June 2023 · Final Project of Course *AI for Science*

AI for physics: *ab-initio* methods for quantum many-body systems

Xiao-Yu Ouyang (欧阳霄宇), Wen-Hao He (何文昊), Shuo Zhou (周烁)

Introduction Solving Wave Function in Schrödinger Equation with Neural Networks

We explore in depth the numerous possibilities of integrating Artificial Intelligence with Physics (**AI+Physics**), specifically Quantum Mechanics. We elucidate the substantial power machine learning possess in solving a core problem in quantum mechanics—**the Schrödinger equation**.

Our investigation begins with a summary of two pertinent research paradigms, followed by the presentation of one such paradigm where machine learning is employed to solve quantum mechanical wave functions. Here **neural-networks-based-wavefunction** have the potential to bring about a paradigm shift in the efficiency and precision of the ***ab-initio* Quantum Monte Carlo (QMC)** methods, a rapidly developing and crucial direction in recent years. We analyze and delve into the frontier research in this area over the past five years, including but not limited to **FermiNet** and **PauliNet**, thoroughly examining the key factors that enhance their performance and providing our insightful observations.

DFT Density Functional Theory

Density Functional Theory (DFT) is a method in quantum mechanics used to compute the electronic structure of systems. Its principle posits that the ground state can be determined by optimizing an energy functional of the **electron density** and avoid using wave function,

$$E = E[n(\vec{r})]$$

The 1st paradigm in AI+physics: Using the labeled energy data from DFT is the first common categories in deep learning for calculating quantum systems. The data is trained for accurate force field in molecular dynamics (MD) simulation.

2 Paradigms

Algorithm	DFT	QMC
Precision	Low	High
Complexity	$O(N^3)$, low	$O(e^N)$, (full configuration)
Function optimized	Charge Density	Wave Function
Error-inducing ansatz	Exchange Correlation Term	Jastrow Factor & Backflow Transform
Applicable systems	Weak correlated	Strong correlated
Paradigm of "AI+"	Used as labeled data for training force field	DNNs represent wave function, complexity of $O(N^4)$
ML type	Supervised	Unsupervised

Quantum Monte Carlo QMC

In DFT, ansatz functionals must be used, which induce errors. Quantum Monte Carlo (QMC) is a more precise group of algorithms that use stochastic methods to compute the wave function and energy in Schrodinger equation. It optimizes the wave function $|\psi\rangle$ to gain the energy of the system,

$$E = \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle}$$

The 2nd paradigm in AI+physics: Using the unlabeled wave function is the second common categories in deep learning for calculating quantum systems. The data is trained for accurate ground state energy and wavefunction.

Preliminary: DeepWF

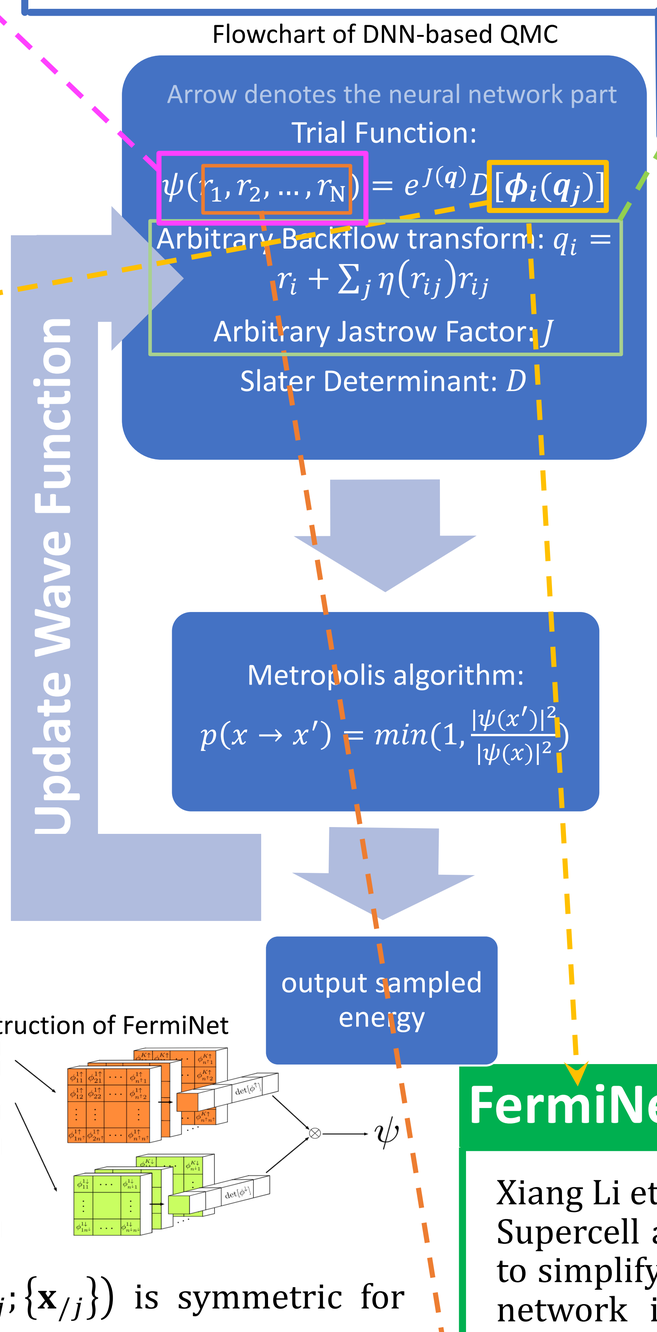
QMC+AI Algo.

The first instructive work trying to handle this problem is DeepWF method from Zhang, E et al. in 2018 [1]. They directly decompose the wavefunction by DNNs,

$$\psi(\mathbf{r}) = S(\mathbf{r})A^{\uparrow}(\mathbf{r}^{\uparrow})A^{\downarrow}(\mathbf{r}^{\downarrow})$$

Where symmetric $S(\mathbf{r})$ can be viewed as a DNN version of Jastrow factor J, and $A^{\uparrow/\downarrow}(\mathbf{r}^{\uparrow/\downarrow})$ substitute for the Slater determinant D, built of a series of antisymmetric blocks, $A(\mathbf{r}) = \prod_{i\neq j} a(\mathbf{r_i}, \mathbf{r_j})$, where the author inforce the antisymmetry by decomposing into 2 DNNs, $a(\mathbf{r_i}, \mathbf{r_j}) = \mathrm{Net}(\mathbf{r_i}, \mathbf{r_j}, r_{ij}) - \mathrm{Net}(\mathbf{r_j}, \mathbf{r_i}, r_{ij})$. In this way no more artificial term is chosen, and the amount of computation decreases a lot to only $O(N^4)$.

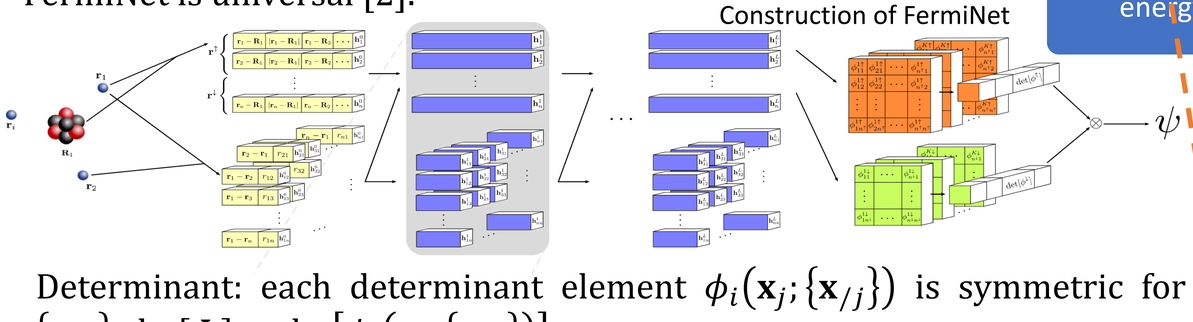
In QMC, some ansatzs are used in wavefunction when sampling energy. The wavefunction is written as $\psi(x) = e^{J(q)}D(q)$, where $D(q) = \det[\phi_i(q_j)]$ is the Slater determinant of single electron orbital WF, which ensures the Pauli exclusion principle that Fermions are antisymmetric when exchanging. $J(q)$ is the arbitrary **Jastrow factor**, and we take the arbitrary **backflow transformation** of the coordinate $q_i = r_i + \sum_j \eta(r_{ij})r_{ij}$. However, the limited parameter space for optimizing greatly restrict the accuracy of QMC's solution to Schrodinger equation. For exactly accurate results, we need to take WF as a linear combination of $O(e^N)$ nums of determinants, which is extremely consuming.



FermiNet Wavefunction DNNs

David Pfau etc. first applied FermiNet method (by DeepMind) to quantum chemistry in 2020[2]. This method greatly improves the accuracy of traditional QMC by using neural network as wave function ansatz, which is called FermiNet. FermiNet is more flexible than traditional determinant of single electron wave functions and models higher-order electron-electron interactions compactly. FermiNet representation avoids a finite basis with theoretical guarantee that a single FermiNet is universal [2].

Construction of FermiNet



Determinant: each determinant element $\phi_i(\mathbf{x}_j; \{\mathbf{x}_{/j}\})$ is symmetric for $\{\mathbf{x}_{/j}\}$, $\det[\mathbf{\Phi}] = \det[\phi_i(\mathbf{x}_j; \{\mathbf{x}_{/j}\})]$.

Permutation-equivariant layer: a way to construct $\phi_i(\mathbf{x}_j; \{\mathbf{x}_{/j}\})$

$$\mathbf{h}_i^{\ell+1\alpha} \leftarrow \tanh\left(\mathrm{matmul}\left(\mathbf{V}^l, \mathbf{f}_i^{\ell\alpha}\right) + \mathbf{b}^l\right) + \mathbf{h}_i^{\ell\alpha}$$

$$\mathbf{f}_i^{\ell\alpha} = \left(\mathbf{h}_i^{\ell\alpha}, \frac{1}{n^{\uparrow}}\sum_{j=1}^{n^{\uparrow}} \mathbf{h}_j^{\ell\uparrow}, \frac{1}{n^{\downarrow}}\sum_{j=1}^{n^{\downarrow}} \mathbf{h}_j^{\ell\downarrow}, \frac{1}{n^{\uparrow}}\sum_{j=1}^{n^{\uparrow}} \mathbf{h}_{ij}^{\ell\alpha\uparrow}, \frac{1}{n^{\downarrow}}\sum_{j=1}^{n^{\downarrow}} \mathbf{h}_{ij}^{\ell\alpha\downarrow}\right)$$

optimization strategy: instead of gradient descent, they used stochastic reconfiguration method[5], which approximates a imaginary time evolution in one small step:

$$\mathrm{Grad} = F^{-1}\nabla_\theta\langle E_l\rangle, \text{ and } F_{ij} = \mathrm{Re}\left[\left\langle\frac{\partial \ln\psi^*}{\partial\theta_i}\frac{\partial\ln\psi}{\partial\theta_j}\right\rangle - \left\langle\frac{\partial\ln\psi^*}{\partial\theta_i}\right\rangle\left\langle\frac{\partial\ln\psi}{\partial\theta_j}\right\rangle\right]$$

PauliNet Fewer DNNs to train

In 2020, J. Hermann et al. proposed PauliNet[7], which is relatively a less computationally damanding method than FermiNet. In contrast, PauliNet doesn't use DNNs to directly represent the wave function, but use DNNs to represent the **Jastrow factor** J and the **backflow transformation f** in the many-body wave function, which are arbitrary ansatz that may cause error in conventional QMC. The other structure of the determinant and single-body orbit WF in QMC are remained. That is, J and $\mathbf{f}$ are encoded in DNNs with trainable parameters θ:

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = \mathrm{e}^{\gamma(\mathbf{r})+J_\theta(\mathbf{r})}\sum_p c_p \det\left[\tilde{\varphi}^{\uparrow}_{\boldsymbol{\theta},\mu_p i}(\mathbf{r})\right]\det\left[\tilde{\boldsymbol{\varphi}}^{\downarrow}_{\boldsymbol{\theta},\mu_p i}(\mathbf{r})\right]$$

$$\tilde{\boldsymbol{\varphi}}_{\mu i}(\mathbf{r}) = \varphi_\mu(\mathbf{r}_i) f_{\boldsymbol{\theta},\mu i}(\mathbf{r})$$

Where $\varphi_\mu(\mathbf{r}_i)$ is the HF-optimized one-electron molecular orbitals, which are then used as an input to PauliNet and are modified during training only by the backflow transformation and the previous Jastrow factor.

Thus the number of parameters that PauliNet needs to train is much fewer than that in FermiNet, which saves a lot of computational cost. Compared to DeepWF, PauliNet remains more physical formations in QMC and only uses two artificial term as net, so it's 10th more accurate than DeepWF.

FermiNet: Solid FermiNet on periodic lattice

Xiang Li etc. combined periodic distant feature with FermiNet in 2022[3]. Supercell approximation with periodic boundary condition is employed to simplify the boundary. Bloch function together with a periodic neural network input ensures the periodic feature. The periodic input is constructed by tow steps: 1. electron coordinates modulo spatial period 2. calculate periodic distant matrices following scheme of scheme of Whitehead et al.[4].

FermiNet: ECP Refined with effective core potential

FermiNet and PauliNet are both $\mathcal{O}(N^4)$ scailing with the system size N and suffer a large prefactor. Thus, X. Li et al. proposed to leverage the effective core potential method to substitute the electronic-interaction term in Born-Oppenheimer Approximation[6]:

$$\hat{V}_{\mathrm{ECP}} = \sum_{v=1}^{n_v} V_{\mathrm{loc}}(r_v) + \sum_{v=1}^{n_v}\sum_{l=0}^{l_{\max}} V_l(r_v)\sum_{m=-l}^{m=l} |lm\rangle\langle lm|$$

The main workflow can be divided into three phases: pretrain, train, and inference. With the ECP employed, the gradient formula for energy optimization is slightly modified as follows:

$$\mathrm{Grad} = \mathbb{E}_{\psi^2(\mathbf{r})}\left[\left(E(\mathbf{r}) - \mathbb{E}_{\psi^2(\mathbf{r})}[E(\mathbf{r})]\right)\nabla_\theta \log|\psi|\right]$$

$$E(\mathbf{r}) = \psi^{-1}(\mathbf{r})\hat{H}\psi(\mathbf{r}) + E_{\mathrm{nl}}(\mathbf{r})$$

Contribution: Most of the work was accomplished by Xiao-Yu Ouyang, including the creation and layout of the poster, the allocation of tasks within the work framework, and the drafting of all main text content excluding that pertaining to FermiNet. Wen-Hao He and Shuo Zhou provided assistance in completing the sections of the main text related to FermiNet.

Reference: [1] J. Han, L. Zhang, and W. E, Journal of Computational Physics **399**, 108929 (2019).
[2] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Phys. Rev. Res. **2**, 033429 (2020).
[3] X. Li, Z. Li, and J. Chen, Nat Commun **13**, 7895 (2022).
[4] W. S. Davidson and T. G. Flynn, Biochem J **177**, 595 (1979).
[5] S. Sorella, Phys. Rev. Lett. **80**, 4558 (1998).
[6] X. Li, C. Fan, W. Ren, and J. Chen, Phys. Rev. Res. **4**, 013021 (2022).
[7] J. Hermann, Z. Schätzle, and F. Noé, Nat. Chem. **12**, 891 (2020).