# Solving Fermi-Hubbard-type models by tensor representations of backflow corrections

Yu-Tong Zhou [1,2,3] Zheng-Wei Zhou,[1,2,3,*] and Xiao Liang [4,5,†]

[1]*CAS Key Laboratory of Quantum Information, University of Science and Technology of China, Hefei 230026, China*
[2]*Synergetic Innovation Center of Quantum Information and Quantum Physics, University of Science and Technology of China, Hefei 230026, China*
[3]*Hefei National Laboratory, University of Science and Technology of China, Hefei 230088, China*
[4]*Pittsburgh Supercomputing Center, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*
[5]*Department of Physics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA*

The quantum many-body problem is an important topic in condensed matter physics. To efficiently solve the problem, several methods have been developed to improve the representation ability of wave functions. For the Fermi-Hubbard model under periodic boundary conditions, current state-of-the-art methods are neural network backflows and the hidden fermion Slater determinant. The backflow correction is an efficient way to improve the Slater determinant of free particles. In this work we propose a tensor representation of the backflow-corrected wave function; we show that for the spinless $t$-$V$ model, the energy precision is competitive or even lower than current state-of-the-art fermionic tensor network methods. For models with spin, we further improve the representation ability by considering backflows on fictitious particles with different spins, thus naturally introducing nonzero backflow corrections when the orbital and the particle have opposite spins. We benchmark our method on molecules in the STO-3G basis and the Fermi-Hubbard model with periodic and cylindrical boundary conditions. We show that the tensor representation of backflow corrections achieves competitive or even lower-energy results than current state-of-the-art neural network methods.

## I. INTRODUCTION

Exotic physical phenomena emerge when a large number of microscopic particles interact with each other. Understanding phenomena such as superconductivity, quantum spin liquids, and the quantum Hall effect requires solving the quantum many-body problem with high accuracy. However, solving the problem is challenging because the Hilbert space of the solution grows exponentially with respect to the size of the problem.

Several methods have been developed, but there are still limitations. For example, exact diagonalization has high accuracy, but the problem size is limited [1]. The density-matrix renormalization group (DMRG) can solve one-dimensional or quasi-one-dimensional systems [2], but the accuracy is not satisfactory for two-dimensional systems. Quantum Monte Carlo has no limitation on dimensions and has high precision, but the computational complexity is too high for systems with the "sign problem" [3]. The projected entangled-pair state (PEPS) can solve the two-dimensional system under the open boundary condition (OBC) with high accuracy; however, the computational complexity is high, especially for the periodic boundary condition (PBC) [4–6]. Recently, neural networks (NNs) have shown potential for representing quantum many-body states [7–23]. To solve Fermi-Hubbard-type models,

one approach is the Jordan-Wigner transformation on the Hamiltonian that treats the problem like solving a spin model [15,16]. Another way is improving the single-particle Slater determinants (SDs) by NN backflow transformations [17–19] or multiplying the SD by a NN Jastrow factor [20]. The state-of-the-art wave function named the hidden fermion Slater determinant (HFSD) considers hidden fermionic particles and calculates the determinant of an enlarged matrix [21].

Solving the ground state of the Fermi-Hubbard model near 1/8 doping is important for understanding the mechanism of superconductivity; however, the ground state is challenging to solve [24–28]. In mean-field theory, the Hamiltonian is in a quadratic form, and the ground state is the Hartree-Fock (HF) state. The wave function of the HF state is a Slater determinant, which is an exact representation for particles without interactions. For particles with interactions, the exact representation is very challenging. One way to improve the representation ability is to add a Jastrow factor before the Slater determinant; many-body correlations are contained in the Jastrow function. The backflow correlation improves the representation ability by adding the positions of other particles to the single-particle orbital [18,29–31]. Backflow corrections on wave functions have been widely used in quantum chemistry [17,19]; however, for strongly correlated many-body systems such as Fermi-Hubbard models, the precision is not sufficiently high [18,21].

Although adding variational parameters can increase the state representation ability of the variational wave function [17–19], more variational parameters lead to higher

---

*zwzhou@ustc.edu.cn
†xiaoliang.physics@gmail.com

245107-1

optimization difficulty, which is not beneficial for achieving the ground energy [13]. In this work, we propose an efficient way to increase the parameter number of the backflow-corrected wave function using tensor representation. Namely, each dimension of the tensor is an independent degree of freedom in the backflow-corrected wave function. We show that the tensor representation can achieve competitive or even lower-energy results compared to state-of-the-art fermionic PEPS (fPEPS) results for the spinless $t$-$V$ model [5]. For models with spin, we improve the representation ability by considering backflow corrections on fictitious particles with different spins, which leads to nonzero backflow corrections when the particle and the orbital have different spins. We numerically demonstrate that our method can achieve energy precision competitive with state-of-the-art Restricted-Boltzmann-Machine (RBM) results for molecules in the STO-3G basis [15] and energy precision competitive with or even better than state-of-the-art NN backflow [18] and HFSD [21] results for the Fermi-Hubbard model.

This paper is organized as follows: Sec. II A recalls backflow corrections on wave functions and introduces our method of backflow corrections when the particle and the orbital have different spins. Section II B introduces the tensor representation of the backflow-corrected wave function. Section III presents the numerical results of our methods and comparisons with other state-of-the-art methods: the spinless $t$-$V$ model in Sec. III B, molecules in the STO-3G basis in Sec. III C, and the Fermi-Hubbard model in Sec. III D. Finally, the paper is concluded in Sec. IV.

## II. METHODS

### A. Backflow corrections of wave functions

The backflow correction is defined on the fictitious coordinate $\mathbf{r}_i^B$, which not only depends on the position $\mathbf{r}_\alpha$ but also depends on the positions of other particles [29]:

$$\mathbf{r}_\alpha^B = \mathbf{r}_\alpha + \sum_\beta \eta_{\alpha\beta}[\mathbf{S}](\mathbf{r}_\beta - \mathbf{r}_\alpha), \qquad (1)$$

where $\mathbf{r}_\alpha$ are actual particle positions and $\eta_{\alpha\beta}[\mathbf{S}]$ are variational parameters depending on the many-body state $|\mathbf{S}\rangle$ to create a return flow of particles.

The backflow-corrected single-particle orbital for a spin $\phi_{k,\sigma}^B$ is constructed by a linear combination of eigenstates of the mean-field Hamiltonian $\phi_{k,\sigma}$ [18,30,31]:

$$\phi_{k\sigma}^B(\mathbf{r}_{i,\sigma}) = \phi_{k\sigma}(\mathbf{r}_{i,\sigma}) + \sum_j c_{ij}[\mathbf{S}]\phi_{k\sigma}(\mathbf{r}_{j,\sigma}), \qquad (2)$$

where $c_{ij}[\mathbf{S}]$ is a variational coefficient depending on the many-body configuration $|\mathbf{S}\rangle$. The orbital is $\phi_{k\sigma}(\mathbf{r}_{i,\sigma}) = \langle 0|\hat{c}_{i,\sigma}|\phi_{k\sigma}\rangle$, where $\hat{c}_{i,\sigma}$ is the annihilation operator on the $i$th site with spin value $\sigma$.

Based on Eq. (2), backflow corrections are performed on positions with a spin identical to that of the orbital; meanwhile, the wave function is the product of two Slater determinants:

$$w_1(\mathbf{S}) = \det[M^{B,\uparrow}]\det[M^{B,\downarrow}], \qquad (3)$$

with the element of the Slater matrix being

$$M_{ik}^{B,\sigma} = \phi_{k\sigma}^B(r_{i\sigma}). \qquad (4)$$

Backflow corrections in the formation of Eqs. (3) and (4) have been successfully used for the $t$-$t'$ Hubbard model with a large interaction of $U$ [30,31], which mimics the effect of the virtual hopping, thus leading to the correct superexchange energy.

To improve the representation ability of the Hamiltonian with couplings of spins, in addition to the original backflow corrections, here we introduce an improvement by considering backflow corrections of the fictitious coordinate for one spin $\mathbf{r}_{i,\sigma_i}^B$ depending on positions of other particles with opposite spins:

$$\mathbf{r}_{i,\sigma_i}^B = \mathbf{r}_{i,\sigma_i} + \sum_j \eta_{ij}[\mathbf{S}] \sum_{\sigma_j=\pm 1}(\mathbf{r}_{j,\sigma_j} - \mathbf{r}_{i,\sigma_i}); \qquad (5)$$

therefore, the backflow correction for one particle at position $\mathbf{r}_i$ with spin $\sigma_i$ is performed on fictitious particles at positions $\mathbf{r}_j$ with different spins of $\sigma_j$.

Meanwhile, the backflow-corrected single-particle orbital for a spin is constructed similarly to Eq. (2), except for the summation on particle spins $\sigma_j$:

$$\phi_{k\sigma_k}^B(\mathbf{r}_{i,\sigma_i}) = \phi_{k\sigma_k}(\mathbf{r}_{i,\sigma_i}) + \sum_j c_{ij}[\mathbf{S}] \sum_{\sigma_j=\pm 1} \phi_{k\sigma_k}(\mathbf{r}_{j,\sigma_j}), \qquad (6)$$

where $c_{ij}[\mathbf{S}]$ are variational coefficients depending on the many-body configuration $|\mathbf{S}\rangle$. The total spin is conserved as $(N_\uparrow - N_\downarrow)/2$, where $N_\uparrow$ ($N_\downarrow$) is the particle number for spin up (down), as the orbital $\phi_{k\sigma_k}(\mathbf{r}_{i,\sigma_i})$ is zero when $\sigma_k \neq \sigma_i$ and nonzero when $\sigma_k = \sigma_i$.

The wave function is represented by the Slater determinant of an $N \times N$ matrix,

$$w_2(\mathbf{S}) = \det[M^B], \qquad (7)$$

where $M_{ik}^B = \phi_{k\sigma_k}^B(\mathbf{r}_{i,\sigma_i})$ and $N = N_\uparrow + N_\downarrow$ is the total particle number.

In this paper, backflow corrections of Eqs. (2) and (6) are denoted as BW1 and BW2, respectively. BW1 requires a particle spin equal to the orbital spin in order to achieve nonzero backflow correction terms. However, in the case of BW2, because of the summation on $\sigma_j$, there are nonzero backflow correction terms for arbitrary configurations of the orbital spin $\sigma_k$ and the particle spin $\sigma_i$.

Therefore, the major difference between BW1 and BW2 is in the matrix in the Slater determinant; the matrix is depicted in Fig. 1. In the matrix, the particle and the orbital have identical spin in submatrices 1 and 4; meanwhile, in submatrices 2 and 3 the particle and the orbital have opposite spins. For BW1, submatrices 2 and 3 are undefined, and no backflow corrections are performed. For BW2, there are nonzero backflow corrections in all submatrices.

### B. Tensor representation of backflow corrections

We evaluate the state representation abilities of BW1 and BW2 by using tensor representations of backflow-corrected wave functions. Each dimension of the tensor is an independent degree of freedom in the backflow-corrected wave function denoted by Eqs. (2) and (6).
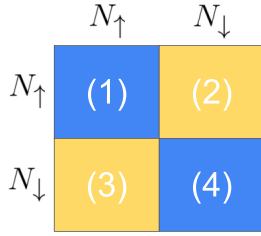
FIG. 1. The matrix in the Slater determinant for $N$ particles, with $N_\uparrow$ and $N_\downarrow$ being the particle numbers for spin up and spin down, respectively. The spin on the horizontal (vertical) axis denotes the spin of the particles (orbitals).

The tensor has representation ability beyond what is provided by the linear combination of eigenstates of the mean-field Hamiltonian, as it includes many more variational parameters. Compared to the backflow corrections for specific configurations of $\mathbf{s}(\mathbf{r}_i)$ and $\mathbf{s}(\mathbf{r}_j)$ [30,31], it includes all possible configurations. Furthermore, it can represent higher-order correlations, such as two-body correlations in the calculation of self-energy based on the diagrammatic perturbation expansion [31,32].

Here we assume coefficients $c_{ij}$ in Eqs. (2) and (6) depend on local configurations $\mathbf{s}(\mathbf{r}_i)$ and $\mathbf{s}(\mathbf{r}_j)$ instead of $|\mathbf{S}\rangle$ for simplicity [30,31]. Independent degrees of freedom in both BW1 and BW2 are the position $\mathbf{r}_{i,\sigma}$, the orbital number of $\phi_{k\sigma}$, the summation index $j$, and configurations $\mathbf{s}(\mathbf{r}_i)$ and $\mathbf{s}(\mathbf{r}_j)$. Therefore, the total dimension of the tensor is

$$[M, N, d, Q, d], \quad (8)$$

where the dimension $M$ equals the site number and the dimension $N$ equals the total particle number. The first and second $d$'s are for configurations $\mathbf{s}(\mathbf{r}_i)$ and $\mathbf{s}(\mathbf{r}_k)$, respectively. $Q$ denotes the index of $i$ as well as indexes in the summation of $j$ considered in either Eq. (2) or Eq. (6).

When considering backflow corrections of nearest neighbors of the $i$th particle, the matrix element in the Slater determinant is assigned by indexing the tensor:

$$M_{ik}^B = g[i, k, \mathbf{s}(\mathbf{r}_i), i, \mathbf{s}(\mathbf{r}_i)] + \sum_{\langle q, i \rangle} g[i, k, \mathbf{s}(\mathbf{r}_i), q, \mathbf{s}(\mathbf{r}_q)], \quad (9)$$

where $g$ is the tensor representation of the wave function, with the dimension defined in Eq. (8). The forward calculation generates the wave-function coefficient $w(\mathbf{S})$ defined by Eq. (3) for BW1 and Eq. (7) for BW2.

For the backward calculation of $w(\mathbf{S})$, the gradient with respect to one matrix element $M_{ik}^B$ is

$$\frac{\partial \det M^B}{\partial M_{ik}^B} = C_{ik}, \quad (10)$$

where the cofactor $C_{ik}$ is an element of an $N \times N$ matrix, defined as the determinant of a matrix obtained by eliminating row $i$ and column $j$ from the original matrix. Expanding the matrix $M^B$ along one column or one row with Laplace expansion, we have

$$\det M^B \cdot \delta_{ik} = \sum_{j=1}^N C_{ij} M_{kj}^B, \quad \det M^B \cdot \delta_{ik} = \sum_{j=1}^N C_{ji} M_{jk}^B, \quad (11)$$

namely,

$$CM^B = M^B C = \det M^B \cdot I. \quad (12)$$

Therefore, the backward of parameter $g$ is given by

$$\frac{\partial w(\mathbf{S})}{\partial g[i, k, \mathbf{s}(\mathbf{r}_i), q, \mathbf{s}(\mathbf{r}_q)]} = \text{inv}(M^B)_{ik} \cdot w(\mathbf{S}). \quad (13)$$

In the variational Monte Carlo (VMC), forward and backward calculations of the wave-function coefficient $w(\mathbf{S})$ are necessary when collecting one MC sample. The forward is achieved by calculating the Slater determinant of the $N \times N$ matrix, and the backward is achieved by inverting the matrix. The complexity of calculating either the determinant or the inverse of matrix $M^B$ is $O(N^3)$.

## III. NUMERICAL INVESTIGATIONS

In this section we numerically demonstrate that backflow corrections under the tensor representation have strong representation abilities. We benchmark on three types of models: (1) the spinless fermionic $t$-$V$ model on the square lattice under the OBC, (2) several molecules in the STO-3G basis, and (3) the spinful Fermi-Hubbard model on rectangular lattices with the PBC and cylindrical boundary condition (CBC).

For the spinless $t$-$V$ model, backflow corrections can achieve state-of-the-art energy results compared to those of the PEPS. For molecules in the STO-3G basis, BW2 has better precision than BW1, and energies obtained with BW2 are competitive to state-of-the-art results. For the Fermi-Hubbard model, both BW1 and BW2 achieve competitive or even lower-energy results compared to state-of-the-art methods like NN backflow and the HFSD. Furthermore, energies obtained with BW2 are lower compared to those of BW1 on finite-size lattices.

### A. Optimization methods

The wave function is first optimized using the VMC method and then further optimized by a Lanczos step. In VMC, the energy and the $\alpha$th parameter's gradient are evaluated through the Markov chain Monte Carlo (MCMC) process [5,6]:

$$E = \langle E_{\text{loc}} \rangle, \quad G^\alpha = 2\langle E_{\text{loc}} O_{\text{loc}}^\alpha \rangle - 2\langle E_{\text{loc}} \rangle \langle O_{\text{loc}}^\alpha \rangle, \quad (14)$$

where the local energy is $E_{\text{loc}}(\mathbf{S}) = \sum_{\mathbf{S}'} \frac{w(\mathbf{S}')}{w(\mathbf{S})} \langle \mathbf{S}' | \hat{H} | \mathbf{S} \rangle$, $O_{\text{loc}}(\mathbf{S})^\alpha = \frac{1}{w(\mathbf{S})} \frac{\partial w(\mathbf{S})}{\partial \alpha}$, and $\langle \cdots \rangle$ denotes the average of MCMC samples.

The variational parameters are updated according to the gradient descent method. Here we adopt only the first-order gradient descent due to the low optimization difficulty of the tensor representation. Because of the limited MC sample number, we take the sign of the gradient and apply a constant step size $\delta$: $\alpha' = \alpha - \delta \text{sgn}(G^\alpha)$. Such a parameter updating

scheme was successfully used to optimize high-dimensional tensors like the PEPS [5,33,34].

A Lanczos step further improves the representation ability of the wave function $|\Psi_{p=0}\rangle$ by considering an additional wave function $|\Psi_{p=0}^\perp\rangle$ orthogonal to $|\Psi_{p=0}\rangle$ [13,35]:

$$|\Psi_{p=1}\rangle = A|\Psi_{p=0}\rangle + B|\Psi_{p=0}^\perp\rangle, \tag{15}$$

where $A$ and $B$ are parameters to be determined and $|\Psi_{p=0}\rangle$ is the wave function obtained after the VMC. The orthogonal wave function is built by $|\Psi_p^\perp\rangle = \frac{1}{\sigma_p}(\hat{H} - E_p)|\Psi_p\rangle$, where the energy expectation $E_p = \langle\Psi_p|\hat{H}|\Psi_p\rangle$ and the variance $\sigma_p^2 = \langle\Psi_p|(\hat{H} - E_p)^2|\Psi_p\rangle$. Because it is necessary to calculate $\langle\Psi_{p=0}|\hat{H}^2|\Psi_{p=0}\rangle$ in the Lanczos optimization, the complexity for one Lanczos optimization is $O(N^5)$. After the Lanczos optimization is completed, the complexity for calculating one coefficient for the Lanczos optimized wave function $w_{p=1}(S)$ is $O(N^4)$.

### B. Spinless $t$-$V$ model

The Hamiltonian of the spinless $t$-$V$ model reads

$$\hat{H} = -t\sum_{\langle i,j \rangle}(\hat{c}_i^\dagger \hat{c}_j + \text{H.c.}) + V\sum_{\langle i,j \rangle}\hat{n}_i\hat{n}_j, \tag{16}$$

where $t$ is the hopping strength and $V$ is the interaction strength between nearest neighbors. $\hat{c}_i^\dagger$ ($\hat{c}_i$) creates (destroys) a particle on the $i$th site, and the particle number operator $\hat{n}_i = \hat{c}_i^\dagger\hat{c}_i$. We set $t = 1$ through our investigations. The maximal occupation per site is $n_i = 1$ for the spinless $t$-$V$ model. We investigate the half-filling case $n = 1/2$, so that the particle number is half of the total site number on the $10 \times 10$ square lattice.

We benchmark on lattices with the OBC to compare the results with the state-of-the-art fPEPS method [5]. The fPEPS is optimized by the imaginary-time-evolution method called the simple update and then by the VMC [5,6,33,34]. The complexity of calculating a $w(S)$ is $O(D^3 D_c^3)$, where $D$ is the bond dimension and $D_c$ is the bond truncation dimension. As reported in a previous study [5], for the $t$-$V$ model with $V = 2$, $D_c \approx 3D$ is well converged, and $D = 8$ gives a $1 \times 10^{-4}$ relative energy error compared to extrapolated infinite $D$. However, $D_c = 3D$ is not well converged when $V = 0$ [5]. Because of the high complexity with respect to $D$ and $D_c$ as well as the high requirement of $D_c/D$, it is very challenging to optimize fPEPS with a large bond dimension. Because $D = 8$ gives high precision and $D_c \approx 3D$ is well converged when $V = 2$, here we use $D = 8$ and $D_c = 4D$ as the reference. For our benchmark cases, the fPEPS takes longer than our BW method when calculating one $w(S)$. Furthermore, it is impossible to perform Lanczos optimizations for fPEPS.

For the spinless model, backflow corrections are performed for one orbital based on Eq. (1). Energy comparisons for different interaction strengths $V$ are denoted in Table I. In Table I, results for the HF state are achieved by representing the spinless HF orbital $\phi_k(\mathbf{r}_i)$ by a tensor with the dimension of $[M, N]$, where $M$ is the site number and $N$ is the total particle number. $p = 0$ is the result obtained with the tensor

TABLE I. Comparison of energies (per site) for the spinless $t$-$V$ model on a $10 \times 10$ square lattice under the OBC; the total particle number is 50. $p = 0$ ($p = 1$) denotes the wave function before (after) one Lanczos step. Reference energies are obtained with the fPEPS method [5].

| $V$ | HF state | $p = 0$ | $p = 1$ | fPEPS |
|------|----------|---------|---------|---------|
| 0.45 | −0.6103 | −0.6132 | −0.6134 | −0.6129 |
| 1 | −0.4561 | −0.4617 | −0.4620 | −0.4620 |
| 2 | −0.2961 | −0.2997 | −0.2999 | −0.2999 |

representation of backflow corrections, and $p = 1$ is the result of one Lanczos step for the $p = 0$ wave function. Backflow corrections are considered on nearest neighbors of $\mathbf{r}_i$; namely, the dimension $Q$ defined in Eq. (8) includes site $\mathbf{r}_i$ as well as its nearest neighbors, and therefore, $Q = 5$. The parameter number of fPEPS is $O(2ND^4)$, where $N$ is the site number. The parameter number based on the tensor dimension defined in Eq. (8) is 100 000; however, the parameter number of fPEPS is approximately 819 200, and thus, our method has many fewer parameters than fPEPS.

Compared to the $p = 1$ results, the relative errors of the HF state are of the magnitude of $10^{-3}$ for $V = 0.45$ and $10^{-2}$ for $V = 1, 2$; meanwhile, backflow corrections decrease relative errors to a magnitude of $10^{-4}$ for all cases. As shown in Table I, both $p = 0$ and $p = 1$ have energy precisions competitive to that of the fPEPS.

The energy convergence of VMC for the spinless $t$-$V$ model with $V/t = 1$ is depicted in Fig. 2. The initial energy of −0.4561 is from the HF state. After the HF state is converged, we continue the optimization by adding backflow corrections, within the initial energy from the HF state. The energy converges smoothly after backflow corrections are added. The parameter updating step size is $\delta = 5 \times 10^{-4}$, and the MC sample number for each step is 128 000. The interval between two MC samples is the total site number. The converged energy per site is −0.4617, while the reference energy obtained with the fPEPS is −0.4620.
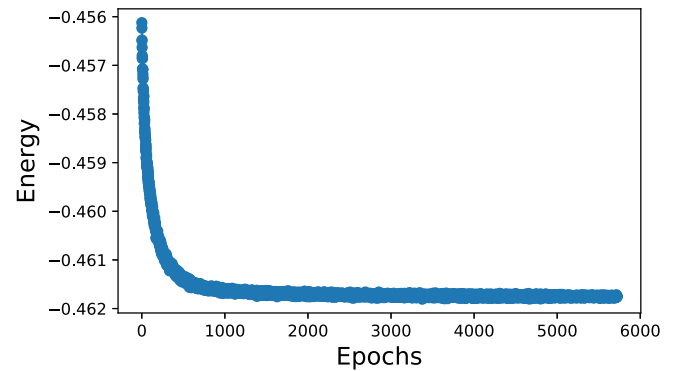


FIG. 2. The energy convergence of the first-order gradient descent for the spinless $t$-$V$ model with $V/t = 1$ on the $10 \times 10$ lattice under the OBC. The MC sample number is 128 000, and the parameter updating step size $\delta = 5 \times 10^{-4}$. The converged energy per site is −0.4617, while the reference energy obtained by fPEPS is −0.4620.

TABLE II. Comparison of energies achieved using several methods for molecules in the STO-3G basis. $M$ is the equivalent site number, and $N$ is the particle number. All energy results achieved with BW1 and BW2 are evaluated by $p = 0$ wave functions. Energy results from CCSD(T) and RBM are from the literature [15].

| Molecule | $M$ | $N$ | BW1 | BW2 | CCSD(T) | RBM |
|----------|-----|-----|------|------|---------|------|
| $H_2O$ | 14 | 10 | −75.0211 | −75.0222 | −75.0231 | −75.0232 |
| $NH_3$ | 16 | 10 | −55.5197 | −55.5276 | −55.5281 | −55.5277 |
| $C_2$ | 20 | 12 | −74.6816 | −74.6876 | −74.6876 | −74.6892 |
| $N_2$ | 20 | 14 | −107.6629 | −107.6747 | −107.6738 | −107.6767 |

### C. Molecules in the STO-3G basis

The Hamiltonian for molecules in the second quantization form is

$$\hat{H} = \sum_{ij} t_{ij} \hat{c}_i^\dagger \hat{c}_j + \sum_{ijkm} u_{ijkm} \hat{c}_i^\dagger \hat{c}_k^\dagger \hat{c}_j \hat{c}_m, \qquad (17)$$

where the label $i$ denotes the fermionic mode, with $t_{ij}$ being the one-body interaction and $u_{ijkm}$ being the two-body interaction. $\hat{c}_i^\dagger$ ($\hat{c}_i$) creates (destroys) a particle on the $i$th fermionic mode. The structure of a molecule in STO-3G is obtained from the literature [15], and we use the software package PYSCF [36] to generate coefficients of $t_{ij}$ and $u_{ijkm}$, with a maximum iteration number of 500. We first optimize a HF state for each molecule without two-body interactions, then continue the optimization by adding backflow corrections from the HF state.

Here we treat spins by considering additional lattice sites; thus, a HF orbital with spin $\phi_{k\sigma}(\mathbf{r}_{i,\sigma})$ is represented by a tensor with dimensions of $[M, N]$, where $M$ is the equivalent site number and $N$ is the total particle number. The equivalent site number is twice the total orbital number because of the spin on site. Configurations on each equivalent site are occupation and nonoccupation; thus, $d = 2$. For backflow corrections, we consider backflow terms from all equivalent sites; thus, the dimension of the tensor is defined by Eq. (8) with $Q = M$.

A challenge of solving molecules is the locality of the ground state in the total Hilbert space, as the ground state is nearly classical. Therefore, the optimization easily gets stuck on local minima, especially when a large MC sample number is used in the initial optimization steps. To avoid local optimizations, we stop the optimization after hundreds of optimization steps, then continue the optimization by enforcing MCMC starting from the configuration based on the Pauli exclusion principle. For each case, we start from 44 800 MC samples in initial optimization steps. After 1000 optimizations, we use roughly 70 000 MC samples for each optimization step, with the interval between two MC samples equal to the number of equivalent sites.

Energy comparisons are denoted in Table II. Each energy of both BW1 and BW2 is obtained using the minimum value of the last 20 optimization steps. From Table II, energy precision achieved with BW1 is lower than that achieved by BW2, and energy results obtained with BW2 are competitive with state-of-the-art results obtained with coupled-cluster theory with single, double, and partially triple excitations [CCSD(T)] and RBM [15]. From Table II, RBM achieves better energy

precision than BW. RBM is efficient for molecules in the STO-3G basis because the wave function is easy to represent and does not demand the full expressibility of BW.

### D. Fermi-Hubbard model

The Hamiltonian of the Fermi-Hubbard model is

$$\hat{H} = -t \sum_{\langle ij \rangle, \sigma} (\hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + \text{H.c.}) + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \qquad (18)$$

where $t$ is the hopping strength and $U$ is the strength of on-site interactions. $\hat{c}_{i\sigma}^\dagger$ ($\hat{c}_{i\sigma}$) creates (destroys) a particle of spin $\sigma$ on the $i$th site, and the particle number operator $\hat{n}_{i\sigma} = \hat{c}_{i\sigma}^\dagger \hat{c}_{i\sigma}$. For the Hubbard model with spin, double occupations are allowed. We set $t = 1$ through our investigations. In each optimization step, there are roughly 44 800 MC samples for calculating gradients, and the interval between two samples is the lattice size.

We first demonstrate the advantage of our single-tensor representation by comparing it to the original backflow. In this work, we represent the original form of the wave function defined in either BW1 or BW2 by two separate tensors. The first tensor represents coefficients $c_{ij}$ with the dimensions of $[M, Q, d, d]$, where $M$ is the site number, $Q = 5$ denotes the position $\mathbf{r}_i$ as well as its nearest neighbors, and $d = 4$ denotes the degrees of freedom per site. The second tensor represents the HF orbital $\phi_k(\mathbf{r}_{i,\sigma})$ with the dimensions of $[M, N, 2]$, where $N$ is the total particle number and the dimension of 2 denotes the spin.

To represent the backflow-corrected wave function by a single tensor, the dimension of the tensor is $[M, N, 2, d/2, Q, d]$, where $d = 4$ denotes the degrees of freedom per site. In the tensor, the dimension of 2 is for $\sigma_i$, and the dimension of $d/2$ denotes whether there is double occupation on $\mathbf{r}_i$. In comparison to the dimension defined in Eq. (8), we divide the first $d$ in order to distinguish the double occupation. $Q = 5$ denotes the position $\mathbf{r}_i$ as well as its nearest neighbors.

Figure 3 denotes energy comparisons between the original backflow and the single-tensor representation on the $4 \times L$ lattice with the PBC; the filling of the Hubbard model is $n = 0.875$. Each energy result is evaluated with the $p = 0$ wave function. From Fig. 3, for either BW1 or BW2, the original backflow has much worse energy precision than the single-tensor representation. For either the original backflow or the single-tensor representation, BW2 has better energy precision than BW1. The HF wave function is represented by a tensor with the dimensions of $[M, N, 2]$, and it reaches $-0.5330$ for $L = 8$, $-0.5398$ for $L = 12$, and $-0.5658$ for $L = 16$. Thus, backflow corrections achieve better energy precision than the HF state. Table III compares the energies of BW1, BW2, and other state-of-the-art methods on square lattices under the PBC. For results obtained with BW1 and BW2 in Table III, we first optimize a HF state under $U = 0$ and then add backflow corrections and continuing optimizations; each energy value is evaluated with the $p = 1$ wave function using 48 000 MC samples, and the interval between two MC samples is the lattice size.

For cases of half filling $n = 1$, the reference energies are from auxiliary-field quantum Monte Carlo (AFQMC) [27]. On the $6 \times 6$ lattice, the relative error is $1.7 \times 10^{-2}$ for BW1
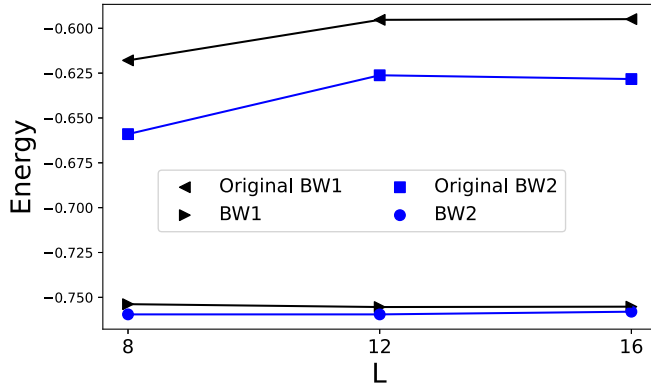
FIG. 3. Comparison of energies (per site) of two kinds of tensor representations of backflow corrections for the Fermi-Hubbard model with $n = 0.875$ and $U = 8$ on the $4 \times L$ lattice with the PBC. "Original BW" denotes the coefficient $c_{ij}$, and the HF orbitals are represented by two separate tensors. "BW" denotes the backflow-corrected wave function is represented by a single tensor. Each energy is evaluated by the $p = 0$ wave function.
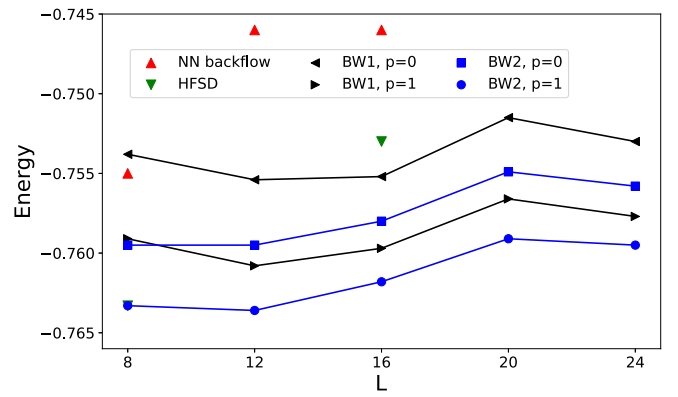


FIG. 4. Energy comparisons for the Fermi-Hubbard model on $4 \times L$ lattices under the PBC and $U = 8$. The filling is $n = 0.875$. Red upward-pointing triangles denote NN backflow, green downward-pointing triangles denote HDFS. BW1 energy results of $p = 0$ ($p = 1$) are denoted by left-pointing (right-pointing) triangles. BW2 energy results of $p = 0$ ($p = 1$) are denoted by squares (circles).

and $4.0 \times 10^{-3}$ for BW2. On the $8 \times 8$ lattice, the relative error is $1.4 \times 10^{-2}$ for BW1 and $4.2 \times 10^{-3}$ for BW2. On the $10 \times 10$ lattice, the relative error is $1.4 \times 10^{-2}$ for BW1 and $4.5 \times 10^{-3}$ for BW2. On the $10 \times 10$ lattice, the relative error is $1.4 \times 10^{-2}$ for BW1 and $4.5 \times 10^{-3}$ for BW2. From Table II, BW2 has better energy precision than BW1.

For the more challenging case of $1/8$ doping, $n = 0.875$, Ref. [1] and Ref. [2] in Table III are from NN backflow [18] and the HFSD [21], respectively. To clearly compare the energies, energies for filling $n = 0.875$ on $4 \times L$ lattices with the PBC are depicted in Fig. 4. For the $p = 0$ results, BW1 achieves satisfactory energies compared to NN backflow; however, the energy precision is not competitive with the HFSD on the $4 \times 8$ lattice. However on the $4 \times 16$ lattice, BW1 achieves lower energy than the HFSD. For all lattice sizes, BW2 achieves better precision than BW1, and BW2 achieves lower energies than those from NN backflow. A Lanczos step can significantly improve the energy precision for all cases. For $p = 1$ cases, both BW1 and BW2 achieve

better energy precision than NN backflow, and BW2 achieves better energy precision than the HFSD.

Compared to the DMRG energy of $-0.7659$ for a system size of $4 \times \infty$ (open, PBC) [18], our energy of $-0.7595$ for $4 \times 24$ is still higher. Because only nearest neighbors of $\mathbf{r}_i$ are considered in our backflow corrections, our energy precision can be improved by considering further backflow corrections. In addition to comparisons of energies, we compare the spin correlation functions of BW1 and BW2 for the half filling $n = 1$. The spin correlation function is defined as [27]

$$C(x, y) = \langle \mathbf{S}(0, 0)\mathbf{S}(x, y) \rangle, \quad (19)$$

where $\mathbf{S}(x, y)$ is the spin operator with coordinates $(x, y)$ [27]. Translational symmetry is preserved statistically; thus, the reference point (0,0) is averaged on the whole lattice to reduce statistical error. Figure 5 depicts the spin correlation functions for lattice sizes of $6 \times 6$, $8 \times 8$, and $10 \times 10$ under $U = 8$. From Fig. 5, long-range order is clearly seen in both BW1 and BW2; however, BW1 gives larger absolute values of correlation functions compared to BW2. The magnetization is defined as the spin structure factor:

$$M_2^2 = \frac{1}{N} \sum_{i=1}^{N} (-1)^{x_i + y_i} C(x_i, y_i). \quad (20)$$

For BW1, magnetizations for $6 \times 6$, $8 \times 8$, and $10 \times 10$ are 0.1554, 0.1524, and 0.1534, respectively. For BW2, magnetizations for $6 \times 6$, $8 \times 8$, and $10 \times 10$ are 0.0975, 0.1003, and 0.0751, respectively. Compared to references in the literature [27], values of $C(x, y)$ and $M_2^2$ obtained with BW2 are more reasonable than those obtained with BW1. For $n = 0.875$, we compare the spin densities of BW1 and BW2, as depicted in Fig. 6. In Fig. 6, the spin density is defined as the average spin value at each site $\langle \mathbf{S}_i \rangle$. Because of the PBC, the spin density is ideally uniform, and the spin density is ideally zero because the total spin of the ground state is zero.

On the $4 \times 8$ lattice, from Figs. 6(a) and 6(b), the ground state obtained with BW2 has a more uniform spin density

TABLE III. Comparison of energies (per site) of the Fermi-Hubbard model on rectangular lattices with the PBC and $U = 8$. Reference energies for $n = 1$ are from the AFQMC [27]. For $n = 0.875$, values for Ref. [1] are from NN backflow [18], and those for Ref. [2] are from the HFSD [21]. All energy results for BW1 and BW2 are evaluated using $p = 1$ wave functions.

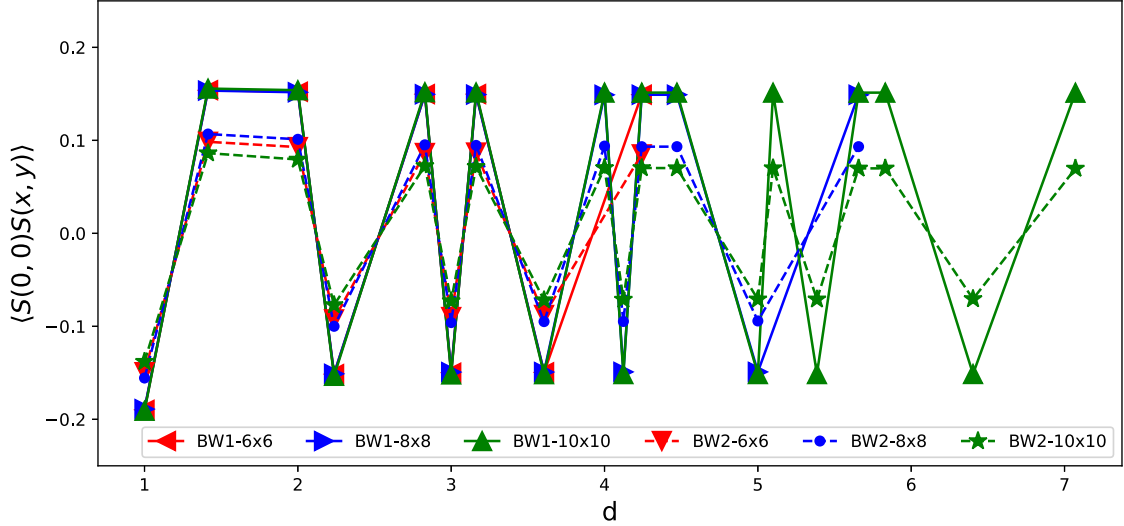| $n$ | Lattice size | BW1 | BW2 | Ref. [1] | Ref. [2] |
|---|---|---|---|---|---|
| 1 | $6 \times 6$ | $-0.5186$ | $-0.5257$ | $-0.5278(5)$ | |
| | $8 \times 8$ | $-0.5188$ | $-0.5241$ | $-0.5263(8)$ | |
| | $10 \times 10$ | $-0.5181$ | $-0.5230$ | $-0.5254(8)$ | |
| 0.875 | $4 \times 8$ | $-0.7591$ | $-0.7633$ | $-0.755$ | $-0.7633$ |
| | $4 \times 12$ | $-0.7608$ | $-0.7636$ | $-0.746$ | |
| | $4 \times 16$ | $-0.7597$ | $-0.7618$ | $-0.746$ | $-0.753$ |
| | $4 \times 20$ | $-0.7566$ | $-0.7591$ | | |
| | $4 \times 24$ | $-0.7577$ | $-0.7595$ | | |

FIG. 5. Comparison of spin correlation functions $\langle S(0,0)S(x,y)\rangle$ for BW1 (solid line) and BW2 (dashed line) for different lattice sizes; the filling is $n = 1$, and $U = 8$. The relative distance is defined as $d = \sqrt{x^2 + y^2}$; all results are evaluated with $p = 1$ wave functions.

than that obtained with BW1. On the $4 \times 24$ lattice, Figs. 6(c) and 6(d) depict the spin density for BW1 and BW2, respectively. From Fig. 6, the ground state achieved with BW2 has a more uniform spin density than that achieved with BW1. It is notable that the ground energy achieved with BW2 is only $2.4 \times 10^{-3}$ lower than that achieved with BW1; thus, BW2 has more representation ability than BW1. Based on the results for the PBC, we investigate BW2 in cases in which ground states are supposed to have stripe orders, such as rectangular lattices under the CBC in the literature [26,28]. The boundary conditions are open along the shorter boundary ($x$) and periodic along the longer boundary ($y$). To break the degeneracy from the translational symmetry, a pinning field is applied on both shorter boundaries: $v_{i\uparrow} = -v_{i\downarrow} = (-1)^{i_x + i_y} v_0$ for $i_y = 1$ and $i_y = L_y$.
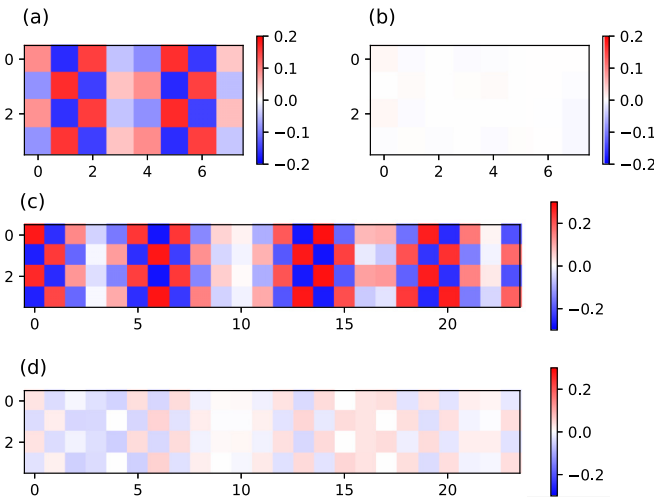


FIG. 6. The spin density of the Fermi-Hubbard model with filling $n = 0.875$ and $U = 8$ under the PBC, evaluated using the $p = 1$ wave function. On the $4 \times 8$ lattice, the spin density obtained with (a) BW1 and (b) BW2. The spin density on the $4 \times 24$ lattice obtained with (c) BW1 and (d) BW2.

Figures 7(a)–7(d) depict the spin density and the hole density under the CBC, evaluated for the $p = 1$ wave function. Figures 7(a) and 7(b) depict the $4 \times 16$ lattice for filling $n = 0.875$ and $U = 8$, with pinning field strength $v_0 = 0.25$. The reference energy from DMRG is $-0.7713$, and BW2 achieves an energy of $-0.7640$ for $p = 0$ and $-0.7678$ for $p = 1$. The relative energy error is $9.4 \times 10^{-3}$ for $p = 0$ and $4.5 \times 10^{-3}$ for $p = 1$. Figures 7(c) and 7(d) denote the $4 \times 20$ lattice, filling $n = 0.9$, and $U = 6$, with pinning field strength $v_0 = 0.5$. BW2 achieves a ground energy of $-0.8485$ for $p = 0$ and $-0.8516$ for $p = 1$. Compared to the energy reported by DMRG ($-0.8352$), BW2 is $1.6 \times 10^{-2}$ lower for $p = 0$ and $1.9 \times 10^{-2}$ lower for $p = 1$. The stripe patterns depicted in Figs. 7(a) and 7(c) match those from both AFQMC and DMRG [26,28].

Furthermore, we benchmark BW2 on lattices as large as $8 \times 16$, with the PBC in both directions, filling $n = 0.875$, and $U = 8$. To break the degeneracy from the translational symmetry, a pinning field is applied on both shorter boundaries with field strength $v_0 = 0.25$. With the interval between two MC samples being the lattice size, it takes roughly 2 min for one optimization step with a total of 44 800 MC samples on 128 AMD EPYC 7742 CPU cores. The converged energies are $-0.7748$ for $p = 0$ and $-0.7784$ for $p = 1$. Figures 7(e) and 7(f) denote the spin density evaluated for the $p = 1$ wave function and the hole density, respectively. The spin density pattern matches that on the $4 \times 16$ lattice, which demonstrates the valid state representation ability of BW2 for large lattices.

## IV. CONCLUSIONS

We showed that tensor representations of backflow corrections after a Lanczos optimization have sufficient representation abilities to achieve state-of-the-art ground energies. Because the tensor representation is easy to optimize, first-order gradient descent is feasible.

For systems with spins, the representation ability can be further improved by considering backflow corrections on different spins, and we naturally introduced nonzero backflow
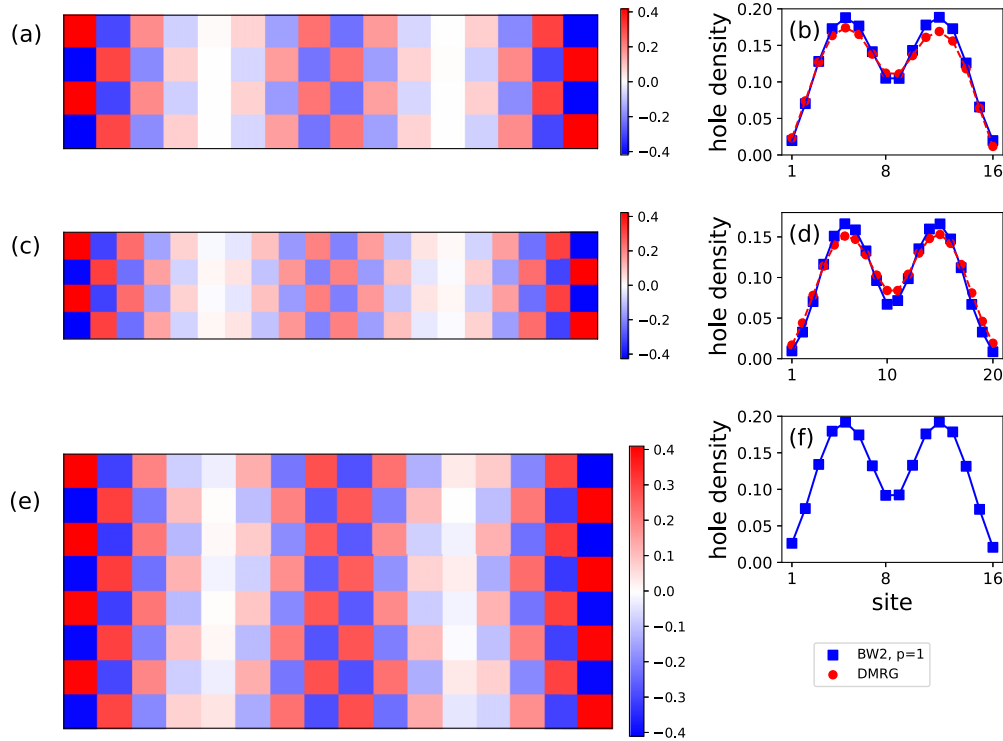
FIG. 7. The spin density and the hole density achieved with BW2 on rectangular lattices under (a)–(d) the CBC and (e) and (f) the PBC, with a pinning field applied on both short edges. (a) and (b) depict the $4 \times 16$ lattice, with filling $n = 0.875$. (c) and (d) depict the $4 \times 20$ lattice, with filling $n = 0.9$. (e) and (f) depict the $8 \times 16$ lattice, with filling $n = 0.875$.

corrections where the particle and the orbital have opposite spins. By numerical demonstrations for molecules in the STO-3G basis and the finite-size Fermi-Hubbard model, we showed that BW2 has a better representation ability than BW1. Furthermore, we showed that BW2 is capable of solving systems with large lattice sizes.

For the Fermi-Hubbard model, compared to the NN backflow [18], the input of the NN is the total many-body configuration $|\mathbf{S}\rangle$. In either BW1 or BW2, backflow terms are limited to nearest neighbors of the position $\mathbf{r}_i$, and the energy precision can be improved by considering backflow terms with further distances [21]. The HFSD considers a matrix larger than $N \times N$ in the Slater determinant by introducing additional hidden particles. Here in either BW1 or BW2, the size of the matrix is kept as $N \times N$, and the enhanced representation ability is achieved by additional degrees of freedom introduced by backflow correction terms in the $N \times N$ matrix.

In our tensor representations, backflow corrections were performed between two sites; however, the representation ability can be, in principle, improved by considering higher-order correlations, and thus, considering backflow

corrections from the perspective of the Green's function is feasible. The representation ability can be improved by increasing variational parameters; thus, the application of neural networks based on the tensor representation is feasible. We hope our work gives some insight into developing numerical methods for solving quantum many-body systems.

[1] A. Lüscher and A. M. Läuchli, Exact diagonalization study of the antiferromagnetic spin-1/2 Heisenberg model on the square lattice in a magnetic field, Phys. Rev. B **79**, 195102 (2009).

[2] 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).

[3] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[4] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, Rev. Mod. Phys. **93**, 045003 (2021).

[5] S.-J. Dong, C. Wang, Y. Han, G.-C. Guo, and L. He, Gradient optimization of fermionic projected entangled pair states on directed lattices, Phys. Rev. B **99**, 195153 (2019).

[6] W.-Y. Liu, J. Hasik, S.-S. Gong, D. Poilblanc, W.-Q. Chen, and Z.-C. Gu, Emergence of gapless quantum spin liquid from deconfined quantum critical point, Phys. Rev. X **12**, 031039 (2022).

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

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

[9] 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).

[10] 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).

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

[12] 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).

[13] X. Liang, M. Li, Q. Xiao, J. Chen, C. Yang, H. An, and L. He, Deep learning representations for quantum many-body systems on heterogeneous hardware, Mach. Learn.: Sci. Technol. **4**, 015035 (2023).

[14] 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).

[15] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for ab-initio electronic structure, Nat. Commun. **11**, 2368 (2020).

[16] N. Yoshioka, W. Mizukami, and F. Nori, Solving quasiparticle band spectra of real solids using neural-network quantum states, Commun. Phys. **4**, 106 (2021).

[17] 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).

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

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

[20] J. Stokes, J. R. Moreno, E. A. Pnevmatikakis, and G. Carleo, Phases of two-dimensional spinless lattice fermions with first-quantized deep neural-network quantum states, Phys. Rev. B **102**, 205122 (2020).

[21] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, Proc. Natl. Acad. Sci. USA **119**, e2122059119 (2022).

[22] J.-G. Liu, L. Mao, P. Zhang, and L. Wang, Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits, Mach. Learn.: Sci. Technol. **2**, 025011 (2021).

[23] D. Luo, Z. Chen, K. Hu, Z. Zhao, V. M. Hur, and B. K. Clark, Gauge-invariant and anyonic-symmetric autoregressive neural network for quantum lattice models, Phys. Rev. Res. **5**, 013216 (2023).

[24] J. P. F. LeBlanc *et al.*, Solutions of the two-dimensional Hubbard model: Benchmarks and results from a wide range of numerical algorithms, Phys. Rev. X **5**, 041041 (2015).

[25] B.-X. Zheng *et al.*, Stripe order in the underdoped region of the two-dimensional Hubbard model, Science **358**, 1155 (2017).

[26] M. Qin, H. Shi, and S. Zhang, Coupling quantum Monte Carlo and independent-particle calculations: Self-consistent constraint for the sign problem based on the density or the density matrix, Phys. Rev. B **94**, 235119 (2016).

[27] M. Qin, H. Shi, and S. Zhang, Benchmark study of the two-dimensional Hubbard model with auxiliary-field quantum Monte Carlo method, Phys. Rev. B **94**, 085103 (2016).

[28] H. Xu, H. Shi, E. Vitali, M. Qin, and S. Zhang, Stripes and spin-density waves in the doped two-dimensional Hubbard model: Ground state phase diagram, Phys. Rev. Res. **4**, 013239 (2022).

[29] Y. Kwon, D. M. Ceperley, and R. M. Martin, Effects of back-flow correlation in the three-dimensional electron gas: Quantum Monte Carlo study, Phys. Rev. B **58**, 6800 (1998).

[30] L. F. Tocchio, F. Becca, A. Parola, and S. Sorella, Role of backflow correlations for the nonmagnetic phase of the $t$-$t'$ Hubbard model, Phys. Rev. B **78**, 041101(R) (2008).

[31] L. F. Tocchio, F. Becca, and C. Gros, Backflow correlations in the Hubbard model: An efficient tool for the study of the metal-insulator transition and the large-$U$ limit, Phys. Rev. B **83**, 195138 (2011).

[32] J. M. Luttinger and J. C. Ward, Ground-state energy of a many-fermion system. II, Phys. Rev. **118**, 1417 (1960).

[33] W.-Y. Liu, S. Dong, C. Wang, Y. Han, H. An, G.-C. Guo, and L. He, Gapless spin liquid ground state of the spin-1/2 $J_1$-$J_2$ Heisenberg model on square lattices, Phys. Rev. B **98**, 241109(R) (2018).

[34] W.-Y. Liu, Y.-Z. Huang, S.-S. Gong, and Z.-C. Gu, Accurate simulation for finite projected entangled pair states in two dimensions, Phys. Rev. B **103**, 235155 (2021).

[35] 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).

[36] Q. Sun *et al.*, Recent developments in the PySCF program package, J. Chem. Phys. **153**, 024109 (2020).