


Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions

Di Luo and Bryan K. Clark

*Institute for Condensed Matter Theory and Department of Physics,
University of Illinois at Urbana-Champaign, Illinois 61801, USA*

 (Received 25 August 2018; revised manuscript received 12 January 2019; published 4 June 2019)

Obtaining an accurate ground state wave function is one of the great challenges in the quantum many-body problem. In this Letter, we propose a new class of wave functions, neural network backflow (NNB). The backflow approach, pioneered originally by Feynman and Cohen [Phys. Rev. **102**, 1189 (1956)], adds correlation to a mean-field ground state by transforming the single-particle orbitals in a configuration-dependent way. NNB uses a feed-forward neural network to learn the optimal transformation via variational Monte Carlo calculations. NNB directly dresses a mean-field state, can be systematically improved, and directly alters the sign structure of the wave function. It generalizes the standard backflow [L. F. Tocchio *et al.*, Phys. Rev. B **78**, 041101(R) (2008)], which we show how to explicitly represent as a NNB. We benchmark the NNB on Hubbard models at intermediate doping, finding that it significantly decreases the relative error, restores the symmetry of both observables and single-particle orbitals, and decreases the double-occupancy density. Finally, we illustrate interesting patterns in the weights and bias of the optimized neural network.

DOI: [10.1103/PhysRevLett.122.226401](https://doi.org/10.1103/PhysRevLett.122.226401)

Introduction.—A key question in strongly correlated quantum systems is to obtain an approximation for the ground state wave function. This is especially important for Fermion systems in two or more dimensions where only approximate or exponentially costly methods for evaluating observables of quantum systems exist. Early attempts for writing down variational Fermion wave functions, such as Slater determinants [1] and BCS wave functions [2], focused on finding the ground state of a mean field Hamiltonian that best matched the interacting ground state. Since these early attempts, more sophisticated wave functions have been developed that dress these mean-field starting points, including Slater Jastrow [3,4], Slater Jastrow backflow [5–7], and iterative backflow [8], which has recently been described as a nonlinear network [9]. These wave functions have the advantage that the mean-field starting point can directly incorporate the basic physics of the problem.

Instead of starting from a dressed mean field, many other classes of wave functions are parametrized by a tuning parameter D , which interpolates from a trivial state at small D to a universal wave function spanning the entire Hilbert space at an exponential D . Examples of such wave functions include matrix-product states [10,11], other forms of tensor networks [11–13], Huse-Elser states [14–16], and string-bond states [17]. Recently, wave functions based on neural network primitives, such as restricted Boltzmann machines (RBM) and feed forward neural network (FNN), have been introduced with similar universal properties [18–42]: as the number of hidden neurons increases, the neural network state can represent all probability distribution, although this

may require complex weights to represent the sign structure of Fermion wave functions. A recent attempt to incorporate RBM into Fermion states by using the RBM as a more general Jastrow [23] shows promise, but was still restricted to the sign structure of the underlying mean-field *Ansatz*. Even though general neural networks could alter the sign structure, it may struggle with capturing the underlying mean-field physics both in terms of the number of neurons required as well as optimization.

In this Letter, we propose a new class of wave functions, the neural network backflow (NNB), which dresses a mean-field wave function, can make changes to the sign structure directly, and can be systematically improved by increasing the number of hidden neurons and is able to be made theoretically exact in the limit of enough neurons. To accomplish this we use a feed-forward neural network (FNN), not in the standard approach of returning a wave function amplitude, but instead to transform the single particle orbitals in a configuration dependent way; these orbitals are then used in the mean-field wave function. Wave functions with configuration-dependent orbitals are known as a backflow wave function [5–9,43–55].

Background.—Mean field theory—approximating the ground state of a quartic Hamiltonian by the ground state, ψ_{MF} , of a quadratic Hamiltonian—is a powerful first step to understanding correlated quantum systems. Various mean fields lead to different types of ground states including Slater determinants,

$$\psi_{\text{SD}}(\mathbf{r}) = \det[M^{\text{SD},\uparrow}] \det[M^{\text{SD},\downarrow}]; \quad (1)$$

$$M_{ik}^{\text{SD},\sigma} = \phi_{k\sigma}(r_{i\sigma}) \quad (2)$$

and Bogoliubov–de Gennes wave functions,

$$\psi_{\text{BDG}}(\mathbf{r}) = \det[\Phi] \quad (3)$$

$$\Phi_{ij} = \sum_{k,l=1}^N \phi_{k\uparrow}(r_{i,\uparrow}) S_{kl} \phi_{l\downarrow}(r_{j,\downarrow}), \quad (4)$$

where $\phi_{k\sigma}$ is the k th single particle orbital and $r_{i,\sigma}$ is the position of the i th particle of spin σ . Equation (1) only takes the occupied orbitals, while Eq. (3) is summed over both occupied and unoccupied orbitals.

Mean field states are uncorrelated by construction. The simplest way to capture correlation physics is through the introduction of a Jastrow giving $\psi_{\text{Jastrow}}(\mathbf{r}) = \exp[-U(\mathbf{r})]\psi_{\text{MF}}(\mathbf{r})$, where $U(\mathbf{r})$ is an arbitrary function. In this Letter, we always use a charge Jastrow $U(\mathbf{r}) = \frac{1}{2} \sum_{i,j} v_{ij} n_i n_j$, where n_i is the charge density, and v_{ij} is the variational parameters. While Jastrow factors can introduce many-body correlations, they can't modify the mean field's sign structure. One approach to add additional sign-structure modifying correlation is through a backflow correction [6–9,43–55], which introduces correlations by having the single-particle orbitals act on a configuration-dependent quasiparticle position. On the lattice, the backflow approach instead uses a configuration-dependent mean field [6,44,45,47]—i.e., the quadratic Hamiltonian or single-particle orbitals $\phi_{k\sigma}^b(r_i; \mathbf{r})$ depend not only on the position r_i but on all other electron positions \mathbf{r} ; our NNB builds on top of this formulation of backflow.

Neural network backflow.—The NNB uses a FNN to modify the single particle orbitals for a spin σ ,

$$\phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) = \phi_{k\sigma}(r_{i,\sigma}) + a_{ki,\sigma}^{\text{NN}}(\mathbf{r}), \quad (5)$$

where each value of $a_{ij,\sigma}^{\text{NN}}$ is represented by an output neuron of the FNN. We use one neural net for each of $\sigma \in \{\uparrow, \downarrow\}$. This is to be contrasted with the standard backflow [44] parametrization,

$$\begin{aligned} \phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) &= \phi_{k\sigma} + \sum_j \eta_{ij,\sigma} \phi_{k\sigma}(r_{j,\sigma}) \\ \eta_{ij,\sigma} &= t D_i H_j \theta_{|i-j|,\sigma}, \end{aligned} \quad (6)$$

with $D_i = n_{i,\uparrow} n_{i,\downarrow}$, $H_i = (1 - n_{i,\uparrow})(1 - n_{i,\downarrow})$. $\theta_{1,\sigma}$ and $\theta_{2,\sigma}$ are the only nonzero variational parameters.

Interestingly, the backflow transformation of Eq. (6) can be represented as a neural network for $a_{ij,\sigma}^{\text{NN}}(\mathbf{r})$ with three hidden layers and a linear number of neurons; an explicit construction will be given in the next section. This ensures that there exists a three layer neural network, which is at least as good as the standard backflow transformation.

We consider two NNB wave functions, Ψ_{SN} and Ψ_{PN} , implemented on top of a Slater Determinant and BCS pairing wave functions, respectively. The neural nets used in these wave functions are similar, although Ψ_{SN} has outputs that only correspond to the occupied orbitals, while the outputs of Ψ_{PN} correspond to all the orbitals. In addition, for Ψ_{SN} there are only two neural nets (one for each of the spin-up and spin-down orbitals), while for Ψ_{PN} there is an additional neural net used to generate a system dependent $S_{kl}(\mathbf{r})$. This is implemented by letting $S_{kl}(\mathbf{r}) = S_{kl} + d_{kl}^{\text{NN}}(\mathbf{r})$, where $d_{kl}^{\text{NN}}(\mathbf{r})$ is represented by an FNN (which in this Letter is always fixed to 16 hidden neurons) that inputs the system configuration \mathbf{r} and outputs the symmetric matrix correction d_{kl}^{NN} . Notice that Ψ_{PN} is trivially a superset of Ψ_{SN} .

Although various architectures can be used, we adopt a three-layer fully connected FNN for each of the functions $a_{ki,\sigma}^{\text{NN}}$ and d_{kl}^{NN} (see Fig. 1). The input layer has $2N$ neurons with neuron i (neuron $i + N$) outputting 1 if there is spin up (spin down) on site i and -1 otherwise, where N is the total system size. The hidden layer contains mN hidden neurons for constant m with rectifier linear units (ReLU) [56] activation functions. The output layer then contains $O(N^2)$ neurons specifying the values of the respective functions. Gradients are computed in the standard way using variational Monte Carlo calculations (see Supplemental Material [57] Sec. II), which requires evaluating the derivative of the wave function with respect to the weights and bias in the neural network. Derivatives for FNN are typically taken using back propagation. Because the wave function is a determinant of a matrix generated by the neural-network output, we evaluate this full derivative by envisioning this determinant as an additional final layer

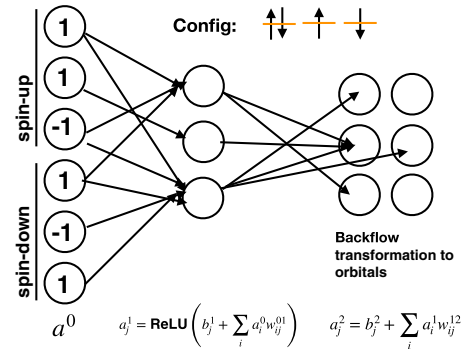


FIG. 1. Cartoon of (spin-up) neural network being used in this work for a^{NN} (other transformations are similar) with input shown for the configuration displayed for four electrons on three sites. Every layer is fully connected with arrows, but only a fraction of them are shown for image clarity. The input layer is set by the configuration. Parameters b_j and w_{ij} are bias and weights to be optimized. The output layer is a $n_{\text{up-electrons}} \times n_{\text{sites}}$ (i.e., 2×3) matrix, which will be the backflow transformation added to the single particle orbitals.

of the neural network and then performing back propagation including this layer. This ensures that the cost of computing all the derivatives is of the same order as the evaluation of the wave function (see Supplemental Material [57] Sec. III). Optimization is performed by stepping each parameter in the direction of the gradient with a random magnitude, which helps us avoid shallow local minima [58], or by the RMSPROP method [59].

The computational complexity of the NNB implemented with a single layer of $O(mN)$ hidden neurons scales as $O(mN^4)$ per sweep (i.e., after N electrons move) for forward and backward propagation and $O(N^4)$ per sweep for the evaluation of the mean-field determinant. This is similar to the scaling of standard backflow.

Explicit construction of standard backflow.—In this section, we provide an explicit construction that represents the standard backflow transformation in the form of Eq. (6) written as a NNB.

In Eq. (6), $\eta_{ij,\sigma} = tD_i H_j \theta_{|i-j|,\sigma} = tn_{i,\uparrow} n_{i,\downarrow} h_{j,\uparrow} h_{j,\downarrow} \theta_{|i-j|,\sigma}$, where $h_{j,\sigma} = 1 - n_{j,\sigma}$, and $\theta_{1,\sigma}$ and $\theta_{2,\sigma}$ are the only nonzero variational parameters. We first demonstrate that $\eta_{ij,\sigma}$ can be presented by a two layer neural network with input layer as $(\sigma_1, \dots, \sigma_N, \sigma_{N+1}, \dots, \sigma_{2N})$ where $\sigma_i = 2n_{i,\uparrow} - 1$ and $\sigma_{i+N} = 2n_{i,\downarrow} - 1$. By construction, $n_{i,\sigma}$ and $h_{j,\sigma}$ take a value of 0 or 1 so that $n_{i,\uparrow} n_{i,\downarrow} h_{j,\uparrow} h_{j,\downarrow}$ is 1 if and only if $n_{i,\uparrow} = n_{i,\downarrow} = h_{j,\uparrow} = h_{j,\downarrow} = 1$. Therefore, $t\theta_{|i-j|,\sigma} D_i H_j = t\theta_{|i-j|,\sigma} n_{i,\uparrow} n_{i,\downarrow} h_{j,\uparrow} h_{j,\downarrow}$ is equivalent to $\text{ReLU}[t\theta_{|i-j|,\sigma}(n_{i,\uparrow} + n_{i,\downarrow} + h_{j,\uparrow} + h_{j,\downarrow} - 3)]$, which is the same as $\text{ReLU}[t\theta_{|i-j|,\sigma}(\sigma_i/2 + \sigma_{i+N}/2 - \sigma_j/2 - \sigma_{j+N}/2 - 1)]$. As a result, for each $\eta_{ij,\sigma}$, we associate it with a hidden neuron, such that the weights connecting it to $\sigma_i, \sigma_{i+N}, \sigma_j, \sigma_{j+N}$ are $t\theta_{|i-j|,\sigma}/2, t\theta_{|i-j|,\sigma}/2, -t\theta_{|i-j|,\sigma}/2, -t\theta_{|i-j|,\sigma}/2$, respectively, the bias is $-t\theta_{|i-j|,\sigma}$, and the activation function is ReLU. In general, for more complicated backflow [6,45,47] with terms $n_{i,\sigma} h_{i,-\sigma} n_{j,-\sigma} h_{j,\sigma}$, $n_{i,\sigma} n_{i,-\sigma} n_{j,-\sigma} h_{j,\sigma}$, and $n_{i,\sigma} h_{i,-\sigma} h_{j,\sigma} h_{j,-\sigma}$, where σ is the spin index, we can use more hidden neurons and represent it in the same way.

After we have the neural network construction for the standard $\eta_{ij,\sigma}$, the term $a_{ki}^{\text{NN}} = \sum_j \eta_{ij,\sigma} \phi_{k\sigma}(r_{j,\sigma})$ in Eq. (5) can be realized through an extra layer taking the outputs $\eta_{ij,\sigma}$ to a neuron representing a_{ki} ; where the weight is given by the single particle orbital values $\phi_{k\sigma}(r_{j,\sigma})$, there is no bias and the activation function is the identity. This

construction shows that the standard backflow parameterization is thus a subset of our three-layer NNB.

Results.—We have benchmarked the quality of our NNB on a number of systems including Hubbard models at various sizes and doping (all at $U/t = 8$) as well as a frustrated magnet, the Heisenberg model on the Kagome lattice. In the main text we focus primarily on the Hubbard model at $n = 0.875$ (primarily on the 4×4 lattice) leaving the additional benchmarks as Supplemental Material [57] (Sec. IV., [60–62]) The Hubbard Hamiltonian is

$$H = -t \sum_{i\sigma} (c_{i\sigma}^\dagger c_{i+1\sigma} + \text{H.c.}) + \sum_i U n_{i\uparrow} n_{i\downarrow}, \quad (7)$$

where we use $U/t = 8$. We compare the results to an optimized unrestricted (i.e., different single particle orbitals for spin up and spin down) Slater determinant (Ψ_{SO}) as well as a backflow BDG wave function (Ψ_{PB}), which transforms the single particle orbitals of each spin by Eq. (6). The formulation and the variational parameters of each wave function *Ansatz* are summarized in Table I. The parameters that aren't optimized, such as the initial set of orbitals $\{\phi_{k\sigma}(r_{i,\sigma})\}$ are obtained for Ψ_{PB} and Ψ_{SN} by optimizing a restricted Slater-Jastrow wave function while Ψ_{PN} uses orbitals taken from the free hopping Hamiltonian (in practice the nature of the neural net allows for a direct change to the orbitals by altering the bias on the final layer).

The relative error of the energy of NNB is 1.4% and 0.66% after variance extrapolation (see Supplementary Material [57] Sec. I), which is significantly better than the standard wave functions [see Fig. 2(left)]. We examine the effect of the number of hidden neurons mN [see Fig. 2(right)]. We find that at a small hidden neuron number, Ψ_{PN} is much better than Ψ_{SN} , but this advantage eventually largely disappears at a large neuron number, suggesting that a backflow parametrized with small neural networks can compensate for the missing pairing in a Slater determinant. Surprisingly, in the regime we've probed, both NNB have energies linear with respect to $1/m$ in spite of the fact that in the $m \rightarrow \infty$ limit, they both must become exact as the FNN could simply put the exact amplitude $\Psi(R)$ on one element of the diagonal [63], one on the rest of the diagonal, and zero everywhere else.

In addition, we have studied the NNB at $4 \times L$ for $L = \{4, 8, 12, 16\}$ comparing against the DMRG energy (PBC, open) for the $4 \times \infty$ system of $-0.7659 \pm 4 \times 10^{-5}$ per site

TABLE I. Wave function *Ansätze*

Method	Backflow transformation	Mean field	Variational functions
ψ_{SO}	not applicable	Eq. (1)	$\phi_{k\uparrow}(r_{i,\uparrow}), \phi_{k\downarrow}(r_{i,\downarrow}), v_{ij}$
Ψ_{SN}	$\phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) = \phi_{k\sigma}(r_{i,\sigma}) + a_{ki,\sigma}^{\text{NN}}(\mathbf{r})$	Eq. (1)	$a_{ki,\sigma}^{\text{NN}}(\mathbf{r}), v_{ij}$
Ψ_{PB}	$\phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) = \phi_{k\sigma}(r_{i,\sigma}) + \theta_{1\sigma} \sum_j t D_i H_j \phi_{k\sigma}(r_{j,\sigma}) + \theta_{2\sigma} \sum_j t D_i H_j \phi_{k\sigma}(r_{j,\sigma})$	Eq. (3)	$\theta_{1\sigma}, \theta_{2\sigma}, S_{kl}, v_{ij}$
Ψ_{PN}	$\phi_{k\sigma}^b(r_{i,\sigma}; \mathbf{r}) = \phi_{k\sigma}(r_{i,\sigma}) + a_{ki,\sigma}^{\text{NN}}(\mathbf{r}); S_{kl}(\mathbf{r}) = S_{kl} + d_{kl}^{\text{NN}}(\mathbf{r})$	Eq. (3)	$a_{ki,\sigma}^{\text{NN}}(\mathbf{r}), d_{kl}^{\text{NN}}(\mathbf{r}), v_{ij}$

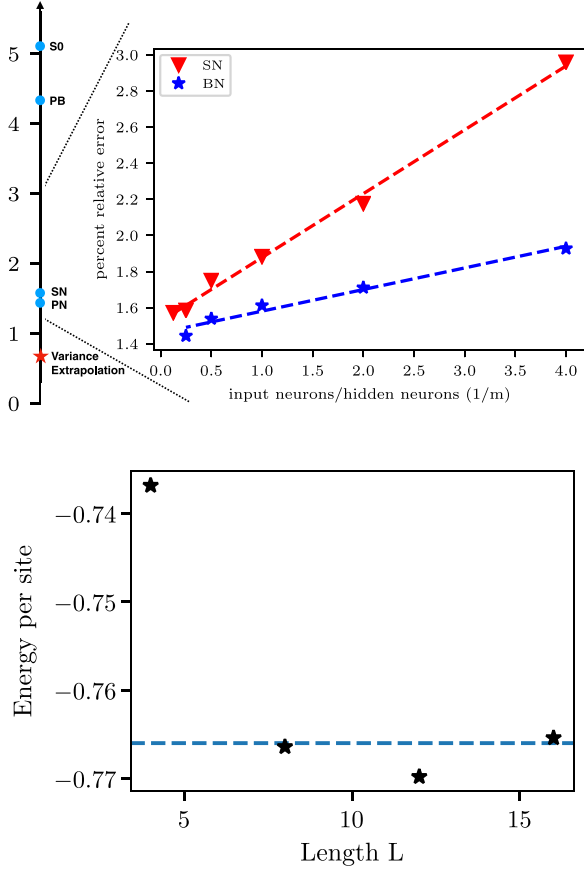


FIG. 2. Top left: Percentage relative error from the exact ground-state energy of Eq. (7) ($E = -11.868$ [64]) on 4×4 Hubbard model at $U/t = 8$, $n = 0.875$, for various classes of wave functions. The star is the variance extrapolation result of Ψ_{PN} (see Supplemental Material [57] Sec. I). Top right: Percentage relative energy error, $(E_{\text{exact}} - E_{\text{NNB}})/E_{\text{exact}} \times 100\%$, as a function of $1/m$ for NNB. Statistical error bars are shown, but are smaller than the marker size. Bottom: Variance extrapolated energy per site for Hubbard model with $U/t = 8$, $n = 0.875$ with system size $L \times 4$ for $L = 4, 8, 12, 16$. The dash line is the DMRG energy per site ($-0.7659 \pm 4 \times 10^{-5}$) for system size $\infty \times 4$ (PBC, open) [65].

[65]. We find our result very comparable [see Fig. 2(bottom)] to the DMRG result, especially for system sizes that are commensurate with a wavelength of eight stripes [65].

We also investigate how the neural network backflow wave function affects the observables of our system. We see the expectation of double occupancy decreases as $1/m$ (see Supplemental Material [57] Sec. I), and the spin and charge densities become significantly more symmetric as the number of hidden neurons increases (see Fig. 3).

To understand the role of neural network in backflow transformation, we investigate how the neural network backflow transformation modify the orbitals. In Fig. 4(right), we notice that, although the backflow transformation on spin-up orbitals and spin-down orbitals are performed by two different neural networks, they produce

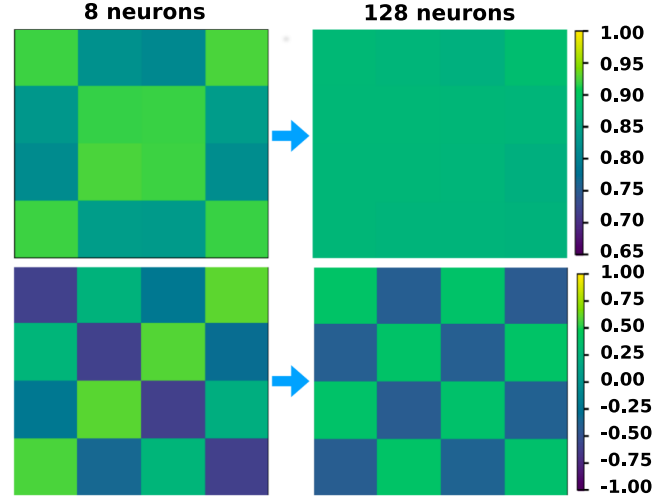


FIG. 3. Charge density (top) and spin density (bottom) from Ψ_{PN} with 8 hidden neurons (left) and 128 hidden neurons (right) on 4×4 Hubbard model at $U/t = 8$, $n = 0.875$.

similar backflow transformed orbitals and roughly preserve the symmetry of spin up and spin down for a given configuration. This is different from the optimized unrestricted Slater determinant Ψ_{S0} , which breaks the spin-up and spin-down symmetry significantly [see Fig. 4(left)].

One feature of using a NNB is the ability to alter the sign structure of the wave function. Here we consider the amount the sign changes between Ψ_{SN} with 16 hidden neurons and Ψ_{S0} by evaluating the integral

$$\frac{\int |\Psi_{\text{S0}}(x)|^2 \text{sgn}(\Psi_{\text{SN}}(x)) \text{sgn}(\Psi_{\text{S0}}(x)) dx}{\int |\Psi_{\text{S0}}(x)|^2 dx}, \quad (8)$$

which is approximately 0.815, giving a 9% difference between the signs.

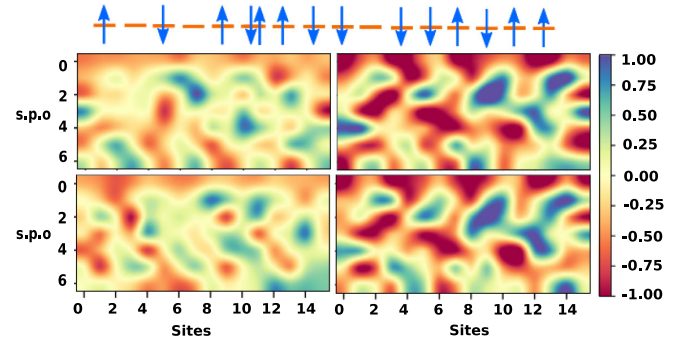


FIG. 4. Spin-up (top) and spin-down (bottom) single particle orbital (s.p.o.) for Ψ_{S0} (left) and Ψ_{SN} (right) with 256 hidden neurons on 4×4 Hubbard model at $U/t = 8$, $n = 0.875$, where the s.p.o.s are evaluated at the (1D reshaped) spin-configuration shown. For the s.p.o., row is orbital index and the column is position index.

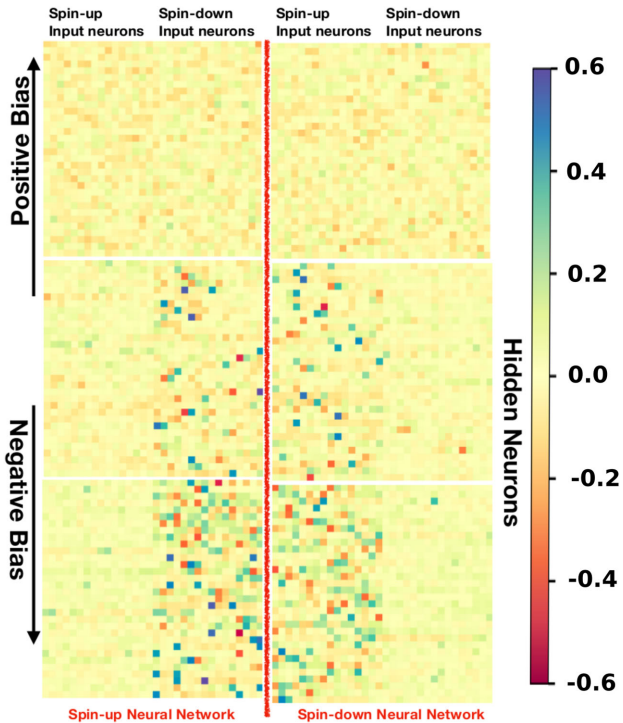


FIG. 5. Weights between the input layer and hidden layer for Ψ_{SN} with 256 hidden neurons for the spin-up (left) and spin-down (right) neural networks on 4×4 Hubbard model at $U/t = 8, n = 0.875$. Hidden neurons are ordered by their bias and shown are neurons 1–32 (top), 96–128 (middle), and 224–256 (bottom).

Furthermore, we open up the Ψ_{SN} neural network for $m = 8$ and analyze the weight between the input layer and the hidden layer, which represents the features that the neural network learns from input. In Fig. 5, we plot these weights for both the spin-up and spin-down neural networks. Interestingly the spin-up neural network primarily has large weights connected to the spin-down configurations, while the spin-down neural network primarily has large weights connected to the spin-up configurations. This allows the neural network to introduce correlation between spin-up and spin-down configurations. Another observation is that more neurons tend to take a large weight in negative bias, and a small weight in positive bias.

Conclusion.—In this Letter, we utilize the generality of artificial neural networks and the physical insight from backflow to develop a new class of wave function *Ansatz*, the neural network backflow wave function, for strongly correlated Fermion systems on lattice. It achieves good performance for Hubbard model at nontrivial filling. We also show improvement on a kagome Heisenberg model in the supplement. While this Letter has focused on a Fermion system on the lattice, the NNB is straightforward to generalize to frustrated spin systems as well as the continuum. In the latter case, the input could be represented as a lexicographically ordered set of particle locations. Our

Letter provides a new approach toward combining machine learning methodology with dressed mean-field variational wave functions, which allows us to take simultaneous advantage of their respective strengths.

This project is part of the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (Grants No. OCI-0725070 and ACI-1238993) and the State of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. This material is based upon work supported by the U.S. Department of Energy, Office of Science under Award No. FG02-12ER46875. Di acknowledges useful discussion with Ryan Levy, Dmitrii Kochkov, Eli Chertkov, Yusuke Nomura, and Giuseppe Carleo.

- [1] J. C. Slater, *Phys. Rev.* **34**, 1293 (1929).
- [2] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957).
- [3] R. Jastrow, *Phys. Rev.* **98**, 1479 (1955).
- [4] L. Mitáš and R. M. Martin, *Phys. Rev. Lett.* **72**, 2438 (1994).
- [5] R. P. Feynman and M. Cohen, *Phys. Rev.* **102**, 1189 (1956).
- [6] L. F. Tocchio, F. Becca, A. Parola, and S. Sorella, *Phys. Rev. B* **78**, 041101(R) (2008).
- [7] Y. Kwon, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **48**, 12037 (1993).
- [8] M. Taddei, M. Ruggeri, S. Moroni, and M. Holzmann, *Phys. Rev. B* **91**, 115106 (2015).
- [9] M. Ruggeri, S. Moroni, and M. Holzmann, *Phys. Rev. Lett.* **120**, 205302 (2018).
- [10] S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [11] G. Vidal, *Phys. Rev. Lett.* **91**, 147902 (2003).
- [12] G. Vidal, *Phys. Rev. Lett.* **99**, 220405 (2007).
- [13] F. Verstraete, M. M. Wolf, D. Perez-Garcia, and J. I. Cirac, *Phys. Rev. Lett.* **96**, 220601 (2006).
- [14] D. A. Huse and V. Elser, *Phys. Rev. Lett.* **60**, 2531 (1988).
- [15] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, *New J. Phys.* **11**, 083026 (2009).
- [16] K. H. Marti, B. Bauer, M. Reiher, M. Troyer, and F. Verstraete, *New J. Phys.* **12**, 103008 (2010).
- [17] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Phys. Rev. Lett.* **100**, 040501 (2008).
- [18] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [19] X. Gao and L.-M. Duan, *Nat. Commun.* **8**, 662 (2017).
- [20] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. X* **7**, 021021 (2017).
- [21] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, *Phys. Rev. X* **8**, 011006 (2018).
- [22] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *Phys. Rev. B* **97**, 085104 (2018).
- [23] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, *Phys. Rev. B* **96**, 205152 (2017).
- [24] G. Carleo, Y. Nomura, and M. Imada, *Nat. Commun.* **9**, 5322 (2018).
- [25] N. Freitas, G. Morigi, and V. Dunjko, [arXiv:1803.02118](https://arxiv.org/abs/1803.02118).
- [26] Z. Cai and J. Liu, *Phys. Rev. B* **97**, 035116 (2018).

- [27] K. Choo, G. Carleo, N. Regnault, and T. Neupert, *Phys. Rev. Lett.* **121**, 167204 (2018).
- [28] J. Han, L. Zhang, and E. Weinan, [arXiv:1807.07014](https://arxiv.org/abs/1807.07014).
- [29] A. Rocchetto, E. Grant, S. Strelchuk, G. Carleo, and S. Severini, *npj Quantum Inf.* **4**, 28 (2018).
- [30] H. Saito, *J. Phys. Soc. Jpn.* **87**, 074002 (2018).
- [31] H. Saito and M. Kato, *J. Phys. Soc. Jpn.* **87**, 014001 (2018).
- [32] Y. Huang and J. E. Moore, [arXiv:1701.06246](https://arxiv.org/abs/1701.06246).
- [33] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. B* **96**, 195145 (2017).
- [34] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *Phys. Rev. B* **97**, 085104 (2018).
- [35] S. R. Clark, *J. Phys. A* **51**, 135301 (2018).
- [36] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- [37] R. Kaubruegger, L. Pastori, and J. C. Budich, *Phys. Rev. B* **97**, 195136 (2018).
- [38] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, *Phys. Rev. B* **98**, 104426 (2018).
- [39] B. Jónsson, B. Bauer, and G. Carleo, [arXiv:1808.05232](https://arxiv.org/abs/1808.05232).
- [40] L. Pastori, R. Kaubruegger, and J. C. Budich, *Phys. Rev. B* **99**, 165123 (2019).
- [41] S. Lu, X. Gao, and L. M. Duan, *Phys. Rev. B* **99**, 155136 (2019).
- [42] D. Kochkov and B. K. Clark, [arXiv:1811.12423](https://arxiv.org/abs/1811.12423).
- [43] M. Holzmann, D. M. Ceperley, C. Pierleoni, and K. Esler, *Phys. Rev. E* **68**, 046707 (2003).
- [44] L. F. Tocchio, F. Becca, and C. Gros, *Phys. Rev. B* **83**, 195138 (2011).
- [45] F. Becca, L. F. Tocchio, and S. Sorella, *J. Phys. Conf. Ser.* **145**, 012016 (2009).
- [46] G. Carleo, S. Moroni, F. Becca, and S. Baroni, *Phys. Rev. B* **83**, 060411(R) (2011).
- [47] L. F. Tocchio, A. Parola, C. Gros, and F. Becca, *Phys. Rev. B* **80**, 064419 (2009).
- [48] N. D. Drummond, P. L. Ros, A. Ma, J. R. Trail, G. G. Spink, M. D. Towler, and R. J. Needs, *J. Chem. Phys.* **124**, 224104 (2006).
- [49] L. F. Tocchio, F. Becca, and C. Gros, *Phys. Rev. B* **81**, 205109 (2010).
- [50] Y. Kwon, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **58**, 6800 (1998).
- [51] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, *Phys. Rev. E* **74**, 066701 (2006).
- [52] M. A. Lee, K. E. Schmidt, M. H. Kalos, and G. V. Chester, *Phys. Rev. Lett.* **46**, 728 (1981).
- [53] K. E. Schmidt, M. A. Lee, M. H. Kalos, and G. V. Chester, *Phys. Rev. Lett.* **47**, 807 (1981).
- [54] R. M. Panoff and J. Carlson, *Phys. Rev. Lett.* **62**, 1130 (1989).
- [55] M. Holzmann, B. Bernu, and D. M. Ceperley, *Phys. Rev. B* **74**, 104510 (2006).
- [56] B. Xu, N. Wang, T. Chen, and M. Li, [arXiv:1505.00853](https://arxiv.org/abs/1505.00853).
- [57] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.122.226401> for detailed calculation and additional data.
- [58] J. Lou and A. W. Sandvik, *Phys. Rev. B* **76**, 104432 (2007).
- [59] S. Ruder, [arXiv:1609.04747](https://arxiv.org/abs/1609.04747).
- [60] A. M. Läuchli, J. Sudan, and R. Moessner, [arXiv:1611.06990](https://arxiv.org/abs/1611.06990).
- [61] B. K. Clark, J. M. Kinder, E. Neuscamman, G. K.-L. Chan, and M. J. Lawler, *Phys. Rev. Lett.* **111**, 187205 (2013).
- [62] J. P. F. LeBlanc *et al.* (Simons Collaboration), *Phys. Rev. X* **5**, 041041 (2015).
- [63] G. Cybenko, *Math. Control Signals Syst.* **2**, 303 (1989).
- [64] E. Dagotto, A. Moreo, F. Ortolani, D. Poilblanc, and J. Riera, *Phys. Rev. B* **45**, 10741 (1992).
- [65] B.-X. Zheng, C.-M. Chung, P. Corboz, G. Ehlers, M.-P. Qin, R. M. Noack, H. Shi, S. R. White, S. Zhang, and G. K.-L. Chan, *Science* **358**, 1155 (2017).