graph networks, [arXiv:2110.06390](https://arxiv.org/abs/2110.06390) (2021).
- [69] R. R. P. Singh and D. A. Huse, Three-Sublattice Order in Triangular- and Kagomé-Lattice Spin-Half Antiferromagnets, *Phys. Rev. Lett.* **68**, 1766 (1992).
- [70] X.-Y. Song, C. Wang, A. Vishwanath, and Y.-C. He, Unifying description of competing orders in two-dimensional quantum magnets, *Nat. Commun.* **10**, 4254 (2019).
- [71] X.-Y. Song, Y.-C. He, A. Vishwanath, and C. Wang, From Spinon Band Topology to the Symmetry Quantum Numbers of Monopoles in Dirac Spin Liquids, *Phys. Rev. X* **10**, 011033 (2020).
- [72] A. Wietek, S. Capponi, and A. M. Läuchli, Quantum electrodynamics in 2+ 1 dimensions as the organizing principle of a triangular lattice antiferromagnet, [arXiv:2303.01585](https://arxiv.org/abs/2303.01585) (2023).
- [73] M. Reh, M. Schmitt, and M. Gärtner, Optimizing design choices for neural quantum states, *Phys. Rev. B* **107**, 195115 (2023).
- [74] P. Kovési, Good colour maps: How to design them, [arXiv:1509.03700](https://arxiv.org/abs/1509.03700) (2015).
- [75] A. Chen and M. Heyl, Efficient optimization of deep neural quantum states toward machine precision, [arXiv:2302.01941](https://arxiv.org/abs/2302.01941) (2023).
- [76] H. Chen, D. G. Hendry, P. E. Weinberg, and A. Feiguin, Systematic improvement of neural network quantum states using Lanczos, in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, Massachusetts, 2023).
- [77] E. R. Gagliano, E. Dagotto, A. Moreo, and F. C. Alcaraz, Correlation functions of the antiferromagnetic Heisenberg model using a modified Lanczos method, *Phys. Rev. B* **34**, 1677 (1986).
- [78] T. Vieijra, C. Casert, J. Nys, W. De Neve, J. Haegeman, J. Ryckebusch, and F. Verstraete, Restricted Boltzmann Machines for Quantum States with Non-Abelian or Anyonic Symmetries, *Phys. Rev. Lett.* **124**, 097201 (2020).
- [79] T. Vieijra and J. Nys, Many-body quantum states with exact conservation of non-Abelian and lattice symmetries through variational Monte Carlo, *Phys. Rev. B* **104**, 045123 (2021).