https://doi.org/10.26599/AIR.2024.9150030 CAAI Artificial Intelligence Research

Variational Quantum Monte Carlo Solution of the Many-Electron Schrödinger Equation Based on Deep Neural Networks

Huiping Su[1], Hongbo Gao[1, 2, 3] ✉, Xinmiao Wang[1], Xi He[1], and Da Shen[1]

ABSTRACT

The solution of Schrödinger equation generates the quantisation properties of the system, which can completely describe the quantum behaviour of microscopic particles in a physical system. However, solving the Schrödinger equation is a non-deterministic polynomial time (NP)-hard problem. Numerical methods for solving the Schrödinger equation require careful selection of basis configurations, and the complete basis set has high accuracy but its complexity increases exponentially in the number of particles. Therefore, it is crucial to find an effective numerical method. To solve this problem, this paper present a deep learning architecture, VMCNet, using the powerful computational efficiency of neural networks to improve the speed of numerical computation. Moreover, this paper proposes a suitable wavefunction ansatz, witch determines the accuracy of neural networks, that achieves more accurate energy solutions of the many-electron Schrödinger equation compared to the conventional single-determinant variational molecular Monte Carlo. We introduce non-local and local blocks to represent quantum mechanical interactions, which contain more physical information about the particle system. The experimental results show that VMCNet can cover more than 93% of the correlation energy of the main group elements (Be–Ne) for the monoatomic system and more than 90% of the correlation energy for the diatomic system.

KEYWORDS

Schrödinger equation; deep learning; energy; Monte Carlo

The ability to solve the Schrödinger equation for a given system allows prediction of its behavior, with important applications in atomic, molecular, solid-state physics, nuclear physics, chemistry, and other fields. It can only be solved analytically in simple systems, such as hydrogen atoms, but in recent decades ab initio quantum chemistry methods have been developed.[1] The term ab initio indicates that the calculation is from first principles and that no empirical data is used. Coupled-cluster with single, double, and triple perturbative excitations (CCSD(T)) is regarded as the gold standard of quantum chemistry methods.[2] Density functional theory (DFT) is at present the best approach for the investigation of reactive chemistry in medium to large systems.[3] While the ability to solve the Schrödinger equation is crucial for predicting the behavior of various systems in fields ranging from atomic physics to chemistry, traditional ab initio methods like CCSD(T) and DFT, though accurate, face significant computational challenges. These methods, while foundational in quantum chemistry, are often constrained by computational feasibility, especially for larger, more complex systems. This limitation highlights the need for innovative approaches that balance computational efficiency with accuracy. Many approaches in quantum chemistry represent the many-electron wave function as a linear combination of Slater derterminants. The difficulty is that solving the Schrödinger equation requires the knowledge of the many-body wave function in the many-body Hilbert space, which typically has an exponentially large size in the number of particles. So the number of possible Slater determinants increases rapidly with the system size, leading to a trade-off between accuracy and computational cost.

Quantum Monte Carlo (QMC) is a way to provide a reliable solution (or an accurate approximation) of the quantum many-body problem, which is not as strongly limited by size considerations and is readily adapted to parallel computers.[4, 5] The precision of QMC is determined by how accurately the exact wave function can be approximated. The wavefunction ansatz can provide a more compact wave function representation. Typically, QMC calculations are carried out in the fixed-node approximation using a Slater–Jastrow ansatz, which is written as a product of a truncated linear combination of Slater determinants and an exponentiated Jastrow function. The latter term captures close-range correlations and depends explicitly on interparticle distances. Additionally, the backflow transformation[6] in which the coordinates of the electrons are modified by a backflow displacement allows the orbitals to depend on the positions of nearby electrons. For continuous-space many-electron problems in three dimensions, the default choice is the Slater–Jastrow–backflow ansatz. There remains a great deal of flexibility in the form of backflow displacement, but recent work[7–11] shows that a neural network is a powerful tool for representing backflow transformations.

Over the past decade, techniques borrowed from the field of machine learning (ML) have had a great impact on quantum chemistry, involving solving the electronic Schrödinger equation using deep learning based on the many-electron wave function

1 Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China
2 Institute of Advanced Technology, University of Science and Technology of China, Hefei 230088, China
3 School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
Address correspondence to Hongbo Gao, ghb48@ustc.edu.cn

ansatz[8, 9], prediction of electronic energies[12–18], electron densities[19–21], molecular orbitals[3, 22], force fields[23–25], and empirical corrections of semiempirical, density functional, or Hartree–Fock (HF) energies to higher-level theories[26, 27]. This work is of particular interest in solving the electronic Schrödinger equation using deep learning. The Jastrow function and backflow transformations are essential to capture the interaction between nucleus-electron and electron-electron, and the neural network is a general function approximator and can provide a more flexible method to approximate the interaction of many-body quantum systems. Most ML is supervised learning, so their accuracy is limited by the quality of the data used to train them. By contrast, we use the Slater determinant as the starting point for our ansatz, the direct representation of correlated wavefunctions with neural networks and their unsupervised training via the variational principle.

There are already some ML methods for solving the Schrödinger equation. The Δ-ML[28] approach is renowned for its capability to obtain energy at an expensive level of theory using an ML-learned additive correction term, given the energy at the baseline theory. This method exemplifies the use of machine learning to refine and enhance predictions based on baseline quantum chemical calculations. DimeNet[29], on the other hand, introduces a novel approach to model local interactions between atoms by including angular information in the feature updates. This method acknowledges the critical role of angular relationships in determining atomic interactions, thereby providing a more nuanced representation of molecular structures. VMCNet is similar to PauliNet[8], but the latter only encodes local interaction information between particles, which makes it less accurate than VMCNet. The architecture of Ferminet[9] is simple, but in addition to antisymmetry, it does not introduce other physical knowledge of the wave function, but replaces it with numerous parameters to be optimized. This leads to higher computational cost. SpookyNet[30] also contains non-local interaction, but only considers interactions between atoms and is independent of coordinates.

VMCNet, the centerpiece of this study, signifies a major leap in ab initio methods. By employing a deep neural network (DNN) for the wavefunction ansatz, VMCNet surpasses the conventional Slater–Jastrow-backflow variational method in both accuracy and computational efficiency. This advancement is not just incremental but represents a paradigm shift in how we approach the quantum many-body problem. It is an *ab initio* quantum chemistry method and does not rely on values derived from ab initio calculations. The main contributions made in this paper can be summarized in the following four points.

(1) We demonstrate the superior computational efficiency of neural networks in solving the Schrödinger equation. This approach significantly reduces the computational time required for complex quantum systems, addressing a major limitation in traditional quantum chemistry computations.

(2) Our development of VMCNet, a novel deep learning architecture, marks a significant advancement in wavefunction ansatz. Unlike traditional single-determinant variational methods, VMCNet achieves markedly higher accuracy in energy solutions for many-electron systems, thus providing a more reliable tool for quantum chemical analysis.

(3) By introducing non-local and local blocks, VMCNet captures a more comprehensive range of quantum mechanical interactions, encompassing a wider array of physical information about particle systems. This innovation allows for a more detailed and accurate representation of particle interactions than previously possible.

(4) VMCNet's ability to achieve higher accuracy in modeling monatomic and diatomic systems, using only atomic coordinates and charges, represents a significant reduction in data dependency. This makes VMCNet not only more efficient but also more applicable in scenarios where data availability is limited.

The organization of the rest paper is as follows: Section 1 introduce Slater–Jastrow–backflow Variational Monte Carlo (VMC) method and the proposed method; Section 2 presents experiments and analysis of experimental results; Section 3 concludes the paper.

1 Proposed Method

VMCNet extends the Slater determinant by generalizing single-electron orbitals to include universal exchangeable nonlinear interactions for all electrons. In this paper, only the single determinant method is considered limited by computer resource. As a deep-learning QMC approach, Section 1.1 will first introduce Slater–Jastrow–backflow VMC method, and Section 1.2 shows the detail structure of VMCNet.

1.1 Variational quantum Monte Carlo

The mathematical description of the quantum state of an isolated quantum system can be represented by a wave function $\Psi(X)$, where X denotes a many-body configuration and contains particle coordinates and electron spin states. The time-independent Schrödinger equation, which predicts that wave functions can form standing waves, has the mathematical form[31] of

$$\hat{H}\Psi = E\Psi \tag{1}$$

where $\hat{H}$ is the Hamiltonian, containing the differential operator and E is the energy eigenvalue, i.e., the system ground state energy. The energy can also be expressed in the form of the expectation value of the energy $\Psi(X)$[32]:

$$E[\Psi] = \frac{\langle \Psi|\hat{H}|\Psi\rangle}{\langle \Psi|\Psi\rangle} = \frac{\int \Psi(X)\hat{H}\Psi(X)\mathrm{d}X}{\int \Psi(X)\Psi(X)\mathrm{d}X} \tag{2}$$

The difficulty of Schrödinger equation lies in solving the integral of Eq. (2). VMC is a quantum Monte Carlo method that applies variational methods to approximate the ground state of a quantum system. It interprets $\frac{|\Psi(X,a)|^2}{\int|\Psi(X,a)|^2\mathrm{d}X}$ as a probability distribution function, where a is the parameter, which is used to estimate the system energy. The optimal values of the parameters a is then found upon minimizing the total energy of the system using the variational principle for the total electronic energy.

$$E = \min_{\Psi} E[\Psi] \leqslant \min_{a} E[\Psi_a] \tag{3}$$

So sample the probability distribution function, and evaluate the energy expectation value $E[\Psi_a]$. Once $E[\Psi_a]$ is known for a given set of variational parameters a, then optimization is performed in order to minimize the energy and obtain the best possible representation of the ground state wave function[33].

1.2 VMCNet architecture

VMCNet projects particle information, including nuclei and electrons, as well as their coordinates and distances, into a high-dimensional feature space by embedding to calculate their interactions. In order to balance efficiency and accuracy, VMCNet incorporates the multireference HF method as a baseline. We use a distance featurization inspired by the DimeNet architecture with

a little modification. In addition to encoding local information based on distance, VMCNet also encodes non-local interactions by introducing a network that is only related to coordinates, while most architectures only consider local interactions and lose edge information. Figure 1 contains a schematic of the network.

VMCNet work in the Born-Oppenheimer approximation[31] and the trial wavefunctions are assumed to be real-valued. We consider a system of M nuclei with atomic numbers $\boldsymbol{Z}=(Z_1, Z_2,\ldots,Z_M|Z_i\in\mathbb{N})$, Cartesian coordinates set $\boldsymbol{R}=(\boldsymbol{R}_1,\boldsymbol{R}_2,\ldots,\boldsymbol{R}_M|\boldsymbol{R}_i\in\mathbb{R}^3)$ and N electrons with Cartesian coordinates set $\boldsymbol{r}=(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_{n^\uparrow},\ldots,\boldsymbol{r}_{n^\uparrow+n^\downarrow}|\boldsymbol{r}_i\in\mathbb{R}^3)$, where $N=n^\uparrow+n^\downarrow$, $n^\uparrow$ denotes up-spin electrons and $n^\downarrow$ denotes down-spin electrons. In VMCNet, The wavefunction ansatz $\Psi_\theta(\boldsymbol{r};\boldsymbol{R})$ of the groundstate wave function $\Psi(\boldsymbol{r};\boldsymbol{R})$ is represented by DNN as

$$\Psi_\theta(\boldsymbol{r};\boldsymbol{R})=e^{\gamma(\boldsymbol{r})}\sum_k \omega_k \det\left[\boldsymbol{\Phi}_\theta^{k,\uparrow}(\boldsymbol{r};\boldsymbol{R})\right]\det\left[\boldsymbol{\Phi}_\theta^{k,\downarrow}(\boldsymbol{r};\boldsymbol{R})\right] \tag{4}$$

where θ denote the set of trainable parameters in the model, $\omega_k\in\mathbb{R}$ is a trainable weight, term $\gamma(\boldsymbol{r})$ represents the cusp correction[34], and $\det[\boldsymbol{\Phi}_\theta^{k,\uparrow}]$, $\det[\boldsymbol{\Phi}_\theta^{k,\downarrow}]$ are Slater-like determinants for up- and down-spin electrons, respectively. The elements of the Slater matrix are the values of each orbital φ_i (in rows) evaluated at each electron's coordinates $\boldsymbol{r}_j$ (in columns). We derive the one-electron molecular orbitals starting with multireference HF calculations using the 6-311g basis set and VMCNet initialize orbitals using PySCF[35] to implement self-consistent field calculations. Then through the calculation of VMCNet, each electron position $\boldsymbol{r}_j$ is replaced by the collective coordinate to capture the correlation energy.

VMCNet contains two parts: the embedding block (Fig. 1 blue box) and the interaction block (Fig. 2), the latter being subdivided into the non-local interaction block, the local interaction block and the out block. VMCNet projects each particle information into an F-dimensional feature space and iteratively updates electron embeddings with particle coordinates and distances. In the current layer l, the electrons and nucleus embeddings are denoted as $\boldsymbol{X}^{(l)}=(\boldsymbol{x}_1^{(l)},\ldots,\boldsymbol{x}_{n^\uparrow}^{(l)},\ldots,\boldsymbol{x}_{n^\uparrow+n^\downarrow}^{(l)})$, and $\boldsymbol{Y}=(\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_M)$, respectively. Using the Born-Oppenheimer hypothesis, the position of the nucleus is fixed, and its feature vectors are only automatically updated during the backpropagation process, and only the feature vectors of the electrons are updated during the entire forward propagation process. At layer $l+1$, for the ith electron, the local interaction block includes the embeddings $\boldsymbol{x}_j^{(l)}$ of its neighbors, distances r_{ij} and the distance between each electron and nucleus r_{il} as input. The input to the non-local block is similarly the $\boldsymbol{x}_j^{(l)}$, but contains coordinates $\boldsymbol{r}$ and $\boldsymbol{R}$ instead of absolute distance. Taking absolute distances and coordinates as input eliminates the need to include a separate Jastrow factor after the determinant, and accuracy is also improved over distance-only models.

The primary goal is to ensure that the represented wavefunction satisfies the anti-symmetry property, which means that swapping the two input variables only changes the sign of the trial wavefunction, which the matrix determinant can achieve.

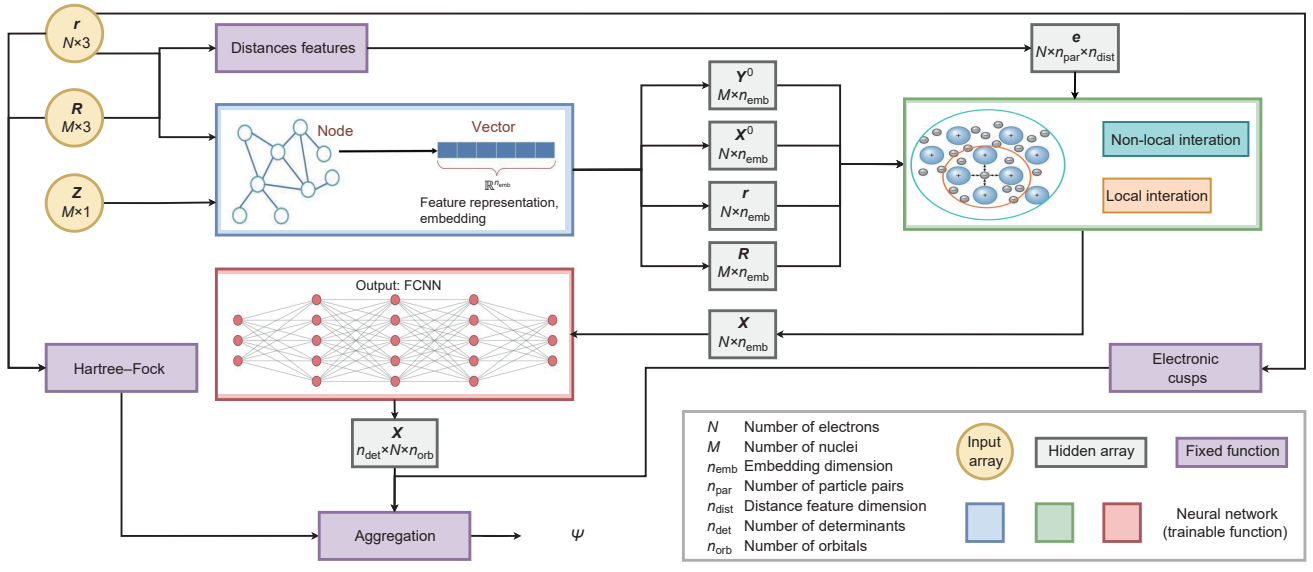

Fig. 1 VMCNet architecture. Global architecture: The electron and nuclear coordinates, r and R, are transformed through several layers, and the wave function Ψ at that position is given as output. Blue box: Embedding block generates the initial message embeddings. Green box: Three interaction blocks, consisting of non-local, local and out blocks, update embeddings via directional message passing to capture the interaction between particles. Red box: Output block compresses feature vectors and generates determinant elements.

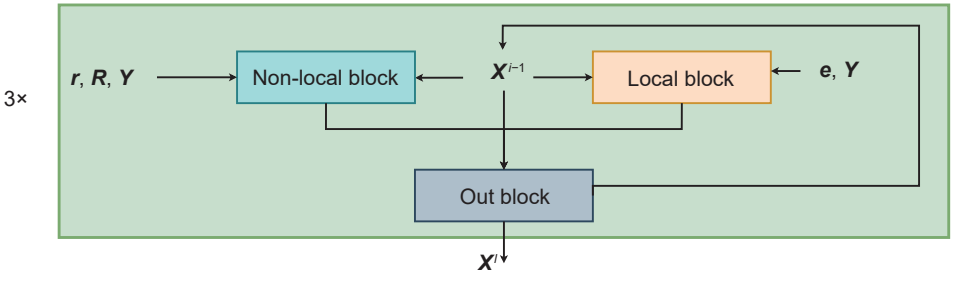

Fig. 2 *Interaction* block. It consists of three parts: non-local block, local block, and output block. This block encodes the interactions between particles and outputs the information uniformly to a high-dimensional representation of the electron.

VMCNet introduces a more sophisticated approach. Unlike simpler architectures such as FermiNet, which does not introduce additional physical knowledge of the wave function beyond antisymmetry, VMCNet incorporates this condition through its advanced neural network architecture. Another necessary condition that the wavefunction needs to satisfy is the Kato cusp condition, meaning that the wavefunction is continuous, and the first partial derivative is bounded but discontinuous at the Coulomb-type singular points of the potential. VMCNet's architecture, which includes both local and non-local interaction blocks, is specifically designed to better capture these intricate interactions. The local block, inspired by SchNet, encodes nuclear charge and coordinates information, enabling VMCNet to model the influence of the local environment on electrons more effectively than methods that only consider either electron-nucleus or electron-electron interactions in isolation. Let $r_{iI}(r_{ij})$ be the electron-nuclear (electron-electron) distance, Z_I be the nuclear charge of the I-th nucleus, and the wavefunction at the origin satisfies

$$\begin{aligned}\lim_{r_{iI}\to 0}\frac{\partial\langle\Psi\rangle}{\partial r_{iI}} &= -Z_I\Psi(r_{iI}=0)\\ \lim_{r_{ij}\to 0}\frac{\partial\langle\Psi\rangle}{\partial r_{ij}} &= \frac{1}{2}\Psi(r_{ij}=0)\end{aligned} \tag{5}$$

where $\langle\Psi\rangle$ is the spherical average of the many-body wavefunction. To ensure nuclear cusp conditions, VMCNet apply the technique proposed by Ma et al.[36] to modify molecular orbitals and for electronic cusp conditions, VMCNet applied PauliNet's method to enforce.

1.3 Coordinate embedding

In order to use neural networks to extract features and find relationships between original data, we need to project the original data (particle coordinates and distances) into a high-dimensional space. This embedding is based on Electronic SchNet[24] and PauliNet. Coordinate data can reflect the similarity of particles to a certain extent. If the coordinate characteristics of two particles are similar, their interaction is strong, otherwise, it is weak. In order to preserve the similarity of the initial coordinates, VMCNet uses the parameter sharing feature of the convolutional neural network to project the 3-dimensional coordinates into the F-dimensional feature space:

$$\begin{aligned}\boldsymbol{r} &= \overline{\boldsymbol{W}}^{e}_{\theta}(\boldsymbol{r})\\ \boldsymbol{R} &= \overline{\boldsymbol{W}}^{n}_{\theta}(\boldsymbol{R})\end{aligned} \tag{6}$$

where $\overline{\boldsymbol{W}}^{e}_{\theta}$ and $\overline{\boldsymbol{W}}^{n}_{\theta}$ represent two different two-layer convolutions, which means that the electron coordinate set $\boldsymbol{r}$ and the nuclear coordinate set $\boldsymbol{R}$ have undergone the same transformation respectively, and both use only the activation function $\sigma=\frac{1}{\beta}\log\left(\frac{1+e^{\beta x}}{2^{\beta}}\right)$ (shifted softplus[8]) and bias in the first layer. Now each element of $\boldsymbol{r(R)}$ is an F-dimensional vector. In VMCNet, it is sufficient to take F as 64, which is smaller than PauliNet's 128 and Ferminet's 256.

Inspired by Dimenet and PauliNet, the distance feature is defined by

$$e_m(d)=d^2e^{-d}\cdot\frac{\sin(\pi md/c)}{d/c} \tag{7}$$

with the spherical Bessel functions of the first kind, where d is the distance between the two particles, $m\in[1,2,\ldots,n_r]$, c is a cutoff parameter. The number of radius basis functions $n_r=16$ is sufficient, which is much less than Gaussian basis functions. The initial value of m is a positive integer, but not a fixed value, and is set to be updated during backpropagation. Equation (7) also guarantees that the value and differential of the distance feature at zero distance are equal to zero to preserve the cusp conditions.

1.4 Non-local block

The purpose of this block is to express electron-electron, electron-nucleus interactions. Many frameworks describing chemical systems in recent years have only considered local interactions, and although they have yielded good results, some key information may be missed. We found that VMCNet yielded 3% higher correlation energy accuracy than PauliNet with only non-local interactions (Fig. 3), demonstrating that nonlocal effects can play an important role in capturing correlation energy. Energy contributions should always depend on the presence of other atoms in the structure. Consequently, it is difficult for models without nonlocal interactions to predict the correct asymptotic behavior for all systems simultaneously and simply increasing the cutoff is not enough to guarantee the capture of all particles' contribution to the energy.

Every electron interacts with every other particle in the system, so its feature vector should include the effect of every other particle on it. VMCNet uses a linear combination of particle feature vectors to represent the effect on the current particle. The interactions between particles with similar characteristics should be stronger and the influence from high to low can be represented by the weight from large to small. Therefore, each update of the electron feature can be written as a weighted sum of all particle features. Figure 4a contains a schematic of non-local block. This is consistent with the role of self-attention[37], which is to calculate the similarity weight, and then do the weighted sum of the input.

We have expressed the coordinates and information of each particle as a high-dimensional vector using embedding. At layer $l+1$, for the ith electron, the non-local interaction block includes the embeddings $\boldsymbol{X}^{(l)}$, coordinates $\boldsymbol{r}$ and $\boldsymbol{R}$. Taking coordinates as input eliminates the need to include a separate Jastrow factor after the determinant, and accuracy is also improved over distance-only models.

We calculate the weights representing the magnitude of the interaction. In layer l, for electron i, the weight between electron and electron is

$$\begin{aligned}\boldsymbol{v}^{(l,\pm)}_{\mathrm{non,i}} &:= \boldsymbol{w}^{(l,\pm)}_{\mathrm{non},\theta}\left(\boldsymbol{x}^{(l)}_i\right),\\ \boldsymbol{h}^{(l,\pm)}_{\mathrm{non,i}} &:= \boldsymbol{h}^{(l,\pm)}_{\mathrm{non},\theta}\left(\boldsymbol{r}_i\odot\boldsymbol{x}^{(l)}_i\right),\\ \boldsymbol{g}^{(l,\pm)}_{\mathrm{non,i}} &:= \boldsymbol{g}^{(l,\pm)}_{\mathrm{non},\theta}\left(\boldsymbol{r}_i\odot\boldsymbol{x}^{(l)}_i\right),\\ c^{(l,\pm)}_i &:= \mathrm{softmax}\left(\frac{\boldsymbol{h}^{(l,\pm)}_{\mathrm{non,i}}\left(\boldsymbol{g}^{(l,\pm)}_{\mathrm{non}}\right)'}{8}\right)\end{aligned} \tag{8}$$

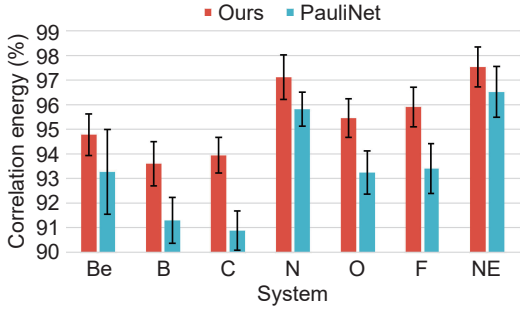

Fig. 3 Percentages of the correlation energy and standard errors of VMCNet and PauliNet. Error bars represent standard errors. In atomic units, VMCNet's standard error range is between 0.000 762(6) and 0.003 292 (3), while PauliNet's range is between 0.001 082(7) and 0.004 126(1).

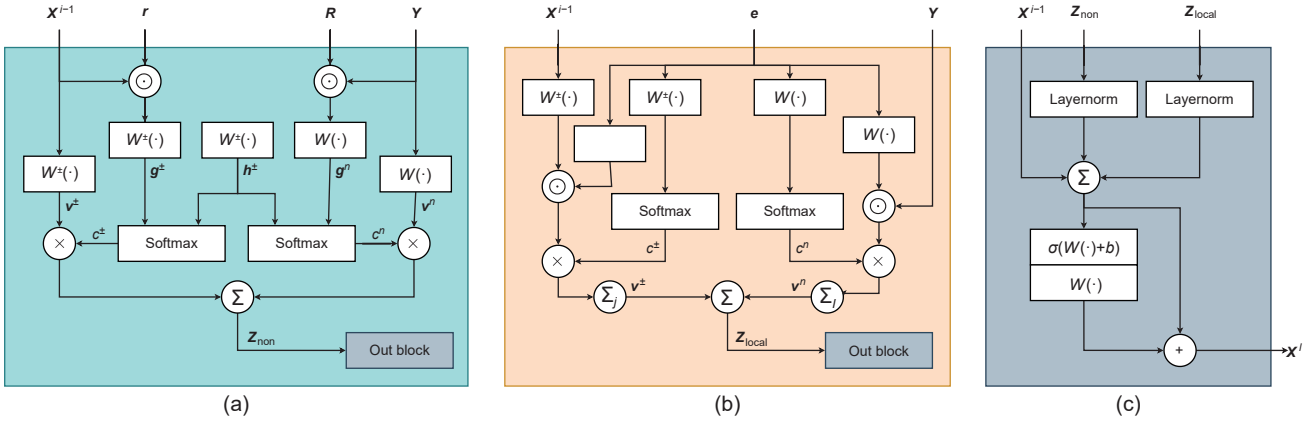

Fig. 4 Interaction blocks. (a): Non-local block encodes the interaction of all particles using particle messages X, Y, and coordinate embeddings r, R. (b): Local block uses the neighboring messages X, Y and the distance representations e to enhance the encoding of local information. (c): Out block aggregates the resulting feature vectors.

and the weight between electron and nucleus is

$$\begin{aligned}
\boldsymbol{v}_{\mathrm{non,j}}^{(l,n)} &:= \boldsymbol{w}_{\mathrm{non},\theta}^{(l,n)}\left(\boldsymbol{y}_j\right), \\
\boldsymbol{g}_{\mathrm{non,j}}^{(l,n)} &:= \boldsymbol{g}_{\mathrm{non},\theta}^{(l,n)}\left(\boldsymbol{R}_j \odot \boldsymbol{y}_j\right), \\
c_i^{(l,n)} &:= \mathrm{softmax}\left(\frac{\boldsymbol{h}_{\mathrm{non,i}}^{(l,\pm)}\left(\boldsymbol{g}_{\mathrm{non}}^{(l,n)}\right)'}{8}\right)
\end{aligned} \tag{9}$$

where $(+)/(-)$ denotes the same/opposite spin electrons, (n) denotes the nuclei, "non" denotes non-local block, $\odot$ denotes element-wise multiplication, superscript $(')$ means matrix transpose, $\boldsymbol{w}_{\mathrm{non},\theta}^{(l)}$, $\boldsymbol{h}_{\mathrm{non},\theta}^{(l)}$, and $\boldsymbol{g}_{\mathrm{non},\theta}^{(l)}$ are trainable functions represented by ordinary fully connected DNNs, $\boldsymbol{v}_{\mathrm{non}}^{(l)}$, $\boldsymbol{h}_{\mathrm{non}}^{(l)}$, and $\boldsymbol{g}_{\mathrm{non}}^{(l)} \in \mathbb{R}^F$ are intermediate variables, and $c_i^{(l)} \in \mathbb{R}$ is the weight. The denominator in the normalization calculation is 8 because F is taken to be 64. After information exchange, the information aggregation method is

$$\boldsymbol{z}_{\mathrm{non,i}}^{(l)} := c_i^{(l,+)}\boldsymbol{v}_{\mathrm{non}}^{(l,+)} + c_i^{(l,-)}\boldsymbol{v}_{\mathrm{non}}^{(l,-)} + c_i^{(l,n)}\boldsymbol{v}_{\mathrm{non}}^{(l,n)} \tag{10}$$

where $\boldsymbol{z}_{\mathrm{non}}^{(l)}$ is the message received by the electron feature vectors in the iteration of non-local block.

1.5 Local block

The local block is concerned with the local information inside the system. SchNet, as a graph convolutional neural network, encodes nuclear charge and coordinates information to model molecular energy. In VMCNet, we take Schnet's ideas, with a little modification, to represent the influence of the local environment on electrons (Fig. 4b). The input to the local block is similarly the embeddings $\boldsymbol{x}_j^{(l)}$ of its neighbors, but contains distances r_{ij} and the distance between each electron and nucleus r_{iI} instead of coordinates. Here, the message $\boldsymbol{z}_{\mathrm{local}}^{(l)}$ is still divided into three channels: same-spin electrons, opposite-spin electrons and the nuclei,

$$\begin{aligned}
c_{ij}^{(l,\pm)} &:= \mathrm{softmax}\left[\boldsymbol{g}_{\mathrm{local},\theta}^{(l,\pm)}\left(\boldsymbol{e}\left(r_{ij}\right)\right)\right] \\
c_{iI}^{(l,n)} &:= \mathrm{softmax}\left[\boldsymbol{g}_{\mathrm{local},\theta}^{(l,n)}\left(\boldsymbol{e}\left(r_{iI}\right)\right)\right] \\
\boldsymbol{v}_{\mathrm{local,i}}^{(l,\pm)} &:= \sum_{j\neq i}^{\pm} c_{ij}^{(l,\pm)} \times \left[\boldsymbol{h}_{\mathrm{local},\theta}^{(l,\pm)}\left(\boldsymbol{e}\left(r_{ij}\right)\right) \odot \boldsymbol{w}_{\mathrm{local},\theta}^{(l)}\left(\boldsymbol{x}_j^{(l)}\right)\right] \\
\boldsymbol{v}_{\mathrm{local,i}}^{(l,n)} &:= \sum_{I} c_{iI}^{(l,n)} \times \left[\boldsymbol{h}_{\mathrm{local},\theta}^{(l,n)}\left(\boldsymbol{e}\left(r_{iI}\right)\right) \odot \boldsymbol{Y}\right] \\
\boldsymbol{z}_{\mathrm{local,i}}^{(l)} &:= \boldsymbol{v}_{\mathrm{local,i}}^{(l,+)} + \boldsymbol{v}_{\mathrm{local,i}}^{(l,-)} + \boldsymbol{v}_{\mathrm{local,i}}^{(l,n)}
\end{aligned} \tag{11}$$

where the notation is similar to that in the non-local block and "local" denotes local block. In Eq. (11), $j \in N_{\backslash i}$, which represents the set that does not contain the currently considered electron i within the cut-off radius.

1.6 Out block

After going through the non-local and local blocks, the electronic feature vector goes through the out block for final aggregation (Fig. 4c),

$$\begin{aligned}
\boldsymbol{z}_i^{(l)} &= \boldsymbol{x}_i^{(l)} + \mathrm{LN}_1\left(\boldsymbol{z}_{\mathrm{non,i}}^{(l)}\right) + \mathrm{LN}_2\left(\boldsymbol{z}_{\mathrm{local,i}}^{(l)}\right) \\
\boldsymbol{x}_i^{(l+1)} &= \boldsymbol{z}_i^{(l)} + \mathbf{out}\left(\boldsymbol{z}_i^{(l)}\right)
\end{aligned} \tag{12}$$

where LN denotes layer normalization and **out** represents two fully connected DNNs. Now each electronic high-dimensional feature vector contains information about all particles.

The determinant part of the Slater–Jastrow network amounts to removing the interaction block from VMCNet. After compressing the feature vectors into the number of orbitals using three fully connected DNNs, VMCNet calculates correlation energy using PauliNet's backflow transformation.

2 Experiment and Analysis

Here we evaluate the performance of the VMCNet on a variety of electronic structures. VMCNet minimizes the total electronic energy (VMC) via the variational principle. Without the need for external data, self-consistent results are obtained by iteratively sampling the electron density. For hyperparameter choices and training setup see Table 1.

2.1 Less determinants, more accurate solutions

To demonstrate the expressive power of the VMCNet, We first investigate the results of methods VMCNet, PauliNet, conventional single-determinant VMC (VMC-SD) and multi-determinant VMC (VMC-MD) methods in calculating the correlation energies of the first row atoms (from Be to Ne). Where

Table 1 Hyperparameters of all models used in this work.

Hyperparameter	Value	Significance
F	64	Dimensionality of feature space
n_r	16	Number of radial basis functions
c	10 Å	Cutoff radius for interactions in the neural network

VMCNet and PauliNet are single-determinant Ansatz in terms of Slater–Jastrow–backflow network. Figure 5 compares the VMCNet results with numbers already available in Refs. [38, 39]. It shows percentages of the correlation energy retrieved for each atom within four methods. VMCNet recovers more than 93% electron correlation energy for the monoatomic system, in all cases far beyond the accuracy of PauliNet (90%). VMCNet covers 97.13(8)% of the correlation energy for the Ne atom, 96.71(8)% for the *N* atom. The numbers in parentheses indicate the statistical uncertainty in the last digit shown. VMCNet also has an overwhelming superiority in the case of a single-determinant Ansatz, especially in Be, B, C, N atomic systems. VMCNet is only surpassed by VMD-MD, but the latter result is numerically unstable (Fig. 6). Furthermore, the VMC optimizations were performed using 5×10^4 statistically independent particle configurations, corresponding to hundreds to thousands of determinants. It is worth noting that as a single determinant method, neural networks have higher accuracy than traditional methods, and PauliNet's paper reaches higher accuracy for some systems, but they use six determinants. VMCNet, on the contrary, can achieve higher accuracy by using only the single determinant.

Inspired by the recently proposed PauliNet, VMCNet contains more physical information about the wave function and has higher accuracy. Figure 3 shows the percentages of the correlation energy and standard errors of VMCNet and PauliNet under the same calculation condition. The bar graphs represent the distribution of the correlation energy results for the two methods, and the lines with endpoints represent the standard error ranges. VMCNet has smaller standard errors on almost all single-atom

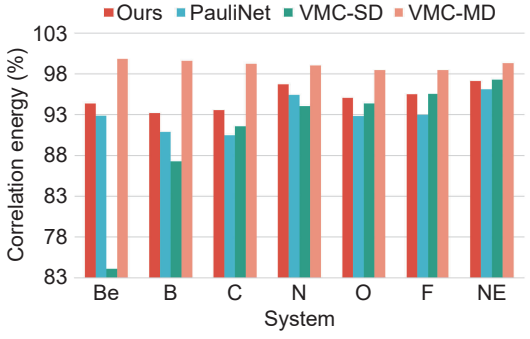

Fig. 5 Comparison of the VMCNet against the Slater–Jastrow–backflow Ansatz. VMCNet covers 93%–98% of the correlation energy using only the single determinant, while PauliNet only covers 90%–97% and VMC-SD only covers 74%–98%. The results of VMCNet and PauliNet[8] are completed under the same calculation condition. The rest of the reference results are taken from single-determinant VMC(VMC-SD)[38] and multi-determinant VMC(VMC-MD)[39].

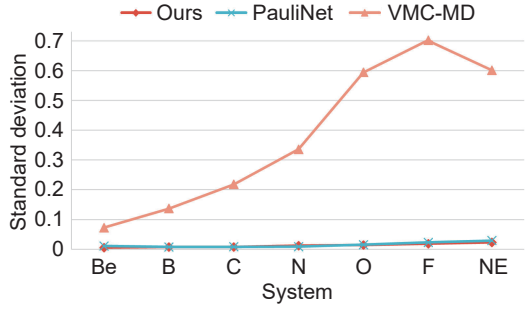

Fig. 6 Standard deviations for VMCNet, PauliNet, and VMC-MD. All Standard deviations are in atomic units. Neural networks reach smaller standard deviation and higher precision. The reference results are taken from VMC-MD[39].

systems. For Be atom, The standard errors of VMCNet is 0.000 762(6), while PauliNet is 0.001 525(7). Apart from this, VMCNet has a smaller standard error range. The range of VMCNet's standard error is from 0.000 762(6) to 0.003 292(3), while PauliNet's range is from 0.001 082(7) to 0.004 126(1). In the absence of Jastrow Ansatz, VMCNet uses non-local blocks to capture more physical information. The results show that VMCNet has higher accuracy and error range, which indicates that non-local block can capture interaction information better than Jastrow Ansatz.

Figure 6 reports the standard deviations for VMCNet, PauliNet, VMC-MD. The results show that it has higher accuracy than the conventional single determinant VMC, and the standard error is smaller than the similar PauliNet. The variance is reduced by approximately several dozens times when a neural network is introduced to a Slater–Jastrow–backflow wave function. The improvement is particularly large for Be, O, and F, where the variance drops by factors of approximately 100, 40, and 35, respectively. In comparison, standard deviations from VMC using a conventional Slater–Jastrow–backflow Ansatz for first-row atoms[39] are uniformly worse than the VMCNet, despite using at least an order of magnitude more determinants. This reflects that although VMC-MD has higher accuracy, the flexibility of neural networks can capture the static correlation more stable. In general, for a given system, the variance decreases as the variable energy decreases. This is affected by the complexity of the particle structure. For Be, VMCNet reaches the standard deviation of 0.005 39(3), while PauliNet reaches 0.010 78(9). Both are in the order of 10^{-2}.

2.2 Equilibrium geometries

Table 2 gives the VMCNet energies and percentages of the correlation energy retrieved for each of the diatomic molecules studied with the same single-determinant VMCNet and training hyperparameters. The reference nonrelativistic energies E_{exact} are taken from Ref. [40]. On LiH, the VMCNet error was within chemical accuracy, defined as 1 kcal/mol (1.594 milli-Hartrees), which is the typical standard for a quantum chemical calculation to be considered "correct". In comparison, energies from PauliNet using a Slater–Jastrow–backflow Ansatz for first-row atoms[8] are uniformly worse than the VMCNet on LiH, Li_2, Be_2, and N_2, despite using higher dimensional feature vectors. VMCNet reaches high accuracy with only single determinants. The VMC-based VMCNet generalizes well to a wide variety of atoms and diatomic. There are many possible causes for the decline in the percent of correlation energy recovered for diatomic molecules like B_2, C_2. It may be that the VMCNet has issues with unpaired spins systems. In fact, our results are with a fixed-width network, while the total number of basis functions grows with the system size for coupled cluster, meaning the coupled cluster Ansatz becomes more expressive for larger systems while the VMCNet stays fixed. The results from the single-determinant network should be a lower bound on the performance of a Slater ansatz with a given number of determinants. This paper reports only the single determinant, and the smaller standard errors make VMCNet with more configurations and determinants should have more competitive results. This will be part of future work.

Figure 7 shows the optimization progress over time for many of these systems. The error in the correlation energy decreases monotonously as the training progresses from the initial HF baseline level to the final reported values. In most cases, the learning curve has not fully reached the platform state, which shows that our neural network ansatz is highly expressive and

Table 2 Groundstate energy at equilibrium geometry for diatomics. Also included are Hartree–Fock energies E_{HF} calculated using 6-311g, the exact energies E_{exact}[40], and the percentage of the correlation energy recovered at the VMCNet level (VMCNet-corr%) and PauliNet level (PauliNet-corr%). All energies are in atomic units and the numbers in parentheses indicate the statistical uncertainty in the last digit shown. The best results are in bold.

Molecule	VMCNet	PauliNet	E_{HF}	E_{exact}	VMCNet-corr%	PauliNet-corr%
LiH	−8.070 63(3)	−8.069 76(2)	−7.984 64(1)	−8.070 548	**100.10(0)%**	99.08(5)%
Li_2	−14.992 44(1)	−14.991 25(6)	−14.869 81(2)	−14.9954	**97.64(4)%**	96.70(0)%
Be_2	−29.324 07(8)	−29.322 64(8)	−29.130 67(0)	−29.338 54(5)	**93.04(0)%**	92.35(3)%
B_2	−49.386 11(6)	−49.387 36(7)	−49.069 22(2)	−49.415(2)	91.59(4)%	**91.95(5)%**
C_2	−75.870 26(2)	−75.871 44(2)	−75.362 78(0)	−75.923(5)	90.50(5)%	**90.71(6)%**
N_2	−109.494 92(6)	−109.492 18(0)	−108.894 17(1)	−109.5423	**92.69(1)%**	92.26(7)%

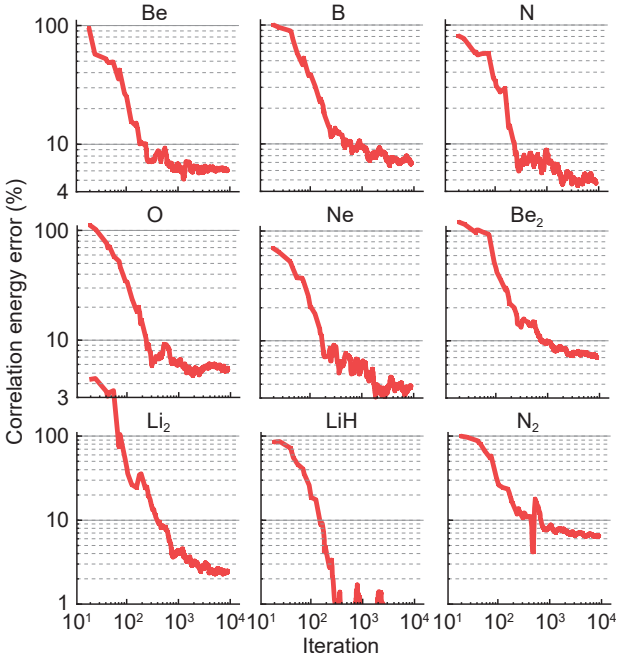

Fig. 7 Optimization progress for some monatomic and diatomic systems. Exponential moving average is applied to the energy at each iteration. For clarity, the median energy over the last 10% of iterations is shown.

shows that higher accuracy can be achieved using more computational resources. This suggests that deep learning can be an effective tool to reduce the large amount of determinant sampling in other VMC methods that act directly on fixed orbital determinants[41, 42] and is expected to further increase its accuracy. By introducing more physical information, VMCNet greatly increases the flexibility of ansatz. Actually, non-block blocks can capture more physical information about the system to cover more correlation energy.

2.3 Ablation experiments

There are many free parameters in the VMCNet architecture that must be chosen to maximize accuracy for a given amount of computation. VMCNet takes particle coordinates into account in the input. The non-local block is introduced to capture global interaction information. And there is additional out block for information aggregation. We compared the accuracy of the VMCNet with and without these features on monatomic and diatomic systems in Table 3. Some configurations, B, Li_2 and N_2, without the non-local block and out block are numerically unstable and diverged. The out block, an extra layer of information convergence, reduces the error by half. When including global features, the error is further reduced. This shows

Table 3 Performance of the VMCNet on monatomic and diatomic systems with features removed. Both non-local block and out block are critical. All numbers are relative to Refs. [40, 43]. mE_h means milli-Hartrees, "O" means without and "W" means with.

Configuration	ΔE (mE_h)			
	Be	B	Li_2	N_2
O-non-local&O-out	0.11	2.05	17.50	12.73
O-non-local&W-out	0.63	0.63	0.67	9.67

that global features and information aggregation contribute towards stability and accuracy.

3 Conclusion

We have designed VMCNet, a DNN representation of electronic wavefunctions ansatz, and shown that it can enable high-accuracy quantum chemistry calculations. The VMCNet Neural Network, which models electronic degrees of freedom and nonlocal interactions, makes the simple and straightforward single-determinant VMC method competitive for equilibrium geometries. The much higher flexibility of neural networks allows us not to have to choose the base set or worry about basis-set extrapolation, a crucial error reduction point in computational quantum chemistry. VMCNet includes physically motivated induction biases and encoding complex interaction information of particle systems through neural networks allows a variational approach to reach higher accuracy. VMCNet's self-iterative training mode makes it immune to the impact of data set quality, which is crucial for other neural networks, and there is no need to carefully select the pre-training data set.

Although VMCNet currently operates using a single-determinant approach, there is potential for incorporating multi-determinant expansions. This could theoretically lead to even higher accuracy, especially in systems with complex electronic correlations. Future work should focus on realizing multi-determinant expansion and extending it to larger, more complex systems.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Nos. U20A20225 and U2013601), Anhui Province Natural Science Funds for Distinguished Young Scholar (No. 2308085J02), the Science and Technology Innovation 2030 - "New Generation Artificial Intelligence" Major Project (No. 2022ZD0116305), Innovation Leading Talent of Anhui Province TeZhi plan, the Natural Science Foundation of Hefei, China (No. 202321), the CAAI-Huawei Mind Spore Open Fund (No.

CAAIXSJLJJ-2022-011A), the State Key Laboratory of Advanced Design and Manufacturing Technology for Vehicle (No. 32215010), and the State Key Laboratory of Intelligent Green Vehicle and Mobility (No. KFY2417).

Article History

Received: 24 April 2023; Revised: 15 December 2023; Accepted: 6 February 2024

References

[1] R. A. Friesner, *Ab initio* quantum chemistry: Methodology and applications, *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 19, pp. 6648–6653, 2005.

[2] R. J. Bartlett and M. Musiał, Coupled-cluster theory in quantum chemistry, *Rev. Mod. Phys.*, vol. 79, no. 1, pp. 291–352, 2007.

[3] K. T. Schütt, M. Gastegger, A. Tkatchenko, K. R. Müller, and R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions, *Nat. Commun.*, vol. 10, no. 1, p. 5024, 2019.

[4] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, *Rev. Mod. Phys.*, vol. 73, no. 1, pp. 33–83, 2001.

[5] B. M. Austin, D. Y. Zubarev, and W. A. Lester Jr, Quantum Monte Carlo and related approaches, *Chem. Rev.*, vol. 112, no. 1, pp. 263–288, 2012.

[6] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, Inhomogeneous backflow transformations in quantum Monte Carlo calculations, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, vol. 74, no. 6 Pt 2, p. 066701, 2006.

[7] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, *Phys. Rev. Lett.*, vol. 122, no. 22, p. 226401, 2019.

[8] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic Schrödinger equation, *Nat. Chem.*, vol. 12, no. 10, pp. 891–897, 2020.

[9] 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. Research*, vol. 2, no. 3, p. 033429, 2020.

[10] L. Otis and E. Neuscamman, Optimization stability in excited-state-specific variational Monte Carlo, *J. Chem. Theory Comput.*, vol. 19, no. 3, pp. 767–782, 2023.

[11] C. Roth, A. Szabó, and A. MacDonald, High accuracy variational Monte Carlo for frustrated magnets with deep neural networks, arXiv preprint arXiv: 2211.07749, 2022.

[12] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.*, vol. 98, no. 14, p. 146401, 2007.

[13] M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, *Phys. Rev. Lett.*, vol. 108, no. 5, p. 058301, 2012.

[14] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, and K. R. Müller, Machine learning of accurate energy-conserving molecular force fields, *Sci. Adv.*, vol. 3, no. 5, p. e1603015, 2017.

[15] M. Welborn, L. Cheng, and T. F. Miller Ⅲ, Transferability in machine learning for electronic structure via the molecular orbital basis, *J. Chem. Theory Comput.*, vol. 14, no. 9, pp. 4772–4779, 2018.

[16] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, Alchemical and structural distribution based representation for universal quantum machine learning, *J. Chem. Phys.*, vol. 148, no. 24, p. 241717, 2018.

[17] O. T. Unke and M. Meuwly, PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges, *J. Chem. Theory Comput.*, vol. 15, no. 6, pp. 3678–3693, 2019.

[18] M. Reh and M. Gärttner, Variational Monte Carlo approach to partial differential equations with neural networks, *Mach. Learn.: Sci. Technol.*, vol. 3, no. 4, p. 04LT02, 2022.

[19] A. Grisafi, A. Fabrizio, B. Meyer, D. M. Wilkins, C. Corminboeuf, and M. Ceriotti, Transferable machine-learning model of the electron density, *ACS Cent. Sci.*, vol. 5, no. 1, pp. 57–64, 2019.

[20] L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, and K. Burke, Kohn-sham equations as regularizer: Building prior knowledge into machine-learned physics, *Phys. Rev. Lett.*, vol. 126, no. 3, p. 036401, 2021.

[21] R. Nagai, R. Akashi, and O. Sugino, Completing density functional theory by machine learning hidden messages from molecules, *NPJ Comput. Mater.*, vol. 6, no. 1, p. 43, 2020.

[22] T. Husch, J. Sun, L. Cheng, S. J. R. Lee, and T. F. Miller Ⅲ, Improved accuracy and transferability of molecular-orbital-based machine learning: Organics, transition-metal complexes, non-covalent interactions, and transition states, *J. Chem. Phys.*, vol. 154, no. 6, p. 064108, 2021.

[23] J. S. Smith, O. Isayev, and A. E. Roitberg, ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost, *Chem. Sci.*, vol. 8, no. 4, pp. 3192–3203, 2017.

[24] K. T. Schütt, H. E. Sauceda, P. J. Kindermans, A. Tkatchenko, and K. R. Müller, SchNet-A deep learning architecture for molecules and materials, *J. Chem. Phys.*, vol. 148, no. 24, p. 241722, 2018.

[25] R. Zubatyuk, J. S. Smith, J. Leszczynski, and O. Isayev, Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network, *Sci. Adv.*, vol. 5, no. 8, p. eaav6490, 2019.

[26] Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, and T. F. Miller Ⅲ, OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features, *J. Chem. Phys.*, vol. 153, no. 12, p. 124111, 2020.

[27] A. S. Christensen, S. K. Sirumalla, Z. Qiao, M. B. O'Connor, D. G. A. Smith, F. Ding, P. J. Bygrave, A. Anandkumar, M. Welborn, F. R. Manby, et al, OrbNet Denali: A machine learning potential for biological and organic chemistry with semi-empirical cost and DFT accuracy, *J. Chem. Phys.*, vol. 155, no. 20, p. 204103, 2021.

[28] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, Big data meets quantum chemistry approximations: The Δ-machine learning approach, *J. Chem. Theory Comput.*, vol. 11, no. 5, pp. 2087–2096, 2015.

[29] J. Klicpera, J. Gro, and S. Günnemann, Directional message passing for molecular graphs. arXiv preprint arXiv: 2003.03123, 2020.

[30] O. T. Unke, S. Chmiela, M. Gastegger, K. T. Schütt, H. E. Sauceda, and K. R. Müller, SpookyNet: Learning force fields with electronic degrees of freedom and nonlocal effects, *Nat. Commun.*, vol. 12, no. 1, p. 7273, 2021.

[31] M. Born and W. Heisenberg, Zur quantentheorie der molekeln (in German), in *Original Scientific Papers Wissenschaftliche Originalarbeiten*, W. Blum, H. Rechenberg, and H. P. Dürr Eds. Berlin, Heidelberg, Germany: Springer, 1985, pp. 216–246.

[32] M. H. Kalos, *Monte Carlo Methods in Quantum Problems*. Dordrecht, the Netherlands: Springer, 2012.

[33] S. Sorella, Wave function optimization in the variational Monte Carlo method, *Phys. Rev. B*, vol. 71, no. 24, p. 241103, 2005.

[34] T. Kato, On the eigenfunctions of many-particle systems in quantum mechanics, *Commun. Pure Appl. Math.*, vol. 10, no. 2, pp. 151–177, 1957.

[35] 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, PySCF: the Python-based simulations of chemistry framework, *Wires Comput. Mol. Sci.*, vol. 8, no. 1, p. e1340, 2018.

[36] A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs, Scheme for adding electron-nucleus cusps to Gaussian orbitals, *J. Chem. Phys.*, vol. 122, no. 22, p. 224322, 2005.

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *Proc. 31st Conf. Neural Information Processing Systems* (*NIPS*

2017), Long Beach, CA, USA, 2017.

[38] M. D. Brown, J. R. Trail, P. López Ríos, and R. J. Needs, Energies of the first row atoms from quantum Monte Carlo, *J. Chem. Phys.*, vol. 126, no. 22, p. 224110, 2007.

[39] P. Seth, P. L. Ríos, and R. J. Needs, Quantum Monte Carlo study of the first-row atoms and ions, *J. Chem. Phys.*, vol. 134, no. 8, p. 084105, 2011.

[40] C. Filippi and C. J. Umrigar, Multiconfiguration wave functions for quantum Monte Carlo calculations of first-row diatomic molecules, *J. Chem. Phys.*, vol. 105, no. 1, pp. 213–226, 1996.

[41] G. H. Booth, A. J. W. Thom, and A. Alavi, Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space, *J. Chem. Phys.*, vol. 131, no. 5, p. 054106, 2009.

[42] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for ab-initio electronic structure, *Nat. Commun.*, vol. 11, no. 1, p. 2368, 2020.

[43] S. J. Chakravorty, S. R. Gwaltney, E. R. Davidson, F. A. Parpia, and C. F. p Fischer, Groundstate correlation energies for atomic ions with 3 to 18 electrons, *Phys. Rev. A*, vol. 47, no. 5, pp. 3649–3670, 1993.