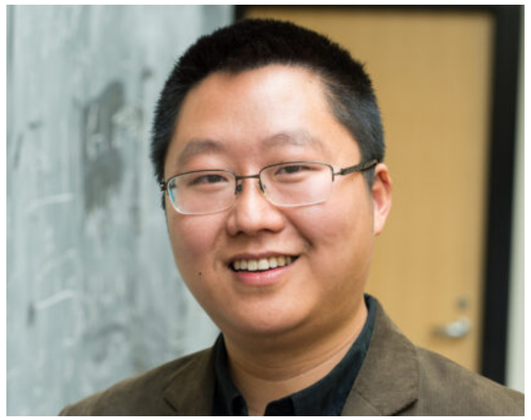
Solving Spinful Fermi Systems with a Self-Attention Ansatz

Alex Avdoshkin

Based on arXiv:2510.18621
Accepted to PRB

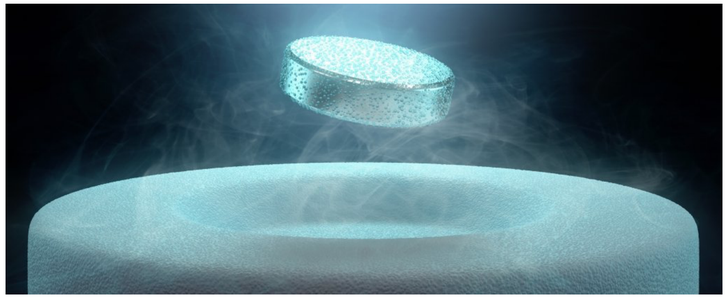


Max Geier

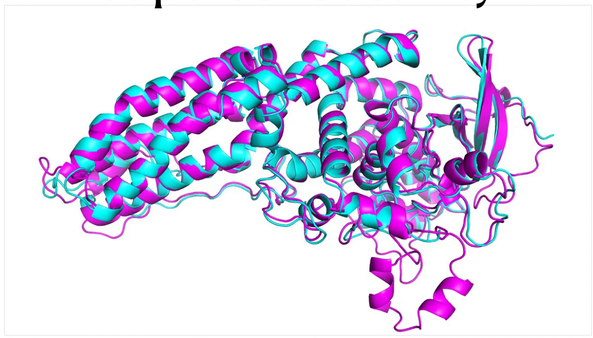
Liang Fu

SubMIT User's group meeting, May 5 2026

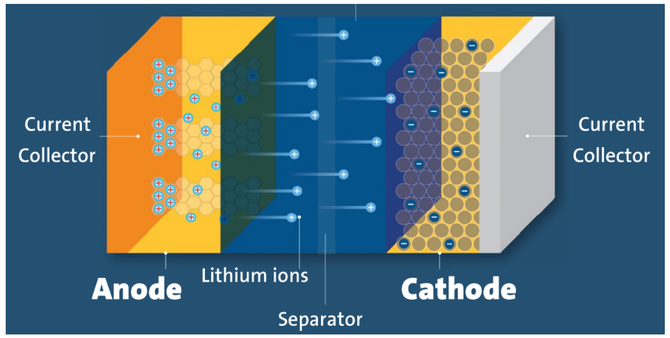
Why solve Schrödinger's equation?



Superconductivity



Proteins



Batteries

The chemical space is vast and only a tiny fraction can be probed in the lab.

Electronic structure is at the core of our ability to simulate materials, chemicals and drugs.

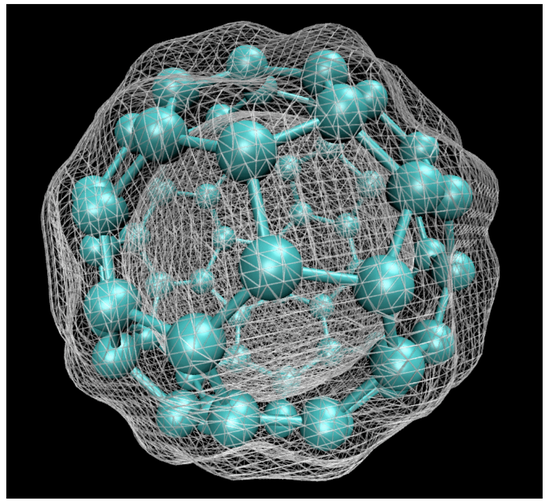
We need reliable (and fast) first principles methods to predict electronic structure

Problem statement

Stationary Schrödinger's equation: $\hat{H}\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$

Energy operator: $\hat{H} = \sum_i \frac{1}{2}\frac{\partial^2}{\partial \mathbf{r}_i^2} + \sum_i V_1(\mathbf{r}_i) + \sum_{i<j} V_2(\mathbf{r}_i - \mathbf{r}_j)$

Pauli principle: $\Psi(\ldots, \mathbf{r}_i, \ldots, \mathbf{r}_j, \ldots) = -\Psi(\ldots, \mathbf{r}_j, \ldots, \mathbf{r}_i, \ldots)$



Electronic structure

Credit: wikipedia

The lowest E - ground state

Variational principle: $|\Psi_0\rangle = \operatorname{argmin}_{|\Psi\rangle} \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle}$

Variational Monte Carlo

Step 1. Specify the **variational wavefunction** Ψ_θ - we use a neural network

Step 2. Define an **energy (loss) estimator** - Markov chain Monte Carlo (MCMC) for us

Step 3. Chose and optimizer to **update the parameters** (weights and biases): $\theta \rightarrow \theta'$

Variational Monte Carlo

Step 1. Specify the **variational wavefunction** Ψ_θ - we use a neural network

Step 2. Define an **energy (loss) estimator** - Markov chain Monte Carlo (MCMC) for us

Step 3. Chose and optimizer to **update the parameters** (weights and biases): $\theta \rightarrow \theta'$

The wavefunction

Slater determinant: $$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \det \begin{pmatrix} \varphi_1(\mathbf{r}_1) & \cdots & \varphi_1(\mathbf{r}_N) \\ \cdots & \cdots & \cdots \\ \varphi_N(\mathbf{r}_1) & \cdots & \varphi_N(\mathbf{r}_N) \end{pmatrix}$$

Feature vector $\mathbf{f}_i^0 = \mathbf{r}_i$ Embed the features $\mathbf{h}_i^0 = W^0 \mathbf{f}_i^0$

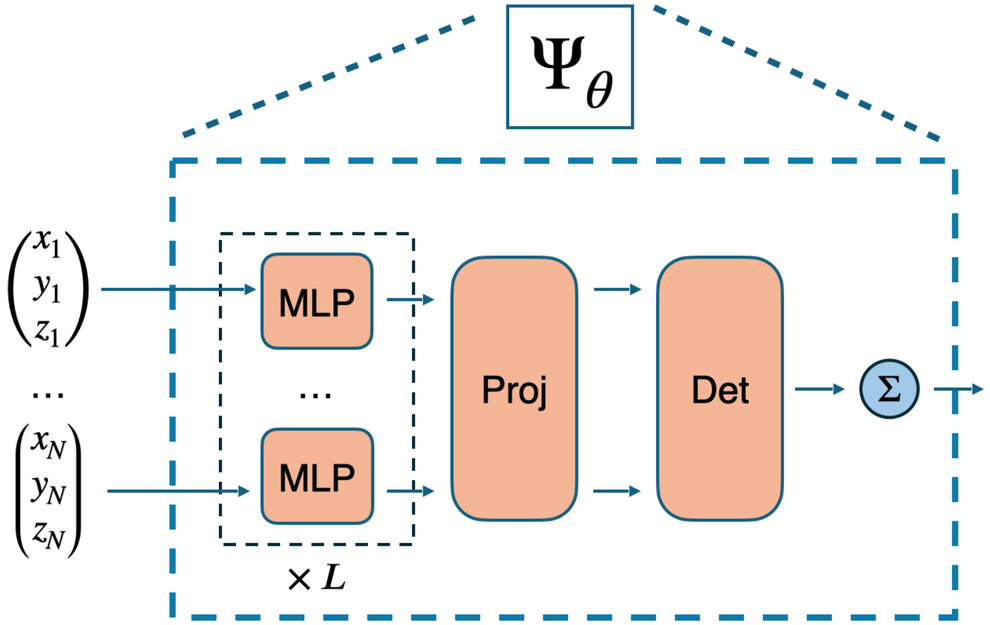
Apply MLP

$$\mathbf{h}_i^{l+1} = \mathbf{h}_i^l + \tanh(W^{l+1}\mathbf{h}_i^l + \mathbf{b}_i^{l+1})$$

Project

$$\varphi_j(\xi_i) = \mathbf{w}_j \cdot \mathbf{h}_i + i\mathbf{w}'_j \cdot h_i,$$

Take det after



Self-Attention

Generalized Slater determinant:

$$\Psi(\mathbf{r}_1, \ldots, \mathbf{r}_N) = \det \begin{pmatrix} \varphi_1(\mathbf{r}_1, \mathbf{r}_{/1}) & \cdots & \varphi_1(\mathbf{r}_N, \mathbf{r}_{/N}) \\ \cdots & \cdots & \cdots \\ \varphi_N(\mathbf{r}_1, \mathbf{r}_{/1}) & \cdots & \varphi_N(\mathbf{r}_N, \mathbf{r}_{/N}) \end{pmatrix}$$

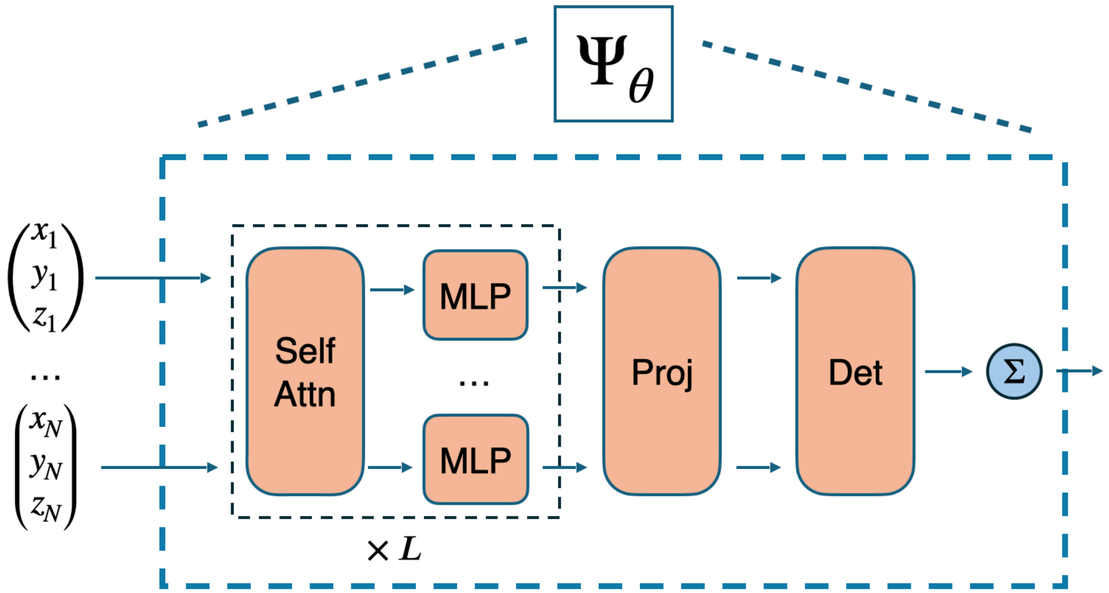
Now apply SelfAttn between MLPs

$$\mathrm{SelfAttn}_i^l = \frac{1}{\mathcal{N}} \sum_{j=1}^{N} \exp\left(\frac{(W_q \mathbf{f}_j^l) \cdot (W_k \mathbf{f}_i^l)}{\sqrt{d_{\mathrm{Attn}}}} \right) (W_v \mathbf{f}_j^l)$$

$$\mathbf{f}_i^{l+1} = \mathbf{h}_i^l + W_o \cdot \mathrm{SelfAttn}_i^l$$

$$\mathbf{h}_i^{l+1} = \mathbf{f}_i^{l+1} + \tanh(W^{l+1} \mathbf{f}_i^{l+1} + \mathbf{b}_i^{l+1})$$

Universal Approximation Theorem



Pfau et al., Phys Rev. Research 2, 033429 (2020)

Inclusion of spin

New state vector $\{\mathbf{r}_i, s_i\}$ where $\mathbf{r}_i \in \mathbb{R}^d$ and $s_i \in \{-1, 1\}$

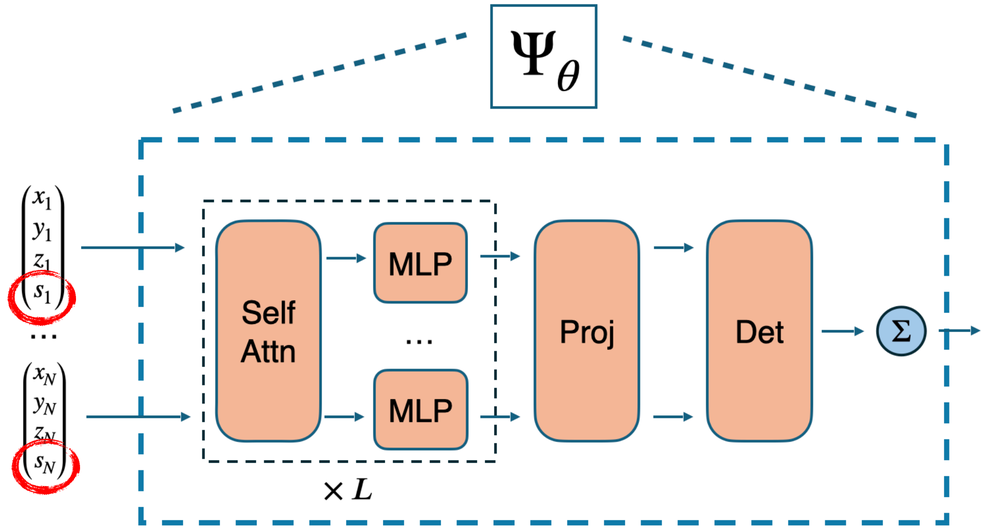
Pauli principle:

$$\Psi(\ldots, \{\mathbf{r}_i, s_i\}, \ldots, \{\mathbf{r}_j, s_j\}, \ldots) = -\Psi(\ldots, \{\mathbf{r}_j, s_j\}, \ldots, \{\mathbf{r}_i, s_i\}, \ldots)$$

The energy operator is now a matrix:

$$\hat{H} = \left(\sum_i \frac{1}{2} \frac{\partial^2}{\partial \mathbf{r}_i^2} + \sum_i V_1(\mathbf{r}_i) \right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} B_z(\mathbf{r}) & B_x(\mathbf{r}) - iB_y(\mathbf{r}) \\ B_x(\mathbf{r}) + iB_y(\mathbf{r}) & -B_z(\mathbf{r}) \end{pmatrix}$$

To construct the wavefunction, we append the spin to the positions



Variational Monte Carlo

Step 1. Specify the **variational wavefunction** Ψ_θ - we use a neural network

Step 2. Define an **energy (loss) estimator** - Markov chain Monte Carlo (MCMC) for us

Step 3. Chose and optimizer to update the parameters (weights and biases): $\theta \rightarrow \theta'$

Energy (loss) evaluation

$$E_\theta = \frac{\langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} = \frac{\sum_{\{\boldsymbol{\xi}\}} \Psi^*_\theta(\{\boldsymbol{\xi}\}) \hat{H}(\{\boldsymbol{\xi}\}) \Psi_\theta(\{\boldsymbol{\xi}\})}{\sum_{\{\boldsymbol{\xi}\}} |\Psi_\theta(\{\boldsymbol{\xi}\})|^2}$$

$$\boldsymbol{\xi} = \{\mathbf{r}, s\}$$

$$E_\theta = \mathbb{E}_{\{\boldsymbol{\xi}\} \sim |\Psi_\theta(\{\boldsymbol{\xi}\})|^2} [E_{\mathrm{loc},\theta}(\{\boldsymbol{\xi}\})]$$

We run Markov chain Monte Carlo $\{\boldsymbol{\xi}\} \rightarrow \{\boldsymbol{\xi}'\}$ to sample $|\Psi_\theta(\{\boldsymbol{\xi}\})|^2$

We track a collection of configurations - the “batch”

Energy (loss) evaluation

$$E_\theta = \frac{\langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} = \frac{\sum_{\{\xi\}} \Psi^*_\theta(\{\boldsymbol{\xi}\}) \hat{H}(\{\boldsymbol{\xi}\}) \Psi_\theta(\{\boldsymbol{\xi}\})}{\sum_{\{\xi\}} |\Psi_\theta(\{\boldsymbol{\xi}\})|^2}$$

$\sim \exp(N)$

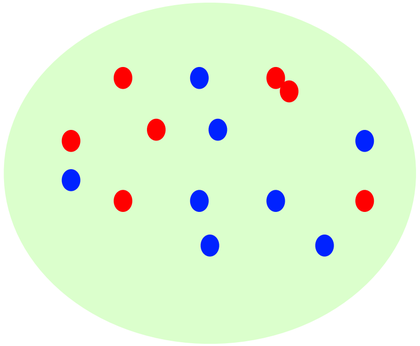
$$E_\theta = \mathbb{E}_{\{\boldsymbol{\xi}\} \sim |\Psi_\theta(\{\boldsymbol{\xi}\})|^2} [E_{\mathrm{loc},\theta}(\{\boldsymbol{\xi}\})]$$

We run Markov chain Monte Carlo $\{\boldsymbol{\xi}\} \rightarrow \{\boldsymbol{\xi}'\}$ to sample $|\Psi_\theta(\{\boldsymbol{\xi}\})|^2$

We track a collection of configurations - the “batch”

Markov chain Monte Carlo

Configuration: $\{\boldsymbol{\xi}\} =$



- spin up

- spin down

Accept update $\{\boldsymbol{\xi}\} \to \{\boldsymbol{\xi}'\}$ with probability $p = \begin{cases} |\psi'/\psi|^2, & |\psi'| < |\psi| \\ 1, & |\psi'| \geq |\psi| \end{cases}$

Position update

$$w(\mathbf{r}_i \to \mathbf{r}_i') = \frac{1}{2\pi\sigma^2} e^{-|\mathbf{r}_i - \mathbf{r}_i'|^2/2\sigma^2}$$

Spin update

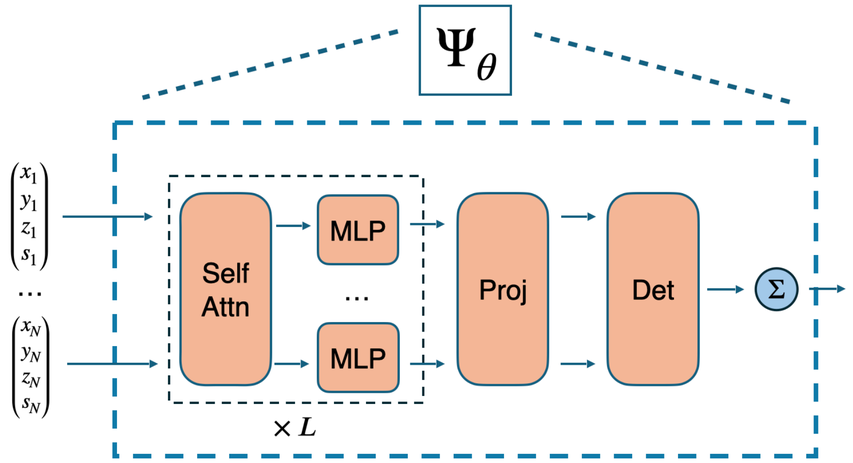
$w(\sigma_i \to -\sigma_i) = p$

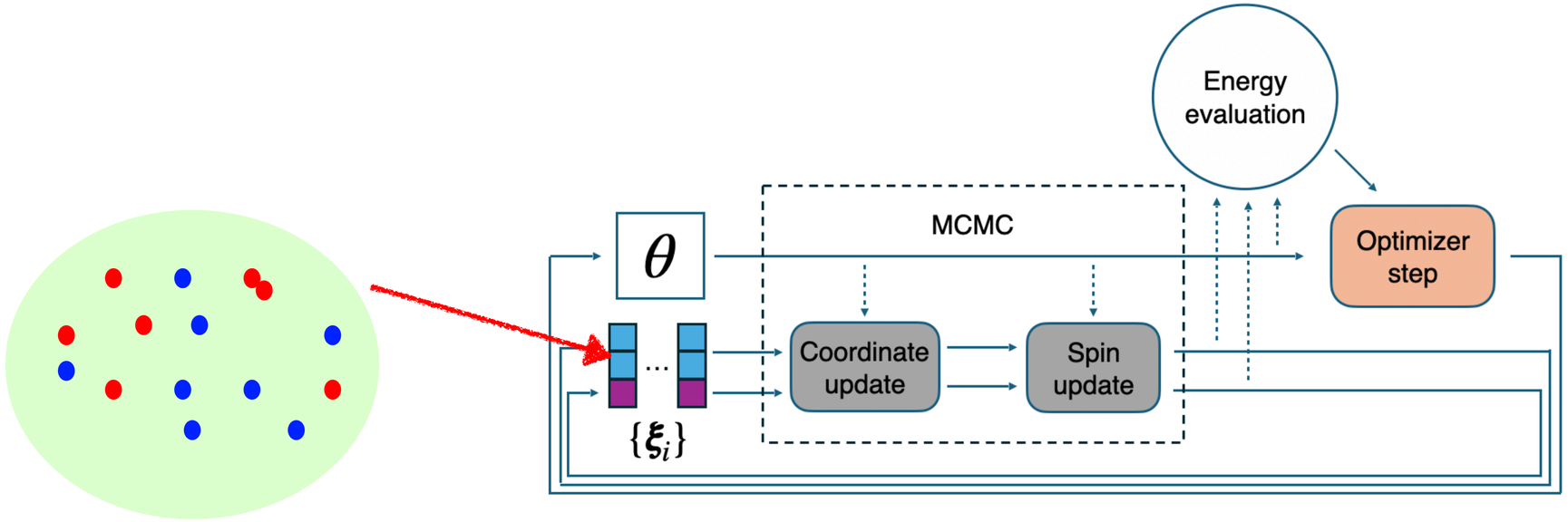
Recap: training loop

Step 1. Specify the **variational wavefunction**

Step 2. Define an **energy (loss) estimator**

Step 3. Chose and optimizer to **update the parameters**

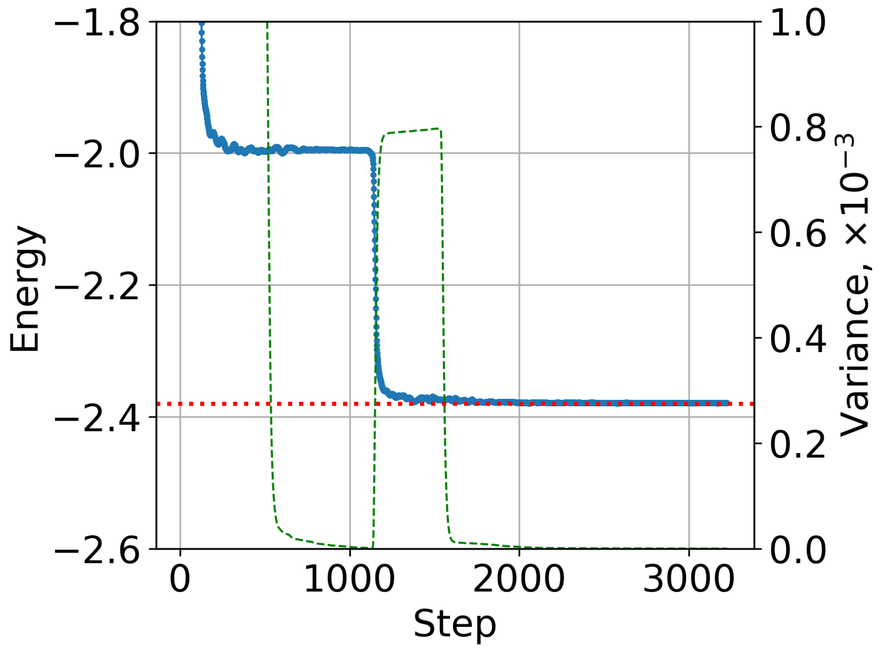




Spinful systems

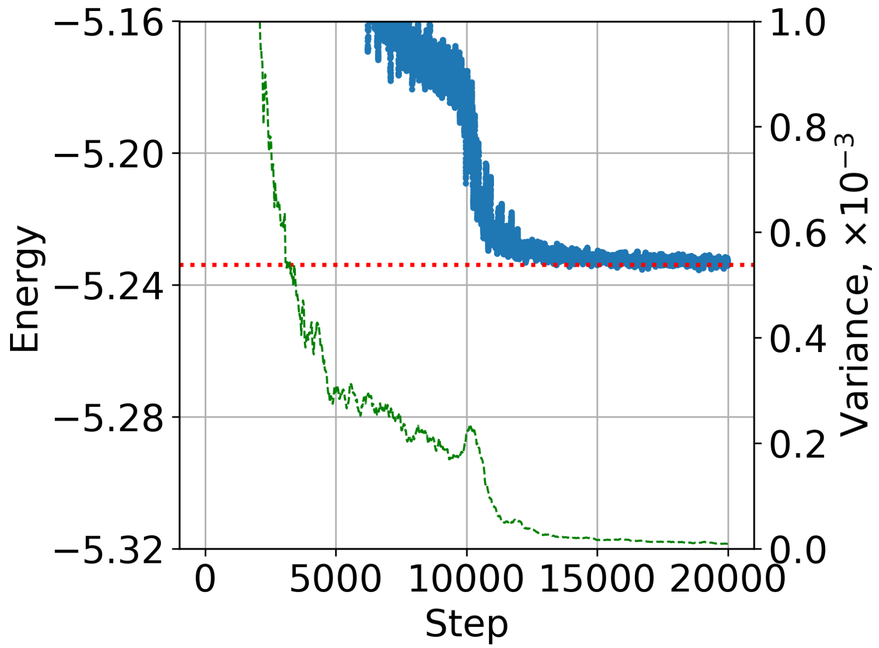
Specially varying Zeeman field

$$H_{\text{spin}} = \sum_i \mathbf{B}(\mathbf{r}_i) \cdot \boldsymbol{\sigma}_i,$$



Rashba spin-orbit coupling

$$H_{\text{SOC}} = p_x\sigma_y - p_y\sigma_x\,.$$



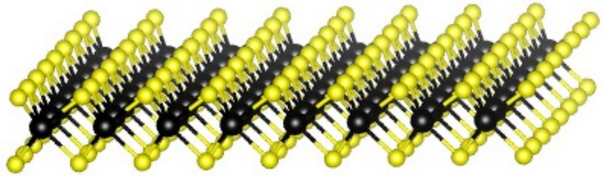
Red - exact energy, blue - training curve, green - energy variance

arXiv:2510.18621

Twisted bilayer TMDs

Effective Hamiltonian with pseudo-spin from layer DoF:

$$H = \sum_i H_0^{(i)} + \sum_{i<j} V_{Coulomb}(\mathbf{r}_i - \mathbf{r}_j)$$



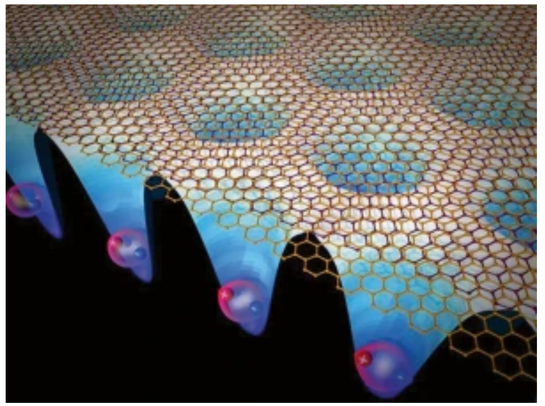
A TMD layer

Single particle Hamiltonian with Skyrmion texture:

$$H_0^{(i)} = \frac{\nabla_i^2}{2m^*} + \mathbf{J}(\mathbf{r}_i) \cdot \boldsymbol{\sigma} + V(\mathbf{r}_i)$$

Skyrmion texture from plane waves:

$$\mathbf{J}(\mathbf{r}_i), V(\mathbf{r}_i) \sim \cos(\mathbf{g} \cdot \mathbf{r}_i + \phi)$$



The effective potential

arXiv:2510.18621

Usage of SubMIT

I did a lot of prototyping on GTX 1080 GPUs
They are usually immediately available
I could do up to ~10 particles in O(1) hour

For going beyond ~13 particles, it is nice to use the V100 GPUs
Usually available within the day

arXiv:2510.18621

Summary

NN-VMC is an emergent first principles method for quantum mechanics that can achieve very high accuracy and has N^4 compute cost

From the ML perspective: the data is self-generated, ...

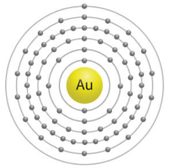
Unlike other scalable methods, it has not shown limitations in what physics/phases it can represent. All of practical quantum problems seem to be solvable by scaling!

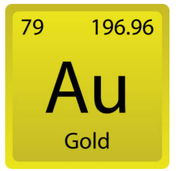
It can now tackle spinful systems (which is most of them)

We are currently using it to solve problem in condensed matter for which there was no scalable method available: superconductivity, topological phases

Back-up

Spinful systems



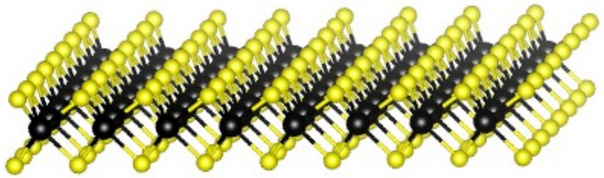


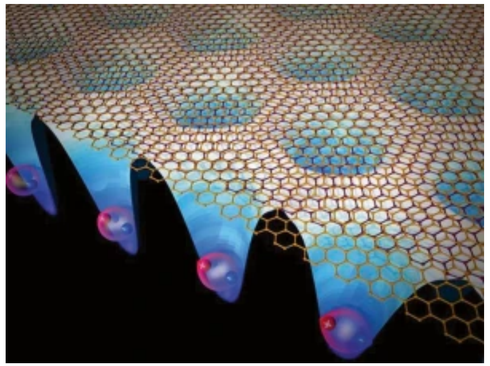
- Relativistic corrections in quantum chemistry, relevant for atoms with Z > 50-40 (it makes gold yellow and mercury liquid). Spin-orbit coupling is one of them

- Magnetism + effects of the magnetic field

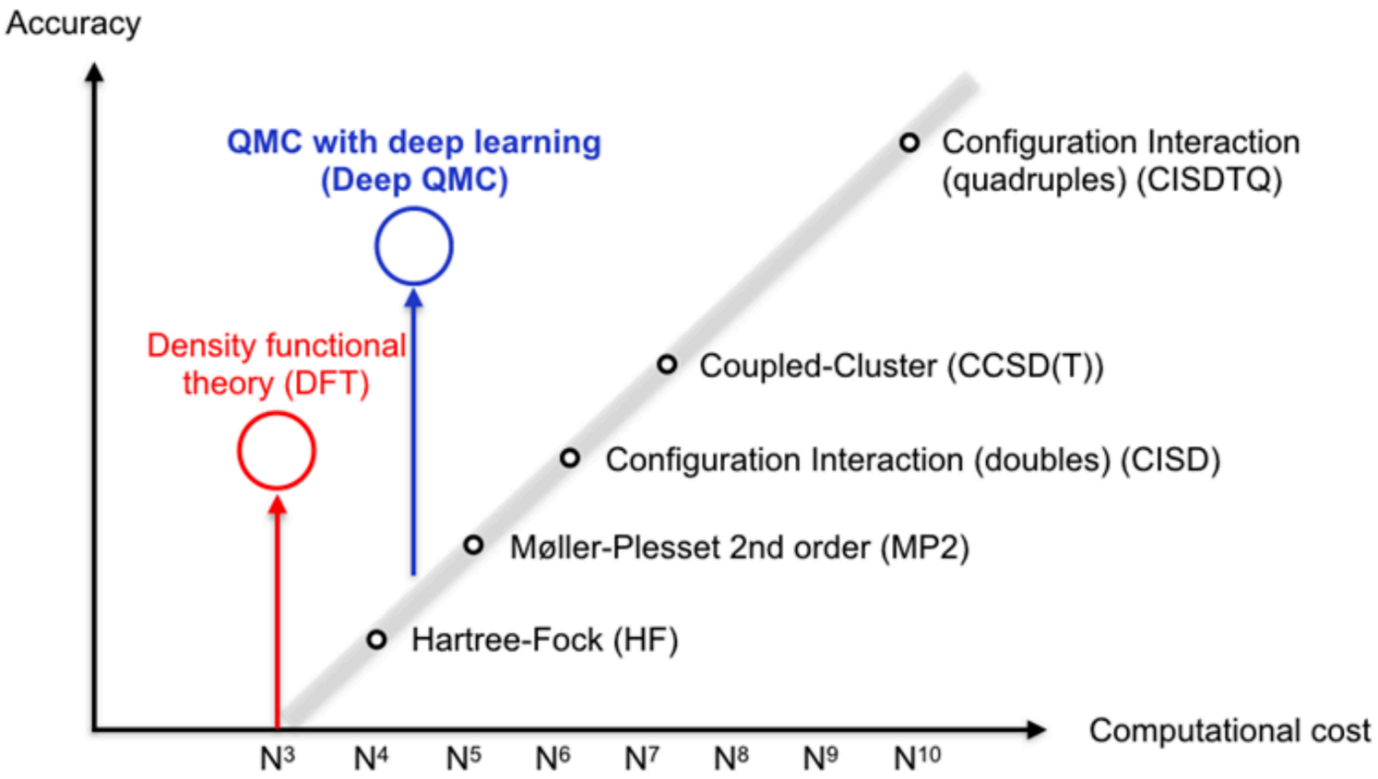
- Superconductivty

- Layered 2D materials: twisted transition metal dichalcogenide bilayers





Existing methods



Credit: Frank Noe

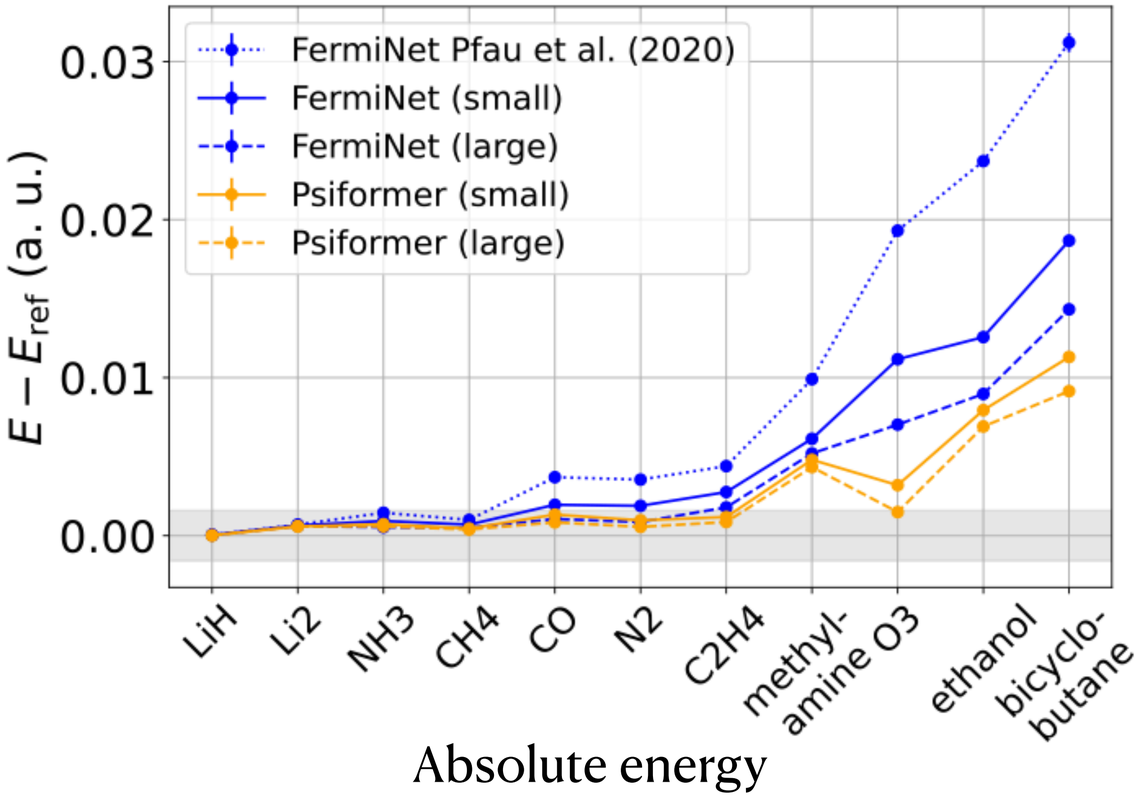
Currently, DFT is the workhorse of computational quantum chemistry.
It delivers reasonable accuracy for many applications with N^3 scaling.

However, it fails for a range of important problems such as bond breaking, transitional metal compounds, superconductors and magnets.

A new scalable algorithm is needed there!

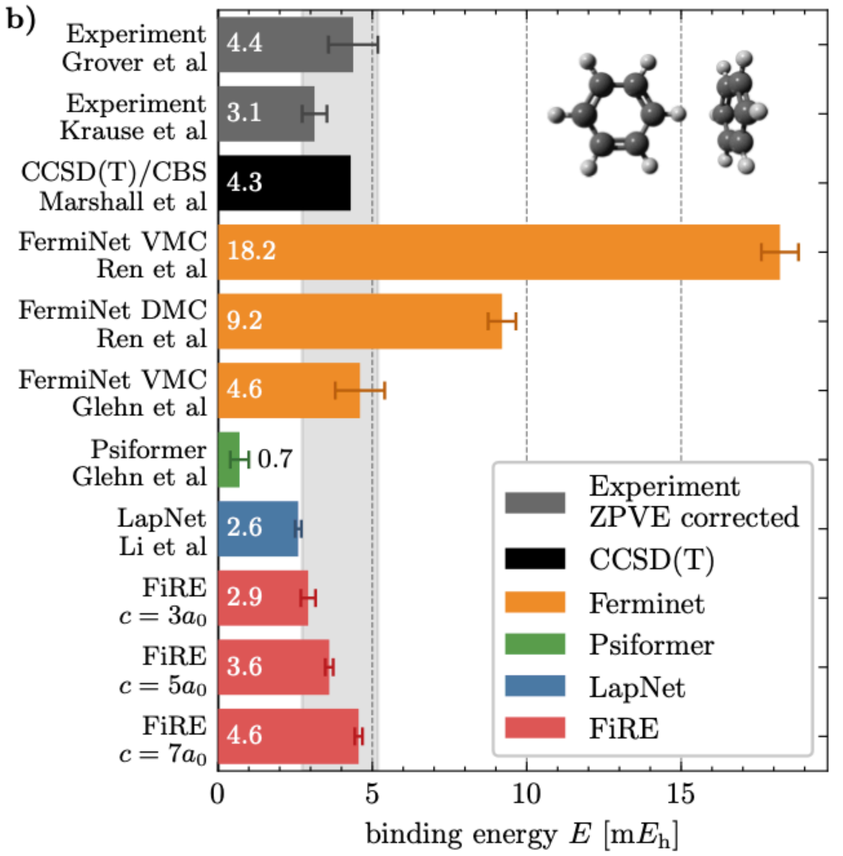
Chemistry results

Chemical accuracy for molecules



Absolute energy

von Glehn, et al, arXiv:2211.13672



Scherbela et al, arXiv:2504.06087