Quantum-Enhanced Neural Networks for Quantum Many-Body Simulations

Zongkang Zhang,* Ying Li, and Xiaosi Xu†
Graduate School of China Academy of Engineering Physics, Beijing 100193, China

(Received 30 April 2025; revised 10 December 2025; accepted 6 May 2026; published 28 July 2026)

Neural quantum states (NQS) have gained prominence in variational quantum Monte Carlo methods in approximating ground-state wavefunctions. Despite their success, they face limitations in optimization, scalability, and expressivity in addressing certain problems. In this work, we propose a quantum-neural hybrid framework that combines parametrized quantum circuits with neural networks to model quantum many-body wavefunctions. This approach combines the efficient sampling and optimization capabilities of autoregressive neural networks with the enhanced expressivity provided by quantum circuits. Numerical simulations demonstrate the scalability and accuracy of the hybrid ansatz in spin systems and quantum chemistry problems. Our results reveal that the hybrid method achieves notably lower relative energy compared to stand-alone NQS. These findings underscore the potential of quantum-neural hybrid methods for tackling challenging problems in quantum many-body simulations.

DOI: 10.1103/2jpn-jh3x

I. INTRODUCTION

The Schrödinger equation lies at the heart of understanding quantum many-body systems across fields such as condensed matter physics, quantum chemistry, and materials science. However, solving it exactly is limited by the exponential growth of the Hilbert space with system size, prompting the development of approximate methods. Among these, variational approaches have proven effective [1–3], though their accuracy largely depends on the ability to represent the wavefunction accurately [4].

Artificial neural networks have emerged as powerful tools for addressing complex problems. The neural quantum state (NQS) [5–10] framework uses a neural network as a variational ansatz to approximate the wavefunction of a quantum system, often more efficiently than traditional approaches like tensor network states [11,12], matrix product states [13,14], and Jastrow-Slater wavefunctions [8]. Despite its promise, the NQS approach is not without limitations. Optimization of neural networks often encounters barren plateaus and local minima [15–17]. Furthermore, scaling to larger systems entails significant computational costs [18]. Additionally, the classical nature of neural networks imposes constraints on their expressivity, limiting their ability to accurately capture the quantum correlations present in many-body wavefunctions [18–23].

In recent years, quantum computing has experienced rapid advancements, presenting new opportunities for addressing certain classically hard problems. One of the principal potential applications of quantum computing is solving quantum many-body problems [24–29]. However, the capabilities of current noisy intermediate-scale quantum devices are restricted to small-scale problems due to hardware limitations and the absence of fault-tolerant error correction [30,31]. Nonetheless, quantum states encoded on such devices inherently capture quantum correlations more efficiently than classical representations, offering potential advantages when integrated with classical approaches.

The interplay between quantum computing and neural networks has demonstrated significant potential in a variety of fields, including classification [32], image processing [33], and optimization [34]. Hybrid methods often outperform their classical counterparts in specific tasks, showcasing a unique advantage [35]. Moreover, numerous studies have illustrated how classical neural networks can assist quantum computing by optimizing quantum circuits [36,37], performing postprocessing [38,39], mitigating noise [10], learning quantum data [40–43], or improving sampling techniques [44,45].

However, relatively few efforts have focused on the reciprocal relationship—how quantum computing can enhance the capabilities of neural networks in solving challenging quantum many-body problems. Notably, quantum machine learning (QML) has demonstrated provable advantages in expressivity, learnability, and trainability, while exhibiting reduced requirements for memory, samples, and parameters [46–59]. In this context, Ref. [60] employs a quantum algorithm to sample the Gibbs distribution determined by an NQS, Ref. [61] introduces a hybrid embedding scheme simulating a subspace with quantum circuits, and Ref. [62] designs a quantum-assisted sampling protocol for NQS training.

In this work, we explore a different angle and propose a hybrid framework that directly integrates parametrized quantum

*Contact author: zongkang.zhang.phys@gmail.com
†Contact author: xsxu@gscaep.ac.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

circuits (PQCs) with neural networks to model the wavefunction of quantum many-body systems. We combine the quantum circuit with a classical wavefunction following the method in Ref. [63]. This approach allows for joint optimization of parameters within both the quantum circuit and the neural network, providing the flexibility to adopt sequential or simultaneous optimization strategies tailored to specific problems. Our results demonstrate that this hybrid ansatz not only accelerates convergence but also achieves a notably lower relative error compared to purely classical neural networks, highlighting the potential of quantum-enhanced neural quantum states for overcoming the limitations of existing methods.

II. HYBRID QUANTUM-NEURAL WAVEFUNCTION

The overview of our hybrid framework is illustrated in Fig. 1. The total variational wavefunction $|\Psi\rangle$ is modeled as a combination of quantum and classical components, serving as the ansatz in the variational quantum Monte Carlo (VMC) method. The complex amplitude of a configuration s is defined as

$$\langle s|\Psi\rangle = \langle s|\phi(\theta)\rangle \times \langle s|\varphi(\lambda)\rangle, \tag{1}$$

where $\phi(\theta)$ and $\varphi(\lambda)$ represent the quantum and classical wavefunctions, respectively. Here, $s = (s_1, s_2, \ldots, s_{N_q})$ is a binary sequence, where $s_i = 0$ (1) represents spin-up (spin-down) in quantum spin systems or unoccupied (occupied) sites in electronic structure problems.

For the quantum computing part, we have

$$\ln \langle s|\phi(\theta)\rangle = f[s; U_1(\theta_1)] + if[s; U_2(\theta_2)], \tag{2}$$

where the PQCs $U_1(\theta_1)$ and $U_2(\theta_2)$ are used to generate the amplitude and phase, respectively. Define $f[s; U(\theta)] = \sum_{i=1}^{Nq} c_i \langle s|U^\dagger(\theta) Z_i U(\theta)|s\rangle$, where N_q is the number of qubits and c_i denotes the coefficient associated with the ith qubit. To estimate $f[s; U(\theta)]$, we apply $U(\theta)$ to the initial state $|s\rangle$ and measure all the qubits in the Z basis. All the expectations $\langle Z_i\rangle$ can be measured simultaneously, and the process is repeated for sufficient circuit shots M. Then, $f[s; U(\theta)]$ is the weighted sum of these expectations. Note that except the gate parameters θ, the coefficients c_i are also learnable. In the context of QML, such a model is also referred to as a quantum-neural network (QNN) [48,54]. It is shown that QNNs can offer advantages over classical neural networks in expressive power and training ability [49,50,58]. Recently, quantum generative models have been demonstrated to efficiently learn and sample from classically intractable distributions [47,51].

The classical neural network part is expressed as

$$\langle s|\varphi(\lambda)\rangle = \sqrt{p(s;\lambda_1)}\, e^{i\gamma(s;\lambda_2)}, \tag{3}$$

where the probability $p(s;\lambda_1)$ is the probability distribution generated by a transformer neural network [10,65], and the phase $\gamma(s;\lambda_2)$ is obtained separately from an FNN. Transformer networks can efficiently capture complex qubit correlations and enable fast, parallel sampling. In contrast, nonautoregressive models rely on sequential Markov chain generation with algorithms like Metropolis-Hastings, requiring burn-in and thinning that reduce accuracy.

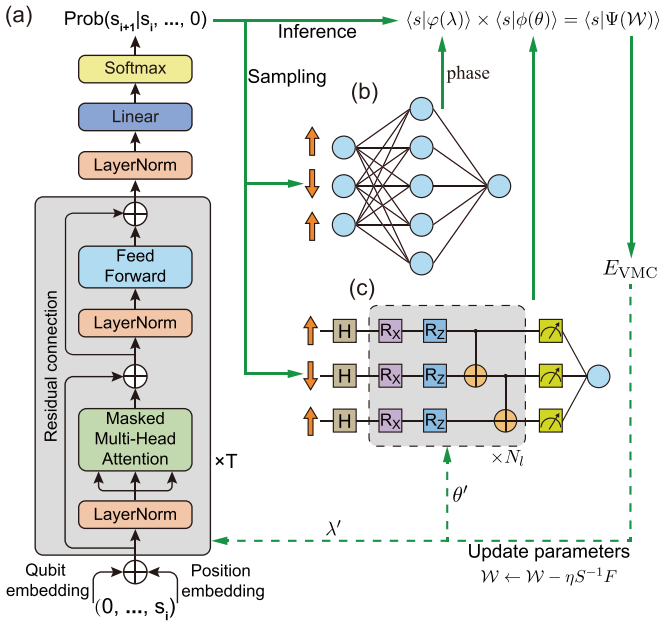

FIG. 1. Overview and workflow of the quantum-enhanced neural networks. (a) The architecture of the transformer neural network. (b) The sketch of the feedforward neural network (FNN). (c) The parametrized quantum circuit, where the rotation gates are defined as $R_X(\theta) = e^{-i\theta X/2}$ and $R_Z(\theta) = e^{-i\theta Z/2}$, with θ representing the variational parameter to be optimized. Our algorithm works as follows: First, the transformer neural network samples a batch of configurations. Then, the neural networks infer, and the PQCs measure, the amplitude and phase for a given configuration. Next, the variational ground-state energy, as well as the gradient vector and the quantum Fisher matrix, is estimated. Finally, the parameters inside the PQCs and neural networks are optimized based on the stochastic reconfiguration method or the Adam optimizer. This process is repeated until convergence or reaching the maximum number of iterations (details on the architecture of transformer and the optimization process can be found in Supplemental Material [64]).

III. VMC WITH TRANSFORMER-GUIDED IMPORTANCE SAMPLING

The VMC method minimizes the Rayleigh quotient of a given Hamiltonian H by optimizing a parametrized wavefunction $|\Psi(\mathcal{W})\rangle$ [66], where the complete parameter set is denoted as $\mathcal{W} = \{\lambda, \theta, c\}$ in our case. Given access to the complex amplitude $\langle s|\Psi(\mathcal{W})\rangle$ for arbitrary configuration s, the ground-state energy E_g can be estimated as $E_{\mathrm{VMC}} = \mathrm{Re}\langle E_{\mathrm{loc}}(s)\rangle$, where $\langle\bullet\rangle = \mathbb{E}_{s\sim\mathbb{P}(s)}[\bullet]$, $\mathbb{P}(s) = |\langle s|\Psi\rangle|^2/\langle\Psi|\Psi\rangle$, and the local energy $E_{\mathrm{loc}}(s) = \sum_{s'}\langle s|H|s'\rangle\langle s'|\Psi\rangle/\langle s|\Psi\rangle$. Here, s is sampled according to the probability distribution $\mathbb{P}(s)$ (see Supplemental Material [64] for details on VMC).

For the hybrid quantum-neural wavefunction defined in Eq. (1), the configuration must now be sampled from the probability distribution $\mathbb{P}(s) = |\langle s|\phi\rangle|^2|\langle s|\varphi\rangle|^2/\langle\Psi|\Psi\rangle$. This process requires performing Markov chain Monte Carlo (MCMC) sampling, which necessitates that the quantum computer estimates $|\langle s|\phi\rangle|^2$ for configuration s that may ultimately be discarded during the sampling process. To address this inefficiency, we employ importance sampling as a workaround.

The original sampling probability is decomposed as

$$\mathbb{P}(s) = p(s)\omega(s), \tag{4}$$

where $p(s) = |\langle s|\varphi\rangle|^2 / \sum_{s'} |\langle s'|\varphi\rangle|^2$ and the importance weight $\omega(s) = |\langle s|\phi\rangle|^2 / \mathbb{E}_{s\sim p(s)}[|\langle s|\phi\rangle|^2]$. Instead of sampling based on $\mathbb{P}(s)$, we can perform importance sampling using $p(s)$, with $\omega(s)$ serving as a modification factor. The transformer neural network can learn the sampling probability $p(s;\lambda)$ and directly generate a batch of configurations $\{s_{(b)}\}_{b=1}^{B}$, where B is the sample size. Besides transformers, other autoregressive neural networks also offer more efficient sampling than that of MCMC. Then, $\omega(s)$ can be estimated on a quantum computer as

$$\omega(s_{(b)}) = \frac{|\langle s_{(b)}|\phi\rangle|^2}{\frac{1}{B}\sum_{b=1}^{B} |\langle s_{(b)}|\phi\rangle|^2}. \tag{5}$$

The variational energy becomes

$$E^H_{\mathrm{VMC}} = \mathrm{Re}\langle \omega(s) E_{\mathrm{loc}}(s)\rangle, \tag{6}$$

where $\langle \bullet \rangle = \mathbb{E}_{s\sim p(s)}[\bullet]$ and $E_{\mathrm{loc}}(s) = \sum_{s'} \langle s|H|s'\rangle \langle s'|\phi\rangle \langle s'|\varphi\rangle / (\langle s|\phi\rangle \langle s|\varphi\rangle)$. We can estimate E^H_{VMC} by computing the average of $\omega(s)E_{\mathrm{loc}}(s)$:

$$\hat{E}^H_{\mathrm{VMC}} = \frac{1}{B}\mathrm{Re}\left[\sum_{b=1}^{B} \omega(s_{(b)}) E_{\mathrm{loc}}(s_{(b)})\right]. \tag{7}$$

When solving molecules, we further add a penalty term to prevent the transformer from only sampling the dominant configurations (see Supplemental Material [64]).

During the optimization process, we can jointly optimize parameters in both the quantum circuit and the neural network, or adopt a sequential scheme in which only the parameters of one component (the neural network or the quantum circuit) are optimized at a time. This reduces the number of parameters involved in each optimization step and improves training stability. Therefore, the hybrid quantum-neural wavefunction can achieve better performance by integrating the strengths of quantum computing in capturing complex correlations and machine learning in resource utilization, computational precision, and robustness to noise.

IV. SAMPLING ERROR CONTROL

In importance sampling, the weight factor affects the variance of the estimator. Large variations in the weight factor can lead to increased fluctuations in the results. From Eq. (7), the variance of the energy is

$$\mathrm{Var}\left(E^H_{\mathrm{VMC}}\right) = \frac{1}{B}\mathrm{Var}[\omega(s)\mathrm{Re}(E_{\mathrm{loc}}(s))]. \tag{8}$$

Our algorithm benefits from the property of vanishing variance in VMC as the wavefunction ansatz [Eq. (1)] converges to the eigenstate of H. However, if the difference between the target distribution $\mathbb{P}(s)$ and auxiliary distribution $p(s)$ is large, the importance weights may become highly unstable, leading to inaccurate estimates.

To reduce sampling error, the range of the weight factor $\omega(s)$ can be controlled using two proposed methods. The first approach involves initializing the coefficients c_i and quantum circuit parameters θ_i to values close to zero. Under this condition, $f[s;U(\theta)] \approx 0$ for arbitrary s ensuring that $\omega(s)$ starts near 1. We observed that this method performed well in numerical simulations.

The second method employs the tanh activation function for control. Specifically, the amplitude $|\langle s|\phi\rangle|$ can be changed from $e^{f[s;U_1(\theta_1)]}$ to $e^{a\tanh(f[s;U_1(\theta_1)])}$, where a is a rescaling factor. This ensures that the amplitude remains within the range $[e^{-a}, e^{a}]$. Consequently, according to Eq. (5), even in the extreme condition, $\omega(s)$ is less than e^{4a}. We note that in practical scenarios, $\omega(s)$ is much smaller than this loose upper bound.

V. SAMPLING SPACE REDUCTION

When a quantum system exhibits certain symmetries, the physically valid Hilbert space can be reduced. By modifying the architecture of the transformer, we can constrain the sampling space to explore, while preserving its autoregressive property.

For chemical molecules or lattice models such as the Fermi-Hubbard model, the number of electrons N_e and the total electronic spin S are typically conserved. Let $N_\uparrow$ and $N_\downarrow$ denote the number of spin-up and spin-down electrons, respectively. These quantities are conserved, with $N_\uparrow = N_e/2 + S$ and $N_\downarrow = N_e/2 - S$. There are several fermion-to-qubit mappings available; in this work, we adopt the Jordan-Wigner (JW) transformation. The JW transformation converts the second-quantized electronic Hamiltonian into a qubit Hamiltonian. Specifically, it maps the fermionic indices $j\uparrow$ and $j\downarrow$ to the qubit indices $2j-1$ and $2j$, respectively, where $j = 1, \ldots, N_O$ and N_O is the number of spatial orbitals.

Based on the conservation of $N_\uparrow$ and $N_\downarrow$, we can mask the transformer's output probability to avoid invalid outputs. As in standard autoregressive models, we prepend a virtual start-of-sequence token $s_0 = 0$ so that the transformer can predict the first physical qubit; hence, the conditional inputs take the form $(s_i, \ldots, s_1, 0)$. When i is even, the masked probability that the next qubit is in the $|1\rangle$ state—i.e., the spin orbital $(i/2+1)\uparrow$ is occupied—is

$$p'(s_{i+1} = 1|s_i, \ldots, s_1, 0) = p \times \Theta\left(N_\uparrow - \sum_{k=1}^{i/2} s_{2k-1}\right), \tag{9}$$

where p is the original probability and $\Theta(x)$ denotes the Heaviside step function, defined as $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ for $x \leqslant 0$. Conversely, the masked probability that the next qubit is in the $|0\rangle$ state is

$$\begin{aligned} &p'(s_{i+1} = 0|s_i, \ldots, s_1, 0) \\ &\quad = p \times \Theta\left[(N_O - N_\uparrow) - \sum_{k=1}^{i/2}(1 - s_{2k-1})\right]. \end{aligned} \tag{10}$$

It is similar when i is odd. After this step, the masked probabilities should further be normalized by dividing $p'(s_{i+1}=1)$ and $p'(s_{i+1}=0)$ by the normalization factor $p'(s_{i+1}=1) + p'(s_{i+1}=0)$. By considering symmetries, the dimension of the sample space is reduced from 2^{2N_O} to $\binom{N_O}{N_\uparrow}\binom{N_O}{N_\downarrow}$.

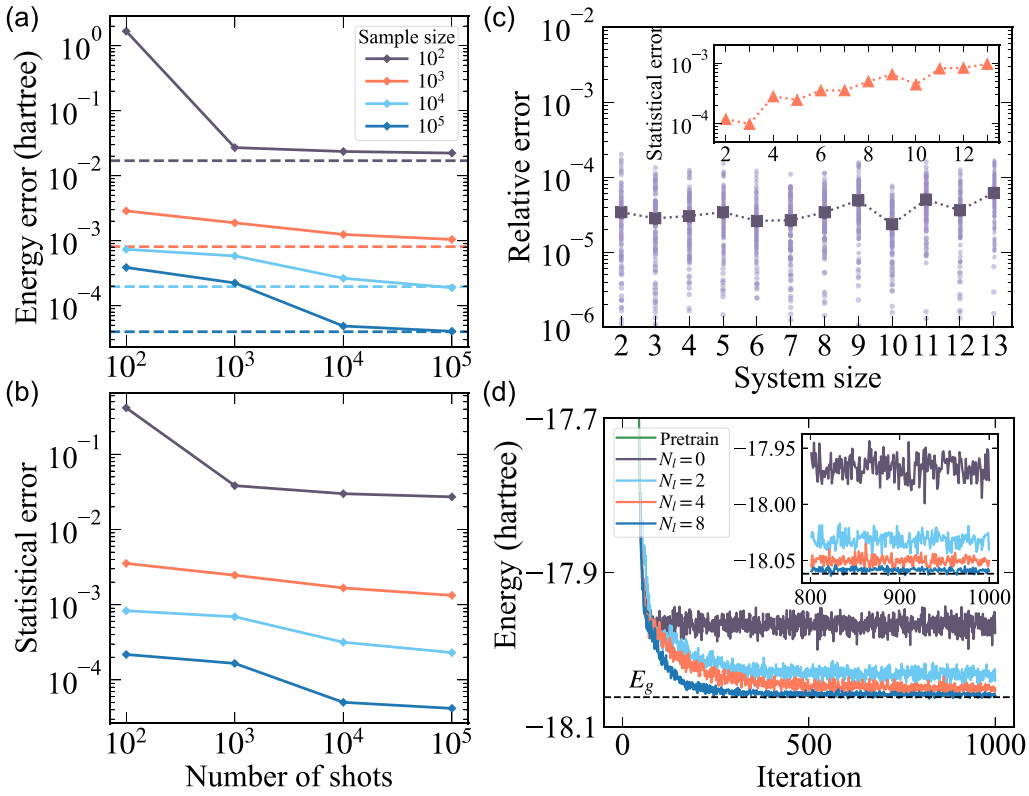

FIG. 2. Performance of the hybrid quantum-neural algorithm for an AFH chain with PBC. (a), (b) Energy and statistical errors for a ten-spin system using various sample sizes and shot numbers. Dashed lines indicate energy estimates without shot noise. (c) Relative energy errors $|(E^H_{\mathrm{VMC}} - E_g)/E_g|$ and statistical errors for 2–13 spins, with sample size and shot number fixed at 10^4. For panels (a)–(c), errors are averaged over 100 random seeds; statistical error is the standard deviation. (d) Sequential optimization with quantum circuits of different depths for a ten-spin system. The neural network is pretrained to provide an initial approximation of the energy, followed by optimization of quantum circuits with different layer counts to refine the results. The sample size is 10^4, and measurements are noiseless. The transformer uses $d = 8$, $h = 4$, $T = 2$, and the FNN has two hidden layers, each with 32 neurons. The quantum circuits use four layers with circular entanglement in panels (a) and (b) and eight layers with linear entanglement in panels (c) and (d) (see Supplemental Material [64] for details).

VI. RESULTS

In this section, we present numerical simulations of our algorithm. We address the problem of finding the ground state for two classes of physical systems: spin models and chemical molecules. A representative spin model is the antiferromagnetic Heisenberg (AFH) model, with its Hamiltonian given by

$$H_{\mathrm{AFH}} = J \sum_{\langle i,j \rangle} \left(\sigma_i^X \sigma_j^X + \sigma_i^Y \sigma_j^Y + \sigma_i^Z \sigma_j^Z\right), \tag{11}$$

where J is the interaction strength and σ_i^X, σ_i^Y, and σ_i^Z are Pauli matrices at the ith spin. For our numerical simulations, we consider a one-dimensional AFH model with periodic boundary conditions (PBCs) and set $J = 1$. The total spin of its ground state is 0 (1/2) for even (odd) spins, which also largely reduces the sampling space.

The second-quantized electronic Hamiltonian of chemical molecules is given as [28]

$$H_e = \sum_{ij} h_{ij} c_i^\dagger c_j + \frac{1}{2} \sum_{ijkl} h_{ijkl} c_i^\dagger c_j^\dagger c_k c_l, \tag{12}$$

where $c_j^\dagger$ and c_j are the fermionic creation and annihilation operators, respectively. The coefficients h_{ij} and h_{ijkl} represent the one-body and two-body integrals, respectively. The fermionic operators are mapped to Pauli operators by the JW transformation.

In our algorithm, statistical errors arise from both Monte Carlo sampling and the finite number of shots in quantum circuits, affecting final accuracy. Using the error control methods, we can systematically reduce statistical errors. In Figs. 2(a) and 2(b), we show the energy and statistical errors of our algorithm for various sample sizes B and shot numbers M. As the sample size and shot number increase, the accuracy of the results improves, while the variance decreases accordingly. The total measurement cost of quantum circuits scales as $\mathcal{O}(BM)$. In our simulations, good performance is observed for $B, M = \mathcal{O}(10^3$–$10^4)$, and results converge to shot-noise-free simulations when $M = \mathcal{O}(10^5)$.

Next, we evaluate the scalability of the algorithm by computing the ground states of an AFH chain with 2–13 spins. Except for the increasing input dimension, the hyperparameters of the neural network and quantum circuit stay unchanged. Figure 2(c) shows the relative energy and statistical errors of the results for different system sizes. As the system size increases, the final relative errors are around 3.6×10^{-5} and show only a slight increase, demonstrating good scalability.

By employing sequential optimization, we can fix the neural network parameters and investigate the expressive power of quantum circuits with varying numbers of layers. The optimization process can be divided into two stages. We first optimize the NQS described by Eq. (3) to obtain a good approximation of the ground state, and then fix its parameters and refine the ansatz with a PQC. Figure 2(d) presents the results for various quantum circuit layers. When $N_l = 0$, the quantum circuit contains solely of a single layer of rotation gates, resulting in almost no improvement. As N_l increases from 2 to 8, the energy error continuously decreases, reaching approximately 10^{-3}. This trend demonstrates that deeper circuits improves expressivity. Furthermore, as N_l increases, fluctuations significantly decrease, as the ansatz better approximates the ground state and benefits from the vanishing-variance property of VMC.

To illustrate how quantum computing enhances NQS expressivity, we compare our hybrid quantum-neural method with stand-alone NQS. We take the nitrogen N_2 molecule in the STO-3G basis set as an example and reduce the number of spin orbitals from 20 to 12 using the one-particle reduced density matrix. Our results demonstrate that when the neural network is limited—either by architecture or by a limited number of parameters—a PQC can effectively improve performance. Figure 3(a) presents the energy curves for N_2 at 1.5 Å under varying hyperparameter settings. An NQS with $d = 14$ achieves an average energy error of 7.1×10^{-3} hartree, which exceeds the "golden standard" coupled cluster singles and doubles with perturbative triples (CCSD(T)) yet does not reach chemical accuracy threshold (1.6×10^{-3} hartree). However, adding a ten-layer PQC, the hybrid method surpasses the chemical accuracy after 2.5×10^3 iterations and reaches an average error of 5.3×10^{-4} hartree. A 20-layer PQC can further reduce the error to 2.0×10^{-4} hartree. For comparison, a larger NQS—a transformer with $d = 16$—achieves 2.5×10^{-3} hartree, which approaches chemical accuracy but still falls short of the hybrid method. Furthermore, replacing the PQC with an FNN containing more parameters yields a similar trend. These findings demonstrate that the hybrid method outperforms stand-alone NQS in both convergence speed and accuracy, despite using far fewer parameters. Similar trends hold for the AFH model and LiH molecule (see Supplemental Material [64]).

Figure 3(b) compares the performance of the NQS and our hybrid method in calculating the potential energy surface of N_2 with interatomic distances from 0.8 to 2.1 Å. The NQS achieves an energy error smaller than that of CCSD(T), but fails to reach chemical accuracy at large interatomic distances, whereas adding a shallow PQC reduces the error by about two orders of magnitude, well within chemical accuracy. Here, to highlight the contribution of quantum circuits, we intentionally used neural networks with a constrained number of parameters.

Although larger networks can improve expressivity, it comes at the cost of increased overhead in inference, sampling, and optimization subroutines. Moreover, NQS have been found to struggle with learning the intricate sign (or phase) structure of strongly interacting spin [19,21,22,90] and fermionic systems [23,91]. Although the Marshall sign rule and backflow transformations can alleviate this issue [88,92], this necessitates NQS architectures tailored to specific Hamiltonians, which remains largely empirical. In contrast, PQCs are inherently capable of capturing sign structures [93], thereby mitigating the sign problem in quantum Monte Carlo methods [94] and offering a valuable complement to classical trial states.

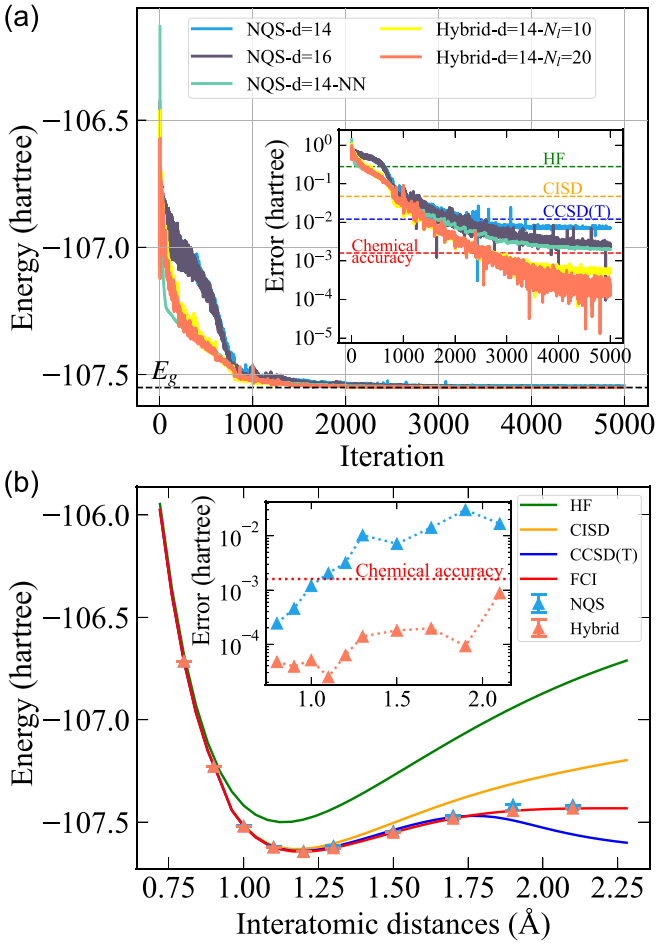

FIG. 3. Performance of the NQS and the hybrid quantum-neural method for a 12-qubit N_2 molecule. The blue curve represents the performance of the NQS with embedding dimension $d = 14$, while the yellow and orange curves show the results of our hybrid method with a 10- or 20-layer quantum circuit, respectively. The green curve replaces the PQCs with FNNs of two hidden layers with 40 neurons. The gray curve represents a larger NQS model, where the transformer uses $d = 16$, $h = 4$ heads, and $T = 3$ blocks, and the FNN uses two hidden layers of 96 neurons each (see Supplemental Material [64] for detailed hyperparameters). (a) Energy convergence curve at an interatomic distance of 1.5 Å. (b) Potential energy surface. For each interatomic distance, energy is averaged over the final 100 iterations; error bars show standard deviation. The inset in panels (a) and (b) displays the energy error (log scale), with chemical accuracy shown by a red dashed line. Results from classical methods such as Hartree-Fock (HF), configuration interaction singles and doubles (CISD), CCSD(T), and full configuration interaction (FCI) are also drawn.

VII. CONCLUSION AND OUTLOOK

In this work, we introduce a hybrid quantum-neural method for solving quantum many-body problems in discrete systems, such as spin models and second-quantized chemical

molecules. The approach is built on a hybrid quantum-neural wavefunction, integrating the wavefunction of a transformer-based neural quantum state with a parametrized quantum circuit. By employing an importance sampling technique that uses the transformer as a direct sampler, the method improves efficiency and addresses the limitations of traditional MCMC sampling. To further enhance performance, we use two variance reduction strategies: (1) constraining the wavefunction amplitude using activation functions and (2) leveraging problem-specific symmetries to refine the transformer's sampling probabilities, effectively reducing the sampling space and improving accuracy.

Numerical simulations demonstrate that adding a PQC significantly enhances the NQS's expressivity, achieving higher accuracy with fewer parameters. Additionally, demonstrations show systematic error decrease with increased sample size, quantum circuit depth, and the number of circuit shots. The method also exhibits favorable scalability, with errors increasing modestly as system size grows. It is worth noting that, since the computational overhead for simulating quantum circuits is exponential, numerical simulations are typically restricted to systems of modest size. In principle, the hybrid method can scale naturally on quantum computers, thereby addressing the scalability limitations inherent in NQS.

This work opens several avenues for future exploration. While we employ a general hardware-efficient PQC in this study, designing problem-specific PQCs tailored to particular systems could yield larger benefits. Similarly, expanding beyond transformers and feedforward neural networks to explore more advanced and deeper NQS architectures may extend this approach to larger and more complex physical systems, where the hybrid quantum-neural method could provide greater advantages. Additionally, investigating the noise resilience of this method on quantum hardware will be important for realizing the full potential of this approach in practical applications.

ACKNOWLEDGMENTS

We implement the neural networks in PyTorch [95], construct the quantum circuits with PennyLane [96], and generate the molecular Hamiltonians using OpenFermion [97,98]. The authors acknowledge discussions with Ji Chen. This work was supported by Innovation Program for Quantum Science and Technology (Grant No. 2023ZD0300200), National Natural Science Foundation of China (Grants No. 12225507 and No. 12088101), and NSAF (Grants No. U1930403, No. U2330201, and No. U2330401).

DATA AVAILABILITY

The data and code that support the findings of this article are openly available at GitHub [99].

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

[2] A. Roggero, A. Mukherjee, and F. Pederiva, Quantum Monte Carlo with coupled-cluster wave functions, Phys. Rev. B **88**, 115138 (2013).

[3] A. Mahajan and S. Sharma, Symmetry-projected Jastrow mean-field wave function in variational Monte Carlo, J. Phys. Chem. A **123**, 3911 (2019).

[4] D. Wu, R. Rossi, F. Vicentini, N. Astrakhantsev, F. Becca, X. Cao, J. Carrasquilla, F. Ferrari, A. Georges, M. Hibat-Allah, *et al.*, Variational benchmarks for quantum many-body problems, Science **386**, 296 (2024).

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

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

[7] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, Phys. Rev. Lett. **122**, 226401 (2019).

[8] 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).

[9] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, Nat. Chem. **12**, 891 (2020).

[10] E. R. Bennewitz, F. Hopfmueller, B. Kulchytskyy, J. Carrasquilla, and P. Ronagh, Neural error mitigation of near-term quantum simulations, Nat. Mach. Intell. **4**, 618 (2022).

[11] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Neural-network quantum states, string-bond states, and chiral topological states, Phys. Rev. X **8**, 011006 (2018).

[12] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, Equivalence of restricted Boltzmann machines and tensor network states, Phys. Rev. B **97**, 085104 (2018).

[13] O. Sharir, A. Shashua, and G. Carleo, Neural tensor contractions and the expressive power of deep neural quantum states, Phys. Rev. B **106**, 205136 (2022).

[14] D. Wu, R. Rossi, F. Vicentini, and G. Carleo, From tensor-network quantum states to tensorial recurrent neural networks, Phys. Rev. Res. **5**, L032001 (2023).

[15] A. Nielsen, *Neural Networks and Deep Learning* (Determination Press, San Francisco, CA, 2015).

[16] I. Goodfellow, *Deep Learning* (MIT Press, Cambridge, MA, 2016).

[17] H. Lange, A. V. de Walle, A. Abedinnia, and A. Bohrdt, From architectures to applications: A review of neural quantum states, Quantum Sci. Technol. **9**, 040501 (2024).

[18] J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W. M. C. Foulkes, D. Pfau, G. Carleo, and F. Noé, *Ab initio* quantum chemistry with neural-network wavefunctions, Nat. Rev. Chem. **7**, 692 (2023).

[19] Z. Cai and J. Liu, Approximating quantum many-body wave functions using artificial neural networks, Phys. Rev. B **97**, 035116 (2018).

[20] 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).

[21] 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).

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

[23] G. Passetti, D. Hofmann, P. Neitemeier, L. Grunwald, M. A. Sentef, and D. M. Kennes, Can neural quantum states learn volume-law ground states? Phys. Rev. Lett. **131**, 036502 (2023).

[24] A. Y. Kitaev, Quantum measurements and the Abelian stabilizer problem, arXiv:quant-ph/9511026.

[25] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[26] G. H. Low and I. L. Chuang, Optimal Hamiltonian simulation by quantum signal processing, Phys. Rev. Lett. **118**, 010501 (2017).

[27] L. Lin and Y. Tong, Near-optimal ground state preparation, Quantum **4**, 372 (2020).

[28] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[29] Y. Alexeev, M. Amsler, M. A. Barroca, S. Bassini, T. Battelle, D. Camps, D. Casanova, Y. J. Choi, F. T. Chong, C. Chung, *et al.*, Quantum-centric supercomputing for materials science: A perspective on challenges and future directions, Future Gener. Comput. Syst. **160**, 666 (2024).

[30] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[31] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[32] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Supervised learning with quantum-enhanced feature spaces, Nature (London) **567**, 209 (2019).

[33] X.-W. Yao, H. Wang, Z. Liao, M.-C. Chen, J. Pan, J. Li, K. Zhang, X. Lin, Z. Wang, Z. Luo, W. Zheng, J. Li, M. Zhao, X. Peng, and D. Suter, Quantum image processing and its application to edge detection: Theory and experiment, Phys. Rev. X **7**, 031041 (2017).

[34] A. Abbas, A. Ambainis, B. Augustino, A. Bärtschi, H. Buhrman, C. Coffrin, G. Cortiana, V. Dunjko, D. J. Egger, B. G. Elmegreen, *et al.*, Challenges and opportunities in quantum optimization, Nat. Rev. Phys. **6**, 718 (2024).

[35] A. Callison and N. Chancellor, Hybrid quantum-classical algorithms in the noisy intermediate-scale quantum era and beyond, Phys. Rev. A **106**, 010101 (2022).

[36] S.-X. Zhang, C.-Y. Hsieh, S. Zhang, and H. Yao, Differentiable quantum architecture search, Quantum Sci. Technol. **7**, 045023 (2022).

[37] K. Nakaji, L. B. Kristensen, J. A. Campos-Gonzalez-Angulo, M. G. Vakili, H. Huang, M. Bagherimehrab, C. Gorgulla, F. Wong, A. McCaskey, J.-S. Kim, *et al.*, The generative quantum eigensolver (GQE) and its application for ground state search, arXiv:2401.09253.

[38] S.-X. Zhang, Z.-Q. Wan, C.-K. Lee, C.-Y. Hsieh, S. Zhang, and H. Yao, Variational quantum-neural hybrid eigensolver, Phys. Rev. Lett. **128**, 120502 (2022).

[39] S.-X. Zhang, J. Miao, and C.-Y. Hsieh, Variational post-selection for ground states and thermal states simulation, Quantum Sci. Technol. **10**, 015028 (2025).

[40] S. Czischek, M. S. Moss, M. Radzihovsky, E. Merali, and R. G. Melko, Data-enhanced variational Monte Carlo simulations for Rydberg atom arrays, Phys. Rev. B **105**, 205108 (2022).

[41] 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).

[42] M. S. Moss, S. Ebadi, T. T. Wang, G. Semeghini, A. Bohrdt, M. D. Lukin, and R. G. Melko, Enhancing variational Monte Carlo simulations using a programmable quantum simulator, Phys. Rev. A **109**, 032410 (2024).

[43] H. Lange, G. Bornet, G. Emperauger, C. Chen, T. Lahaye, S. Kienle, A. Browaeys, and A. Bohrdt, Transformer neural networks and quantum simulators: A hybrid approach for simulating strongly correlated systems, Quantum **9**, 1675 (2025).

[44] G. Torlai, G. Mazzola, G. Carleo, and A. Mezzacapo, Precise measurement of quantum observables with neural-network estimators, Phys. Rev. Res. **2**, 022060(R) (2020).

[45] Y. Yang, Z. Zhang, A. Wang, X. Xu, X. Wang, and Y. Li, Maximizing quantum-computing expressive power through randomized circuits, Phys. Rev. Res. **6**, 023098 (2024).

[46] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[47] X. Gao, Z.-Y. Zhang, and L.-M. Duan, A quantum machine learning algorithm based on generative models, Sci. Adv. **4**, eaat9004 (2018).

[48] M. Cerezo, G. Verdon, H.-Y. Huang, L. Cincio, and P. J. Coles, Challenges and opportunities in quantum machine learning, Nat. Comput. Sci. **2**, 567 (2022).

[49] Y. Liu, S. Arunachalam, and K. Temme, A rigorous and robust quantum speed-up in supervised machine learning, Nat. Phys. **17**, 1013 (2021).

[50] Y. Du, M.-H. Hsieh, T. Liu, and D. Tao, Expressive power of parametrized quantum circuits, Phys. Rev. Res. **2**, 033125 (2020).

[51] H.-Y. Huang, M. Broughton, N. Eassa, H. Neven, R. Babbush, and J. R. McClean, Generative quantum advantage for classical and quantum problems, arXiv:2509.09033.

[52] H.-Y. Huang, R. Kueng, and J. Preskill, Information-theoretic bounds on quantum advantage in machine learning, Phys. Rev. Lett. **126**, 190505 (2021).

[53] E. R. Anschuetz, H.-Y. Hu, J.-L. Huang, and X. Gao, Interpretable quantum advantage in neural sequence learning, PRX Quantum **4**, 020338 (2023).

[54] Y. Du, M.-H. Hsieh, T. Liu, S. You, and D. Tao, Learnability of quantum neural networks, PRX Quantum **2**, 040337 (2021).

[55] Z. Yu, Q. Chen, Y. Jiao, Y. Li, X. Lu, X. Wang, and J. Yang, Non-asymptotic approximation error bounds of parameterized quantum circuits, Adv. Neural Inf. Process. Syst. **37**, 99089 (2024).

[56] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, *et al.*, Quantum advantage in learning from experiments, Science **376**, 1182 (2022).

[57] V. Dunjko, J. M. Taylor, and H. J. Briegel, Quantum-enhanced machine learning, Phys. Rev. Lett. **117**, 130501 (2016).
[58] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, The power of quantum neural networks, Nat. Comput. Sci. **1**, 403 (2021).
[59] R. Sweke, J.-P. Seifert, D. Hangleiter, and J. Eisert, On the quantum versus classical learnability of discrete distributions, Quantum **5**, 417 (2021).
[60] R. Xia and S. Kais, Quantum machine learning for electronic structure calculations, Nat. Commun. **9**, 4195 (2018).
[61] S. Barison, F. Vicentini, and G. Carleo, Variational embeddings for many body quantum systems, arXiv:2309.08666.
[62] M. Sajjan, V. Singh, and S. Kais, Polynomially efficient quantum enabled variational Monte Carlo for training neural-network quantum states for physico-chemical applications, npj Quantum Inf. (2026), doi:10.1038/s41534-026-01230-1.
[63] F. Metz, G. Pescia, and G. Carleo, Simulating continuous-space systems with quantum-classical wave functions, arXiv:2409.06415.
[64] See Supplemental Material at http://link.aps.org/supplemental/10.1103/2jpn-jh3x for methodological and numerical details, which includes Refs. [5,10,19,28,65,67–89].
[65] L. L. Viteritti, R. Rende, and F. Becca, Transformer variational wave functions for frustrated quantum spin systems, Phys. Rev. Lett. **130**, 236401 (2023).
[66] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).
[67] 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).
[68] T. D. Barrett, A. Malyshev, and A. Lvovsky, Autoregressive neural-network wavefunctions for *ab initio* quantum chemistry, Nat. Mach. Intell. **4**, 351 (2022).
[69] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets, Nature (London) **549**, 242 (2017).
[70] S. Jerbi, L. J. Fiderer, H. P. Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, Quantum machine learning beyond kernel methods, Nat. Commun. **14**, 517 (2023).
[71] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for *ab-initio* electronic structure, Nat. Commun. **11**, 2368 (2020).
[72] D. R. Hartree, The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods, in *Mathematical Proceedings of the Cambridge Philosophical Society* (Cambridge University Press, Cambridge, 1928), Vol. 24, pp. 89–110.
[73] J. C. Slater, A simplification of the Hartree-Fock method, Phys. Rev. **81**, 385 (1951).
[74] R. Jastrow, Many-body problem with strong forces, Phys. Rev. **98**, 1479 (1955).
[75] D. M. Ceperley and B. J. Alder, Ground state of the electron gas by a stochastic method, Phys. Rev. Lett. **45**, 566 (1980).
[76] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* (John Wiley & Sons, Hoboken, New Jersey, 2013).
[77] G. D. Purvis, III and R. J. Bartlett, A full coupled-cluster singles and doubles model: The inclusion of disconnected triples, J. Chem. Phys. **76**, 1910 (1982).
[78] 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).
[79] 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).
[80] S. Sorella, Green function Monte Carlo with stochastic reconfiguration, Phys. Rev. Lett. **80**, 4558 (1998).
[81] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference on Learning Representations, San Diego, CA, May 7–9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun (ICLR, 2015).
[82] C.-Y. Park and M. J. Kastoryano, Geometry of learning neural quantum states, Phys. Rev. Res. **2**, 023232 (2020).
[83] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, Nature (London) **323**, 533 (1986).
[84] J. Li, X. Yang, X. Peng, and C.-P. Sun, Hybrid quantum-classical approach to quantum optimal control, Phys. Rev. Lett. **118**, 150503 (2017).
[85] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98**, 032309 (2018).
[86] H. Lange, F. Döschl, J. Carrasquilla, and A. Bohrdt, Neural network approach to quasiparticle dispersions in doped antiferromagnets, Commun. Phys. **7**, 187 (2024).
[87] A. Chen and M. Heyl, Empowering deep neural quantum states through efficient optimization, Nat. Phys. **20**, 1476 (2024).
[88] 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, Commun. Phys. **7**, 260 (2024).
[89] I. Loshchilov and F. Hutter, SGDR: Stochastic gradient descent with warm restarts, in *International Conference on Learning Representations, Toulon, France* (ICLR, 2017).
[90] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, SciPost Phys. **10**, 147 (2021).
[91] E. Ibarra-García-Padilla, H. Lange, R. G. Melko, R. T. Scalettar, J. Carrasquilla, A. Bohrdt, and E. Khatami, Autoregressive neural quantum states of Fermi Hubbard models, Phys. Rev. Res. **7**, 013122 (2025).
[92] Z. Denis, A. Sinibaldi, and G. Carleo, Comment on "can neural quantum states learn volume-law ground states?", Phys. Rev. Lett. **134**, 079701 (2025).
[93] G. Torlai, J. Carrasquilla, M. T. Fishman, R. G. Melko, and M. P. A. Fisher, Wave-function positivization via automatic differentiation, Phys. Rev. Res. **2**, 032060(R) (2020).
[94] W. J. Huggins, B. A. O'Gorman, N. C. Rubin, D. R. Reichman, R. Babbush, and J. Lee, Unbiasing fermionic quantum Monte Carlo with a quantum computer, Nature (London) **603**, 416 (2022).
[95] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, PyTorch: An imperative style, high-performance deep learning library, *Proceedings of the 33rd International Conference on*

Neural Information Processing Systems, Vol. 11 (Curran Associates Inc., Red Hook, NY, 2019).

[96] V. Bergholm, J. A. Izaac, M. Schuld, C. Gogolin, and N. Killoran, PennyLane: Automatic differentiation of hybrid quantum-classical computations, arXiv:1811.04968.

[97] J. R. McClean, N. C. Rubin, K. J. Sung, I. D. Kivlichan, X. Bonet-Monroig, Y. Cao, C. Dai, E. S. Fried, C. Gidney, B. Gimby, *et al.*, OpenFermion: The electronic structure package for quantum computers, Quantum Sci. Technol. **5**, 034014 (2020).

[98] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, *et al.*, Recent developments in the PySCF program package, J. Chem. Phys. **153**, 024109 (2020).

[99] https://github.com/ZongkangZhang/Quantum-enhanced-neural-networks-for-quantum-many-body-simulations.