

Neural network evolution strategy for solving quantum sign structures

Ao Chen,^{1,*} Kenny Choo,² Nikita Astrakhantsev,² and Titus Neupert^{1,2}¹Department of Physics, ETH Zurich, CH-8093 Zurich, Switzerland²Department of Physics, University of Zurich, CH-8057 Zurich, Switzerland

(Received 22 November 2021; accepted 22 March 2022; published 2 May 2022)

Feed-forward neural networks are a novel class of variational wave functions for correlated many-body quantum systems. Here, we propose a specific neural network ansatz suitable for systems with real-valued wave functions. Its characteristic is to encode the all-important rugged sign structure of a quantum wave function in a convolutional neural network with *discrete* output. Its training is achieved through an evolutionary algorithm. We test our variational ansatz and training strategy on two spin-1/2 Heisenberg models, one on the two-dimensional square lattice and one on the three-dimensional pyrochlore lattice. In the former, our ansatz converges with high accuracy to the analytically known sign structures of ordered phases. In the latter, where such sign structures are *a priori* unknown, we obtain better variational energies than with other neural network states. Our results demonstrate the utility of discrete neural networks to solve quantum many-body problems.

DOI: [10.1103/PhysRevResearch.4.L022026](https://doi.org/10.1103/PhysRevResearch.4.L022026)

Introduction. Finding the ground states of frustrated quantum magnets poses a formidable challenge to existing numerical algorithms. For instance, exact diagonalization (ED) methods are limited to small system sizes, while quantum Monte Carlo [1] fails in frustrated models due to the notorious sign problem [2]. The density matrix renormalization group (DMRG) [3,4], despite being very successful in low-dimensional models, struggles to express entanglement in higher-dimensional systems. Complementary to all these are variational approaches, which are, in principle, scalable to larger systems but carry different variational bias [5]. To reduce the latter, many-variable variational Monte Carlo (VMC) has been put forward, either based on pair product wave functions [6] or based on the application of neural networks (NNs) to quantum systems known as the *neural quantum states* (NQSs) approach [7].

Numerous studies aimed to improve the accuracy of the NQS wave functions in order to take full advantage of their expressibility for many-body calculations. For instance, incorporating point group symmetries of the state into the restricted Boltzmann machine, employing the convolutional neural network (CNN) [8] or the group-CNN [9] structure, has led to significant improvements in the performance of NQSs [10–12]. The SU(2) symmetry can also be enforced to improve the approximation quality [13,14]. The large-scale implementation of deep NNs is another way to obtain better variational states [15–17]. Furthermore, combining several NNs [18,19], NNs with traditional Gutzwiller-projected fermionic wave functions [20–22], and

NNs with projected entangled pair states [23] has led to more accurate variational sign structures and better variational solutions.

In spite of all these improvements, representing and finding sign structures are still a major problem in pure NQS methods [24], which prevents NQSs from obtaining desirable variational states in highly frustrated systems. For instance, in three-dimensional systems such as the frustrated Heisenberg pyrochlore model, NQS methods obtained less accurate variational energies than the pair product wave function ansatz [25].

To alleviate this problem of sign structure approximation, we propose an architecture constructed by the combination of a base network and a sign “helper” network. The sign network has discrete output values representing the signs of wave function amplitudes. Such a discrete sign network requires a gradient-free optimization method [26]. In this work, we adopt an evolutionary algorithm—the evolution strategy (ES) [27]—to perform the optimization. We contrast our ansatz with a previously introduced one, in which a helper network that represents a continuous complex phase was introduced [19].

To demonstrate an improvement over existing NQS techniques, we test the performance of our network with ES optimization on the spin-1/2 j_1 - j_2 Heisenberg models on both the two-dimensional (2D) square lattice and the three-dimensional (3D) pyrochlore lattice. Both of them are exemplary open problems in frustrated quantum magnetism. In the former system, we illustrate the working mechanism of the combined network and the advantage of introducing the ES optimization. In the latter, we show that our approach can produce competitive variational states with variational energies that are better than all previous NQS studies [25] and are comparable to the ones obtained within DMRG [28]. Finally, we discuss the optimal NN architecture for learning wave function sign structures and the future prospects of the ES approach.

*chenao.phys@gmail.com

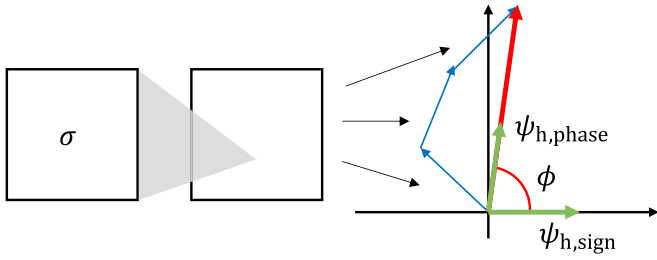


FIG. 1. Architecture of the helper network. The network has a convolutional layer followed by an activation function which returns a phase ϕ . One convolutional layer contains only one channel without bias. The red vector represents the sum of all blue vectors, and the green vectors are the outputs of phase and sign networks.

Sign network with evolution strategy. We use a product of two separate networks to express the variational wave function ψ as

$$\psi(\sigma) = \psi_b(\sigma) \times \psi_h(\sigma), \quad (1)$$

where σ is the spin configuration of a spin-1/2 magnet in a fixed basis, while ψ_b and ψ_h are the outputs of the base and helper networks, respectively. The base network and the symmetrized wave function are introduced in the Supplemental Material [29] (SM; Secs. A and B).

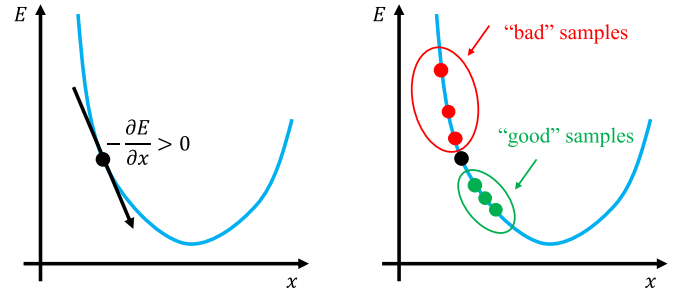
The purpose of introducing an additional helper network is to help the base network to express the sign structure and increase the accuracy of the variational state. The helper network has a one-layer convolutional architecture, as shown in Fig. 1. To interpret the output of the convolutional layer as the sign structure, we apply the approach used in Ref. [19]. Namely, every element in the output of the single convolutional layer is treated as the phase of a unit complex number (phasor), and the sum of all these phasors gives the output phase, i.e.,

$$\phi(\sigma) = \arg\left(\sum_j e^{iv_j(\sigma)}\right), \quad (2)$$

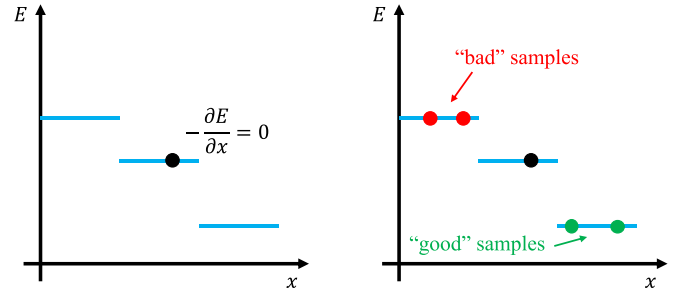
where v_j is the output of the convolutional layer. In the previous NQS studies, the helper network usually expresses the phase of wave functions $\psi_{h,\text{phase}}(\sigma) = e^{i\phi(\sigma)}$, which yields a continuous output [19,25,30].

In this work, we target only ground state wave functions that have real entries. This allows us to find ground states for Hamiltonians which can be written in a basis in which they have exclusively real matrix elements, a class of systems that encompasses many relevant and highly frustrated models, e.g., with Heisenberg and Kitaev spin-spin interactions.¹ In such a case of a real ground state, the imaginary part in $\psi_{h,\text{phase}}(\sigma)$ introduces an unnecessary redundancy which can potentially reduce the performance. In contrast, the output of the sign

¹If a specific complex-valued ground state is desired (e.g. with a well-defined non-time-reversal-symmetric momentum), such a wave function can be built by superposing several ES-trained real-valued wave functions.



(a) Continuous case



(b) Discrete case

FIG. 2. Comparison between gradient descent and evolution strategy (ES). Both of them are applicable in continuous functions, but only ES is applicable in the discrete case.

helper network is given by

$$\psi_{h,\text{sign}}(\sigma) = \text{sgn}[\cos \phi(\sigma)], \quad (3)$$

which maps the phase ϕ to a discrete sign ± 1 and eliminates the redundancy.

We adopt the gradient-free ES method to train the sign network, while the base network and the phase helper network (required for performance comparison) are trained using stochastic reconfiguration (SR) [31]. Compared with other gradient-free methods, the ES has shown outstanding performance in optimizing complicated functions [32]. In Fig. 2 we illustrate the difference between the ES and gradient-based optimization methods such as stochastic gradient descent (SGD). If the energy E is a continuous function of a variable x , SGD moves x in the steepest descent direction $-\frac{\partial E}{\partial x}$. Instead, the ES method generates samples in the vicinity of the current variable value x and estimates the direction of energy descent therewith. If the function is piecewise constant in the variational parameters, which is the case with the sign network introduced in Eq. (3), the ES can be used to optimize the variational wave function, while SGD is no longer applicable.

Numerous variations of the ES optimization method exist. In this Letter, we employ the covariance matrix adaptation evolution strategy (CMA-ES) [27], which consists of the following three steps in each iteration:

(1) Starting from the sign network weights $\mathbf{W}$, generate a set of sample networks with different weights drawn from the multivariate normal distribution

$$\mathbf{X}_n \sim \mathcal{N}(\mathbf{W}, \Sigma) \sim \mathbf{W} + \sqrt{\Sigma} \mathcal{N}(0, \mathbf{1}), \quad (4)$$

where the index n labels sample networks and $\Sigma = \langle (\mathbf{X} - \mathbf{W})(\mathbf{X} - \mathbf{W})^T \rangle$ is the covariance matrix.

(2) Compute the variational energy E_n of sample networks using the reweighting VMC in the SM (Sec. C).

(3) Update the sign network weights as

$$\mathbf{W}' = \mathbf{W} + \sum_{n=1}^N w_n (\mathbf{X}_n - \mathbf{W}) \quad (5)$$

and the covariance matrix as

$$\Sigma' = \sum_{n=1}^N w_n (\mathbf{X}_n - \mathbf{W})(\mathbf{X}_n - \mathbf{W})^T, \quad (6)$$

where N is the number of samples. The samples with lower energy E_n are assigned a higher importance w_n (see the SM, Sec. D, for details).

These iterations are repeated until the variational energy is converged. The PYCMA package [33] is adopted to perform the CMA-ES optimization. Further details of CMA-ES and an analysis of computational complexity are given in the SM (Sec. D).

Numerical experiments. As benchmark examples, we study the spin-1/2 j_1 - j_2 Heisenberg model on the 2D square lattice and on the 3D pyrochlore lattice. Both of these systems are possible hosts to quantum spin liquid states and have recently been the subject of many-variable Monte Carlo and NQS variational studies, which has led to a better understanding of their phase diagram [22,25]. The Heisenberg Hamiltonian is given by

$$H = j_1 \sum_{\langle i,j \rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j + j_2 \sum_{\langle\langle i,j \rangle\rangle} \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j, \quad (7)$$

where $\langle \dots \rangle$ and $\langle\langle \dots \rangle\rangle$ indicate nearest- and next-nearest-neighbor spins on the respective lattice. The maximally frustrated region is located in the vicinity of $j_2/j_1 = 0.5$ for the square lattice [22,34] and at $j_2/j_1 = 0$ for the pyrochlore lattice [25,28,35].

In the case of even lattice dimensions in both directions, the square lattice can be divided into A and B sublattices with a checkerboard pattern or a stripe pattern. In the $j_2/j_1 = 0$ and $j_1/j_2 = 0$ limits, the sign of a wave function of a fixed spin configuration is given by $(-1)^{M_A}$, with M_A being the total number of spins up in the A sublattice for this configuration. In the $j_2/j_1 = 0$ limit, the sublattice is a checkerboard pattern formed by one of the two bipartite components of the lattice. This special case is known as the Marshall sign rule (MSR) [36]. In the $j_1/j_2 = 0$ limit, a sublattice has a stripe pattern consisting of alternating rows or columns.

We start to test our approach in a case where the exact sign structure is known, the 10×10 square lattice at $j_2/j_1 = 0$. The MSR in this limit, despite being easy to express analytically, can be tough for a generic NQS to learn and express [37]. To show how our method approximates the MSR, we subtract the phase Φ of the helper network outputs $\psi_h(\sigma)$ from the phase Φ_{MSR} given by the MSR. In Fig. 3 we show the cumulative distribution function (CDF) of this difference in phases. The sign network trained by the ES leads to a step-function CDF, indicating that our approach finds the exact sign structure. We contrast this with the performance of the

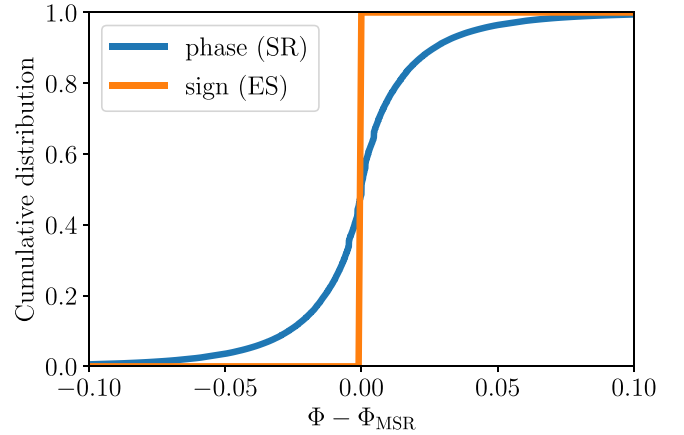


FIG. 3. Cumulative distribution function (CDF) of phase relative to the MSR on the 10×10 square lattice. The sign network learns all signs correctly in 10^6 samples, resulting in the steplike CDF with the step at $\Phi - \Phi_{\text{MSR}} = 0$.

phase helper network in Ref. [19], which expresses a continuous phase instead of a discrete sign. As seen from Fig. 3, this network also approximates MSR but contains significant phase errors.² This showcases that the discrete sign network trained by ES obtains a more accurate sign structure than the conventional continuous-phase networks for comparable network complexity.

We now consider the performance of the sign network in the frustrated region of the square lattice j_1 - j_2 model. To this end, we show the relative error of the energy and depict the sign network weights for various $0.45 \leq j_2/j_1 \leq 0.65$ in Fig. 4. These weights explicitly show patterns corresponding to the checkerboard and stripe patterns described above and represent the simple sign structures at the two extreme bipartite limits of Néel and stripe orders, which means the sign network can adapt to suitable simple sign structures successfully. Nevertheless, the sign network cannot express the remaining frustrated sign structure. This observation extends to more complex sign network architectures with, for instance, multiple channels and was also observed in Ref. [19]. Although the sign network cannot express the full sign structure, the simple sign structures provided by the sign network still lead to a significantly improved performance of the base network compared to the traditional CNN approach in Ref. [11] (see Fig. 4). A similar conclusion also holds for the 10×10 lattice, in which case the base network cannot even be trained to approximate the ground state without the sign network.

In addition to the accuracy of variational states given by the converged optimization, we also study the expected time cost for ES to yield an accurate sign structure. To this end, we perform 100 attempts of helper network training at each j_2/j_1 . A comparison between SR and ES is performed with all NN manipulations run parallelly on a GPU (see the SM, Sec. D, for details of parallelization). Each training attempt is terminated if either its convergence is accurate or it exceeds 40 s,

²To reduce the phase error, a multichannel phase network more complex than the one-channel helper network in Fig. 1 was adopted in Ref. [19].

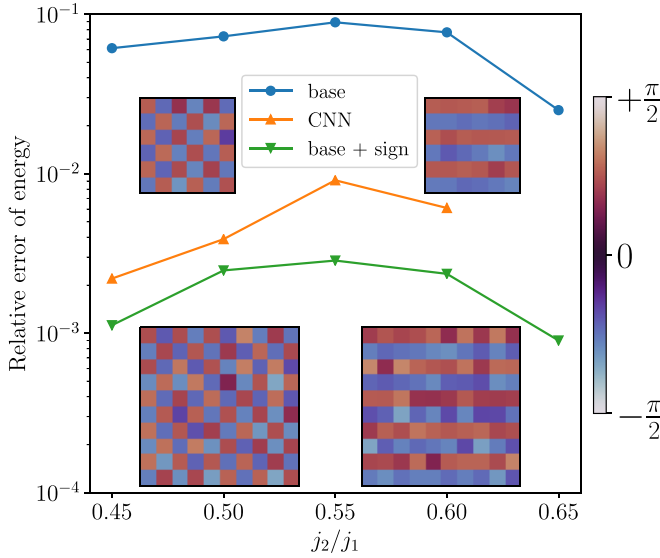


FIG. 4. Relative error of the variational energy for the spin-1/2 j_1 - j_2 Heisenberg model on the 6×6 square lattice. The “base” and “base + sign” points represent the performance of the base network without and with the sign network, respectively. The “CNN” data represent the approach employed in Ref. [11] which uses a complex-valued CNN to express the amplitude and phase with the MSR applied analytically. The energy on larger lattices (e.g., 10×10) is not shown because the base network cannot approximate the ground state without the sign network in this case. The insets show the weights of the sign network at $j_2/j_1 = 0.5$ and $j_2/j_1 = 0.6$ on the 6×6 and 10×10 lattices.

a typical time required to assess the accuracy of the result.³ If the obtained sign structure is not accurate, the algorithm is restarted. In Fig. 5, we present the accurate convergence probability p , the time cost of every converged training t_1 , and the expected total time t , which, including the time cost of inaccurate training attempts, is given by

$$t = \sum_{n=1}^{\infty} (1-p)^{n-1} p n t_0 + \bar{t}_1 = \frac{1-p}{p} t_0 + \bar{t}_1, \quad (8)$$

where $t_0 = 40$ s is the time cost of every inaccurate training and $\bar{t}_1$ is the mean value of t_1 . In Fig. 5(a), the ES optimization converges faster in the highly frustrated region around $j_2/j_1 = 0.5$ compared with SR. As seen from Fig. 5(b), ES is less prone to local minima and more often accurately converges at all frustrated j_2/j_1 values due to its larger searching space compared to SR. The advantages in both aspects, the required time and probability to reach the accurate sign structure, lead to the shorter expected total run time of ES as given by Fig. 5(c). This proves that using the ES technique with discrete signs does not yield significant computational overhead compared to the classical SR optimization.

³The criterion of accurate convergence is the 90% correctness in the prediction of the sign structure (for the phase network trained within SR, the prediction is deemed accurate if the phase error is smaller than $\pi/2$). Many candidate sign structures with similar energies exist at $j_2/j_1 = 0.5$, so it is replaced by $j_2/j_1 = 0.48$ in Fig. 5.

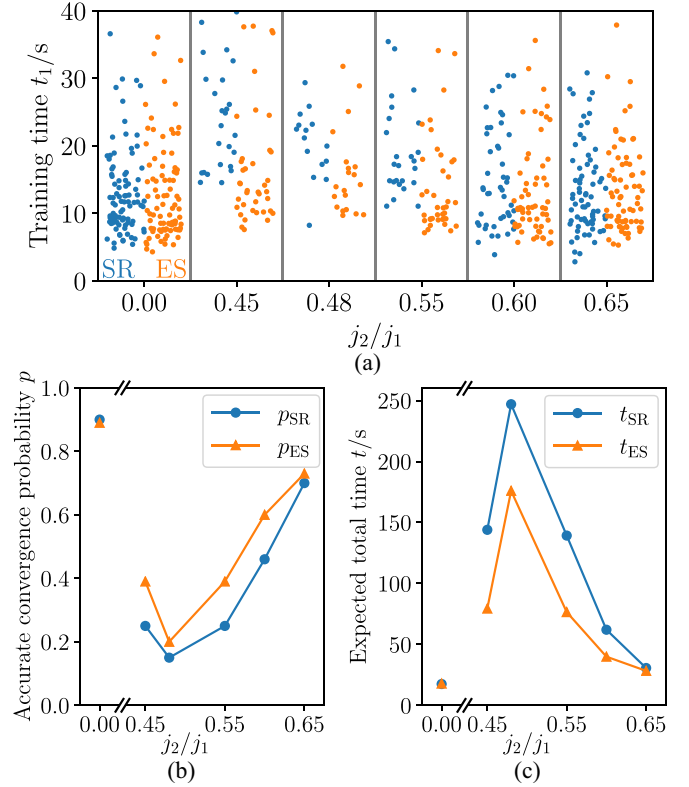


FIG. 5. Comparison of training efficiency between SR and ES. (a) Training time. (b) Accurate convergence probability. (c) Expected total time. The data for all panels are from the same training results on a Tesla-P100 GPU. In (a), every scattered point represents a training that converges within 40 s. In (c), the expected total time t is given by Eq. (8).

The application of the sign network in the 2D square j_1 - j_2 model might be considered a special case since the well-known MSR can be used as a good starting point [10,11] instead of using a sign network trained by ES. To complement this, we study the case of the 3D pyrochlore lattice in which such prior sign structures are unknown. We begin with the performance assessment for our approach on the 4×2^3 ED-accessible pyrochlore lattice cluster. In Table I, we present the relative error of the variational energy and the squared overlap between the variational state and the exact ground state to compare the improvement provided by the

TABLE I. Overlap with the exact ground state and relative error of energy on the 4×2^3 pyrochlore Heisenberg model.

Sign ^a	Amplitude ^a	Helper network	Overlap ²	\Delta E/E _{GS} \sup>
Variational	Exact	No	0.87	
Variational	Exact	Sign network	0.91	
Exact	Variational	No	0.79	
Exact	Variational	Sign network	0.83	
Variational	Variational	No	0.77	6.2×10^{-3}
Variational	Variational	Sign network	0.82	4.1×10^{-3}

^aThe exact sign or exact amplitude represents the variational wave function expressed by the network with signs or amplitudes replaced by the exact ground state values.

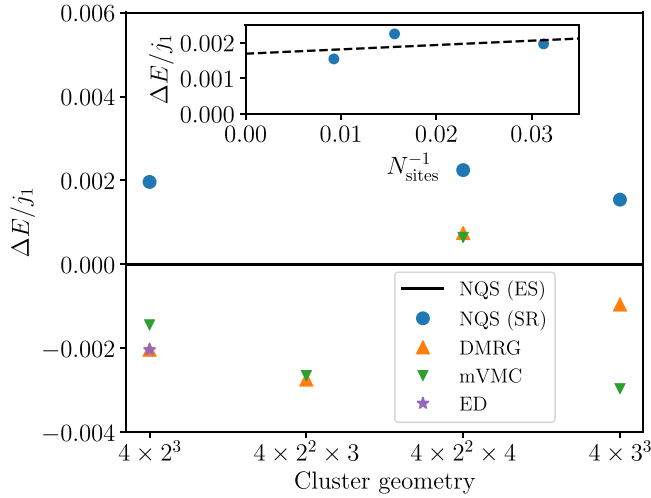


FIG. 6. Energy difference with respect to the ES approach in the pyrochlore Heisenberg model at various lattice sizes. The inset shows the infinite-size extrapolation for the NQS (SR) data. The NQS (SR) approach in Ref. [25] contains two separate CNNs for amplitude and phase. The many-variable variational Monte Carlo (mVMC) and ED data are also given in Ref. [25]. The DMRG data are given in Ref. [28].

sign helper network. The data with variational amplitudes and exact signs represent the variational state with the sign structure replaced by the exact ground state sign structure and hence reflect the accuracy of the variational amplitude. Similarly, the data with variational signs and exact amplitudes reflect the accuracy of the variational sign structure. These results show that the sign helper network not only permits more accurate sign structures but also helps the base network to learn more accurate amplitudes. The improvement in the signs and amplitudes altogether leads to better wave functions, as shown by the energy error and the overlap of the variational wave function. The supplementary data for the performance of different approaches on the 4×2^3 pyrochlore lattice are given in the SM (Sec. E).

In Fig. 6, we compare our approach with other numerical methods in the pyrochlore Heisenberg model at various lattice sizes.⁴ The performance of the base network without a sign helper network is not shown because its variational energy is far higher than other data exhibited in the plot when the lattice size is larger than 4×2^3 . The base network with the sign helper network trained by ES improves the result

of the pure SR method on all available cluster sizes with a nonvanishing energy gain under infinite-size extrapolation. This shows that the ES approach is, in principle, scalable to larger system sizes. Furthermore, our method also approaches the accuracy of DMRG for sizes larger than $4 \times 2^2 \times 3$.⁵ These numerical results show that the combination of the base and sign networks can improve the performance of NQS to a level comparable to other modern numerical methods in highly frustrated systems and can scale the approach to larger volumes.

Discussion. In this work, we introduced a NQS ansatz for quantum many-body spin systems that consists of a continuous part and a discrete part, the latter of which is trained with an evolutionary algorithm. The intrinsic discrete nature of quantum sign structures provides an ideal application scenario for such a derivative-free optimization method. The ES allows us to use the nondifferentiable sign network as part of a variational ansatz, which is a different numerical approach for quantum many-body problems. With all redundant phase factors eliminated and a larger search space, ES exhibits higher variational accuracy at a shorter training time. The success of the nondifferentiable sign network obtained in this Letter hints at the possible prospect of other nondifferentiable variational states, which may give more insights in the study of frustrated systems.

Furthermore, we found that a reliable network architecture to learn the sign structure is the combination of a base network and a sign helper network. The sign network, with an independent sign structure uninfluenced by the amplitude, is successful at expressing basic sign structures such as MSR. On the other hand, the base network also expresses part of the sign structure together with the amplitude, which is indispensable in some frustrated systems, as shown in our numerical experiments. Consequently, future NQS studies could focus on the combined network consisting of a base network and a sign helper network, and the ES method proposed in this Letter provides a reliable approach for optimizing the primary sign structure expressed by the sign network in various situations.

T.N., K.C., and N.A. acknowledge support from the Swiss National Science Foundation (Grant No. PP00P2_176877) and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-StG-Neupert-757867-PARATOP).

⁴The same symmetry group projector and a similar number of parameters compared with Ref. [25] are used to ensure a fair comparison between the two approaches.

⁵The DMRG energy adopted in Fig. 6 is the maximum bond dimension result. Extrapolation of DMRG will outperform NQS and mVMC.

- [1] D. Ceperley and B. Alder, Quantum Monte Carlo, *Science* **231**, 555 (1986).
- [2] M. Troyer and U.-J. Wiese, Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations, *Phys. Rev. Lett.* **94**, 170201 (2005).

- [3] F. Verstraete, V. Murg, and J. Cirac, Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems, *Adv. Phys.* **57**, 143 (2008).
- [4] U. Schollwöck, The density-matrix renormalization group in

- the age of matrix product states, *Ann. Phys. (NY)* **326**, 96 (2011).
- [5] F. Becca and S. Sorella, Variational Monte Carlo, in *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, 2017) pp. 103–130.
- [6] D. Tahara and M. Imada, Variational Monte Carlo method combined with quantum-number projection and multi-variable optimization, *J. Phys. Soc. Jpn.* **77**, 114701 (2008).
- [7] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [8] V. Dumoulin and F. Visin, A guide to convolution arithmetic for deep learning, [arXiv:1603.07285](https://arxiv.org/abs/1603.07285).
- [9] T. Cohen and M. Welling, Group equivariant convolutional networks, *Proceedings of The 33rd International Conference on Machine Learning*, edited by M. F. Balcan, and K. Q. Weinberger, Proceedings of Machine Learning Research (PMLR, New York, New York, USA, 2016), pp. 2990–2999.
- [10] Y. Nomura, Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry, *J. Phys.: Condens. Matter* **33**, 174003 (2021).
- [11] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated J_1 - J_2 model studied with neural network quantum states, *Phys. Rev. B* **100**, 125124 (2019).
- [12] C. Roth and A. H. MacDonald, Group convolutional neural networks improve quantum state accuracy, [arXiv:2104.05085](https://arxiv.org/abs/2104.05085).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] L. Yang, Z. Leng, G. Yu, A. Patel, W.-J. Hu, and H. Pu, Deep learning-enhanced variational Monte Carlo method for quantum many-body physics, *Phys. Rev. Research* **2**, 012039(R) (2020).
- [17] 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 Transactions on Parallel and Distributed Systems* **1** (2022).
- [18] Z. Cai and J. Liu, Approximating quantum many-body wave functions using artificial neural networks, *Phys. Rev. B* **97**, 035116 (2018).
- [19] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, *Phys. Rev. Research* **2**, 033075 (2020).
- [20] F. Ferrari, F. Becca, and J. Carrasquilla, Neural Gutzwiller-projected variational wave functions, *Phys. Rev. B* **100**, 125131 (2019).
- [21] 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).
- [22] 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).
- [23] X. Liang, S.-J. Dong, and L. He, Hybrid convolutional neural network and projected entangled pair states wave functions for quantum many-particle states, *Phys. Rev. B* **103**, 035138 (2021).
- [24] 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).
- [25] 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).
- [26] C. Audet and W. Hare, *Derivative-Free and Blackbox Optimization* (Springer, Gewerbestrasse 11, 6330 Cham, Switzerland, 2017).
- [27] N. Hansen, The CMA evolution strategy: A tutorial, [arXiv:1604.00772](https://arxiv.org/abs/1604.00772).
- [28] I. Hagymási, R. Schäfer, R. Moessner, and D. J. Luitz, Possible Inversion Symmetry Breaking in the $s = 1/2$ Pyrochlore Heisenberg Magnet, *Phys. Rev. Lett.* **126**, 117204 (2021).
- [29] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.4.L022026> for algorithm details and supplemental data.
- [30] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, *Nat. Phys.* **14**, 447 (2018).
- [31] S. Sorella, M. Casula, and D. Rocca, Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods, *J. Chem. Phys.* **127**, 014105 (2007).
- [32] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Pošík, Comparing results of 31 algorithms from the black-box optimization benchmarking bbob-2009, in *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation* (Association for Computing Machinery, Portland, Oregon, USA, 2010), pp. 1689–1696.
- [33] N. Hansen, Y. Akimoto, and P. Baudis, CMA-ES/pycma.
- [34] S.-S. Gong, W. Zhu, D. N. Sheng, O. I. Motrunich, and M. P. A. Fisher, Plaquette Ordered Phase and Quantum Phase Diagram in the Spin- $\frac{1}{2}$ J_1 - J_2 Square Heisenberg Model, *Phys. Rev. Lett.* **113**, 027201 (2014).
- [35] Y. Iqbal, T. Müller, P. Ghosh, M. J. P. Gingras, H. O. Jeschke, S. Rachel, J. Reuther, and R. Thomale, Quantum and Classical Phases of the Pyrochlore Heisenberg Model with Competing Interactions, *Phys. Rev. X* **9**, 011005 (2019).
- [36] W. Marshall and R. E. Peierls, Antiferromagnetism, *Proc. R. Soc. London, Ser. A* **232**, 48 (1955).
- [37] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, Solving frustrated quantum many-particle models with convolutional neural networks, *Phys. Rev. B* **98**, 104426 (2018).