Variational optimization of the amplitude of neural-network quantum many-body ground states

Jia-Qi Wang [1,2] Han-Qing Wu,[3] Rong-Qiang He [1,2,*] and Zhong-Yi Lu [1,2,†]

[1]*Department of Physics, Renmin University of China, Beijing 100872, China*
[2]*Key Laboratory of Quantum State Construction and Manipulation (Ministry of Education),*
Renmin University of China, Beijing 100872, China
[3]*Guangdong Provincial Key Laboratory of Magnetoelectric Physics and Devices, School of Physics,*
Sun Yat-sen University, Guangzhou 510275, China

Neural-network quantum states (NQSs), variationally optimized by combining traditional methods and deep learning techniques, is a new way to find quantum many-body ground states and has gradually become a competitor of traditional variational methods. However, there are still some difficulties in the optimization of NQSs, such as local minima, slow convergence, and sign structure optimization. Here, we split a quantum many-body variational wave function into a multiplication of a real-valued amplitude neural network and a sign structure, and focus on the optimization of the amplitude network while keeping the sign structure fixed. The amplitude network is a convolutional neural network (CNN) with residual blocks, namely a residual network (ResNet). Our method is tested on three typical quantum many-body systems. The obtained ground state energies are better than or comparable to those from traditional variational Monte Carlo methods and density matrix renormalization group. Surprisingly, for the frustrated Heisenberg J_1-J_2 model, our results are better than those of the complex-valued CNN in the literature, implying that the sign structure of the complex-valued NQS is difficult to optimize. We will study the optimization of the sign structure of NQSs in the future.

I. INTRODUCTION

How to obtain the ground state wave function of an interacting many-particle system has always been an important research issue in the field of quantum many-body computation, which is crucial to understanding the physical properties. However, there is no computational method that is efficient and applicable to all quantum many-body systems to date.

In recent years, with the rapid development of artificial intelligence, the related technologies have gradually been used in scientific research, and substantial achievements have been made in climate prediction [1], protein folding [2], etc. In the field of quantum many-body computation [3–5], after the seminal attempt [6] which adopted artificial neural networks as a representation of wave functions of quantum many-body systems, dubbed neural-network quantum states (NQSs), a lot of research have been done [7–20], such as Ferminet [21] for small molecular systems.

Various architectures of neural networks, such as the feedforward neural network, restricted Boltzmann machine (RBM), and convolutional neural network (CNN), have expanded the representation of the wave function of many-body systems, and their performance is comparable to that of traditional methods, such as variational Monte Carlo (VMC) methods [22–26], density matrix renormalization

group (DMRG) [27–29], and quantum Monte Carlo (QMC) methods [23,30]. This is mainly due to the improvement of algorithms and the rapid growth of GPU computing power. In the research of deep learning, the increase of network size can usually bring huge benefits, but also has expensive costs, such as large language models (LLMs) [31] with emergent abilities [32], which have aroused the interest of many researchers recently. The threshold of "large" in LLMs is 50 billion parameters. Its training is extremely expensive. However, increasing model size is just one way to improve the level of deep learning tasks. Other directions include improving the efficiency of optimization algorithms [33], improving the model architectures [34,35], and so on. In this work, the network architecture is redesigned and the corresponding parameters are carefully optimized, and then the potential of the neural network for quantum many-body wave function representation is explored.

In the variational optimization process of the ground state wave function of a two-dimensional frustrated system, it is hard to obtain a high-precision solution. One reason is that the sign structure of the ground state wave function for a frustrated quantum many-body system is very complex [36–39] and is very difficult for machine learning. In the past studies, people focused on replacing the wave function representation with a new form, namely optimizing the sign structure and amplitude of the wave function simultaneously, and expected to obtain a more accurate ground state. Here, we separate the difficulty. The wave function is split into its amplitude part and sign part. And we focus on the optimization of the amplitude, and leave the study of the sign optimization to the future.

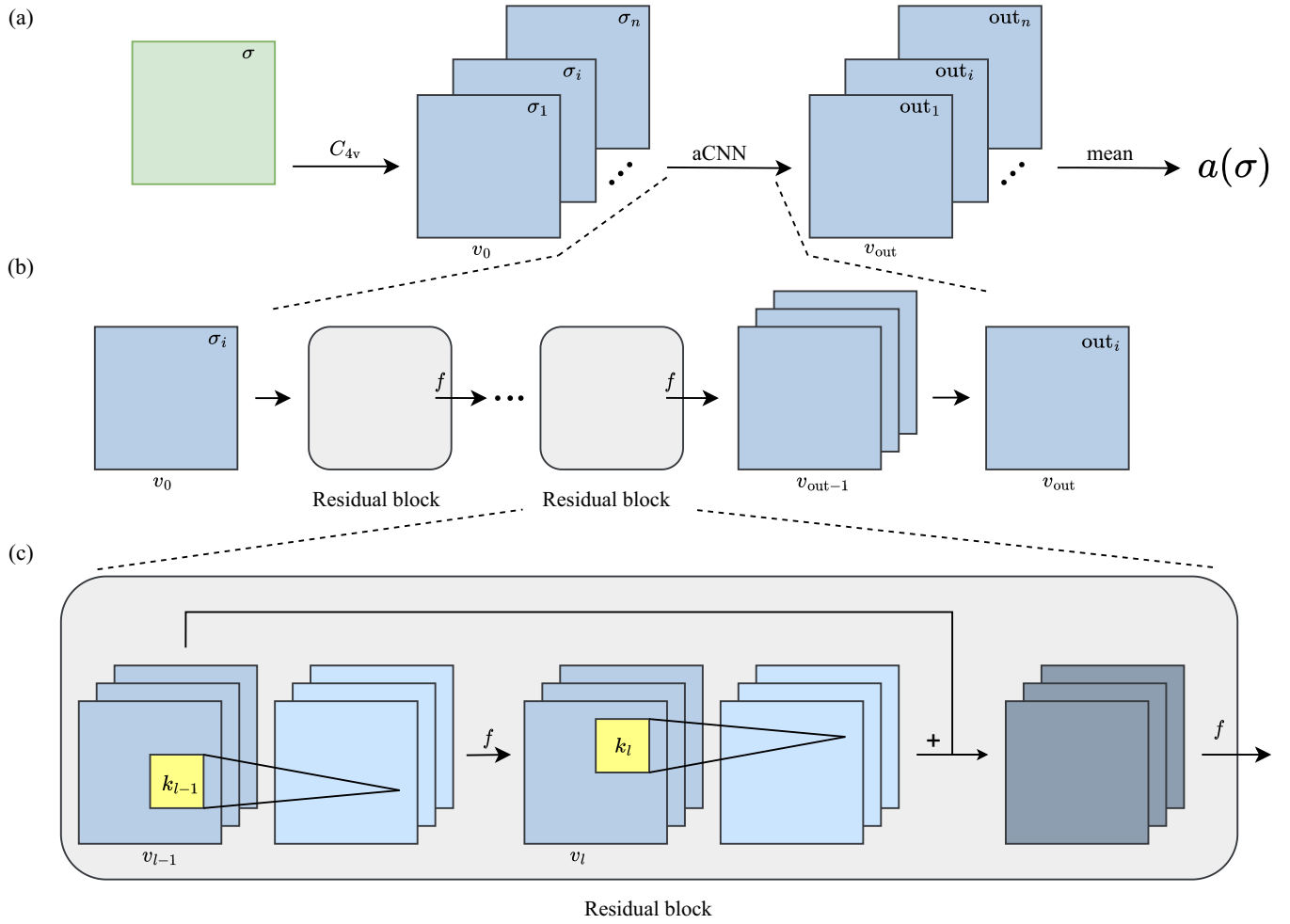*Contact author: rqhe@ruc.edu.cn
†Contact author: zlu@ruc.edu.cn

FIG. 1. The architecture of the aCNN. The input is a spin configuration σ of the system, while the output is $a(\sigma)$ [Eq. (6)]. (a) $\mathcal{C}_{4v}$ symmetrization of the aCNN as a data augmentation. v_0 and v_{out} denote the input and output layers, respectively. (b) A single spin configuration σ_i propagation in the aCNN, where f is an activation function. (c) Residual block, where k_l is the convolution kernel of layer l. The shortcut connection achieves cross-layer information transmission.

II. FRAMEWORK

A. Neural network quantum states

According to the universal approximation theorem [40,41], in principle, a neural network with sufficient depth and width can fit any function. We use deep CNN to represent the amplitude of a quantum many-body wave function. In the research of deep learning, CNN has been widely used in related fields of picture processing. Compared with network architectures such as recurrent neural networks, feedforward neural networks, and RBM, CNN has remarkable advantages in extracting spatial correlation of input information.

The trial wave function is divided into two parts: sign $s(\sigma)$ and amplitude $A(\sigma)$,

$$\psi(\sigma) = s(\sigma) \times A(\sigma), \tag{1}$$

where σ is a spin configuration of quantum spin systems. The sign part is fixed. The amplitude is represented by a deep CNN, whose architecture is shown in Fig. 1(b), which is denoted as amplitude CNN (aCNN). The input of the aCNN is σ, which is generated by Markov chain Monte Carlo sampling according to $|\psi(\sigma)|^2$. In the hidden layers of the

aCNN, a shortcut connection is used for every two layers, which is packed into a residual block. The concept of a shortcut connection was first proposed in the residual network (ResNet) [35], which helps to alleviate the problem of gradient vanishing and exploding encountered in the training process of deep neural networks, and its effectiveness has been widely proved in the research and applications of deep learning. The shortcut connection used in this work is shown in Fig. 1(c), where each residual block contains two consecutive convolution operations and the shortcut data are added before the second activation function acts. The pooling layers adopted by traditional CNNs are discarded here because the pooling operations, which resize the shape of the graph, would destroy the translational symmetry of quantum lattice systems.

The connection of the hidden layer in the aCNN is

$$v_l^{c'} = f\left[\sum_{c=1}^{3} C\left(k_{l-1}^{cc'}, v_{l-1}^{c}, b_{l-1}^{c'}\right) \right], \tag{2}$$

$$v_{l+1}^{c'} = f\left[\sum_{c=1}^{3} C\left(k_{l}^{cc'}, v_{l}^{c}, b_{l}^{c'}\right) + v_{l-1}^{c'} \right], \tag{3}$$

in which v_l^c is the matrix of layer l with input channel index c, c' is the channel index of the output layer of a convolution operation, $v_{l-1}^{c'}$ in Eq. (3) is the shortcut connection, f is the activation function, $\theta = \{k, b\}$ denotes all parameters of the NQS, k_l is convolution kernel, and b_l is the bias of layer l. Function C is the standard convolution operation. The two layers constitute a residual block, and the layer number $l = 2, 4, \ldots$.

The head and tail of the network can be written as

$$v_1^c = f\big[C\big(k_0^c, v_0, b_0^c\big)\big], \tag{4}$$

$$v_{\text{out}} = f\left[\sum_{c=1}^{3} C\big(k_{\text{out}-1}, v_{\text{out}-1}^c, b_{\text{out}-1}\big)\right], \tag{5}$$

where v_0 and v_{out} are the input and output layers, respectively. The term

$$a = \text{mean}(v_{\text{out}}) \tag{6}$$

gives the final output of the aCNN, which is the mean value of all outputs of the last convolutional layer. The operation mean is for symmetry, detailed in Sec. II C.

Then the wave function can be rewritten as

$$\psi_\theta(\sigma) = s(\sigma) \times \exp[a_\theta(\sigma)]. \tag{7}$$

Obviously, the number of the network parameters is related to the depth and width of the aCNN and the size of the convolution kernel. In this work, all calculations use the same aCNN architecture with 30 layers (14 residual blocks) unless otherwise stated. The activation function f is chosen to be ReLU, which maps the data nonlinearly. The middle layer has 3 channels, and the kernel size is 5×5. The kernel size does not limit the range of the spin correlation length. As the spin data propagate in the deep CNN, the long-range spin correlation will be gradually built. The actual amount of computation is in proportion to the number of lattice size. The number of (real-valued) parameters of the aCNN is 6538. All tests in this article were performed on an Nvidia 3090 GPU. Actually, when the size of the lattice is large, the Hilbert space of spin configuration is too large to be gone through, which means that the spin configurations involved in the training process are only a very small subset of the Hilbert space. Subsequent tests show the robustness and generalization ability of the aCNN.

B. Optimization

Theoretically, the approximate ground state of the system can be obtained by optimizing the network's parameters in terms of minimizing the expectation value of energy E_θ of a quantum system. E_θ is used as the loss function of the neural network, and is expressed as

$$\begin{aligned} E_\theta &= \frac{\langle \psi_\theta(\sigma)|H|\psi_\theta(\sigma)\rangle}{\langle \psi_\theta(\sigma)|\psi_\theta(\sigma)\rangle} \\ &= \langle E_{\text{loc}}(\sigma, \theta)\rangle_{p(\sigma,\theta)}, \end{aligned} \tag{8}$$

where $E_{\text{loc}}(\sigma, \theta) = \frac{\langle \sigma|H|\psi_\theta\rangle}{\langle \sigma|\psi_\theta\rangle}$ is the local energy of the system for configuration σ, $p(\sigma, \theta) = \frac{|\psi_\theta(\sigma)|^2}{\sum_\sigma |\psi_\theta(\sigma)|^2}$ is the probability distribution, and $\langle \cdot \rangle_p$ denotes the expectation with distribution p.

The derivative $\nabla_\theta E_\theta$ guides the direction of the optimization. The update of a network's parameters can be expressed as $\delta\theta \propto -\lambda\nabla_\theta E_\theta$, where λ is the learning rate. Here, we adopt the idea of the theory of automatic differentiable Monte Carlo [42] and represent the energy in the following form:

$$\text{Es} = \frac{\left\langle \frac{\psi_\theta^2(\sigma)}{\perp(\psi_\theta^2(\sigma))} \perp (E_{\text{loc}}(\sigma, \theta))\right\rangle_{\perp(p(\sigma,\theta))}}{\left\langle \frac{\psi_\theta^2(\sigma)}{\perp(\psi_\theta^2(\sigma))}\right\rangle_{\perp(p(\sigma,\theta))}}, \tag{9}$$

where ψ denotes $\psi_\theta(\sigma)$. $\perp (x)$ is the detach function, which is defined as $\perp (x) = x$ in forward propagation, and $\frac{\partial \perp (x)}{\partial x} = 0$ in backward propagation. Numerically, the values and first derivative of Eq. (8) and Eq. (9) are equal:

$$\perp (\langle E_{\text{loc}}(\sigma, \theta)\rangle_{p(\sigma,\theta)}) = \perp (\text{Es}), \tag{10}$$

$$\perp (\nabla_\theta\langle E_{\text{loc}}(\sigma, \theta)\rangle_{p(\sigma,\theta)}) = \perp (\nabla_\theta\text{Es}). \tag{11}$$

The form of expectation Eq. (9) can be encoded into the neural network automatic differentiation framework, which avoids taking the derivative of $p(\sigma, \theta)$ and $E_{\text{loc}}(\sigma, \theta)$.

In principle, an effective optimization algorithm is the stochastic reconfiguration (SR) [43] method. However, the SR requires the inverse of a large matrix of the variational parameters, which limits the number of the parameters and hence the expression ability of the neural network. Here, we only discuss the optimization of amplitude, and choose to use the popular Adam algorithm [44] in deep learning, an algorithm with adaptive learning rate and momentum, which performs well in most tasks.

In the test, the amplitude $a_\theta(\sigma)$ is calculated by the aCNN with real-valued θ, which is initialized randomly by the Kaiming initialization [45], and the sign structure is fixed. The amplitude network is optimized until the loss function no longer drops.

C. Symmetry

The models simulated in this work are subject to symmetries $\mathcal{C}_{4v}$ and translation, unless otherwise stated. Among them, because the scanning of the convolutional kernel on each layer is weight sharing, translation symmetry is an intrinsic property of convolutional connections. Rotation and mirror symmetries can be achieved through data augmentation ($\mathcal{C}_{4v}$ group operations), as shown in Fig. 1(a). After forward propagation, the corresponding results from all symmetric operations are averaged to preserve the symmetries. The symmetrized wave function is

$$\psi_\theta(\sigma) = s(\sigma) \times \exp\left[\frac{1}{n}\sum_{i=1}^{n} a\big(\hat{c}_{4v}^{(i)}\sigma, \theta\big)\right], \tag{12}$$

where $\hat{c}_{4v}^{(i)}$ denotes a group-element operation of the $\mathcal{C}_{4v}$ group, and n is the order of the group.

In addition, the ground states of the Heisenberg model and Heisenberg J_1-J_2 model both have a conserved quantity $S_z = 0$. This can be achieved by hard encoding, which keeps the numbers of up and down spins equal in the sampling process of spin configurations.

Symmetry imposition helps to accelerate the decline of the loss function. In the case of limited computing resources,

TABLE I. Ground state energies of the Heisenberg model with $L = 10$, 16, and 24. Comparison of the results of the aCNN with VMC and QMC; the data are excerpted from Ref. [23]. The relative error is $1 - E_{aCNN}/E_{QMC}$.

Method	$L = 10$	$L = 16$	$L = 24$
VMC		−0.66965	−0.66922
QMC	−0.67156	−0.66998	−0.66961
aCNN($\mathcal{C}_{4v}$)	−0.671431(2)	−0.669716(1)	−0.669237(2)
Relative error	0.019%	0.039%	0.056%

although $\mathcal{C}_{4v}$ symmetry achieved by data augmentation will limit the size of configuration batch allowed in a single training step, data augmentation itself does not limit the number of the network's parameters. Reasonable data augmentation is beneficial to the numerical stability of the optimization process. The accuracy benefits of additional symmetry through data augmentation will be shown in the results below.

III. NUMERICAL RESULTS AND DISCUSSION

A. Heisenberg model

In condensed matter physics, as one of the prototypical quantum many-body systems, the two-dimensional Heisenberg model is a testing ground for various analytical and numerical methods for strongly correlated systems, whose Hamiltonian contains only the nearest-neighbor interactions of the spins, which is unfrustrated. Its ground state is doubtlessly Néel antiferromagnetically ordered. The existing method [23] is very accurate in simulating the ground state of this model, and allows the lattice size to be quite large, providing prefect data for benchmark tests.

The Hamiltonian of the Heisenberg model is

$$H_{HM} = \sum_{\langle ij \rangle} \hat{S}_i \cdot \hat{S}_j, \tag{13}$$

where $\hat{S}_i$ is the spin operator on site i, and $\langle ij \rangle$ denotes the nearest-neighbor sites i and j.

In this part, we calculate $L \times L$ lattices with L up to 24 under periodic boundary conditions and compare the results with those obtained by VMC and QMC. The sign structure is fixed to the Marshall sign rule [46]. As shown in Table I, the variational optimization results of the aCNN are better than the results of VMC, but there is still a very small gap compared with QMC. The last row of Table I is the relative error of the energy of the aCNN with respect to that from QMC. The results show that in the direct variational optimization process, the trial wave functions represented by neural networks are better than those of the traditional methods, which may be because of the powerful expression ability of neural networks and the advancement of optimization algorithms.

In addition, the spin structure factor

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

which is related to observable quantities, can be calculated readily, where $\vec{k}$ is a wave vector in **k** space. $\vec{r}_i$ is the location of site i in real space and N is the number of lattice sites. As
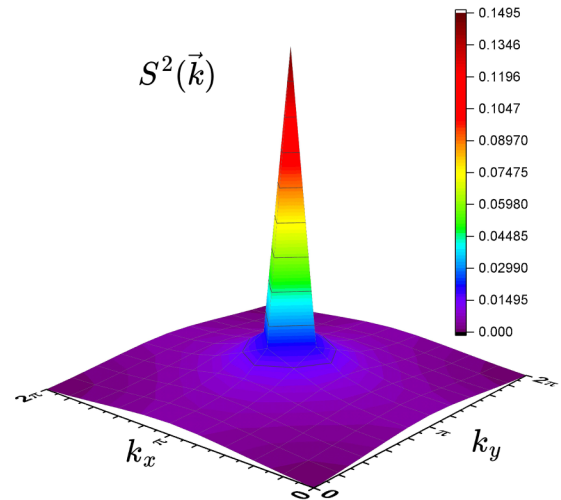

FIG. 2. Spin structure factor $S^2(\vec{k})$ as defined in Eq. (14). The **k**-space distribution of $S^2(\vec{k})$ of 10×10 square lattice of the Heisenberg model, which peaks at $\vec{k} = (\pi, \pi)$. k_x and k_y are the two components of $\vec{k}$.

shown in Fig. 2, $S^2(\vec{k})$ is peaked at $\vec{k} = (\pi, \pi)$, indicating the ground state is antiferromagnetically ordered.

B. Transverse field Ising model

We also test the aCNN for the two-dimensional transverse field Ising model (TFIM) on the square lattice with $L = 8$ and 16. The Hamiltonian is

$$H_{TFIM} = -\sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z - \lambda \sum_i \sigma_i^x, \tag{15}$$

where σ_i^z and σ_i^x are the Pauli operators on site i, and λ is the transverse field.

All tests are simulations with transverse field $\lambda = 3.05$ (close to the critical point $\lambda = 3.04438(2)$ [47] of the model), in periodic boundary conditions. The exact sign structure has been employed, namely, positive definite signs. As shown in Table II, the results of the aCNN show excellent agreement with the extrapolated results from the fully augmented tree tensor network (FATTN; $L = 8$) [48] and fully-augmented matrix product state (FAMPS; $L = 16$) [49]. This test shows that the aCNN works well near the critical point.

TABLE II. Ground state energies of the two-dimensional transverse field Ising model with $L = 8$ and 16. Comparison of the results of the aCNN and the extrapolated results of FATTN [48] and FAMPS [49].

Method	$L = 8$	$L = 16$
aCNN($\mathcal{C}_{4v}$)	−3.241642(2)	−3.239568(2)
FATTN(extrapolated)		−3.23956(2)
FAMPS(extrapolated)	−3.24165	

TABLE III. Ground state energies of the Heisenberg J_1-J_2 model with $L = 10$. Comparison of the results of the aCNN with the complex-valued CNN [50], VMC [24], DMRG [27], RBM+PP [20], and Deep ViT [51] with $J_2/J_1 = 0.4, 0.45, 0.5$, and 0.55. Here, p represents the number of Lanczos steps, and $p = 0$ corresponds to the original variational wave function. VMC($p = \infty$) denotes the variance extrapolated results. In DMRG(4096), 4096 indicates the number of SU(2) states. DMRG(∞) is obtained through straight-line energy extrapolation with DMRG truncation error.

Method	$J_2/J_1 = 0.4$	$J_2/J_1 = 0.45$	$J_2/J_1 = 0.5$	$J_2/J_1 = 0.55$
VMC($p = 0$)	−0.52188(1)	−0.50811(1)	−0.49521(1)	−0.48335(1)
VMC($p = 1$)	−0.52368(1)	−0.50973(1)	−0.49718(1)	−0.48622(1)
VMC($p = 2$)	−0.5240(1)	−0.51001(1)	−0.49755(1)	−0.48693(3)
VMC($p = \infty$)	−0.52429(2)	−0.51017(2)	−0.49781(2)	−0.48766(6)
DMRG(4096)	−0.521487	−0.507193	−0.495044	−0.484890
DMRG(6144)	−0.522043	−0.507677	−0.495301	−0.485239
DMRG(8192)	−0.522391	−0.507976	−0.495530	−0.485434
DMRG(∞)	−0.5253	−0.5110	−0.4988	−0.4880
RBM+PP			−0.497629(1)	
Deep ViT			−0.497634(1)	
Complex-valued CNN	−0.52371(1)	−0.50905(1)	−0.49516(1)	−0.48277(1)
aCNN($\mathcal{C}_{4v}$)	−0.523999(4)	−0.509207(5)	−0.495627(6)	−0.483490(5)
aCNN($\mathcal{C}_4$)	−0.523872(4)			−0.483173(5)

C. Heisenberg J_1-J_2 model

The above two quantum systems have no frustration, and the sign structures of their exact ground states are known, on which our method has a good performance. Here we try our method on a typical frustrated quantum spin system, the Heisenberg J_1-J_2 model. The sign structure of its exact ground states is unknown, and even its ground state properties are in debate. Quantum fluctuation and geometric frustration are key factors. Several magnetic long-range orders compete in its most frustrated region, where a quantum spin liquid may emerge.

The Hamiltonian of the Heisenberg J_1-J_2 model is

$$H_{\text{J1J2}} = J_1 \sum_{\langle ij \rangle} \hat{S}_i \cdot \hat{S}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \hat{S}_i \cdot \hat{S}_j, \qquad (16)$$

where J_1 and J_2 are the coupling strengths of nearest-neighbor sites and next-nearest-neighbor sites (denoted as $\langle\langle ij \rangle\rangle$), respectively. The microscopic details of the model, controlled by the ratio of J_2/J_1, play a decisive role in the type of the ground state and its quantum phase transitions. Here, the model on the square lattice with $L = 10$ has been simulated and $J_2/J_1 = 0.4, 0.45, 0.5$, and 0.55. The results of the aCNN, the complex-valued CNN [50], VMC [24], DMRG [27], RBM+PP [a method using RBM combined with pair-product (PP) states] [20], and the deep vision transformer (Deep ViT) [51] are shown in Table III. The sign structure is fixed to the Marshall sign rule, which is not exact when $J_2/J_1 \neq 0$.

As shown in Table III, the energy obtained by the aCNN is better than that obtained by VMC($p = 0$) and especially by the complex-valued CNN. This shows that the wave function of the aCNN has advantages, and in direct variational optimization, NQS optimized by the deep learning algorithm is a better form than the traditional trial wave function used by VMC. Compared with the complex-valued CNN, the results show that it is difficult to optimize both sign and amplitude for the complex-valued CNN.

Since the mirror reflection symmetry is not used in the simulation with the complex-valued CNN, the test using $\mathcal{C}_4$ symmetry only is also done in this work to eliminate the influence of different symmetries. As an example, the Heisenberg J_1-J_2 model ($L = 10$) with $J_2 = 0.4$ and 0.55 has been simulated. The energy obtained in this work is significantly better than that of the complex-valued CNN, as shown in Table III. The complex-valued CNN uses SR to optimize the network, which in principle can optimize the sign structure and amplitude of the wave function simultaneously. The comparison results show that it seems that the sign structure is not well optimized in the complex-valued CNN since its energy is even worse than that from the aCNN with an incorrect sign structure.

Comparing the network architectures of the aCNN and the complex-valued CNN, the latter network is wide and shallow while the former is narrow and deep relatively, and the former has residual connections additionally. The different architectures have a certain impact on their expression ability. However, it is important to note that the numbers of their parameters are almost equal. The complex-valued CNN contains 3838 complex-valued parameters, while the aCNN contains 6538 real-valued parameters.

The results of VMC and DMRG are listed in Table III. The energies of the aCNN are even better than those of VMC($p = 1$) (with $p = 1$ Lanczos correction), and comparable to that of VMC($p = 2$) when $J_2/J_1 = 0.4$. As the frustration increases, the advantage of the aCNN gradually decreases. When the frustration is the strongest at $J_2/J_1 = 0.55$, the performance of the aCNN deteriorates. This implies that the Marshall sign rule is significantly broken by frustration here. The ground state behavior of the system is most controversial near the strongest frustration point $J_2/J_1 = 0.55$. Because the combined effect of frustration and quantum fluctuation destroys the long-range antiferromagnetic order, the ground state behaves as nonmagnetic. However, the nature of this quantum phase remains controversial. The above results illustrate the importance of sign structure optimization, which will be studied in the future.

In addition, comparing the results of the aCNN($\mathcal{C}_{4v}$) with that of the aCNN($\mathcal{C}_4$) in Table III, it is clear whether the use
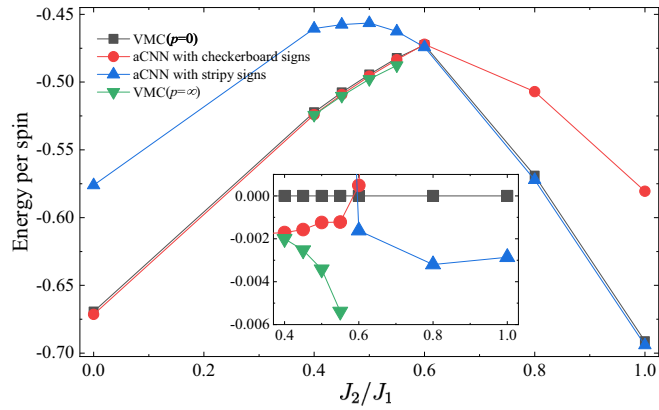
FIG. 3. Ground state energies of the Heisenberg J_1-J_2 model with $L = 10$. The aCNN is optimized with different Marshall sign structures, the checkerboard-patterned (exact for $J_2 = 0$) and the stripe-patterned (exact for $J_1 = 0$). The VMC($p = 0$) results are from Ref. [50]. VMC($p = \infty$) denotes the variance extrapolated results from Ref. [24], which can be regarded as the exact ground state energies. The inset shows $E - E(\text{VMC}(p = 0))$, the relative energies with respect to the VMC($p = 0$) energies, for an enlarged view.

of mirror reflection symmetry is important for the simulation. This shows again that symmetry imposition helps.

D. Applicability of sign structures and sources of error

In the calculations of the Heisenberg J_1-J_2 model discussed earlier, the sign structure is fixed to the Marshall sign rule (checkerboard-patterned), which is not exact when $J_2 \neq 0$. It is essential to understand the applicability of sign structures and to figure out the sources of error in the aCNN.

Here, the Heisenberg J_1-J_2 model ($L = 10$) is simulated with different Marshall sign structures, the checkerboard-patterned (exact for $J_2 = 0$) and the stripe-patterned (exact for $J_1 = 0$). The results are depicted in Fig. 3, together with the results of VMC($p = 0$) (provided by F. Ferrari and F. Becca in Ref. [50]) and VMC($p = \infty$) [24]. VMC($p = \infty$) denotes the variance extrapolated results, which can be regarded as the exact ground state energies.

As depicted in Fig. 3, the ground state energies obtained from the aCNN consistently outperform those from VMC($p = 0$), and are close to the results of VMC($p = \infty$) at $J_2/J_1 = 0.4$, 0.45, 0.5, and 0.55, which indicates that even though the frustration is strong ($J_2/J_1 = 0.5$ and 0.55), the sign structures fixed according to the Marshall sign rule remain highly close to the exact ground state sign structures. This is consistent with the estimation from Refs. [15,52]. In addition, when $J_2/J_1 < 0.6$, the energies obtained by the aCNN with the stripy signs are significantly higher than those from VMC($p = 0$), and when $J_2/J_1 > 0.6$, the energies obtained by the aCNN with the checkerboard signs are significantly higher than those from VMC($p = 0$). This demonstrates that when the sign structure employed by the aCNN diverges significantly from that of the exact ground state, the errors in the results will become very large, underscoring the importance of a proper sign structure in obtaining accurate results.

TABLE IV. Ground state energies of the Heisenberg J_1-J_2 model with $J_2/J_1 = 0.5$ on a 6×6 square lattice and the Heisenberg model on a triangular lattice (HMTL) of 36 sites. The aCNN is optimized with the exact sign structure from ED. Results of LCN are from Ref. [53], and GNN/GNN-2 are from Ref. [54]. The column P/M indicates the number of parameters (P) in millions (M). The relative error is $1 - E_{\text{aCNN}}/E_{\text{ED}}$.

Method	J_1-J_2	P/M	HMTL	P/M
ED	-0.503810		-0.560374	
GNN	$-0.5022(4)$	0.86	$-0.55892(2)$	0.86
GNN-2	$-0.5023(5)$		$-0.55949(2)$	
LCN	$-0.5022(2)$	0.28	$-0.5601(4)$	0.29
aCNN	$-0.503258(4)$	0.0065	$-0.559214(6)$	0.022
Rel. error	0.110%		0.207%	

To pinpoint the actual source of error in the aCNN, we conduct simulations of the Heisenberg J_1-J_2 model with $J_2/J_1 = 0.5$ on a 6×6 square lattice, and the Heisenberg model on a triangular lattice (HMTL) of 36 sites, whose exact ground states are available through exact diagonalization (ED) methods. Here, the aCNN is optimized with the fixed sign structures of the exact ground states.

These two models are both strongly frustrated. In particular, the geometric frustration inherent in the triangular lattice makes the amplitude more complex. This is a challenge for the expressive power of neural networks. To improve the accuracy of the result, a deep neural network with $\mathcal{C}_{6v}$ symmetry, consisting of 100 layers, is employed in the simulation for the HMTL.

The results of the aCNN are shown in Table IV, together with the results of the lattice convolutional network (LCN) [53] and graph neural network (GNN) [54] for comparison. As evident, the aCNN achieves precision levels comparable to those of GNN, GNN-2, and LCN, yet with significantly lesser parameters than each of them. Comparing the relative errors listed in Table I and Table IV, the relative errors for frustrated systems are higher than those for unfrustrated systems. This implies that the presence of strong frustration would result in heightened complexity of the amplitude of the ground state wave function, and deeper neural networks are necessary for such systems.

In this paper, the aCNN is optimized with fixed sign structure. Sometimes there are known sign structures, such as the Heisenberg model and the transverse field Ising model. At other times, approximate ground state wave functions and their sign structures can be obtained through alternative methods, such as mean-field theories, low-parameter VMC [24], or tensor network states [55]. In such cases, using the aCNN could achieve more accurate ground state wave functions.

Furthermore, the ground state wave functions of the models calculated in this paper are all real-valued. When the Hamiltonian contains complex numbers, or when the wave function evolves in real time, the system's wave function will be complex-valued with continuous phases. In those cases, the aCNN needs to be used with a phase network, which is worth further research in the future.

IV. SUMMARY AND OUTLOOK

In summary, we have proposed a quantum many-body wave function, named aCNN, which is a multiplication of a real-valued amplitude neural network (a deep CNN with residual blocks) and a fixed sign structure. The ground state energy has been obtained by variational optimization of the aCNN with the Adam algorithm. Our method is tested on typical quantum many-body systems. The obtained ground state energies are better than or comparable to those from traditional VMC methods, DMRG, and the complex-valued CNN.

The results provide specific evidence to support the viewpoint that the deep aCNN can be a competitive wave function ansatz for describing the amplitude of the ground state wave functions and properties of quantum many-body systems compared with traditional VMC methods. Importantly, this work shows that the aCNN can obtain better energies than the complex-valued CNN with the same number of variational parameters, which implies that it is not easy to optimize both the sign structure and amplitude simultaneously. And it supports the viewpoint that it is worth trying to use different tools to represent the sign structure and amplitude, respectively, suggested also by Ref. [38]. Here we show that CNN is a good choice for the amplitude part.

In the long run, it is worth continuing to explore the use of more advanced artificial intelligence technologies to optimize the ground state wave function of quantum many-body systems, especially quantum frustrated systems and strongly correlated fermionic systems with a sign problem. Particularly, it is worthwhile to explore a way to optimizing the sign structure with NQSs in the future.

The main code and data for the aCNN are available [56].

[1] R. Lam, A. Sanchez-Gonzalez, M. Willson, P. Wirnsberger, M. Fortunato, F. Alet, S. Ravuri, T. Ewalds, Z. Eaton-Rosen, W. Hu, A. Merose, S. Hoyer, G. Holland, O. Vinyals, J. Stott, A. Pritzel, S. Mohamed, and P. Battaglia, GraphCast: Learning skillful medium-range global weather forecasting, Science **382**, 1416 (2023).

[2] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back *et al.*, Highly accurate protein structure prediction with Alphafold, Nature (London) **596**, 583 (2021).

[3] M. Randeria, Introduction to many-body physics, Phys. Today **70**(5), 59 (2017).

[4] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, Science **377**, eabk3333 (2022).

[5] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[6] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[7] Y. Nomura, Investigating network parameters in neural-network quantum states, J. Phys. Soc. Jpn. **91**, 054709 (2022).

[8] W. Zhang, X. Xu, Z. Wu, V. Balachandran, and D. Poletti, Ground state search by local and sequential updates of neural network quantum states, Phys. Rev. B **107**, 165149 (2023).

[9] C. Roth, A. Szabó, and A. H. MacDonald, High-accuracy variational Monte Carlo for frustrated magnets with deep neural networks, Phys. Rev. B **108**, 054410 (2023).

[10] L. L. Viteritti, R. Rende, and F. Becca, Transformer variational wave functions for frustrated quantum spin systems, Phys. Rev. Lett. **130**, 236401 (2023).

[11] Y.-H. Zhang and M. Di Ventra, Transformer quantum state: A multipurpose model for quantum many-body problems, Phys. Rev. B **107**, 075147 (2023).

[12] M. Reh, M. Schmitt, and M. Gärttner, Optimizing design choices for neural quantum states, Phys. Rev. B **107**, 195115 (2023).

[13] 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).

[14] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, Phys. Rev. Res. **2**, 023358 (2020).

[15] A. Szabó and C. Castelnovo, Neural network wave functions and the sign problem, Phys. Rev. Res. **2**, 033075 (2020).

[16] 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).

[17] S.-H. Lin and F. Pollmann, Scaling of neural-network quantum states for time evolution, Phys. Status Solidi B **259**, 2100172 (2022).

[18] A. Chen and M. Heyl, Efficient optimization of deep neural quantum states toward machine precision, arXiv:2302.01941.

[19] 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).

[20] 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).

[21] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks, Phys. Rev. Res. **2**, 033429 (2020).

[22] J. Lou and A. W. Sandvik, Variational ground states of two-dimensional antiferromagnets in the valence bond basis, Phys. Rev. B **76**, 104432 (2007).

[23] A. W. Sandvik and H. G. Evertz, Loop updates for variational and projector quantum Monte Carlo simulations in the valence-bond basis, Phys. Rev. B **82**, 024407 (2010).

[24] W.-J. Hu, F. Becca, A. Parola, and S. Sorella, Direct evidence for a gapless Z_2 spin liquid by frustrating Néel antiferromagnetism, Phys. Rev. B **88**, 060402(R) (2013).

[25] A. W. Sandvik and G. Vidal, Variational quantum Monte Carlo simulations with tensor-network states, Phys. Rev. Lett. **99**, 220602 (2007).

[26] L. Tagliacozzo, G. Evenbly, and G. Vidal, Simulation of two-dimensional quantum systems using a tree tensor network that exploits the entropic area law, Phys. Rev. B **80**, 235127 (2009).

[27] 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).

[28] E. Stoudenmire and S. R. White, Studying two-dimensional systems with the density matrix renormalization group, Annu. Rev. Condens. Matter Phys. **3**, 111 (2012).

[29] L. Wang and A. W. Sandvik, Critical level crossings and gapless spin liquid in the square-lattice spin-1/2 J_1-J_2 Heisenberg antiferromagnet, Phys. Rev. Lett. **121**, 107202 (2018).

[30] A. W. Sandvik, Finite-size scaling of the ground-state parameters of the two-dimensional Heisenberg model, Phys. Rev. B **56**, 11678 (1997).

[31] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu *et al.*, A survey of large language models, arXiv:2303.18223.

[32] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, Emergent abilities of large language models, arXiv:2206.07682.

[33] S. Ruder, An overview of gradient descent optimization algorithms, arXiv:1609.04747.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. **30**, 5998 (2017).

[35] 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 (CVPR)* (IEEE, New York, 2016), pp. 770–778.

[36] C.-Y. Park and M. J. Kastoryano, Expressive power of complex-valued restricted Boltzmann machines for solving nonstoquastic Hamiltonians, Phys. Rev. B **106**, 134437 (2022).

[37] A. Chen, K. Choo, N. Astrakhantsev, and T. Neupert, Neural network evolution strategy for solving quantum sign structures, Phys. Rev. Res. **4**, L022026 (2022).

[38] T. Westerhout, M. I. Katsnelson, and A. A. Bagrov, Many-body quantum sign structures as non-glassy Ising models, Commun. Phys. **6**, 275 (2023).

[39] 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).

[40] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[41] G. V. Cybenko, Approximation by superpositions of a sigmoidal function, Math. Control Signals Syst. **2**, 303 (1989).

[42] S.-X. Zhang, Z.-Q. Wan, and H. Yao, Automatic differentiable Monte Carlo: Theory and application, Phys. Rev. Res. **5**, 033041 (2023).

[43] 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**, 48 (2007).

[44] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[45] K. He, X. Zhang, S. Ren, and J. Sun, Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification, in *Proceedings of the IEEE International Conference on Computer Vision* (IEEE, Piscataway, 2015), pp. 1026–1034.

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

[47] H. W. J. Blöte and Y. Deng, Cluster Monte Carlo simulation of the transverse Ising model, Phys. Rev. E **66**, 066110 (2002).

[48] X. Qian and M. Qin, From tree tensor network to multiscale entanglement renormalization ansatz, Phys. Rev. B **105**, 205102 (2022).

[49] X. Qian and M. Qin, Augmenting density matrix renormalization group with disentanglers, Chin. Phys. Lett. **40**, 057102 (2023).

[50] 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).

[51] R. Rende, L. L. Viteritti, L. Bardone, F. Becca, and S. Goldt, A simple linear algebra identity to optimize large-scale neural network quantum states, arXiv:2310.05715.

[52] A. Voigt, J. Richter, and N. B. Ivanov, Marshall-Peierls sign rule for excited states of the frustrated J_1-J_2 Heisenberg antiferromagnet, Physica A **245**, 269 (1997).

[53] C. Fu, X. Zhang, H. Zhang, H. Ling, S. Xu, and S. Ji, Lattice convolutional networks for learning ground states of quantum many-body systems, in *Proceedings of the 2024 SIAM International Conference on Data Mining (SDM)* (SIAM, 2024), pp. 490–498.

[54] 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.

[55] 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).

[56] The main code and data for the aCNN are available on GitHub at https://github.com/rqHe1/aCNN.