

Neural network wave functions and the sign problem

Attila Szabó  and Claudio Castelnovo 

TCM Group, Cavendish Laboratory, University of Cambridge, Cambridge CB3 0HE, United Kingdom



(Received 14 February 2020; accepted 9 June 2020; published 15 July 2020)

Neural quantum states (NQS) are a promising approach to study many-body quantum physics. However, they face a major challenge when applied to lattice models: convolutional networks struggle to converge to ground states with a nontrivial sign structure. We tackle this problem by proposing a neural network architecture with a simple, explicit, and interpretable phase *Ansatz*, which can robustly represent such states and achieve state-of-the-art variational energies for both conventional and frustrated antiferromagnets. In the latter case, our approach uncovers low-energy states that exhibit the Marshall sign rule and are therefore inconsistent with the expected ground state. Such states are the likely cause of the obstruction for NQS-based variational Monte Carlo to access the true ground states of these systems. We discuss the implications of this observation and suggest potential strategies to overcome the problem.

DOI: [10.1103/PhysRevResearch.2.033075](https://doi.org/10.1103/PhysRevResearch.2.033075)

I. INTRODUCTION

Over the last decade, machine learning has had a profound impact on nearly all aspects of life as well as the physical sciences [1]. In condensed matter physics in particular, using neural networks as *Ansätze* for quantum many-body wave functions has emerged as an exciting application of machine learning techniques to solve challenging problems. Since the first demonstration [2] of the restricted Boltzmann machine (RBM) [3] as a practical wave-function *Ansatz* for obtaining ground states of many-body Hamiltonians using variational Monte Carlo (VMC) techniques, such neural quantum states (NQS), including deep convolutional networks [4–7], have become an important branch of many-body numerical techniques, competitive with, and sometimes even outperforming, state-of-the-art tensor network (TN) methods. NQS approaches are fundamentally appealing because both RBMs [8] and deep neural networks [9] can represent almost any function accurately, without *a priori* limitations like the Monte Carlo sign problem [10] or the area-law entanglement of TNs [11–13]. RBM states have also been used successfully in higher dimensions [2,4] and for chiral topological states [14], both of which pose well-known difficulties to TN techniques.

All promising properties notwithstanding, NQS approaches are not without their own difficulties. In particular, while NQS *Ansätze* can in principle represent nontrivial sign structures, actually learning them appears to pose significant challenges, especially in frustrated systems. This appears less pointedly in RBMs and other shallow, fully connected architectures, which are able to reach low variational energies even for Hamiltonians with a severe sign problem [3,15–20].

By contrast, deep convolutional networks (which are desirable for cutting-edge applications due to their better scalability and explicit translational invariance [21]) quite often fail to converge to ground states with near-zero average signs (e.g., of antiferromagnetic or fermionic systems) [4]. Attempting to learn such states can generate unphysically rough amplitude profiles, resulting in poor convergence or even complete breakdown of the protocol. Successful variational learning of NQS ground states in these cases requires transforming the Hamiltonian to remove its sign problem, severely limiting the usefulness of the method [4,10].¹

The origins of this “sign problem” remain poorly understood, and its existence is counterintuitive given the success of convolutional networks in a range of machine learning applications [24]. Some insight has recently been offered in Ref. [19], which studied the ability of NQS *Ansätze* to reconstruct the exact ground state from partial data in a supervised learning scenario. In certain frustrated phases, a range of architectures show poor generalization properties, especially when it comes to representing their highly nontrivial sign structures. This also impedes the convergence of VMC algorithms, which rely on reconstructing quantum expectation values based on a small sample of the Hilbert space [25]. Further work to understand these phenomena and develop more robust NQS *Ansätze* is therefore crucial to deploy neural network-based VMC algorithms to study many interesting and challenging problems in condensed matter physics.

¹For instance, the sign problem of unfrustrated antiferromagnets can be cured by imposing the Marshall sign rule [22]; the same was used successfully to stabilize the learning of frustrated ground states [4]. To the best of our knowledge, the only convolutional NQS that successfully approached an antiferromagnetic ground state without such preconditioning is that of Ref. [23], which, however, produces variational energies quite far from the state of the art even with a large number of adjustable parameters.

In this paper, we make a contribution to this quest by introducing an NQS architecture, and a corresponding variational optimization protocol, which can reliably find low-energy variational states of antiferromagnetic Hamiltonians without any prior knowledge of the sign structure of the ground-state wave function. Our *Ansatz* has a simple, explicit, and interpretable phase representation, whose convergence properties improve even upon deep convolutional networks. We benchmark our approach on the spin- $\frac{1}{2}$ J_1 - J_2 Heisenberg antiferromagnet (HAFM) on the square lattice, and achieve variational energies comparable to the state of the art [4,26,27] both at the unfrustrated point $J_2 = 0$ and in the fully frustrated quantum spin-liquid phase at $J_2/J_1 = 0.5$.

In the unfrustrated case, our approach is able to learn the expected Marshall sign rule (MSR) with excellent accuracy. This is a crucial improvement over previous VMC protocols based on convolutional NQS, which suffer from the sign problem even in this simpler case [4]. Our approach, therefore, paves the way for systematically studying conventional phases, including critical ones [28], where the area-law entanglement of TNs is a serious impediment [12].

In the frustrated case, we again achieve an excellent variational energy; however, we find a state that obeys the same MSR, even though the true ground state is expected to deviate from it significantly. The existence of such “MSR-like” low-energy variational states, and the ease and stability with which the VMC algorithm homes in on them, highlight the risks of using the energy as the only criterion for assessing the accuracy of variational wave functions and may explain the poor generalisation properties of supervised NQS learning in frustrated regimes [16,19]. We suggest potential improvements toward the end of this work.²

II. OUR APPROACH

We start by separating the amplitudes and phases of the wave function into two neural networks.³ For convenience [4,25,30], we take the networks to represent the logarithm of the wave function $|\psi\rangle$:

$$\log\langle\sigma|\psi\rangle = A(\sigma) + i\Phi(\sigma), \quad (1)$$

where $|\sigma\rangle$ are σ^z basis states, and $A(\sigma)$ and $\Phi(\sigma)$ are two functions (represented as real-valued neural networks) mapping the basis state to the log modulus and phase of its complex amplitude, respectively.

This *Ansatz* is then optimized in two stages. First, the phases are optimized to minimize the variational energy while

²Beyond NQS, Ref. [29] proposed a new “long-range entangled-plaquette state” variational *Ansatz* that appears to achieve both excellent variational energies and the correct phase structure in strongly frustrated one-dimensional systems. It will be interesting to see in the future whether this method remains successful for two-dimensional spin liquids (e.g., the model considered in our work) as well, and what insights its structure may provide for designing better NQS architectures.

³We note that this idea is in line with the observation made in Ref. [19] that the amplitudes and phases of the wave function coefficients in NQS have very different generalization properties, with the former being far easier to generalize than the latter.

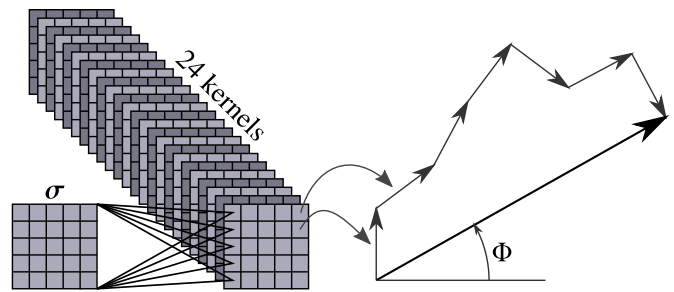


FIG. 1. Single-layer convolutional network used to represent the phase Φ of the wave function. The spins σ are mapped through convolutional kernels spanning the entire lattice (with periodic boundary conditions). Each entry in the images is then taken as the argument of a unit complex number; Φ is given by the argument of their sum.

keeping the amplitudes of all σ^z basis states equal. This allows moving away from the initial guess for Φ (which resembles the ground state of a *ferromagnetic* Hamiltonian) and approaching the correct phases without scrambling the corresponding amplitude profile. Optimal sign structures depend weakly on the imposed amplitudes, a prime example being the MSR for antiferromagnets on bipartite lattices [22], which minimizes the variational energy for *any* given set of amplitudes. Therefore, a well-converged result of this first stage is expected to be a good initial guess in the second one, where A and Φ are optimized simultaneously to approach the true ground state.

To make use of this protocol, however, the phases Φ have to be represented in a way that can approach the true ground-state sign structure starting from an initial guess very far from it. We propose a single-layer convolutional architecture, visualized in Fig. 1, where the activation layer is replaced by summing the convolutional output as phasors:

$$\Phi(\sigma) = \arg \left[\sum_{n,\mathbf{r}} \exp \left(ib_n + \sum_{\mathbf{r}'} iw_{n,\mathbf{r}-\mathbf{r}'} \sigma_{\mathbf{r}'}^z \right) \right], \quad (2)$$

where $w_{n,\mathbf{r}}$ and b_n are the real-valued weights and uniform biases of the convolutional filters, respectively.

Once an appropriate phase structure is found, good variational energies can readily be obtained using any typical neural network architecture to represent $A(\sigma)$. Similar to most machine learning tasks [24], we found that deeper, wider networks generally perform better. In our numerical experiments, a six-layer convolutional network was used, details of which are given in Appendix B 1.

We performed the optimization using stochastic reconfiguration (SR), which approximates the imaginary-time evolution of the initial state [25]. For neural network quantum states, this protocol has been shown to be superior to stochastic gradient descent and other commonly used algorithms in deep learning [31]. SR is described in Appendix A and details of the optimization are given in Appendix B 2. The simulations were implemented using the NETKET library [30].

III. NUMERICAL EXPERIMENTS

We deployed our method to the spin- $\frac{1}{2}$ HAFM on the square lattice with nearest- and next-nearest-neighbor inter-

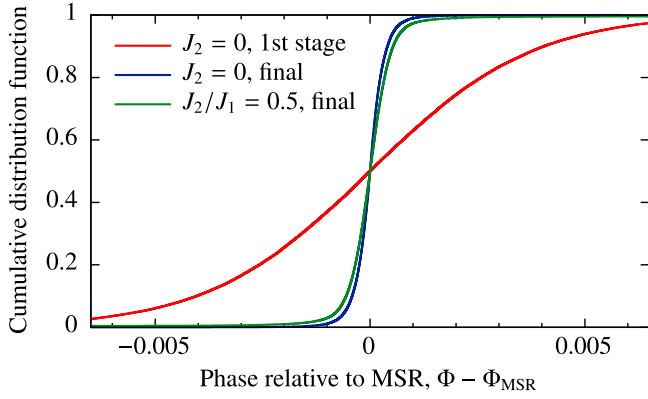


FIG. 2. Cumulative distribution function of Marshall-adjusted phases $\Phi - \Phi_{\text{MSR}}$ for the unfrustrated case $J_2 = 0$, learned before (red) and after (blue) optimizing the amplitudes, and for the final state of the frustrated case $J_2/J_1 = 0.5$ (green). All distributions are very sharply peaked, indicating that Φ is a good approximation of the MSR (5). The standard deviations of the distributions are, in order, 3.3×10^{-3} , 4.0×10^{-4} , and 2.8×10^{-3} .

actions:

$$H = J_1 \sum_{\langle ij \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j, \quad (3)$$

where $J_1, J_2 \geq 0$ and $\langle ij \rangle$ and $\langle\langle ij \rangle\rangle$ refer to nearest- and second-neighbor sites, respectively. We considered a 10×10 lattice with periodic boundary conditions and set $J_1 = 1$ without loss of generality.

Our first benchmark was the nearest-neighbor Hamiltonian $J_2 = 0$. In this case, the sign problem can be cured by rotating all spins on a checkerboard sublattice A of the square lattice by π around the σ^z axis:

$$\sigma^x \rightarrow -\sigma^x, \quad \sigma^y \rightarrow -\sigma^y, \quad \sigma^z \rightarrow \sigma^z, \quad (4)$$

as this makes the coefficients of all off-diagonal terms $\sigma_i^+ \sigma_j^-$ negative. As a result, $\langle \sigma | \prod_{i \in A} \sigma_i^z | \text{GS} \rangle$ is positive (up to an overall phase) for all σ^z basis states $|\sigma\rangle$. In terms of the phase structure $\Phi(\sigma)$, the resulting Marshall sign rule (MSR) can be written as

$$\Phi_{\text{MSR}}(\sigma) = \pi \sum_{i \in A} \frac{1 - \sigma_i^z}{2}. \quad (5)$$

We find that our phase structure *Ansatz* (2) converges reliably to (5) in the first stage of the optimization, as shown in Fig. 2. This is to be expected as the same sign structure attains optimal variational energy for any set of amplitudes [22], including the uniform one used here. Other convolutional networks, by contrast, fail to approach the MSR, which in turn leads to instabilities in the amplitude optimization, as shown in Appendix C.

In the subsequent optimization of amplitudes and phases, we achieved a variational energy of -0.671275 per spin, 2.7×10^{-4} higher than the numerically exact energy given

TABLE I. Variational energies (in units of J_1 per spin) attained in this work compared to other state-of-the-art energies on the same system. Our approach consistently outperforms plain Gutzwiller projected fermionic wave functions (GWF) [26], and achieves similar accuracy to the RBM-enhanced GWF of Ref. [33] and the convolutional networks of Ref. [4] (CNN). The “best” energy is obtained using numerically exact stochastic series expansion for $J_2 = 0$ [32] and Lanczos-corrected GWF for $J_2/J_1 = 0.5$ [26].

	$J_2 = 0$	$J_2/J_1 = 0.5$
Our work	$-0.671275(5)$	$-0.494757(12)$
GWF [26]	$-0.66935(1)$	$-0.49439(1)$
GWF + RBM [33]	$-0.67111(2)$	$-0.49575(3)$
CNN [4]	$-0.67135(1)$	$-0.49516(1)$
Best [26,32]	$-0.671549(4)$	$-0.49755(1)$

by stochastic series expansion [32]. This energy is only slightly above the one attained by the convolutional network of Ref. [4], even though the latter has substantially more variational parameters (7676 compared to our 5145) and is preconditioned with the exact MSR.

We then used our approach to study the fully frustrated phase of the model at $J_2/J_1 = 0.5$ [26,27]. We achieved a variational energy of -0.494757 per spin, 2.8×10^{-3} higher than the best energies obtained by Lanczos iterating a Gutzwiller projected fermionic wave function [26]. Our result again compares favorably with the best NQS-based variational energy in the literature [33], where the corresponding error is 1.8×10^{-3} . The variational energies obtained in both cases, together with relevant benchmarks, are summarized in Table I.

Surprisingly, however, we find that the converged phase structure $\Phi(\sigma)$ in the frustrated case recovers the MSR to a high accuracy, and no bimodality consistent with having both positive and negative amplitudes can be seen (see Fig. 2). This is at odds with the fact that the frustrated Hamiltonian (3) remains nonstoquastic even after the Marshall transformation (4), and as such its average sign should fall below 1 [34]. However, the MSR is expected to remain a relatively good approximation of the exact ground-state sign structure (see Appendix D) and for smaller but finite values of J_2 , it remains exact [35].

To further quantify the quality of the variational wave functions given by our approach, we evaluated their total spin $\langle \vec{S}^2 \rangle$, the expectation value of the parity operator $\mathcal{P} = \prod \sigma^x$, as well as the statistical weight of the five irreducible representations (irreps) of the point group D_4 of the square lattice in the wave function. Since \vec{S}^2 , \mathcal{P} , and the point-group symmetry operators all commute with the Hamiltonian (3), the true ground state is an eigenstate of the former two, and transforms according to precisely one irrep of the latter; furthermore, ground states of HAFMs are normally singlets ($\vec{S}^2 = 0$) and thus have even parity. Deviations from these expectations can be used as a quantitative test of the converged wave functions. These results are shown in Table II; computational details are given in Appendix B 3. We achieve similar figures of $\langle \vec{S}^2 \rangle$ to those of Ref. [4] in both cases, which, together with the very low weight of parity odd states, suggests a small admixture

TABLE II. Average parity $\langle \mathcal{P} \rangle$, total spin $\langle \bar{S}^2 \rangle$, and antiferromagnetic order parameters of, and statistical weight of irreps of the point group D_4 in, the fully converged NQS wave functions in the unfrustrated limit $J_2 = 0$ and for $J_2/J_1 = 0.5$. Both wave functions are predominantly parity even and transform according to the trivial representation A_1 (bold); the weights of states with odd parity and/or different spatial symmetry are ≈ 0.001 and ≈ 0.01 in the two cases. The converged $\langle \bar{S}^2 \rangle$ is similarly larger in the frustrated case, consistent with its worse energy convergence.

	$J_2 = 0$	$J_2/J_1 = 0.5$
Parity $\langle \mathcal{P} \rangle$	0.998373(29)	0.990426(69)
Weight of irrep A_1	0.998645(18)	0.989363(51)
A_2	0.000142(6)	0.000928(15)
B_1	0.000283(8)	0.003335(29)
B_2	0.000167(6)	0.001169(17)
E	0.000763(14)	0.005205(36)
Total spin $\langle \bar{S}^2 \rangle$	0.065(21)	0.581(43)
Stripy o.p. $S^2(\pi, 0)$	0.00498(5)	0.00521(7)
Néel o.p. $S^2(\pi, \pi)$	0.1571(2)	0.0633(2)

of states with high spin quantum numbers.⁴ We also note that discrepancies from the ideal ground state are an order of magnitude larger in the frustrated case by all three measures: this is consistent with the worse energy convergence and sign structure discrepancies of the same. Finally, we evaluated the antiferromagnetic order parameter

$$S^2(\vec{q}) = \frac{1}{N(N+2)} \sum_{i,j} \langle \vec{\sigma}_i \cdot \vec{\sigma}_j \rangle e^{i\vec{q} \cdot (\vec{r}_i - \vec{r}_j)} \quad (6)$$

for $\vec{q} = (\pi, 0)$ and (π, π) , which correspond to stripy and Néel orders, respectively: The results are consistent with those plotted in Ref. [4].

IV. DISCUSSION

We developed a robust and efficient protocol for finding low-energy states with a nontrivial sign structure using convolutional neural quantum states without any prior knowledge on the sign problem of the Hamiltonian. We used an *Ansatz* with two neural networks that represent the amplitudes and phases separately, and optimized it in two stages, first generating an approximate phase structure, from which the entire wave function can readily converge without encountering severe instabilities. We demonstrated our approach by attempting to learn the ground states of the square lattice spin- $\frac{1}{2}$ J_1 - J_2 HAFM both at the unfrustrated point $J_2 = 0$ and at $J_2/J_1 = 0.5$, inside the fully frustrated spin-liquid phase. In both cases, we reached variational energies comparable to the best NQS energies reported in the literature [4]; the difference might be attributed to the smaller number of variational parameters in

⁴ $\langle \bar{S}^2 \rangle \approx 0.6$ in the frustrated case could be consistent with a 70-30 mixture of singlet and triplet states; however, this would yield an average parity of 0.4. While one cannot rule out a large $s = 2$ admixture on these grounds, it is more natural to assume contributions with a range of higher spin quantum numbers.

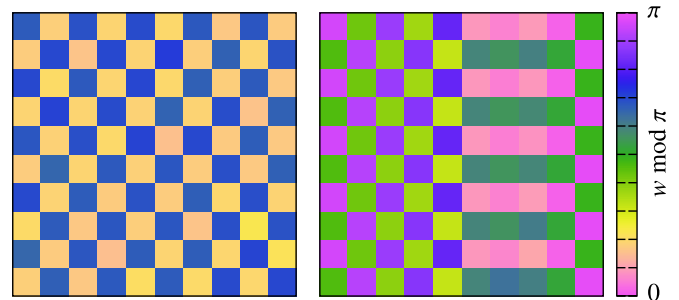


FIG. 3. Weights w_r of a typical convolutional kernel converged for $J_2 = 0$ (left) and $J_2/J_1 = 0.5$ (right). The checkerboard pattern of the former is a direct consequence of the Marshall sign rule (see Appendix E); the latter shows an admixture of a stripy antiferromagnetic pattern, consistent with the MSR for the opposite limit, $J_1 = 0$. (Perceptually uniform color map chosen following Ref. [38].)

our *Ansatz*. Importantly, we used a fully convolutional architecture: This automatically imposes translational invariance, a useful inductive bias that speeds up the convergence to a robust state representation [21] and allows for resolving the lowest-energy states in different symmetry sectors [15]. Furthermore, the convolutional structure reduces the number of variational parameters from the $O(N^2)$ typical for RBMs and other fully connected architectures [20,36] to $O(N)$, which keeps VMC algorithms viable for larger system sizes.

At $J_2 = 0$, our phase structure *Ansatz* (2) learns the Marshall sign rule with better generalization properties than other convolutional networks, both deep and shallow, which is crucial for finding ground states reliably [19]. A possible origin of the underlying inductive bias is the following: $\sigma_i^+ \sigma_j^-$ exchanges a positive and a negative value of σ_r in (1), leading to each phasor changing its phase by $\Delta\phi_{n,r} = 2(w_{n,r-r_i} - w_{n,r-r_j})$, where the dummy variable \mathbf{r} covers the entire lattice. The change of the overall phase Φ is an “average” of these. While the energy of an antiferromagnetic interaction is optimized if $\Delta\Phi = \pi$, this can be realized by a range of distributions of the $\Delta\phi$ centered on π , suggesting that the MSR can be encoded in a robust way by such an architecture. Indeed, the weights w_r produced by VMC in the unfrustrated case show a distinct checkerboard pattern that produces $\Delta\phi \approx \pi$ for all nearest-neighbor pairs, consistent with the MSR (see Fig. 3). By contrast, deep neural networks have an inductive bias for functions that only change significantly upon large-scale changes of the input [24,37]. While this is desirable for most machine learning applications, it is detrimental for learning a nontrivial quantum phase structure.

In the frustrated case, the same approach fails to find the appropriate ground-state sign structure, homing in instead on the MSR. The existence of low-lying variational states with “simple” sign structures deep within frustrated phases parallels the poor generalization of the corresponding ground states in supervised learning scenarios [19], hinting at a possible bias of NQS *Ansätze* toward such states and the corresponding ubiquity of this “residual sign problem.” This issue is compounded with the gradual nature of VMC optimization, which makes it prone to get stuck in such minima. By contrast, tensor network wave functions are typically optimized using singular value decompositions [12], which allows large yet

controlled changes in the wave function, yielding high-quality ground-state estimates for a range of challenging frustrated systems [39]. Developing similar methods for NQS states may thus help overcome this problem.

Beyond the explicit sign structure, the converged variational state in the frustrated case obeys the spin rotation and spatial symmetries of the Hamiltonian (3) substantially less accurately than for the nearest-neighbor model: $\langle \vec{S}^2 \rangle$ as well as the statistical weights of parity-odd states and those that do not transform according to the dominant irrep of the point group are all an order of magnitude higher for the former. While this may be a straightforward consequence of the gapless spin-liquid ground state [26], it might also be necessary to reconcile low variational energies with the MSR and serve as a signature of the residual sign problem. If this is the case, the performance of our approach might be improved substantially by imposing symmetries, either through making the wave function explicitly symmetric [4,15,20,40], or using the variational protocol to project out states that do not have the right symmetry (e.g., ones for which $\langle \vec{S}^2 \rangle \neq 0$).

We also note that the representation of the MSR learned in the two cases is starkly different (see Fig. 3 and Appendix E for a detailed analysis): While the unfrustrated sign structure attains the simplest possible representation of the MSR, stripe features consistent with the MSR of the opposite unfrustrated limit, $J_1 = 0$, appear in the frustrated case. This hints at an (ultimately failed) attempt at learning a “compromise” between the two limits, consistent with our qualitative understanding of frustrated phases. More detailed insight into this learning dynamics may open up the possibility of finding similarly simple *Ansätze* with significantly better generalization abilities.

Finally, we believe that the simplicity, interpretability, and robustness of our phase representation, as well as the insight it affords us about the sign problem of NQS *Ansätze*, make it a useful resource to guide efforts to design novel network architectures and training protocols that will one day reliably learn frustrated ground states with complex phase structures. Successfully learning the MSR also suggests that our method can readily be used for large-scale simulations of conventional phases, including excited states [15,17,20] and gapless or critical systems [28], without the entanglement limitations of tensor networks [13].

ACKNOWLEDGMENTS

We thank K. Choo, A. Feiguin, and T. Neupert for illuminating discussions, and N. Cooper and B. Simons for sharing computational resources with us. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC) Grants No. EP/M007065/1 and No. EP/P034616/1.

APPENDIX A: STOCHASTIC RECONFIGURATION

Stochastic reconfiguration (SR) is a generic method for optimizing parametrized trial wave functions $|\Psi(\{\alpha_k\})\rangle$ so as to minimize their energy with respect to a Hamiltonian H [25,41]. It is commonly used in variational Monte Carlo studies owing to its more reliable convergence compared

to other common optimization protocols, such as stochastic gradient descent [31,41].

The method proceeds by approximating the imaginary-time evolution of the trial wave function using Monte Carlo sampling. Namely, given $|\Psi(\{\alpha_k\})\rangle$, we want to find a new set of the real parameters⁵ $\alpha'_k = \alpha_k + \delta\alpha_k$ such that $|\Psi'\rangle = |\Psi(\{\alpha'_k\})\rangle$ is a good approximation to

$$|\Psi'_{\text{exact}}\rangle = e^{-\eta H} |\Psi(\{\alpha_k\})\rangle \approx (1 - \eta H) |\Psi(\{\alpha_k\})\rangle, \quad (\text{A1})$$

where η is a small positive number playing the role of the learning rate in machine learning language. Since we only want to project out all excited states, the Trotterization error in (A1) is irrelevant. $|\Psi'\rangle$ is optimized by maximizing the overlap of the (unnormalized) wave functions $|\Psi'_{\text{exact}}\rangle$ and $|\Psi'\rangle$:

$$|C|^2 = \frac{\langle \Psi'_{\text{exact}} | \Psi' \rangle \langle \Psi' | \Psi'_{\text{exact}} \rangle}{\langle \Psi'_{\text{exact}} | \Psi'_{\text{exact}} \rangle \langle \Psi' | \Psi' \rangle}. \quad (\text{A2})$$

To linear order in both η and $\delta\alpha_k$, the condition $\partial_{\alpha_k} |C|^2 = 0$ leads to

$$\begin{aligned} \sum_j \delta\alpha_j \text{Re} \left[\frac{\langle \partial_{\alpha_j} \Psi | \partial_{\alpha_k} \Psi \rangle}{\langle \Psi | \Psi \rangle} - \frac{\langle \partial_{\alpha_j} \Psi | \Psi \rangle \langle \Psi | \partial_{\alpha_k} \Psi \rangle}{\langle \Psi | \Psi \rangle^2} \right] \\ = \eta \text{Re} \left[\frac{\langle \Psi | H | \partial_{\alpha_k} \Psi \rangle}{\langle \Psi | \Psi \rangle} - \frac{\langle \Psi | H | \Psi \rangle \langle \Psi | \partial_{\alpha_k} \Psi \rangle}{\langle \Psi | \Psi \rangle^2} \right]. \quad (\text{A3}) \end{aligned}$$

In order to find $\delta\alpha_k$ numerically, we want to rewrite the expectation values in (A3) as Monte Carlo averages with respect to the quantum probability distribution $p(\sigma) = |\langle \sigma | \Psi \rangle|^2 / \langle \Psi | \Psi \rangle$. This can readily be done by inserting a resolution of the identity; for example, we have

$$\begin{aligned} \frac{\langle \Psi | H | \partial_{\alpha_k} \Psi \rangle}{\langle \Psi | \Psi \rangle} &= \sum_{\sigma} \frac{\langle \Psi | H | \sigma \rangle \langle \sigma | \partial_{\alpha_k} \Psi \rangle}{\langle \Psi | \Psi \rangle} \\ &= \sum_{\sigma} \frac{|\langle \sigma | \Psi \rangle|^2}{\langle \Psi | \Psi \rangle} \frac{\langle \Psi | H | \sigma \rangle}{\langle \Psi | \sigma \rangle} \frac{\langle \sigma | \partial_{\alpha_k} \Psi \rangle}{\langle \sigma | \Psi \rangle} \\ &= \sum_{\sigma} p(\sigma) E_{\text{loc}}^*(\sigma) O_k(\sigma), \quad (\text{A4}) \end{aligned}$$

where we introduce

$$O_k(\sigma) = \frac{\langle \sigma | \partial_{\alpha_k} \Psi \rangle}{\langle \sigma | \Psi \rangle} = \partial_{\alpha_k} \log \langle \sigma | \Psi \rangle \quad (\text{A5})$$

and the local energy

$$E_{\text{loc}}(\sigma) = \frac{\langle \sigma | H | \Psi \rangle}{\langle \sigma | \Psi \rangle}. \quad (\text{A6})$$

The expectation value (A4) can now be estimated as the Monte Carlo average of $E_{\text{loc}}^* O_k$ for samples distributed according to $p(\sigma)$; the others follow from analogous considerations,

⁵Equivalent expressions can be derived for trial wave functions that are (piecewise) analytic functions of complex parameters [2,25]. The result is identical to (A7), omitting the real-part signs.

resulting in

$$\sum_j \underbrace{\text{Re cov}(O_j, O_k)}_{S_{kj}} \delta\alpha_j = -\eta \underbrace{\text{Re cov}(E_{\text{loc}}, O_k)}_{f_k}. \quad (\text{A7})$$

Since the covariance matrix S depends entirely on the parametrization of the wave function rather than its energy under the Hamiltonian, it can be thought of as a metric tensor on the parametrized Hilbert space [25]. Equation (A7) is thus analogous to the natural gradient approaches used to stabilize gradient descent in other machine learning contexts [31]. To improve numerical stability, the covariance is calculated as

$$\text{cov}(X, Y) = \langle X^* Y \rangle - \langle X^* \rangle \langle Y \rangle = \langle (X - \langle X \rangle)^* (Y - \langle Y \rangle) \rangle.$$

In our setup, $\log(\sigma|\Psi) = A(\sigma, \{\kappa\}) + i\Phi(\sigma, \{\lambda\})$, where both A and Φ are real-valued functions of real parameters. It follows that O_κ is real for all κ and O_λ is imaginary for all λ , and so all entries of the covariance matrix S connecting a κ and a λ vanish. This allows us to solve (A7) for the $\delta\kappa$ and the $\delta\lambda$ separately, speeding up the algorithm.

Solving (A7) for $\delta\alpha_\kappa$ requires inverting the covariance matrix S , which, while positive semidefinite, tends to be ill conditioned even for a large number of Monte Carlo samples [25]. This can be resolved either by using the pseudoinverse or, more commonly, by adding a small positive constant to the diagonal entries in order to make the matrix invertible. In our case, however, the entries of the S matrix corresponding to the parameters of the amplitude and phase have vastly different values (separated by up to eight orders of magnitude). To keep the optimization of both parts viable, we add 10^{-5} times the average diagonal entry (i.e., the trace divided by the number of parameters) to both blocks of the matrix:

$$\tilde{S}_\kappa = S_\kappa + 10^{-5} \frac{\text{tr } S_\kappa}{\dim S_\kappa} \mathbb{1},$$

and likewise for S_λ .

APPENDIX B: DETAILS OF NUMERICAL EXPERIMENTS

1. Neural network architectures

The amplitude structure $A(\sigma)$ was represented using a six-layer convolutional neural network in all numerical experiments. The layers consist of 8, 7, 6, 5, 4, and 3 10×10 lattice replicas, respectively, which are connected by convolutional filters with real-valued kernels spanning 4×4 sites in periodic boundary conditions, and (real-valued) ReLU activation functions:

$$f_{1\dots 5}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

The amplitude is given by the modulus of the product of all entries in the last convolutional layer. Since the NQS networks implemented in NETKET represent the logarithm of wave functions [30], this is achieved using a final activation function $f_6(x) = \ln|x|$, followed by summing all entries. All convolutional layers before the last one are initialized with Gaussian distributed random numbers of zero mean, and standard deviation chosen so as to preserve the typical magnitude of backpropagated derivatives [42]. The last set of kernels are initialized with a uniform bias of 1.0 and Gaussian distributed

kernel entries with standard deviation 2×10^{-4} . This results in amplitudes uniformly close to 1 upon initialization.

For all data discussed in the main text, the phases $\Phi(\sigma)$ were represented by the phasor sum *Ansatz* (2). We employed 24 lattice replicas (cf. Fig. 1). Similarly good results are achieved using fewer replicas, but the wider network allows for faster and more reliable convergence. The convolutional filters are initialized with Gaussian distributed random numbers of standard deviation 0.043.

2. Optimization protocol

We optimized each wave-function ansatz via stochastic reconfiguration in two stages. First, the phases are optimized with a uniform amplitude distribution (i.e., setting $A \equiv 0$): 10 000 such SR steps with learning rate $\eta = 0.01$ were followed by 10 000 steps with $\eta = 0.05$. Next, both A and Φ were optimized starting from the phase distribution achieved in the first stage: for this, we used 5000 steps with $\eta = 0.001$, 5000 steps with $\eta = 0.01$, and, finally, 50 000 steps with $\eta = 0.05$. The learning rate was increased during the optimization because the imaginary-time evolution emulated by SR results in infinitesimal temperature (and thus energy) reduction close to the ground state. In both stages, the Monte Carlo averages in (A7) were evaluated using 5000 samples obtained via the Metropolis–Hastings algorithm as implemented by NETKET [30].

The convergence of the phasor sum *Ansatz* to the ground state is shown by the blue curves in Fig. 4. The first stage quickly attains an approximately constant minimum of variational energy. We find, however, that any residual optimization speeds up the next stage significantly, which is desirable as a single, simple neural network can be evaluated an order of magnitude faster than the full *Ansatz*. The second stage also reaches a nearly converged variational energy in about 20 000 steps; however, the variational energy is further reduced slightly throughout the procedure.

3. Observable estimation

Once the wave function had converged, the estimates of cumulative distribution functions plotted in Figs. 2 and 5 were generated by drawing 100 000 samples out of the probability distribution $p(\sigma) = |\langle \sigma | \Psi \rangle|^2 / \langle \Psi | \Psi \rangle$ and sorting their phases.

Analogous to the estimation of variational energies, (A6), the expectation value of any operator A can be evaluated as the Monte Carlo average of the local estimates

$$A_{\text{loc}}(\sigma) = \frac{\langle \sigma | A | \Psi \rangle}{\langle \sigma | \Psi \rangle} \quad (\text{B1})$$

with respect to the quantum probability distribution $p(\sigma) = |\langle \sigma | \Psi \rangle|^2 / \langle \Psi | \Psi \rangle$. NETKET provides a facility for evaluating such expectation values within the SR protocol: This was used to estimate the variational energy $\langle \tilde{S}^2 \rangle$ and the spin correlators (6) using 10^6 Monte Carlo samples. We exploited the translational invariance of the wave function to rewrite the latter two as

$$S^2(\vec{q}) = \frac{1}{N+2} \sum_i \langle \vec{\sigma}_i \cdot \vec{\sigma}_0 \rangle e^{i\vec{q} \cdot \vec{r}_i}; \quad (\text{B2})$$

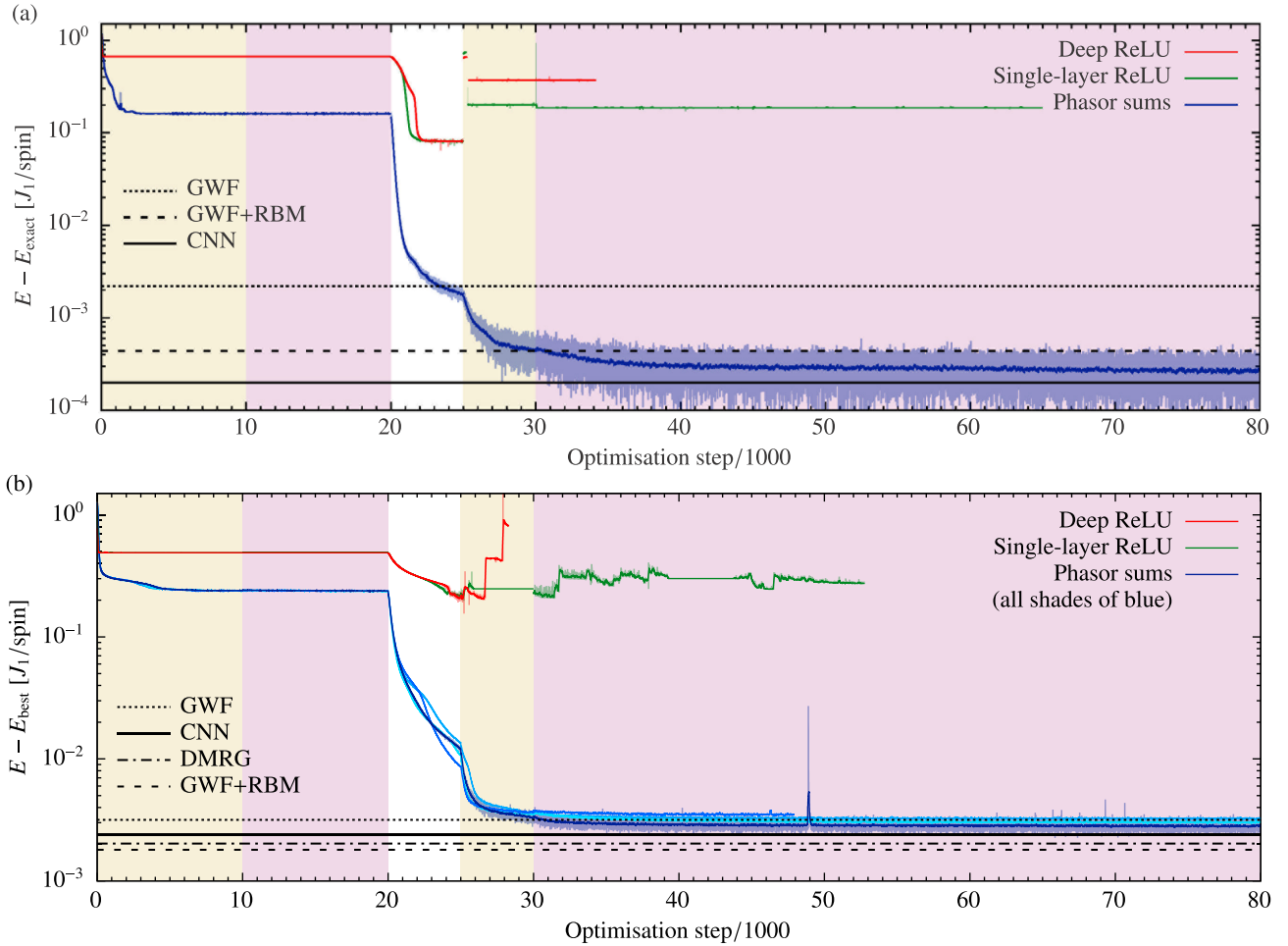


FIG. 4. Convergence of our optimization scheme with various neural network *Ansätze* (described in Appendices B 1 and C) for the nearest-neighbor (top panel) and the $J_2/J_1 = 0.5$ (bottom panel) square-lattice HAFM. The shaded area shows the full spread of energy estimates used by the SR algorithm, the thicker lines show 100-step moving averages. In the frustrated case, the best of four runs (dark blue) is shown in detail together with the moving averages of the others (shades of turquoise). The background shading indicates the learning rate η (white: 0.001, yellow: 0.01, purple: 0.05). The phasor sum *Ansatz* (2) (blue curve) converges reliably to energies close to the true ground state; other *Ansätze* (red and green curves; see Appendix C for details), both shallow and deep, fail to reach either a consistent variational energy, or one close to the ground state. Energies are compared to exact stochastic series expansion for $J_2 = 0$ [32] or Lanczos-corrected Gutzwiller projected fermionic wave functions (GWF) for $J_2/J_1 = 0.5$ [26]. For reference, variational energies are also shown for the convolutional network of Ref. [4] (CNN), plain [26] and RBM-improved [33] GWF, as well as DMRG [27] for the frustrated case.

$$\langle \vec{S}^2 \rangle = N \sum_i \langle \vec{\sigma}_i \cdot \vec{\sigma}_0 \rangle, \quad (\text{B3})$$

where both sums include $i = 0$ (note, however, that $\vec{\sigma}_0 \cdot \vec{\sigma}_0 \equiv \frac{3}{4}$).

We checked furthermore whether the converged wave functions obey parity and point-group symmetries. The parity operator $\mathcal{P} = \prod_i \sigma_i^x$ commutes with all point-group symmetries as well as the Hamiltonian: Therefore, the parity of a wave function is fully characterized by the expectation value $\langle \mathcal{P} \rangle$, without any possibility of symmetry-protected degeneracies. By contrast, the non-Abelian point group D_4 gives rise to such degeneracies, limiting the usefulness of plain symmetry operator expectation values. Instead, we evaluated the statistical weight of eigenstates transforming according to the different

irreps α of D_4 , using the projection operators [43]

$$\hat{P}_\alpha = \frac{d_\alpha}{|D_4|} \sum_{g \in D_4} \chi_\alpha(g) \hat{g}, \quad (\text{B4})$$

where d_α and χ_α are the dimension and characters of the irrep, respectively, and $|D_4| = 8$. The weight of each irrep is given by

$$w_\alpha = \frac{\langle \psi | \hat{P}_\alpha^\dagger \hat{P}_\alpha | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\langle \psi | \hat{P}_\alpha | \psi \rangle}{\langle \psi | \psi \rangle} = \langle \hat{P}_\alpha \rangle,$$

where we used the fact that the projector (B4) is Hermitian and squares to itself. Both w_α and $\langle \mathcal{P} \rangle$ were evaluated using 4×10^6 Monte Carlo samples again using (B1). [We used different samples here, as NETKET does not offer a simple implementation of these symmetry operators, so it proved more expedient to sample $p(\sigma)$ directly.]

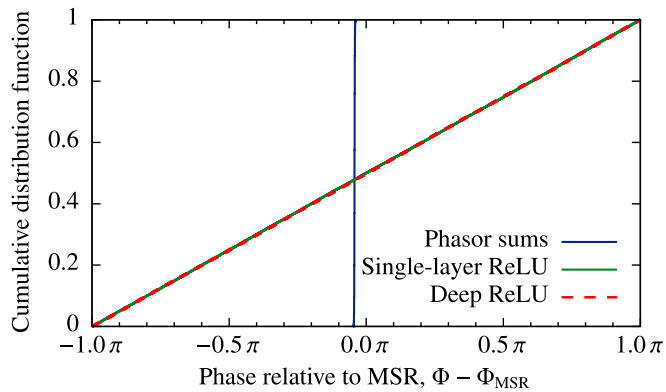


FIG. 5. Cumulative distribution function of the Marshall-adjusted phases $\Phi - \Phi_{\text{MSR}}$ learned by the three phase structure *Ansätze* in the first stage of the optimization in the unfrustrated case $J_2 = 0$. The phasor sum *Ansatz* (2) develops a very narrow distribution, that is, it reproduces the MSR to a good approximation; the other *Ansätze* show an approximately uniform distribution, i.e., a complete failure to learn the sign rule.

Converged variational energies and all other observables are reported in Tables I and II, respectively.

APPENDIX C: COMPARISON OF PHASE STRUCTURE ANSÄTZE

To demonstrate the advantage of our phase structure *Ansatz* over other convolutional networks, we considered two alternative architectures:

(1) 24 convolutional filters spanning the entire lattice, followed by a ReLU activation layer and summation.

(2) The architecture used for the amplitudes, except for the last layer, where the $\ln|x|$ activation is also replaced by ReLU.

Amplitudes are encoded using the same *Ansatz* as described in Appendix B 1. The performance of these architectures under the protocol described in Appendix B 2 is shown by the green and red curves in Fig. 4, respectively. In the first stage, neither of them approach the optimal variational energy found with the phasor sum *Ansatz*; subsequently, the amplitude network also fails to approach the ground state, even though it is capable of representing it closely, as found previously. We also point out that as the Monte Carlo sampling is restarted after changing the learning rate η , the estimates of the variational energy change substantially, leading to discontinuities in Fig. 4. This indicates that the amplitude structure had developed several unphysically strong peaks, which make subsequent Monte Carlo sampling unable to recover the correct wave function.

In the unfrustrated case, we also probe the phase structures learned by the various *Ansätze* by comparing them directly to the exact Marshall sign rule. The distribution of the differences $\Phi - \Phi_{\text{MSR}}$ is shown in Fig. 5. While the architecture used in the main paper learns the MSR to a high accuracy (up to an irrelevant overall phase), the alternatives produce essentially random phases.

TABLE III. Average sign $\langle s \rangle$ and Marshall-adjusted sign $\langle s \rangle_{\text{MSR}}$ of the ground state of the $J_2/J_1 = 0.5$ square-lattice HAFM for 4×4 and 4×6 lattices, calculated by exact diagonalization. The 10×10 lattice is extrapolated from these, assuming that $\langle s \rangle$ decays exponentially in the number of spins [34].

	$\langle s \rangle$	$\langle s \rangle_{\text{MSR}}$
4×4	3.53×10^{-2}	0.9745
4×6	3.87×10^{-3}	0.9650
10×10	$\approx 3 \times 10^{-12}$	≈ 0.88

APPENDIX D: AVERAGE SIGNS IN EXACT DIAGONALIZATION

To estimate the average Marshall-adjusted sign of the true ground state of the 10×10 frustrated model, we obtained the exact ground states for 4×4 and 4×6 lattices using the Lanczos algorithm as implemented in SCIPY [44], and calculated the average sign

$$\langle s \rangle_{\psi} = \left| \sum_{\sigma} p_{\psi}(\sigma) \frac{\langle \sigma | \psi \rangle}{|\langle \sigma | \psi \rangle|} \right| = \left| \frac{\sum_{\sigma} |\langle \sigma | \psi \rangle| \langle \sigma | \psi \rangle}{\sum_{\sigma} |\langle \sigma | \psi \rangle|^2} \right| \quad (\text{D1})$$

for both the original and the Marshall-adjusted ground states $|\text{GS}\rangle$ and $\prod_{i \in A} \sigma_i^z |\text{GS}\rangle$. These average signs are given in Table III, together with an extrapolation to the 10×10 lattice, assuming an exponential decay of both $\langle s \rangle$ [34]. The average sign of the original wave function decays extremely fast and becomes negligibly small for our lattice size. By contrast, the average Marshall-adjusted sign remains close, but clearly distinct from, 1. For a 10×10 lattice, the expected average sign is ≈ 0.88 : Since this is the difference of the statistical weights of positive and negative amplitude states, we expect those to be about 94% and 6%, respectively.

APPENDIX E: KERNELS OF THE GROUND-STATE PHASE STRUCTURES

As discussed in Sec. IV, the change in the phase *Ansatz* (2) upon exchanging an up and a down spin separated by \mathbf{R} is a kind of average of the change $\Delta\phi = 2(w_{n,\mathbf{r}+\mathbf{R}} - w_{n,\mathbf{r}})$ in each elementary phase that enters it. (The factor of 2 is due to representing up and down spins as ± 1 rather than $\pm \frac{1}{2}$.) Since the energy $+J\sigma_i^+ \sigma_j^-$ of an antiferromagnetic interaction is optimized if $\Delta\Phi = \pi$ upon exchanging spins i and j , we expect the corresponding set of $\Delta\phi$ to have a distribution centered upon π , especially in the unfrustrated case, where all interactions in the Hamiltonian can be optimized simultaneously.

In particular, one can represent the Marshall sign rule corresponding to the unfrustrated limit $J_2 = 0$ exactly by requiring that all $\Delta\phi \equiv \pi \pmod{2\pi}$, i.e., $w_{\mathbf{r}+\mathbf{R}} - w_{\mathbf{r}} \equiv \pi/2 \pmod{\pi}$ for all nearest-neighbor pairs, or $\mathbf{R} = \hat{\mathbf{x}}, \hat{\mathbf{y}}$. This is readily achieved by a checkerboard pattern of weights w and $w + \pi/2$, provided both sides of the lattice are of even length. Indeed, the phase of each elementary phasor in this

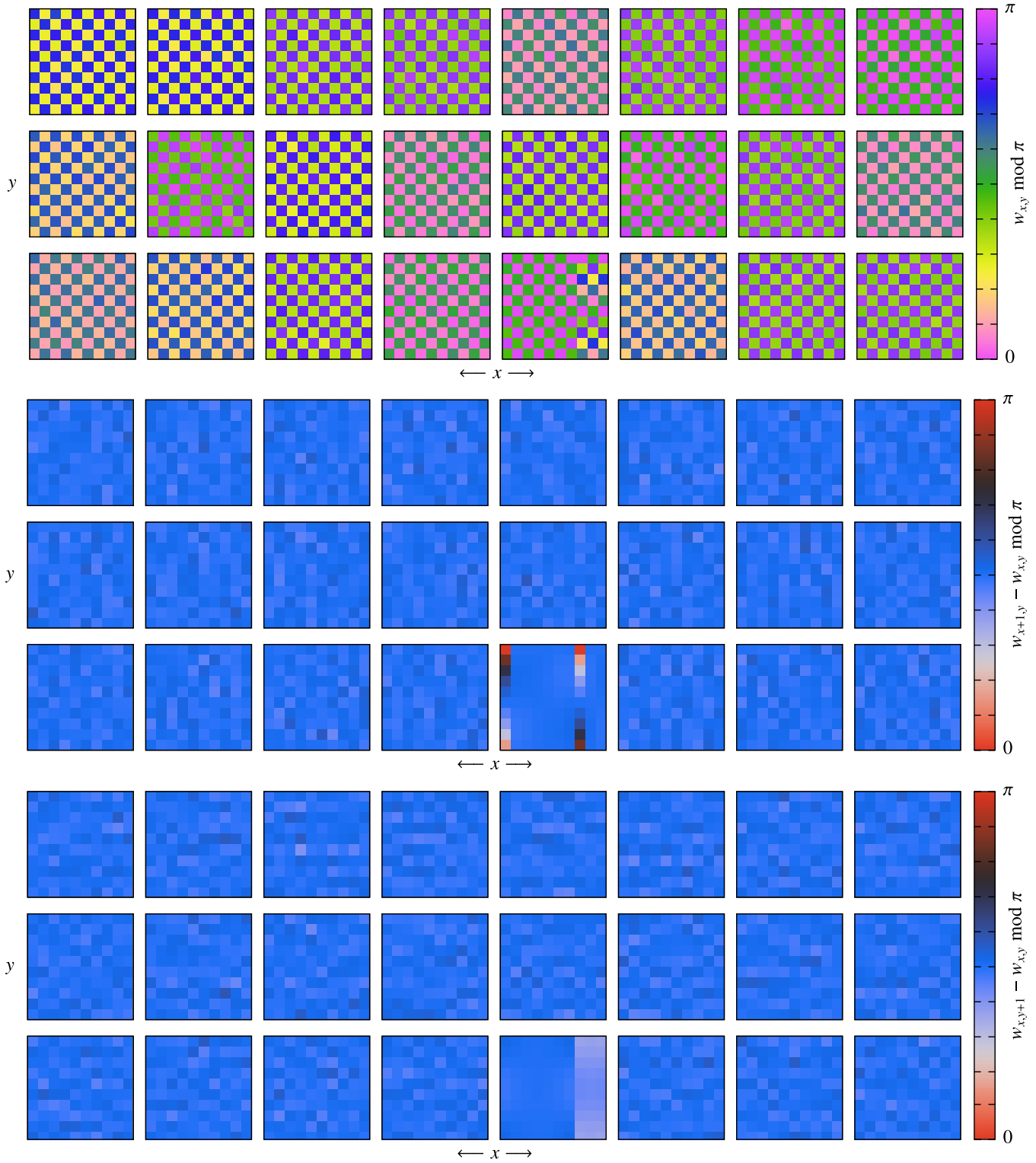


FIG. 6. Weights $w_{n,r}$ of all convolutional kernels n converged for $J_2 = 0$ (top three rows), and the differences $w_{n,r+\hat{x}} - w_{n,r}$ and $w_{n,r+\hat{y}} - w_{n,r}$ (middle and bottom three rows, respectively). All kernels but one show a clear checkerboard pattern, which results in $\Delta\phi \approx \pi$ upon exchanging a neighboring up and down spin, consistent with the Marshall sign rule. In the only exception (fifth kernel of the third row), a “topological fault” spanning three columns appears, causing some $\Delta\phi$ to wind from 0 to 2π : these have little effect on the overall $\Delta\Phi$, and might persist due to a topologically invariant winding number. The kernel shown in Fig. 3 is the first one in the second row. (Perceptually uniform color maps chosen following Ref. [38]).

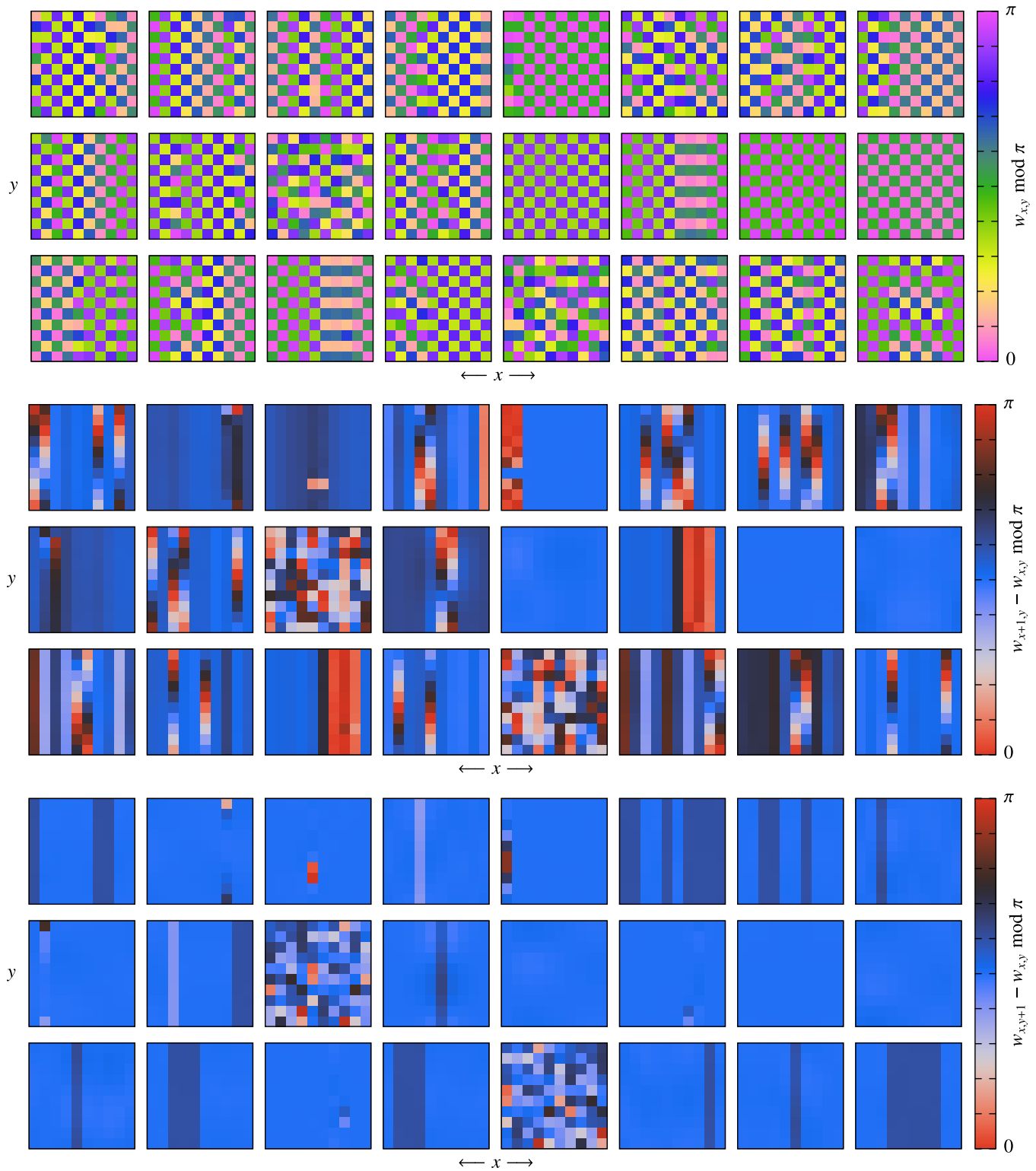


FIG. 7. Weights $w_{n,r}$ of all convolutional kernels n converged for $J_2/J_1 = 0.5$ (top three rows), and the differences $w_{n,r+\hat{x}} - w_{n,r}$ and $w_{n,r+\hat{y}} - w_{n,r}$ (middle and bottom three rows, respectively). The kernels are altogether much less regular than in the unfrustrated case, several of them appearing completely scrambled, and many showing various winding patterns. Nonetheless, the checkerboard pattern corresponding to the MSR is still common. In addition, a stripy pattern with successive rows differing by $\pi/2$ appears, leading to $\Delta\phi \approx 0$ upon exchanging nearest-neighbor up and down spins horizontally but $\Delta\phi \approx \pi$ vertically, in an apparent breaking of rotational symmetry. The kernel shown in Fig. 3 is the sixth one in the second row. (Perceptually uniform color maps chosen following Ref. [38].)

setup is

$$\begin{aligned}
 \phi_{\mathbf{r}} &= \sum_{\mathbf{r}'} w_{n,\mathbf{r}-\mathbf{r}'} \sigma_{\mathbf{r}'}^z = w \sum_{\mathbf{r}'} \sigma_{\mathbf{r}'}^z + \frac{\pi}{2} \sum_{\mathbf{r}' \in A(B)} \sigma_{\mathbf{r}'}^z \\
 &= -\frac{\pi N}{4} + \pi \sum_{\mathbf{r}' \in B(A)} \frac{1 - \sigma_{\mathbf{r}'}^z}{2} \\
 &\equiv \frac{\pi N}{4} + \pi \sum_{\mathbf{r}' \in A} \frac{1 - \sigma_{\mathbf{r}'}^z}{2} \pmod{2\pi}, \quad (\text{E1})
 \end{aligned}$$

where N is the number of lattice sites, A and B are the two checkerboard sublattices of the lattice, and we repeatedly use the fact that $\sum \sigma^z = 0$. In the last line, we also note that the total number of down spins (which is measured by the sums for each sublattice) is $N/2$, which is even; therefore, the parity of the sum for the two sublattices is the same. That is, all terms in (2) have the same phase, which is also consistent with the MSR (5) up to an overall phase $\pi N/4$. It follows that Φ equals this phase, and thus recovers the MSR.

Beyond the nearest-neighbor case, there is a Marshall sign rule corresponding to the other unfrustrated limit $J_1 = 0$, arising from the requirement that $\Delta\Phi = \pi$ upon exchanging next-nearest-neighbor up and down spins. At this point, the two checkerboard sublattices decouple, each being effectively a square lattice with diagonal axes. Using the above construction, appropriate kernels $w_{\mathbf{r}}$ can be found for both of them; however, the offset between the two sublattices is arbitrary, as it only contributes to an unimportant overall phase. Upon reintroducing a weak nearest-neighbor coupling, however, the model develops a stripy antiferromagnetic order [26,27]. Consistently, the sublattices in the convolutional weights w are expected to lock so as to form rows or columns with equal w , shifted from one another by $\pi/2 \pmod{\pi}$.

We now consider the convolutional kernels generated by the variational Monte Carlo protocol both in the unfrustrated limit and at $J_2/J_1 = 0.5$. The weights $w_{n,\mathbf{r}}$ are plotted for each kernel in Figs. 6 and 7, respectively. In both cases, we also plot the differences between horizontally and vertically nearest-neighbor kernel entries.

At $J_2 = 0$, the checkerboard pattern derived for the exact MSR can be seen in almost all kernels to a very good approximation; consistently, $\Delta\phi \approx \pi$ upon exchanging nearest-

neighbor spins, both horizontally and vertically. This suggests that this representation of the MSR is especially stable; presumably, it sits at the bottom of a wide basin of the learning landscape, making it easy to find for optimization algorithms [24,45]. The only surprising feature in Fig. 6 is a kernel that develops a three-column-wide “topological fault,” in which the Marshall-adjusted convolutional weights wind around the vertical direction. This results in a number of $\Delta\phi$ far from the desired π . The large number of kernels, however, allows these to be corrected through slight deviations of the other kernels from the exact MSR. Furthermore, the additional kernels probably play a key role in eliminating such detrimental structures: Unwinding a “topological fault” requires large-scale changes in the individual phasors, which substantially increase the variational energy, unless corrected for by the other kernels. Indeed, our attempts to use a single convolutional kernel in (2) were plagued by robust “topological faults” spanning the entire kernel, leading to convergence far above the ground-state energy.

The kernels obtained in the frustrated case are substantially more complex, with many “topological faults” and some kernels that show no discernible pattern. Many kernels, however, retain the checkerboard pattern consistent with the MSR, and the $\Delta\phi$ upon exchanging nearest-neighbor spins vertically are clearly dominated by values close to π . Both of these are consistent with the fact that the kernels represent the MSR rather than the true frustrated sign structure. Nevertheless, we observe several columns of the stripy pattern consistent with the MSR of the $J_1 = 0$ limit, as discussed above. These result in $\Delta\phi \approx 0$ for horizontal nearest-neighbor exchanges, leading to a much more varied pattern in these differences. It is surprising that this more diverse distribution has no apparent effect on the overall sign structure learned by the network.

Furthermore, the striking difference between $\Delta\phi$ along the horizontal and the vertical directions is not warranted either by any fundamental property of the *Ansatz*, or by the final converged wave function that obeys all point-group symmetries to a high accuracy (see Table II). Nevertheless, it might hint at spontaneous point-group symmetry breaking at higher variational energies that ultimately leads to learning an incorrect sign structure. A more detailed analysis of the learning dynamics is necessary to better understand and overcome any such problems.

-
- [1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [2] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [3] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, *Nat. Phys.* **15**, 887 (2019).
- [4] K. Choo, T. Neupert, and G. Carleo, *Phys. Rev. B* **100**, 125124 (2019).
- [5] K. Choo, A. Mezzacapo, and G. Carleo, *Nat. Commun.* **11**, 2368 (2020).
- [6] O. Sharir, Y. Levine, N. Wies, G. Carleo, and A. Shashua, *Phys. Rev. Lett.* **124**, 020503 (2020).
- [7] L. Yang, Z. Leng, G. Yu, A. Patel, W.-J. Hu, and H. Pu, *Phys. Rev. Research* **2**, 012039(R) (2020).
- [8] N. Le Roux and Y. Bengio, *Neural Comput.* **20**, 1631 (2008).
- [9] G. Cybenko, *Math. Control Signals Syst.* **2**, 303 (1989).
- [10] Z.-X. Li and H. Yao, *Annu. Rev. Condens. Matter Phys.* **10**, 337 (2019).
- [11] M. M. Wolf, F. Verstraete, M. B. Hastings, and J. I. Cirac, *Phys. Rev. Lett.* **100**, 070502 (2008); A. Riera and J. I. Latorre, *Phys. Rev. A* **74**, 052326 (2006).
- [12] R. Orús, *Nat. Rev. Phys.* **1**, 538 (2019).
- [13] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, *Phys. Rev. Lett.* **122**, 065301 (2019).

- [14] R. Kaubruegger, L. Pastori, and J. C. Budich, *Phys. Rev. B* **97**, 195136 (2018).
- [15] K. Choo, G. Carleo, N. Regnault, and T. Neupert, *Phys. Rev. Lett.* **121**, 167204 (2018).
- [16] Z. Cai and J. Liu, *Phys. Rev. B* **97**, 035116 (2018).
- [17] D. Hendry and A. E. Feiguin, *Phys. Rev. B* **100**, 245123 (2019).
- [18] G. Torlai, J. Carrasquilla, M. T. Fishman, R. G. Melko, and M. P. A. Fisher, [arXiv:1906.04654](https://arxiv.org/abs/1906.04654).
- [19] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. I. Katsnelson, and A. A. Bagrov, *Nat. Commun.* **11**, 1593 (2020).
- [20] Y. Nomura, *J. Phys. Soc. Jpn.* **89**, 054706 (2020).
- [21] S. d'Ascoli, L. Sagun, J. Bruna, and G. Biroli, [arXiv:1906.06766](https://arxiv.org/abs/1906.06766).
- [22] W. Marshall and R. E. Peierls, *Proc. R. Soc. London A* **232**, 48 (1955).
- [23] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, *Phys. Rev. B* **98**, 104426 (2018).
- [24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016).
- [25] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).
- [26] W.-J. Hu, F. Becca, A. Parola, and S. Sorella, *Phys. Rev. B* **88**, 060402(R) (2013).
- [27] S.-S. Gong, W. Zhu, D. N. Sheng, O. I. Motrunich, and M. P. A. Fisher, *Phys. Rev. Lett.* **113**, 027201 (2014).
- [28] R. Zen, L. My, R. Tan, F. Hebert, M. Gattobigio, C. Miniatura, D. Poletti, and S. Bressan, [arXiv:2002.02618](https://arxiv.org/abs/2002.02618).
- [29] J. Thibaut, T. Roscilde, and F. Mezzacapo, *Phys. Rev. B* **100**, 155148 (2019).
- [30] G. Carleo, K. Choo, D. Hofmann, J. E. 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, F. Vicentini, and A. Wietek, *SoftwareX* **10**, 100311 (2019).
- [31] C.-Y. Park and M. J. Kastoryano, *Phys. Rev. Research* **2**, 023232 (2020).
- [32] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, *New J. Phys.* **11**, 083026 (2009).
- [33] F. Ferrari, F. Becca, and J. Carrasquilla, *Phys. Rev. B* **100**, 125131 (2019).
- [34] M. Troyer and U. J. Wiese, *Phys. Rev. Lett.* **94**, 170201 (2005).
- [35] J. Richter, N. B. Ivanov, and K. Retzlaff, *Europhys. Lett.* **25**, 545 (1994); A. Voigt, J. Richter, and N. B. Ivanov, *Physica A (Amsterdam)* **245**, 269 (1997).
- [36] D. Sehayek, A. Golubeva, M. S. Albergo, B. Kulchytskyy, G. Torlai, and R. G. Melko, *Phys. Rev. B* **100**, 195125 (2019).
- [37] H. W. Lin, M. Tegmark, and D. Rolnick, *J. Stat. Phys.* **168**, 1223 (2017); G. Valle-Pérez, C. Q. Camargo, and A. A. Louis, [arXiv:1805.08522](https://arxiv.org/abs/1805.08522); G. De Palma, B. T. Kiani, and S. Lloyd, [arXiv:1812.10156](https://arxiv.org/abs/1812.10156).
- [38] P. Kovési, [arXiv:1509.03700](https://arxiv.org/abs/1509.03700).
- [39] U. Schollwöck, *Rev. Mod. Phys.* **77**, 259 (2005).
- [40] T. Vieijra, C. Casert, J. Nys, W. De Neve, J. Haegeman, J. Ryckebusch, and F. Verstraete, *Phys. Rev. Lett.* **124**, 097201 (2020).
- [41] S. Sorella, *Phys. Rev. B* **64**, 024512 (2001).
- [42] K. He, X. Zhang, S. Ren, and J. Sun, [arXiv:1502.01852](https://arxiv.org/abs/1502.01852).
- [43] V. Heine, *Group Theory in Quantum Mechanics*, International Series in Natural Philosophy, Vol. 91 (Pergamon, Oxford, 1960).
- [44] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey *et al.*, *Nat. Methods* **17**, 261 (2020).
- [45] C. Baldassi, F. Pittorino, and R. Zecchina, *Proc. Natl. Acad. Sci. USA* **117**, 161 (2020).