Solving Many-Electron Schrödinger Equation Using Deep Neural Networks

Jiequn Han and Linfeng Zhang[*]

Program in Applied and Computational Mathematics,
Princeton University, Princeton, NJ 08544, USA

Weinan E[†]

Department of Mathematics and Program in Applied and Computational Mathematics,
Princeton University, Princeton, NJ 08544, USA and
Center for Data Science, Peking University, Beijing Institute of Big Data Research, Beijing, 100871, P.R. China

We introduce a new family of trial wave-functions based on deep neural networks to solve the many-electron Schrödinger equation. The Pauli exclusion principle is dealt with explicitly to ensure that the trial wave-functions are physical. The optimal trial wave-function is obtained through variational Monte Carlo and the computational cost scales quadratically with the number of electrons. The algorithm does not make use of any prior knowledge such as atomic orbitals. Yet it is able to represent accurately the ground-states of the tested systems, including He, H_2, Be, B, LiH, and a chain of 10 hydrogen atoms. This opens up new possibilities for solving large-scale many-electron Schrödinger equation.

An accurate quantum mechanical treatment of the interaction between many electrons and ions is the foundation for modeling physical, chemical, and biological systems. Theoretically, these systems are described by the many-electron Schrödinger equation, which consists of the kinetic and Coulomb interaction terms [1]. However, solving the Schrödinger equation accurately for real physical systems has been prohibitively difficult due to the high dimensionality of the Hilbert space involved, the high degree of entanglement produced by the electron-electron and electron-ion interactions, and the notorious Pauli exclusion principle imposed on the wave-function, i.e., the wave-function has to change its sign when two identical electrons exchange places [2].

Developing efficient algorithms for this problem is among the most heroic endeavors in computational science and has achieved remarkable success. An incomplete list of the major methodologies developed so far includes the Hartree-Fock (HF) based methods [3, 4], configuration interaction (CI) based methods [5–8], coupled cluster (CC) based schemes [9–12], Monte Carlo-based approaches [13–20], and the more recently developed density matrix renormalization group (DMRG) theory [21–25] and density matrix embedding theory (DMET) [26–28]. We refer to Refs. [29–31] for a more detailed review of these advances.

Of particular interest to this work is the variational Monte Carlo (VMC) scheme which uses variational principle and Monte Carlo sampling to obtain the best parametrized trial wave-function [13–15, 20]. Naturally the key component is the representation of the trial wave-functions. The most commonly used trial wave-functions typically consist of an anti-symmetric Slater determinant [32] multiplied by a symmetric Jastrow correlation factor [33]. There have been tremendous efforts on improving the nodal surface (a subspace on which the function value equals zero and across which it changes the sign) of the anti-symmetric part of the trial wave-function and the representability of the symmetric part [34–37].

With remarkable successes in many fields such as computer vision and speech recognition, the deep neural network (DNN) has shown great capacity in approximating high-dimensional functions (see e.g. review [38] and the references therein). Furthermore, DNN has been successfully used in solving general high-dimensional partial differential equations [39–42] and certain quantum many-body problems for Bosonic and lattice systems [43–47]. However, there have been few attempts to solve the many-electron Schrödinger equations based on DNN, and this constitutes the main objective of this work.

To achieve this, we develop a general and efficient DNN representation for the many-electron wave-function satisfying the Pauli exclusion principle. The resulted trial wave-function can naturally fit into the framework of VMC to optimize the parameters in our model. As preliminary tests, we show that this DNN-based trial wave-function is able to produce reasonably well ground-state energies for some small systems, such as Be, B, LiH, and a chain of 10 hydrogen atoms (H_{10}). In addition, learning from scratch without any prior knowledge and without resorting to a reference of atomic bases, the DNN-based wave-function is able to reproduce the electronic structures of the tested systems. We call the methodology introduced here the *DeepWF* method. This paper only reports our initial results. There is still a huge room for improvement.

We consider a system of N electrons and M ions, under the Born-Oppenheimer approximation [48]. This system
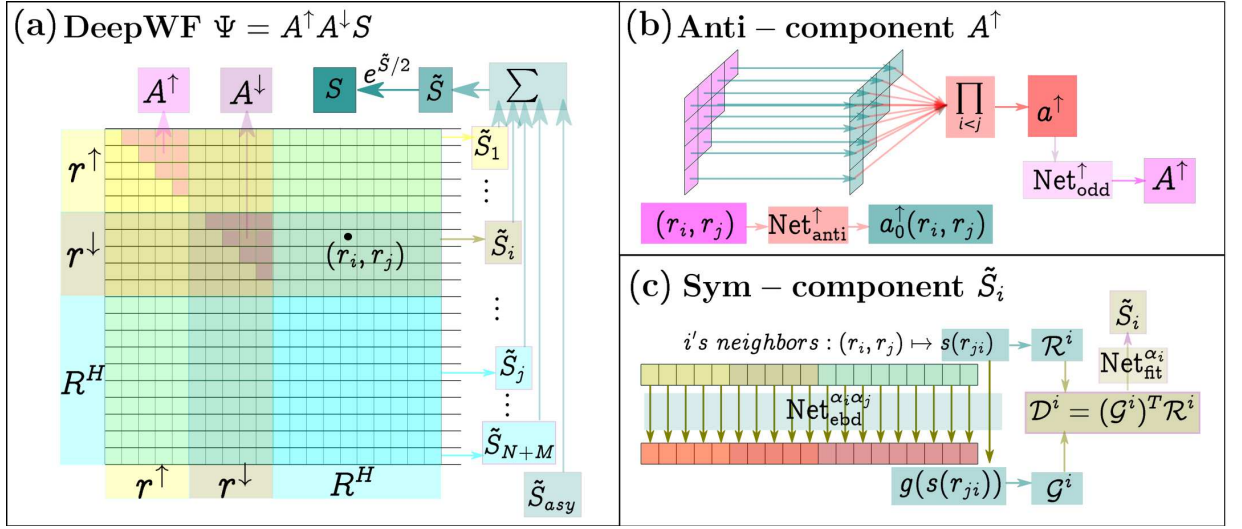
FIG. 1: (color online). Schematic plot of DeepWF, taking H_{10} as an example. $\boldsymbol{r}^\uparrow$ and $\boldsymbol{r}^\downarrow$ denote the positions of 5 electrons of spin-up and 5 of spin-down, respectively, and $\boldsymbol{R}^H$ denotes the positions of 10 hydrogen atoms. (a) the entire wave-function $\Psi(\boldsymbol{r};\boldsymbol{R}) = S(\boldsymbol{r};\boldsymbol{R})A^\uparrow(\boldsymbol{r}^\uparrow)A^\downarrow(\boldsymbol{r}^\downarrow)$, decomposed into symmetric function $S(\boldsymbol{r};\boldsymbol{R})$ and two anti-symmetric functions $A^\uparrow(\boldsymbol{r}^\uparrow), A^\downarrow(\boldsymbol{r}^\downarrow)$, with respect to spin-up and spin-down electrons; (b) anti-symmetric function $A^\uparrow(\boldsymbol{r}^\uparrow)$ for spin-up electrons; (c) symmetric function $\tilde{S}_i(\mathcal{R}^i)$ as the contribution of particle i in the symmetric function $\log(S^2(\boldsymbol{r};\boldsymbol{R}))$. See text for details.

is described by the Hamiltonian

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$$
$$-\sum_{I=1}^{M}\sum_{i=1}^{N}\frac{Z_I}{|\boldsymbol{r}_i - \boldsymbol{R}_I|} + \sum_{I=1}^{M}\sum_{J=I+1}^{M}\frac{Z_I Z_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|}, \quad (1)$$

where $\boldsymbol{r} = (\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ and $\boldsymbol{R} = (\boldsymbol{R}_1, \ldots, \boldsymbol{R}_M)$ are the coordinates of the electrons and the ions, respectively, and Z_I denotes the nuclear charge. Since $\hat{H}$ is spin-independent, it is valid to assume that the first $N_\uparrow$ electrons are of spin-up and the remaining $N_\downarrow = N - N_\uparrow$ electrons are of spin-down [20]. Accordingly, we can write the wave-function in a spin-independent form $\Psi(\boldsymbol{r};\boldsymbol{R})$. Let N_{tp} be the number of ion types in the system, then there are $N_{\text{tp}} + 2$ types of particles in total, taking into account the electrons of spin-up and spin-down separately.

In this work we restrict our attention to the ground state of the system. The variational principle states that the wave-function associated with the ground state minimizes within the required symmetry the following energy functional

$$E[\Psi] = \frac{\int \Psi^*(\boldsymbol{r};\boldsymbol{R})\hat{H}\Psi(\boldsymbol{r};\boldsymbol{R})\,d\boldsymbol{r}}{\int \Psi^*(\boldsymbol{r};\boldsymbol{R})\Psi(\boldsymbol{r};\boldsymbol{R})\,d\boldsymbol{r}}.$$

We now discuss how to represent the wave-function $\Psi(\boldsymbol{r};\boldsymbol{R})$ with DNN. The trial wave-functions are assumed to be real-valued. The primary goal is to ensure that the represented wave-function satisfies the anti-symmetry property. To this end, we decompose the wave-function as $\Psi(\boldsymbol{r};\boldsymbol{R}) = S(\boldsymbol{r};\boldsymbol{R})A^\uparrow(\boldsymbol{r}^\uparrow)A^\downarrow(\boldsymbol{r}^\downarrow)$, where

$\boldsymbol{r}^\uparrow, \boldsymbol{r}^\downarrow$ denote the positions of spin-up and spin-down electrons, respectively. We require $S(\boldsymbol{r},\boldsymbol{R})$ to be symmetric and $A^\uparrow(\boldsymbol{r}^\uparrow), A^\downarrow(\boldsymbol{r}^\downarrow)$ to be both anti-symmetric. As an analogy, one can view $S(\boldsymbol{r};\boldsymbol{R})$ as a Jastrow factor and $A^\uparrow(\boldsymbol{r}^\uparrow), A^\downarrow(\boldsymbol{r}^\downarrow)$ as Slater determinant-like functions. In practice, $\log(\Psi^2(\boldsymbol{r};\boldsymbol{R}))$ is more convenient for implementing VMC. From this perspective, the introduced decomposition of wave-function becomes

$$\log(\Psi^2(\boldsymbol{r};\boldsymbol{R}))$$
$$=\tilde{S}(\boldsymbol{r};\boldsymbol{R}) + \log(|A^\uparrow(\boldsymbol{r}^\uparrow)|^2) + \log(|A^\downarrow(\boldsymbol{r}^\downarrow)|^2), \quad (2)$$

where $\tilde{S}(\boldsymbol{r};\boldsymbol{R}) := \log(S^2(\boldsymbol{r};\boldsymbol{R}))$ is still a general symmetric function. On the other hand, although being symmetric, $\log(|A^\uparrow(\boldsymbol{r}^\uparrow)|^2)$ and $\log(|A^\downarrow(\boldsymbol{r}^\downarrow)|^2)$ have singularities on the nodal surface. Therefore we need to secure this property by representing $A^\uparrow(\boldsymbol{r}^\uparrow)$ and $A^\downarrow(\boldsymbol{r}^\downarrow)$ directly. In the following we discuss how to represent $A^\uparrow(\boldsymbol{r}^\uparrow), A^\downarrow(\boldsymbol{r}^\downarrow)$, and $\tilde{S}(\boldsymbol{r};\boldsymbol{R})$ with DNN and how to learn the optimal parameters using VMC. See Fig. 1 for a schematic illustration.

The basic building block for the anti-symmetric function $A^\uparrow(\boldsymbol{r}^\uparrow)$ is an ansatz

$$\boldsymbol{a}^\uparrow(\boldsymbol{r}^\uparrow) = \prod_{1\leq i < j \leq N_\uparrow} \boldsymbol{a}_0^\uparrow(\boldsymbol{r}_i, \boldsymbol{r}_j), \quad (3)$$

where $\boldsymbol{a}_0^\uparrow$ is a general two-body anti-symmetric function with M_1-dimensional output, i.e., $\boldsymbol{a}_0^\uparrow(\boldsymbol{r}_i, \boldsymbol{r}_j) = -\boldsymbol{a}_0^\uparrow(\boldsymbol{r}_j, \boldsymbol{r}_i)$, and the products are component-wise. In (3) we have assumed $N_\uparrow \geq 2$. Such representation can be viewed as a generalization of the Laughlin wave-

function [49], which describes well the anomalous quantum Hall effect. A related model is the so-called correlator product state (CPS) [37], which works well for lattice systems and has natural connections with some recently proposed neural-network quantum states [50, 51]. In practice, we let $\boldsymbol{a}_0^\uparrow(\boldsymbol{r}_i, \boldsymbol{r}_j) = \mathrm{Net}_{\mathrm{anti}}^\uparrow(\boldsymbol{r}_i, \boldsymbol{r}_j, r_{ji}) - \mathrm{Net}_{\mathrm{anti}}^\uparrow(\boldsymbol{r}_i, \boldsymbol{r}_j, r_{ji})$, where $\mathrm{Net}_{\mathrm{anti}}^\uparrow(\cdot)$ is represented by a DNN. By convention $r_{ji} = |\boldsymbol{r}_{ji}| = |\boldsymbol{r}_j - \boldsymbol{r}_i|$ denotes the Euclidean distance between particles i and j. Next $\boldsymbol{a}^\uparrow$ is fed into another DNN $\mathrm{Net}_{\mathrm{odd}}^\uparrow$, which outputs a scalar and satisfies $\mathrm{Net}_{\mathrm{odd}}^\uparrow(\boldsymbol{x}) = -\mathrm{Net}_{\mathrm{odd}}^\uparrow(-\boldsymbol{x})$. We can further adjust its scale and define $A^\uparrow(\boldsymbol{r}^\uparrow)$ through $\log|A^\uparrow(\boldsymbol{r}^\uparrow)| = f_{\mathrm{odd}}\log|\mathrm{Net}_{\mathrm{odd}}^\uparrow(\boldsymbol{a}^\uparrow(\boldsymbol{r}))|$ where f_{odd} is a positive scalar factor. The sign of $A^\uparrow(\boldsymbol{r}^\uparrow)$ is taken as the same as $\mathrm{Net}_{\mathrm{odd}}^\uparrow(\boldsymbol{a}^\uparrow(\boldsymbol{r}^\uparrow))$ such that $A^\uparrow(\boldsymbol{r}^\uparrow)$ inherits the anti-symmetry. The same construction introduced above are applied to spin-down electrons and yield $A^\downarrow(\boldsymbol{r}^\downarrow)$.

While DNN-based anti-symmetric function has seldom been investigated in the literature, DNN-based symmetric function has gained wide attentions recently in modeling many-body potential energy surface [52–56], free energy surface [57–59], etc. For instance, the Deep Potential-Smooth Edition (DeepPot-SE) model is able to efficiently describe the interatomic potential energy of both finite and extended systems [55]. A crucial step there is to faithfully map the input atomic coordinates onto a symmetry-preserving feature space. Therefore, it shares the symmetric property of the Jastrow factor, and might become a more general Jastrow factor that is able to better capture the many-body correlations between the electrons and ions.

We extend the DeepPot-SE model to construct $\tilde{S}(\boldsymbol{r}; \boldsymbol{R})$. For later convenience we regard $\boldsymbol{R}$ as $(\boldsymbol{r}_{N+1}, \ldots, \boldsymbol{r}_{N+M})$ and let $\bar{\boldsymbol{r}} := (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_{N+M})$. For each particle i, we consider its neighboring environment in terms of the Cartesian coordinates:

$$\mathcal{R}^i = \{\boldsymbol{s}(\boldsymbol{r}_{1i})^T, \cdots, \boldsymbol{s}(\boldsymbol{r}_{ji})^T, \cdots\}^T \in \mathbb{R}^{(N+M-1)\times 4}, \quad (4)$$

where $\boldsymbol{s}(\boldsymbol{r}_{ji}) = (\boldsymbol{r}_{ji}/r_{ji}^2, 1/r_{ji}) \in \mathbb{R}^4$, $j \in \{1, \ldots, N + M\}\backslash\{i\}$, are extended relative coordinates. Then we decompose $\tilde{S}(\boldsymbol{r}; \boldsymbol{R})$ into the sum of contributions from each particle and an extra asymptotic term, i.e.,

$$\tilde{S}(\boldsymbol{r}; \boldsymbol{R}) = \tilde{S}_{\mathrm{asy}}(\bar{\boldsymbol{r}}) + \sum_{i=1}^{N+M} \tilde{S}_i(\mathcal{R}^i). \quad (5)$$

We define $\tilde{S}_i$ in a way that guarantees the symmetric property. We first construct a series of embedding networks $\mathrm{Net}_{\mathrm{ebd}}^{\alpha_i\alpha_j}$, each mapping the extended relative coordinates $\boldsymbol{s}(\boldsymbol{r}_{ji})$ to M_2 components. Here α_i denotes the type of particle i, and the parameters of $\mathrm{Net}_{\mathrm{ebd}}^{\alpha_i\alpha_j}$ depend on the types of both particle i and its neighbor particle j. The embedding matrix $\mathcal{G}^i \in \mathbb{R}^{(N+M-1)\times M_2}$ and the encoded feature matrix $\mathcal{D}^i \in \mathbb{R}^{M_2\times 4}$ of particle i are then defined as

$$(\mathcal{G}^i)_{jk} = (\mathrm{Net}_{\mathrm{ebd}}^{\alpha_i\alpha_j}(\boldsymbol{s}(\boldsymbol{r}_{ji})))_k, \quad (6)$$

$$\mathcal{D}^i = (\mathcal{G}^i)^T \mathcal{R}^i. \quad (7)$$

By definition one can check that the feature matrix $\mathcal{D}^i$ is invariant under the permutation of particles in the same type, and its $M_2 \times 4$ components are finally reshaped into a vector to serve as the input of a fitting network $\mathrm{Net}_{\mathrm{fit}}^{\alpha_i}$ and yield $\tilde{S}_i$.

The remaining term $\tilde{S}_{\mathrm{asy}}(\bar{\boldsymbol{r}})$ should also be symmetric and serve the additional purposes of making the trial wave-function close to some desired asymptotic properties. This term has the form

$$\tilde{S}_{\mathrm{asy}}(\bar{\boldsymbol{r}}) = -\sum_{k=N+1}^{N+M} Z_{k-N}\sum_{i=1}^{N}\frac{r_{ik} + f_{\mathrm{dec}}r_{ik}^2}{r_{ik} + 0.5}$$
$$+ \sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{r_{ij}}{r_{ij}+1}, \quad (8)$$

where f_{dec} is a positive scalar as decay factor. The first purpose is to ensure that $\Psi(\boldsymbol{r}; \boldsymbol{R})$ is square integrable when f_{dec} is large enough such that $\Psi(\boldsymbol{r}; \boldsymbol{R})$ has exponential decay for the electrons that are far away. The second purpose is related to the so-called *cusp* conditions, meaning that the trial wave-function has the correct singular behavior when two particles are close to each other [60]. If we let $\log(\Psi^2(\boldsymbol{r}; \boldsymbol{R})) = \tilde{S}_{\mathrm{asy}}(\bar{\boldsymbol{r}})$, it can be shown from (8) that for $1 \le i < j \le N$, $N + 1 \le k \le N + M$,

$$\frac{1}{\Psi}\frac{\partial\Psi}{\partial r_{ik}}\bigg|_{r_{ik}=0} = -Z_{k-N}, \quad \frac{1}{\Psi}\frac{\partial\Psi}{\partial r_{ij}}\bigg|_{r_{ij}=0} = \frac{1}{2}. \quad (9)$$

This indicates $\Psi(\boldsymbol{r}; \boldsymbol{R})$ meets the cusp conditions of both the electron-nuclear pairs and unlike spin pairs (see e.g., [20, 60] for more details). In numerical computation we find that adding $\tilde{S}_{\mathrm{asy}}(\bar{\boldsymbol{r}})$ makes the training process much more stable, even though the cusp conditions are not strictly satisfied due to the influence of $\tilde{S}_i$ and $A(\boldsymbol{r})$ on the terms in (9).

We denote all the parameters in the DNNs ($\mathrm{Net}_{\mathrm{anti}}^\uparrow$, $\mathrm{Net}_{\mathrm{anti}}^\downarrow$, $\mathrm{Net}_{\mathrm{odd}}^\uparrow$, $\mathrm{Net}_{\mathrm{odd}}^\downarrow$, $\mathrm{Net}_{\mathrm{ebd}}^{\alpha_i\alpha_j}$, and $\mathrm{Net}_{\mathrm{fit}}^{\alpha_i}$), the two scalar factors (f_{odd} and f_{dec}) together as $\boldsymbol{\theta}$ and the associated trial wave-function as $\Psi_{\boldsymbol{\theta}}(\boldsymbol{r})$ (below for convenience we ignore the dependence on $\boldsymbol{R}$, the clamped ion positions). $\boldsymbol{\theta}$ are initialized randomly from a Gaussian distribution without any pre-training on pre-calculated wave-functions. We use VMC to optimize the trial wave-function. Specifically, we keep track of N_{wk} walkers to approximate the squared wave-function through an empirical distribution. The learning process consists of two phases, the sampling phase and the optimization phase, which we proceed with alternatively. In the sampling phase of step t, we run several steps of the Metropolis-Hasting algorithm to update the positions of the walkers,

according to a target probability proportional to $\Psi_{\boldsymbol{\theta}_t}^2(\boldsymbol{r})$. In the optimization phase of step t, we aim to minimize the second moment of the local energy $E_{\text{loc}}(\boldsymbol{r})$ with respect to a fixed reference energy E_{ref}. Here the local energy is defined as $E_{\text{loc}}(\boldsymbol{r}) := \hat{H}\Psi_{\boldsymbol{\theta}}(\boldsymbol{r})/\Psi_{\boldsymbol{\theta}}(\boldsymbol{r})$. Accordingly, the objective function has the explicit form

$$\Omega_{E_{\text{ref}}}(\boldsymbol{\theta}) := \frac{\int \Psi_{\boldsymbol{\theta}}^2(\boldsymbol{r})(E_{\text{loc}}(\boldsymbol{r}) - E_{\text{ref}})^2 \, d\boldsymbol{r}}{\int \Psi_{\boldsymbol{\theta}}^2(\boldsymbol{r}) \, d\boldsymbol{r}}.$$

In numerical computation, in order to reduce the variance when evaluating the objective function, we use a technique called correlated sampling. Assuming the walkers' current positions $\{\boldsymbol{r}^{(1)}, \boldsymbol{r}^{(2)}, \ldots, \boldsymbol{r}^{(N_{\text{wk}})}\}$ approximate the distribution $\Psi_{\boldsymbol{\theta}_t}^2(\boldsymbol{r})$ well, we define a reweighting factor $w_i := \Psi_{\boldsymbol{\theta}}(\boldsymbol{r}^{(i)})/\Psi_{\boldsymbol{\theta}_t}(\boldsymbol{r}^{(i)})$ and rewrite the objective function as

$$\Omega_{E_{\text{ref}}}(\boldsymbol{\theta}) = \frac{\int \Psi_{\boldsymbol{\theta}_t}^2(\boldsymbol{r})(\Psi_{\boldsymbol{\theta}}^2(\boldsymbol{r})/\Psi_{\boldsymbol{\theta}_t}^2(\boldsymbol{r}))(E_{\text{loc}}(\boldsymbol{r}) - E_{\text{ref}})^2 \, d\boldsymbol{r}}{\int \Psi_{\boldsymbol{\theta}_t}^2(\boldsymbol{r})(\Psi_{\boldsymbol{\theta}}^2(\boldsymbol{r})/\Psi_{\boldsymbol{\theta}_t}^2(\boldsymbol{r})) \, d\boldsymbol{r}}$$
$$\approx \frac{\sum_{i=1}^{N_{\text{wk}}} w_i^2 (E_{\text{loc}}(\boldsymbol{r}^{(i)}) - E_{\text{ref}})^2}{\sum_{i=1}^{N_{\text{wk}}} w_i^2}.$$

Note here we take $\boldsymbol{\theta}_t$ as constants and view both w_i and $E_{\text{loc}}(\boldsymbol{r}^{(i)})$ as functions of the parameters $\boldsymbol{\theta}$ given the walker's position $\boldsymbol{r}^{(i)}$. The gradients $\nabla_{\boldsymbol{\theta}}\Omega_{E_{\text{ref}}}(\boldsymbol{\theta})$ are computed by the backpropagation algorithm and used to update the parameters through $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta\nabla_{\boldsymbol{\theta}}\Omega_{E_{\text{ref}}}(\boldsymbol{\theta}_t)$, with η being the learning rate. Motivated by the idea of stochastic gradient descent (SGD), we actually evaluate $\nabla_{\boldsymbol{\theta}}\Omega_{E_{\text{ref}}}(\boldsymbol{\theta})$ with a random batch of data $i \in \mathcal{B} \subseteq \{1, \ldots, N_{\text{wk}}\}$ and update $\boldsymbol{\theta}_t$ with a few steps.

TABLE I: Ground-state energy of several systems obtained by DeepWF. The bond length of H_2, LiH, and H_{10} is 1.4, 1.62, and 1.8 $a.u.$, respectively. The benchmarks for H_2, Be, LiH, and B are taken from the Computational Chemistry Comparison and Benchmark (CCCBDB) DataBase [61], at the level of configuration interaction, singles and doubles (CISD) theory. The benchmark for H_{10} is taken from [31], at the level of multi-reference configuration with the Davidson correction (MRCI+Q) [7]. All the benchmark results are extrapolated to the complete basis set (CBS) limit. The relative difference (Rel. Diff) is reported.

System	DeepWF [$a.u.$]	Benchmark [$a.u.$]	Rel. Diff
H_2	-1.1738	-1.1741	0.26%
He	-2.9039	-2.9029	-0.03%
LiH	-7.8732	-8.0243	1.88%
Be	-14.6141	-14.6190	0.03%
B	-24.2124	-24.6006	1.58%
H_{10}	-5.5685	-5.6655	1.71%

Using the algorithm described above, for a series of benchmark systems, we obtain the optimized parameters $\boldsymbol{\theta}^*$ and run further Monte Carlo simulations for several statistical properties. In Table I, we present the ground state energies of H_2, He, LiH, Be, B, and a chain of 10
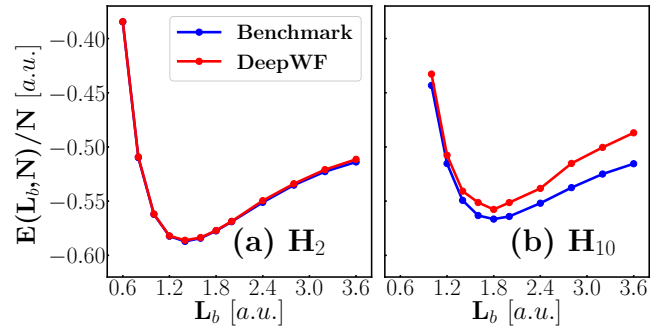

FIG. 2: Potential energy per atom $E(L_b, N)/N$ vs. bond length L_b for (a) H_2 ($N = 2$), and (b) H_{10} ($N = 10$). The benchmark result for H_2 is obtained using the PySCF package [62], at the level of full CI. The benchmark result for H_{10} is taken from [31], at the level of MRCI+Q [7]. Both the benchmark results are extrapolated to the CBS limit.
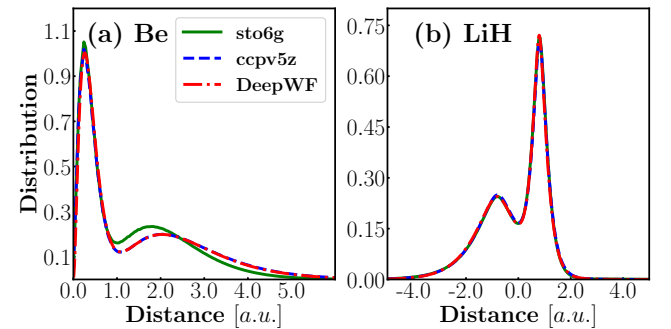

FIG. 3: Spatial distribution functions of electrons in different systems. (a) radial distribution of 4 electrons in a Be atom; (b) axial distribution of 4 electrons in a LiH molecule. The results obtained using the PySCF package [62] with the sto6g and the ccpv5z bases are plotted for comparison.

hydrogen atoms under open boundary conditions. Note that for H_2 and He we don't have the anti-symmetric part in the wave-function. Overall the ground-state energies obtained by DeepWF show great consistency with the benchmarks. However, as the number of electrons increases, the accuracy deteriorates.

In Fig. 2, we plot the potential energy curves of H_2 and H_{10} as a function of bond length. The result for H_2 is very close to the benchmark. For H_{10}, the DeepWF shows a good relative energy, with the correct prediction of local minimum, but is still a bit above the benchmark result.

Finally, Fig. 3 plots the spatial distributions of electrons in different systems. In the case of radial distribution function of electrons in a Be atom (Fig. 3 (a)), it is remarkable that the DeepWF learns from scratch the shell structure of the electrons and shows great consistency with the result obtained using the ccpv5z basis [63]. As a comparison, the sto6g basis [64] is not enough to describe the electronic dispersion. In the case of axial dis-

tribution function of electrons in a LiH molecule (Fig. 3 (b)), DeepWF and the other two methods show excellent agreement.

What is appealing to us is the simplicity of the proposed approach. Obviously there is a huge room for further improvement. In terms of the representation of DeepWF, the symmetric part is relatively general, considering the success of a similar version in representing the inter-atomic potential energy surface. However, the anti-symmetric ansatz, although appealing due to its quadratic scaling, might not be sufficient in representing the electronic correlations caused by the Pauli exclusion rule. Second, in terms of optimization, techniques for accelerating VMC through more efficient sampling (see e.g. [65, 66]) can be directly adapted into our DeepWF method. Optimization method other than the SGD-like correlated sampling method can also be employed. In any case, we hope that ideas presented here will add some ammunition to the heroic endeavor of attempting to solve the many-body Schrödinger equation.

---

* Electronic address: linfengz@princeton.edu
† Electronic address: weinan@math.princeton.edu

[1] P. A. M. Dirac, Proc. R. Soc. Lond. A **123**, 714 (1929).
[2] W. Pauli, Zeitschrift für Physik **31**, 765 (1925).
[3] C. Roothaan, Reviews of modern physics **32**, 179 (1960).
[4] J. Pople and R. Nesbet, The Journal of Chemical Physics **22**, 571 (1954).
[5] J. A. Pople, M. Head-Gordon, and K. Raghavachari, The Journal of chemical physics **87**, 5968 (1987).
[6] H.-J. Werner and P. J. Knowles, The Journal of chemical physics **89**, 5803 (1988).
[7] P. J. Knowles and H.-J. Werner, Chemical Physics Letters **145**, 514 (1988).
[8] T. Shiozaki, G. Knizia, and H.-J. Werner, The Journal of chemical physics **134**, 034113 (2011).
[9] G. D. Purvis III and R. J. Bartlett, The Journal of Chemical Physics **76**, 1910 (1982).
[10] J. Paldus and X. Li, Advances in Chemical Physics **110**, 1 (1999).
[11] R. J. Bartlett and M. Musiał, Reviews of Modern Physics **79**, 291 (2007).
[12] I. Shavitt and R. J. Bartlett, *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory* (Cambridge university press, 2009).
[13] W. L. McMillan, Physical Review **138**, A442 (1965).
[14] D. Ceperley, G. Chester, and M. Kalos, Physical Review B **16**, 3081 (1977).
[15] D. Bressanini and P. J. Reynolds, Advances in chemical physics **105**, 37 (1999).
[16] R. Blankenbecler, D. Scalapino, and R. Sugar, Physical Review D **24**, 2278 (1981).
[17] S. Zhang, J. Carlson, and J. Gubernatis, Physical Review B **55**, 7464 (1997).
[18] S. Zhang and H. Krakauer, Physical review letters **90**, 136401 (2003).
[19] K. Van Houcke, E. Kozik, N. Prokofev, and B. Svistunov, Physics Procedia **6**, 95 (2010).
[20] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Reviews of Modern Physics **73**, 33 (2001).
[21] S. R. White, Physical review letters **69**, 2863 (1992).
[22] S. R. White and R. L. Martin, The Journal of chemical physics **110**, 4127 (1999).
[23] G. K.-L. Chan and M. Head-Gordon, The Journal of chemical physics **116**, 4462 (2002).
[24] G. K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, and S. R. White, The Journal of chemical physics **145**, 014102 (2016).
[25] E. M. Stoudenmire and S. R. White, Physical review letters **119**, 046401 (2017).
[26] G. Knizia and G. K.-L. Chan, Physical review letters **109**, 186404 (2012).
[27] G. Knizia and G. K.-L. Chan, Journal of chemical theory and computation **9**, 1428 (2013).
[28] S. Wouters, C. A. Jiménez-Hoyos, Q. Sun, and G. K.-L. Chan, Journal of chemical theory and computation **12**, 2706 (2016).
[29] J. Grotendorst, John von Neumann Institute for Computing (2000).
[30] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory* (Courier Corporation, 2012).
[31] M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, *et al.*, Physical Review X **7**, 031059 (2017).
[32] J. C. Slater, Physical Review **35**, 210 (1930).
[33] R. Jastrow, Physical Review **98**, 1479 (1955).
[34] C. Umrigar, K. Wilson, and J. Wilkins, Physical Review Letters **60**, 1719 (1988).
[35] C. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, Physical review letters **98**, 110201 (2007).
[36] M. Casula and S. Sorella, The Journal of Chemical Physics **119**, 6500 (2003).
[37] H. J. Changlani, J. M. Kinder, C. J. Umrigar, and G. K.-L. Chan, Physical Review B **80**, 245116 (2009).
[38] Y. LeCun, Y. Bengio, and G. Hinton, nature **521**, 436 (2015).
[39] J. Han, A. Jentzen, and W. E, arXiv preprint arXiv:1707.02568 (2017).
[40] W. E, J. Han, and A. Jentzen, Communications in Mathematics and Statistics **5**, 349 (2017).
[41] J. Berg and K. Nyström, arXiv preprint arXiv:1711.06464 (2017).
[42] Y. Khoo, J. Lu, and L. Ying, arXiv preprint arXiv:1707.03351 (2017).
[43] H. Saito, Journal of the Physical Society of Japan **87**, 074002 (2018).
[44] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[45] X. Gao and L.-M. Duan, Nature communications **8**, 662 (2017).
[46] H. Saito, Journal of the Physical Society of Japan **86**, 093001 (2017).

[47] Z. Cai and J. Liu, Physical Review B **97**, 035116 (2018).

[48] M. Born and R. Oppenheimer, Annalen Der Physik **389**, 1 (1927).

[49] R. B. Laughlin, Physical Review Letters **50**, 1395 (1983).

[50] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Physical Review X **8**, 011006 (2018).

[51] S. R. Clark, Journal of Physics A: Mathematical and Theoretical **51**, 135301 (2018).

[52] J. Behler and M. Parrinello, Physical review letters **98**, 146401 (2007).

[53] J. Han, L. Zhang, R. Car, and W. E, Communications in Computational Physics **23**, 629 (2018).

[54] L. Zhang, J. Han, H. Wang, R. Car, and W. E, Phys. Rev. Lett. **120**, 143001 (2018).

[55] L. Zhang, J. Han, H. Wang, W. A. Saidi, R. Car, and W. E, arXiv preprint arXiv:1805.09003 (2018).

[56] K. Schütt, P.-J. Kindermans, H. E. S. Felix, S. Chmiela, A. Tkatchenko, and K.-R. Müller, in *Advances in Neural Information Processing Systems* (2017) pp. 992–1002.

[57] L. Zhang, H. Wang, and W. E, The Journal of Chemical Physics **148**, 124113 (2018).

[58] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, arXiv preprint arXiv:1802.08549 (2018).

[59] E. Schneider, L. Dai, R. Q. Topper, C. Drechsel-Grau, and M. E. Tuckerman, Physical review letters **119**, 150601 (2017).

[60] T. Kato, Communications on Pure and Applied Mathematics **10**, 151 (1957).

[61] R. D. J. III, NIST Standard Reference Database Number **101** (2018).

[62] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. D. McClain, E. R. Sayfutyarova, S. Sharma, *et al.*, Wiley Interdisciplinary Reviews: Computational Molecular Science **8**, e1340 (2018).

[63] D. E. Woon and T. H. Dunning Jr, The Journal of chemical physics **103**, 4572 (1995).

[64] W. J. Hehre, R. F. Stewart, and J. A. Pople, The Journal of Chemical Physics **51**, 2657 (1969).

[65] R. Lee, G. Conduit, N. Nemec, P. L. Ríos, and N. Drummond, Physical Review E **83**, 066706 (2011).

[66] M. Dewing, The Journal of Chemical Physics **113**, 5123 (2000).