香港中文大學
The Chinese University of Hong Kong

Journal Club: *Ab initio* solution of the many-electron Schrödinger equation with deep neural networks

Huiwen Tan

11 Jul 2025

Ab initio solution of the many-electron Schrödinger equation with deep neural networks

David Pfau,[*,†] James S. Spencer,[*] and Alexander G. D. G. Matthews
DeepMind, 6 Pancras Square, London N1C 4AG, United Kingdom

W. M. C. Foulkes
Department of Physics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

(Received 18 March 2020; accepted 6 August 2020; published 16 September 2020)

Given access to accurate solutions of the many-electron Schrödinger equation, nearly all chemistry could be derived from first principles. Exact wave functions of interesting chemical systems are out of reach because they are NP-hard to compute in general, but approximations can be found using polynomially scaling algorithms. The key challenge for many of these algorithms is the choice of wave function approximation, or Ansatz, which must trade off between efficiency and accuracy. Neural networks have shown impressive power as accurate practical function approximators and promise as a compact wave-function Ansatz for spin systems, but problems in electronic structure require wave functions that obey Fermi-Dirac statistics. Here we introduce a novel deep learning architecture, the Fermionic neural network, as a powerful wave-function Ansatz for many-electron systems. The Fermionic neural network is able to achieve accuracy beyond other variational quantum Monte Carlo Ansatz on a variety of atoms and small molecules. Using no data other than atomic positions and charges, we predict the dissociation curves of the nitrogen molecule and hydrogen chain, two challenging strongly correlated systems, to significantly higher accuracy than the coupled cluster method, widely considered the most accurate scalable method for quantum chemistry at equilibrium geometry. This demonstrates that deep neural networks can improve the accuracy of variational quantum Monte Carlo to the point where it outperforms other *ab initio* quantum chemistry methods, opening the possibility of accurate direct optimization of wave functions for previously intractable many-electron systems.

DOI: 10.1103/PhysRevResearch.2.033429

The Many-Electron Problem in Quantum Chemistry

The time-independent Schrödinger equation of a chemical system with *n* electrons,

$$\hat{H}\psi(\mathbf{x}_1,\ldots,\mathbf{x}_n) = E\psi(\mathbf{x}_1,\ldots,\mathbf{x}_n) \tag{1}$$

where the Hamiltonian is,

$$\begin{aligned}\hat{H} &= \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ne} + \hat{T}_n + \hat{V}_{nn} \\ &= -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 + \sum_{i>j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &\quad -\sum_{iI}\frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} - \sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 + \sum_{I>J}\frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}\end{aligned} \tag{2}$$

Under the Born-Oppenheimer (BO) approximation,

$$\hat{H}_{\mathrm{BO}} = -\frac{1}{2}\sum_i \nabla_i^2 + \sum_{i>j}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{iI}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \quad \text{in a.u.} \tag{3}$$

Variational Principle and Fermi–Dirac Statistics

Since the many-electron Schrödinger equation cannot be solved analytically, the **variational principle** enables numerical approximation of the ground-state wave function,

$$E_{\text{trial}} = \frac{\langle \Psi_{\text{trial}} | \hat{H} | \Psi_{\text{trial}} \rangle}{\langle \Psi_{\text{trial}} | \Psi_{\text{trial}} \rangle} = \frac{\int \Psi^*_{\text{trial}}(\mathbf{X}) \hat{H} \Psi_{\text{trial}}(\mathbf{X})\, d\mathbf{X}}{\int \Psi^*_{\text{trial}}(\mathbf{X}) \Psi_{\text{trial}}(\mathbf{X})\, d\mathbf{X}} \geq E_0 \tag{4}$$

where the $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$.The trial wave function must obey **Fermi-Dirac statistics**, i.e.,

$$\psi(\ldots, \mathbf{x}_i, \ldots, \mathbf{x}_j, \ldots) = -\psi(\ldots, \mathbf{x}_j, \ldots, \mathbf{x}_i, \ldots) \tag{5}$$

The **Slater determinant** provides the simplest way to construct an antisymmetric wave function.

$$\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \frac{1}{\sqrt{n!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \cdots & \phi_n(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \cdots & \phi_n(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_n) & \phi_2(\mathbf{x}_n) & \cdots & \phi_n(\mathbf{x}_n) \end{vmatrix} \tag{6}$$

Wave Function Ansatz

Hartree–Fock:

$$(\mathbf{x}_1,\ldots,\mathbf{x}_n) \xrightarrow[\text{orbital}]{\text{single-electron}} \phi_i(\mathbf{x}_j) \xrightarrow[\text{determinant}]{\text{Slater}} \det[\phi_i(\mathbf{x}_j)] = \psi_{\text{HF}}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$$

Configuration interaction:

$$(\mathbf{x}_1,\ldots,\mathbf{x}_n) \xrightarrow[\text{orbital}]{\text{single-electron}} \phi_i^k(\mathbf{x}_j) \xrightarrow[\text{determinant}]{\text{Slater}} \det[\phi_i^k(\mathbf{x}_j)] \xrightarrow[\text{combination}]{\text{linear}} \psi_{\text{CI}}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$$

Slater–Jastrow:

$$(\mathbf{x}_1,\ldots,\mathbf{x}_n) \xrightarrow[\text{orbital}]{\text{single-electron}} \phi_i^k(\mathbf{x}_j) \xrightarrow[\text{determinant}]{\text{Slater}} \det[\phi_i^k(\mathbf{x}_j)] \xrightarrow[\text{correlation}]{\text{Jastrow factor}} \psi_{\text{SJ}}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$$

Slater–Jastrow–Backflow:

$$(\mathbf{x}_1,\ldots,\mathbf{x}_n) \xrightarrow[\text{orbital}]{\text{single-electron}} \phi_i^k(\mathbf{x}_j+\xi_j) \xrightarrow[\text{determinant}]{\text{Slater}} \det[\phi_i^k(\mathbf{x}_j+\xi_j)] \xrightarrow[\text{correlation}]{\text{Jastrow factor}} \psi_{\text{SJB}}(\mathbf{x}_1,\ldots,\mathbf{x}_n)$$

Two Key Observations of FermiNet

1. Artificial neural networks are very flexible and effective function approximators
   - perhaps deep neural networks can be used to represent $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_n)$
2. Nothing requires the single-electron wave functions (orbitals) in a Slater determinant. The only requirement is that exchanging any two input variables, $\mathbf{x}_i$ and $\mathbf{x}_j$, exchanges two columns
   - this allows for the replacement of single-electron orbitals $\phi_i(\mathbf{x}_j)$ with multi-electron functions

$$\phi_i(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_n) = \phi_i(\mathbf{x}_j; \{\mathbf{x}_{/j}\}) \tag{7}$$

   - and the many-electron wave function remains antisymmetric

$$\begin{aligned}\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) &= \sum_k \omega_k \det[\phi_i^k(\mathbf{x}_j; \{\mathbf{x}_{/j}\})] \\ &= \sum_k \omega_k \begin{vmatrix} \phi_1^k(\mathbf{x}_1; \{\mathbf{x}_{/1}\}) & \phi_2^k(\mathbf{x}_1; \{\mathbf{x}_{/1}\}) & \cdots & \phi_n^k(\mathbf{x}_1; \{\mathbf{x}_{/1}\}) \\ \phi_1^k(\mathbf{x}_2; \{\mathbf{x}_{/2}\}) & \phi_2^k(\mathbf{x}_2; \{\mathbf{x}_{/2}\}) & \cdots & \phi_n^k(\mathbf{x}_2; \{\mathbf{x}_{/2}\}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1^k(\mathbf{x}_n; \{\mathbf{x}_{/n}\}) & \phi_2^k(\mathbf{x}_n; \{\mathbf{x}_{/n}\}) & \cdots & \phi_n^k(\mathbf{x}_n; \{\mathbf{x}_{/n}\}) \end{vmatrix}\end{aligned} \tag{8}$$

Two Key Observations of FermiNet

1. Artificial neural networks are very flexible and effective function approximators
   - perhaps deep neural networks can be used to represent $\psi(\mathbf{x}_1, \ldots, \mathbf{x}_n)$
2. Nothing requires the single-electron wave functions (orbitals) in a Slater determinant. The only requirement is that exchanging any two input variables, $\mathbf{x}_i$ and $\mathbf{x}_j$, exchanges two columns
   - this allows for the replacement of single-electron orbitals $\phi_i(\mathbf{x})$ with multi-electron functions

$$\phi_i(\mathbf{x}_j; \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_n) = \phi_i(\mathbf{x}_j; \{\mathbf{x}_{/j}\}) \tag{7}$$

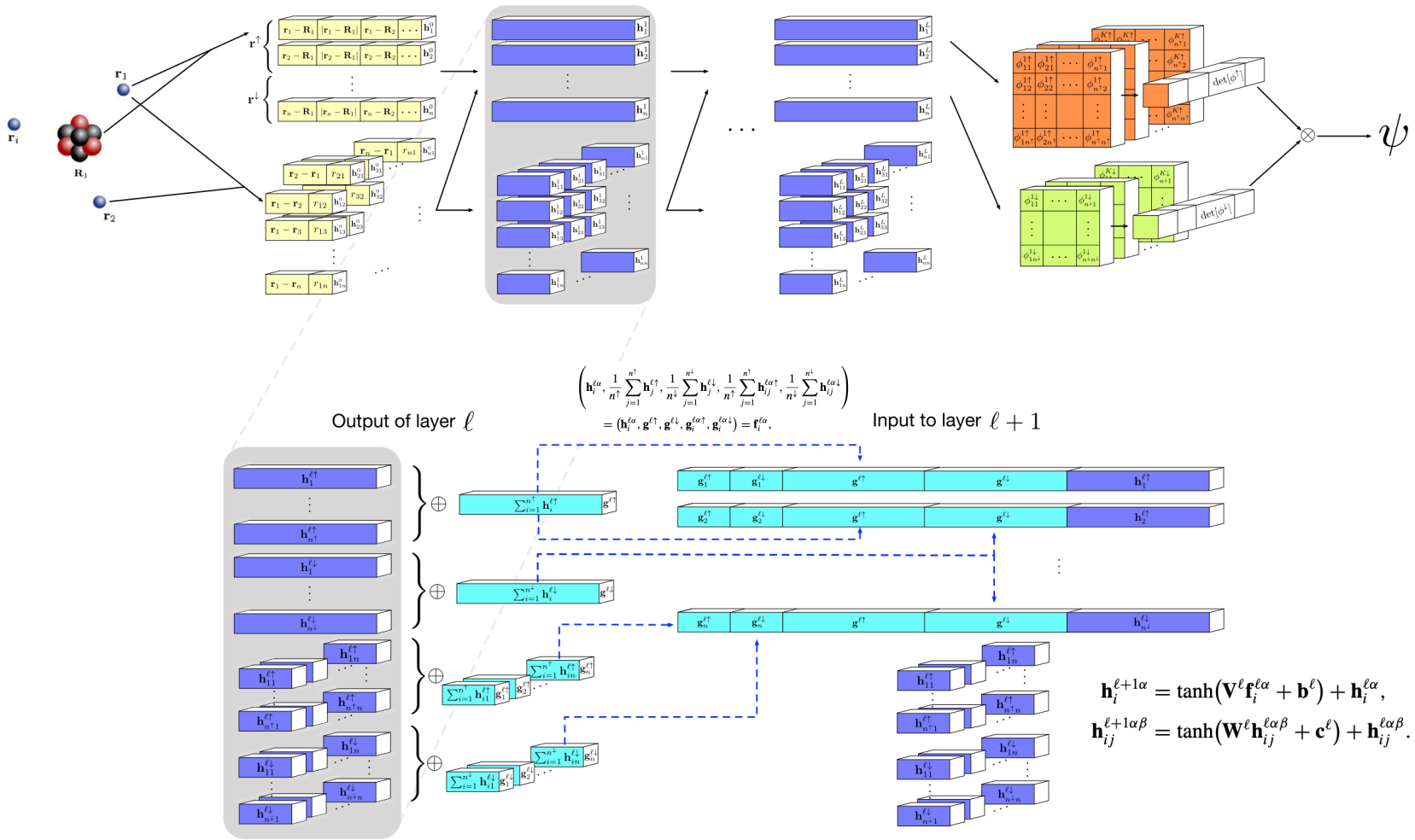
   - and the many-electron wave function remains antisymmetric

$$\psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) = \sum_k \omega_k \det[\phi_i^k(\mathbf{x}_j; \{\mathbf{x}_{/j}\})]$$

$$= \sum_k \omega_k \begin{vmatrix} \phi_1^k(\mathbf{x}_1; \{\mathbf{x}_{/1}\}) & \phi_2^k(\mathbf{x}_1; \{\mathbf{x}_{/1}\}) & \cdots & \phi_n^k(\mathbf{x}_1; \{\mathbf{x}_{/1}\}) \\ \phi_1^k(\mathbf{x}_2; \{\mathbf{x}_{/2}\}) & \phi_2^k(\mathbf{x}_2; \{\mathbf{x}_{/2}\}) & \cdots & \phi_n^k(\mathbf{x}_2; \{\mathbf{x}_{/2}\}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1^k(\mathbf{x}_n; \{\mathbf{x}_{/n}\}) & \phi_2^k(\mathbf{x}_n; \{\mathbf{x}_{/n}\}) & \cdots & \phi_n^k(\mathbf{x}_n; \{\mathbf{x}_{/n}\}) \end{vmatrix} \tag{8}$$

The construction of these permutation-equivariant functions with a neural network is the main innovation of FermiNet

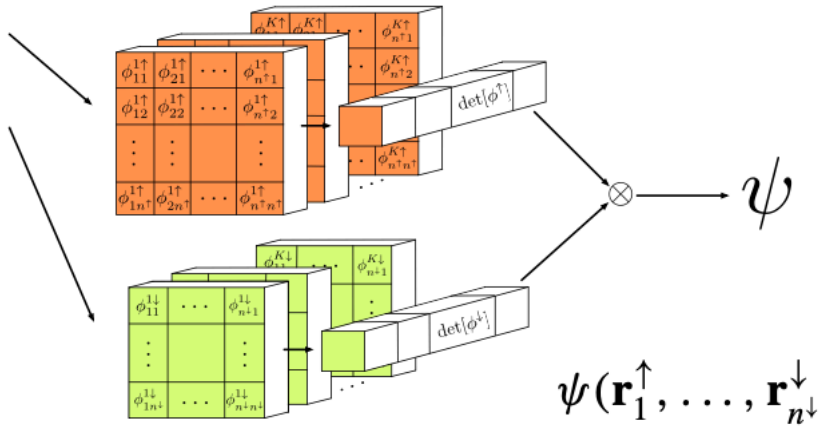
Ferminonic Neural Network Architecture



Require: walker configuration $\{\mathbf{r}_1^\uparrow, \cdots, \mathbf{r}_{n^\uparrow}^\uparrow, \mathbf{r}_1^\downarrow, \cdots, \mathbf{r}_{n^\downarrow}^\downarrow\}$
Require: nuclear positions $\{\mathbf{R}_I\}$
for each electron i, α **do**
  $\mathbf{h}_i^{\ell\alpha} \leftarrow \text{concatenate}(\mathbf{r}_i^\alpha - \mathbf{R}_I, |\mathbf{r}_i^\alpha - \mathbf{R}_I| \; \forall I)$
  $\mathbf{h}_{ij}^{\ell\alpha\beta} \leftarrow \text{concatenate}(\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta, |\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta| \; \forall j, \beta)$
end for
for each layer $\ell \in \{0, L-1\}$ **do**
  $\mathbf{g}^{\ell\uparrow} \leftarrow \frac{1}{n^\uparrow} \sum_i^{n^\uparrow} \mathbf{h}_i^{\ell\uparrow}$
  $\mathbf{g}^{\ell\downarrow} \leftarrow \frac{1}{n^\downarrow} \sum_i^{n^\downarrow} \mathbf{h}_i^{\ell\downarrow}$
  for each electron i, α **do**
    $\mathbf{g}_i^{\ell\alpha\uparrow} \leftarrow \frac{1}{n^\uparrow} \sum_j^{n^\uparrow} \mathbf{h}_{ij}^{\ell\alpha\uparrow}$
    $\mathbf{g}_i^{\ell\alpha\downarrow} \leftarrow \frac{1}{n^\downarrow} \sum_j^{n^\downarrow} \mathbf{h}_{ij}^{\ell\alpha\downarrow}$
    $\mathbf{f}_i^{\ell\alpha} \leftarrow \text{concatenate}(\mathbf{h}_i^{\ell\alpha}, \mathbf{g}^{\ell\uparrow}, \mathbf{g}^{\ell\downarrow}, \mathbf{g}_i^{\ell\alpha\uparrow}, \mathbf{g}_i^{\ell\alpha\downarrow})$
    $\mathbf{h}_i^{\ell+1\alpha} \leftarrow \tanh\left(\text{matmul}(\mathbf{V}^l, \mathbf{f}_i^{\ell\alpha}) + \mathbf{b}^l\right) + \mathbf{h}_i^{\ell\alpha}$
    $\mathbf{h}_{ij}^{\ell+1\alpha\beta} \leftarrow \tanh\left(\text{matmul}(\mathbf{W}^l, \mathbf{h}_{ij}^{\ell\alpha\beta}) + \mathbf{c}^l\right) + \mathbf{h}_{ij}^{\ell\alpha\beta}$
  end for
end for
for each determinant k **do**
  for each orbital i **do**
    for each electron j, α **do**
      $e \leftarrow \text{envelope}(\mathbf{r}_j^\alpha, \{r_i^\alpha - \mathbf{R}_I\})$
      $\phi_i(\mathbf{r}_j^\alpha; \{\mathbf{r}_{/j}^\alpha\}; \{\mathbf{r}^{\bar{\alpha}}\}) = (\text{dot}(\mathbf{w}_i^{k\alpha}, \mathbf{h}_j^{L\alpha}) + g_i^{k\alpha})e$
    end for
  end for
  $D^{k\uparrow} \leftarrow \det\left[\phi_i^{k\uparrow}(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}^\uparrow\}; \{\mathbf{r}^\downarrow\})\right]$
  $D^{k\downarrow} \leftarrow \det\left[\phi_i^{k\downarrow}(\mathbf{r}_j^\downarrow; \{\mathbf{r}_{/j}^\downarrow\}; \{\mathbf{r}^\uparrow\})\right]$
end for
$\psi \leftarrow \sum_k \omega_k D^{k\uparrow} D^{k\downarrow}$

Output Layer of FermiNet

$$\phi_i^{k\alpha}\left(\mathbf{r}_j^{\alpha};\{\mathbf{r}_{/j}^{\alpha}\};\{\mathbf{r}^{\bar{\alpha}}\}\right)=\left(\mathbf{w}_i^{k\alpha}\cdot\mathbf{h}_j^{L\alpha}+g_i^{k\alpha}\right)\times\sum_m \pi_{im}^{k\alpha}\exp\left(-\left|\mathbf{\Sigma}_{im}^{k\alpha}\left(\mathbf{r}_j^{\alpha}-\mathbf{R}_m\right)\right|\right),$$



$$\psi(\mathbf{r}_1^{\uparrow},\ldots,\mathbf{r}_{n^{\downarrow}}^{\downarrow})=\sum_k \omega_k\left(\det\left[\phi_i^{k\uparrow}(\mathbf{r}_j^{\uparrow};\{\mathbf{r}_{/j}^{\uparrow}\};\{\mathbf{r}^{\downarrow}\})\right]\times\det\left[\phi_i^{k\downarrow}(\mathbf{r}_j^{\downarrow};\{\mathbf{r}_{/j}^{\downarrow}\});\{\mathbf{r}^{\uparrow}\};\right]\right).$$

Wave Function Optimization

1. Optimization Objective

$$E[\psi_\theta] = \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \frac{\int \psi_\theta^*(\mathbf{X}) \hat{H} \psi_\theta(\mathbf{X})\, d\mathbf{X}}{\int \psi_\theta^*(\mathbf{X}) \psi_\theta(\mathbf{X})\, d\mathbf{X}} \tag{9}$$

2. The Metropolis-Hastings Markov Chain Monte Carlo (MCMC) is used to sample the configuration based on the probability distribution, $\mathbf{X} \sim |\psi_\theta(\mathbf{X})|^2$. The local energy is,

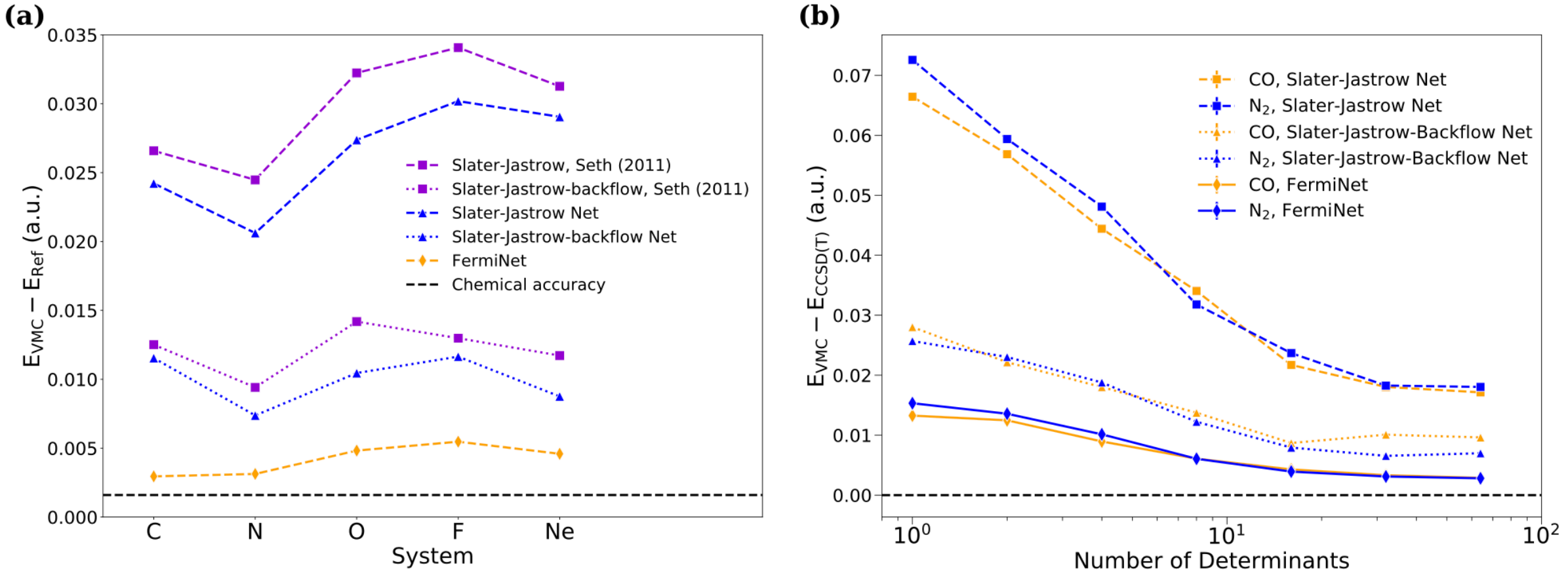
$$E_{\mathrm{L}}(\mathbf{X}) = \psi_\theta^{-1}(\mathbf{X}) \hat{H} \psi_\theta(\mathbf{X}) \tag{10}$$

3. Evaluate the energy expectation from

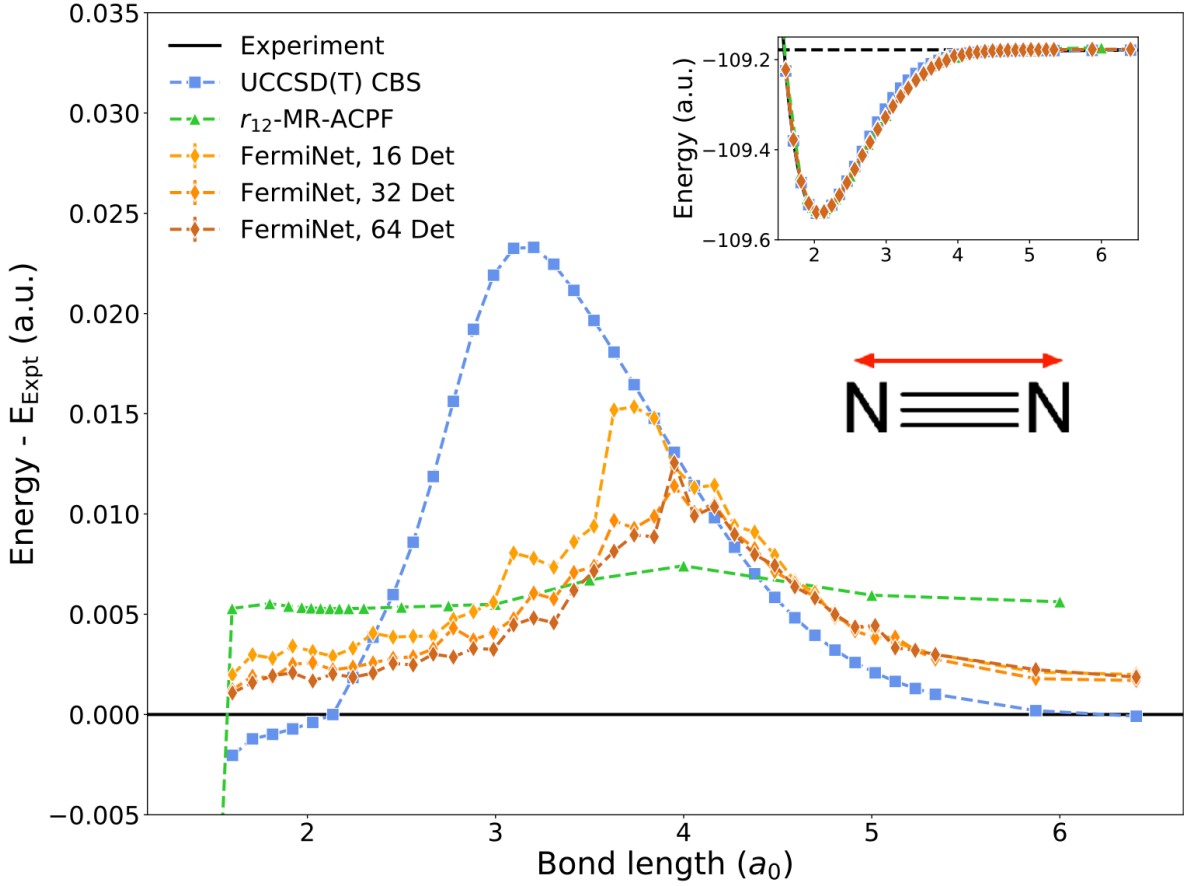
$$E[\psi_\theta] \approx \frac{1}{M} \sum_{i=1}^{M} E_{\mathrm{L}}(\mathbf{X}_i) \tag{11}$$

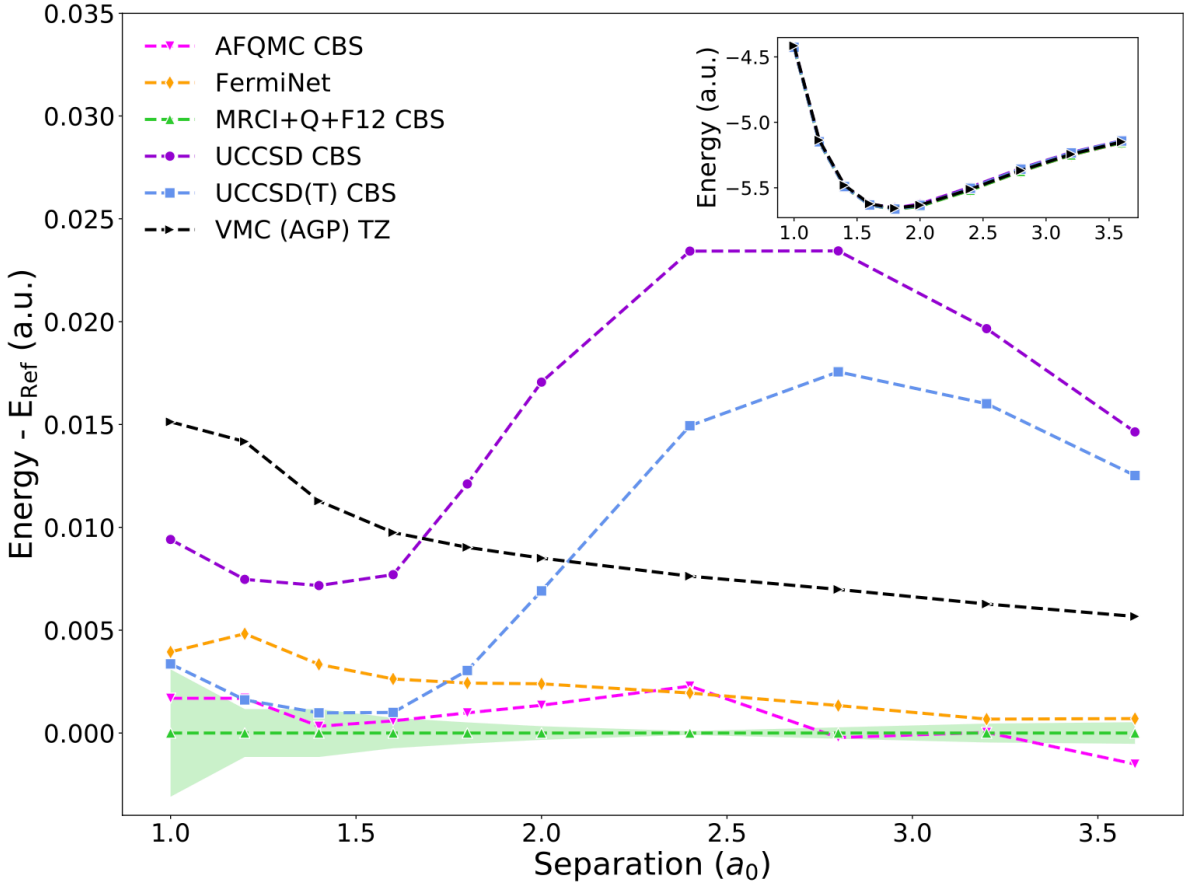
4. Then the Kronecker-factored approximate curvature (KFAC) is used to update parameter θ, until the energy converges

Results – Slater-Jastrow versus FermiNet Ansatz



Results – Nitrogen Molecule and Hydrogen Chain





Conclusion

1. The many-electron Schrödinger equation is intractable due to electron–electron interactions.
2. Traditional wave function methods rely on Slater determinants and approximate correlation via expansion or correction.
3. FermiNet introduces a new paradigm by directly parameterizing the wavefunction with a neural network, while preserving antisymmetric via determinant structure.
4. It achieves chemical accuracy for atoms and small molecules using only a single determinant, outperforming traditional variational quantum Monte Carlo methods.
5. This work bridges quantum chemistry and deep learning, opening a scalable path to *ab initio* solutions using modern ML techniques.

Thank You