Freie Universität Berlin

Neural-network wave functions for quantum chemistry

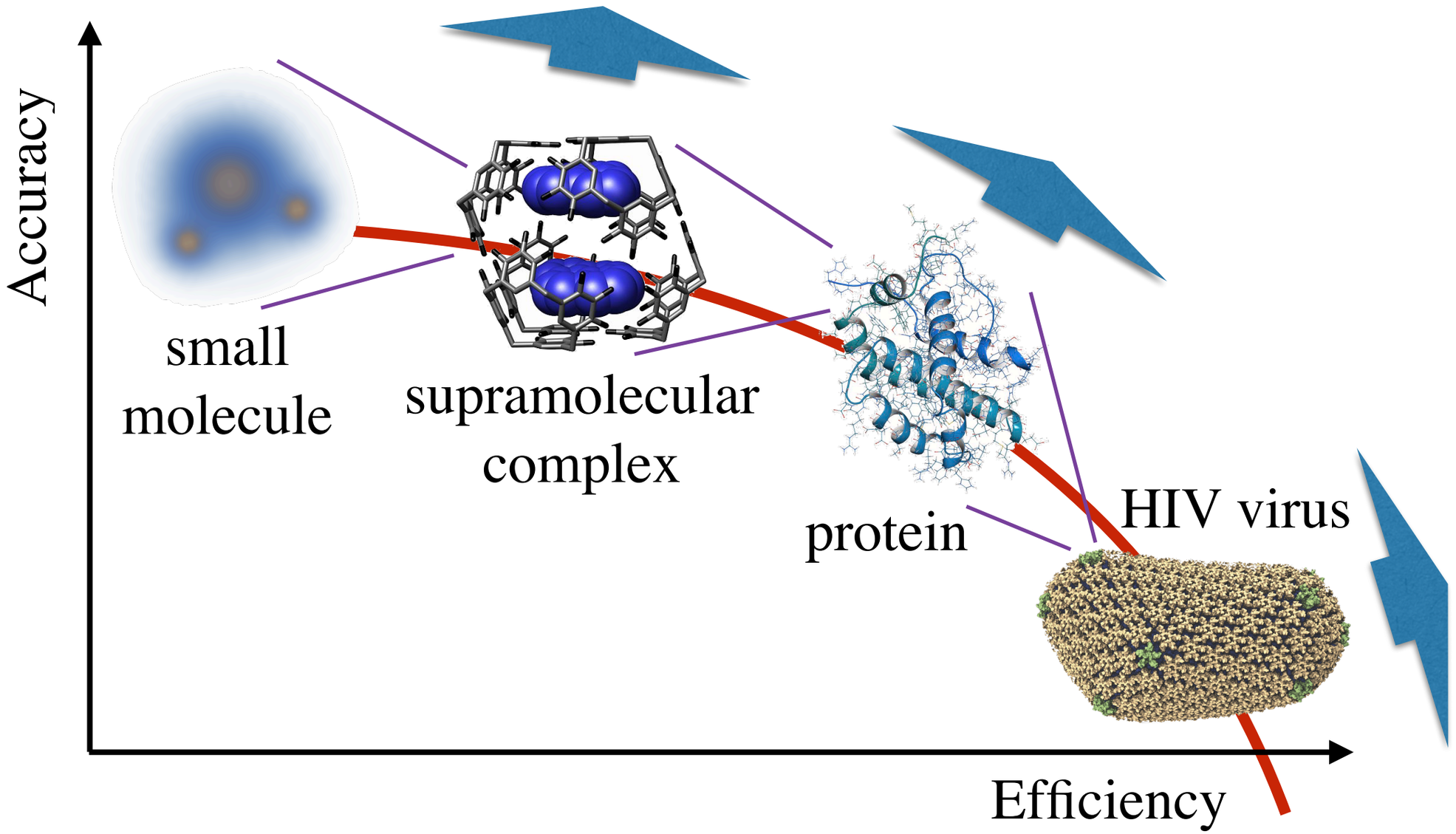
Jan Hermann

FU Berlin, Department of Mathematics, Berlin, Germany
https://jan.hermann.name

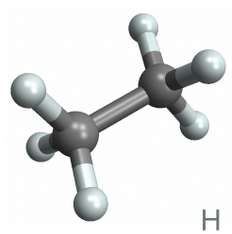
QMMWS4, IPAM, 25 May 2022

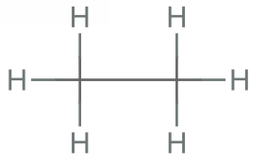
Computational chemistry: Accuracy vs. efficiency

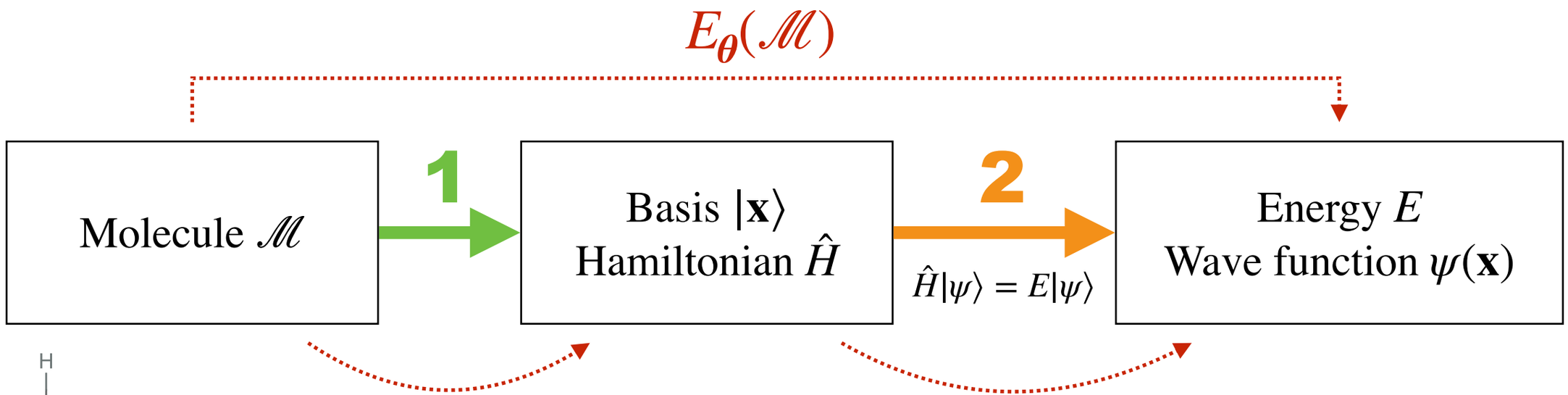
- Electronic Schrödinger equation – exact in principle, impossible to solve exactly in practice



Electronic structure problem







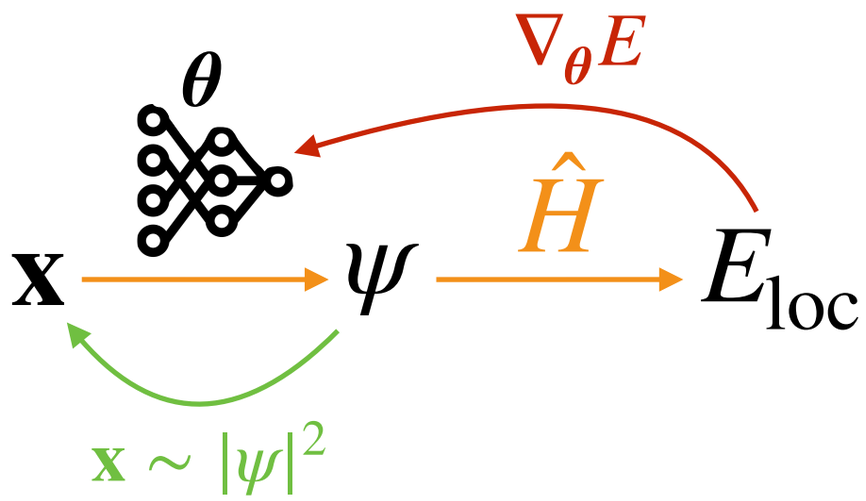
- Ab-initio: exact 1, approximate 2 (*coupled clusters*, *quantum Monte Carlo*)
- Semi-empirical: approximate 1, exact 2 (*density functional theory*, *tight-binding*, *Hubbard model*)
- Machine learning (force fields, density functionals, wave function ansatzes)

Variational quantum Monte Carlo

Schrödinger equation	$\hat{H}\,\lvert\psi\rangle = E\,\lvert\psi\rangle$
Basis states, wave function	$\lvert\psi\rangle = \sum_{\mathbf{x}} \lvert\mathbf{x}\rangle\langle\mathbf{x}\rvert\psi\rangle = \sum_{\mathbf{x}} \psi(\mathbf{x})\,\lvert\psi\rangle$
Monte Carlo integration	$E[\psi] = \dfrac{\langle\psi\rvert\hat{H}\lvert\psi\rangle}{\langle\psi\rvert\psi\rangle} = \mathbb{E}_{\mathbf{x}\sim\lvert\psi(\mathbf{x})\rvert^2}[E_{\mathrm{loc}}(\mathbf{x})]$
Local energy	$E_{\mathrm{loc}}(\mathbf{x}) = \sum_{\mathbf{x}'} \langle\mathbf{x}\rvert\hat{H}\rvert\mathbf{x}'\rangle \dfrac{\psi(\mathbf{x}')}{\psi(\mathbf{x})}$ (zero variance)
Variational principle	$E = \min_{\psi} E[\psi] \le \min_{\boldsymbol{\theta}} E[\psi_{\boldsymbol{\theta}}]$
Gradient descent	$\nabla_{\boldsymbol{\theta}} E[\psi_{\boldsymbol{\theta}}] = \mathbb{E}_{\mathbf{x}}\left[\left(E_{\mathrm{loc}}(\mathbf{x}) - \mathbb{E}_{\mathbf{x}'}[E_{\mathrm{loc}}(\mathbf{x}')]\right)\nabla_{\boldsymbol{\theta}} \ln\lvert\psi_{\boldsymbol{\theta}}(\mathbf{x})\rvert\right]$
Markov-chain sampling	$\mathbf{x}' \sim \exp(-\lVert\mathbf{x}-\mathbf{x}'\rVert/\tau), \quad P(\mathbf{x} := \mathbf{x}') = \lvert\psi(\mathbf{x}')\rvert^2/\lvert\psi(\mathbf{x})\rvert^2$

Deep variational quantum Monte Carlo

```
def fit_wf(
    hamil: Hamiltonian[X],
    wf: Callable[[X], Psi],
    sampler: Iterator[X],
    opt: torch.optim.Optimizer,
    steps: int,
):
    for _ in zip(steps):
        x = next(sampler)
        psi = wf(x)
        E_loc = hamil.local_energy(x, wf)
        loss = (
            (E_loc - E_loc.mean()).detach() * psi.log()
        ).mean()
        loss.backward()
        opt.step()
        opt.zero_grad()
```



First and second quantization

- Quantum states of many-particle systems must be (anti)symmetric for bosons (fermions) with respect to exchange of particles
- Two ways to treat the (anti)symmetry

$N = 3,\ K = 5$

First quantization

$$|\psi\rangle = \sum_{\mathbf{x}_1 \cdots \mathbf{x}_N} \psi(\mathbf{x}_1, \ldots, \mathbf{x}_N)\, |\mathbf{x}_1 \cdots \mathbf{x}_N\rangle$$

coordinates $\mathbf{x}_i$

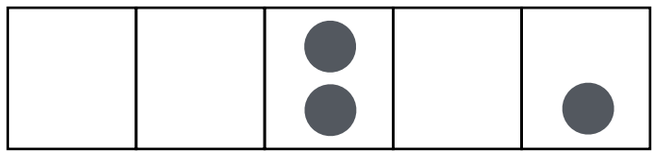
$\mathbf{x} = (2.7, 3.1, 4.9)$

Second quantization (one-particle discrete basis: $\mathbf{x} \to k$)

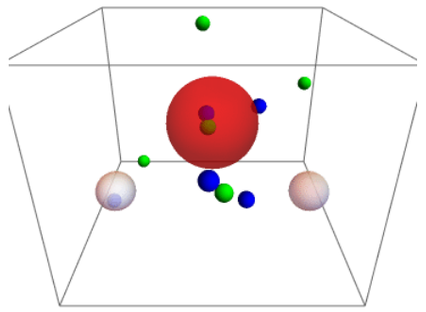
$$|\psi\rangle = \sum_{n_1 \cdots n_K} \psi(n_1, \ldots, n_K)\, |n_1 \cdots n_K\rangle$$

occupation numbers n_i

$\mathbf{n} = (0, 0, 0, 2, 0, 1)$



Electronic Schrödinger equation
(first quantization)

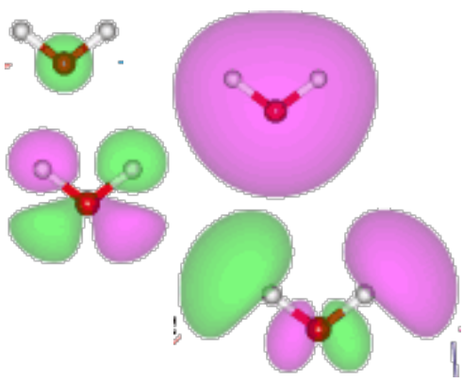


- Electron coordinates, $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$, $3N$-dimensional space
- (Asymmetric) basis states are $|\mathbf{r}\rangle$, antisymmetric wave function $\psi(\mathbf{r})$

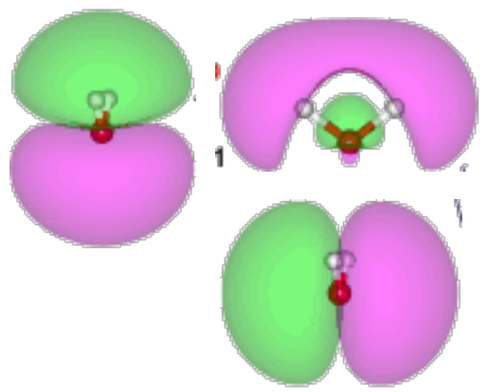
$$\hat{H} := \underbrace{-\frac{1}{2}\Delta}_{\text{kinetic}} + \underbrace{\frac{1}{2}\sum_{ij}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\text{electron–electron}} - \underbrace{\sum_{iI}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}}_{\text{electron–nucleus}}$$

- Molecule specified by atom charges Z_I and coordinates $\mathbf{R}_I$
- Local energy needs evaluation of Laplacian

$$E_{\text{loc}}(\mathbf{r}) = -\frac{1}{2}\frac{\Delta\psi(\mathbf{r})}{\psi(\mathbf{r})} + \frac{1}{2}\sum_{ij}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{iI}\frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}$$



Electronic Schrödinger equation
(second quantization)



- Discrete basis: orthogonal orbitals $\varphi_k(\mathbf{r})$, $k = 1,\ldots, K$
- Occupation numbers $\mathbf{n}$, $n_k = 0, 1$, $\sum_k n_k = N$, $\mathbf{n} \equiv \{k_1, \ldots, k_N\}$
- Antisymmetric basis states are *Slater determinants*, $|\mathbf{n}\rangle = \det \varphi_{k_j}(\mathbf{r}_i)$, (asymmetric) wave function $\psi(\mathbf{n}) \equiv c_{\mathbf{n}}$
- Local energy as a sum over all Slater determinants that differ in occupation of up to two orbitals

$$E_{\mathrm{loc}}(\mathbf{n}) = \sum_{\mathbf{n}':k\to l} h_{kl}\frac{\psi(\mathbf{n}')}{\psi(\mathbf{n})} + \sum_{\mathbf{n}'':k,l\to m,n} V_{klmn}\frac{\psi(\mathbf{n}'')}{\psi(\mathbf{n})}$$

$$h_{kl} = \int \mathrm{d}\mathbf{r}\, \varphi_k(\mathbf{r})\left(-\frac{1}{2}\Delta - \sum_I \frac{Z_I}{|\mathbf{r}-\mathbf{R}_I|}\right)\varphi_l(\mathbf{r})$$

$$V_{klmn} = \frac{1}{2}\int \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\frac{\varphi_k(\mathbf{r})\varphi_l(\mathbf{r}')\varphi_m(\mathbf{r})\varphi_n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$$

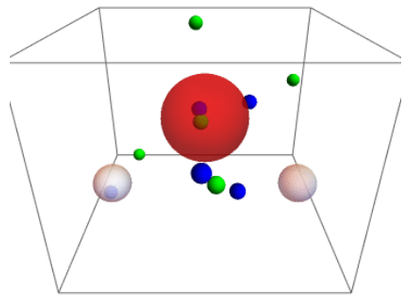
Intermezzo: Electron spin

- Electrons have both spatial and spin coordinates,
  $\mathbf{r}_i \in \mathbb{R}^3, s_i \in \{\uparrow, \downarrow\}$

$$\psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \ldots)$$

- Hamiltonian has no spin, so one consider the spins fixed

$$\psi(\mathbf{r}_1^{\uparrow}, \ldots, \mathbf{r}_{N_{\text{up}}}^{\uparrow}, \mathbf{r}_{N_{\text{up}}+1}^{\downarrow}, \ldots, \mathbf{r}_N^{\downarrow}), \quad s_i = \begin{cases} \uparrow & i \leq N_{\text{up}} \\ \downarrow & i > N_{\text{up}} \end{cases}$$



- Antisymmetry required for same-spin electrons
- Opposite-spin exchange symmetry encodes spin state (singlet, triplet,…)
- Slater determinants $\det \varphi_k(\mathbf{r}_i) = \det \varphi_k^{\uparrow}(\mathbf{r}_i^{\uparrow}) \det \varphi_k^{\downarrow}(\mathbf{r}_i^{\downarrow})$

Pauncz, *Spin eigenfunctions* (Plenum Press, 1979)

Real-space neural-network electronic wave functions

1. Slater determinant

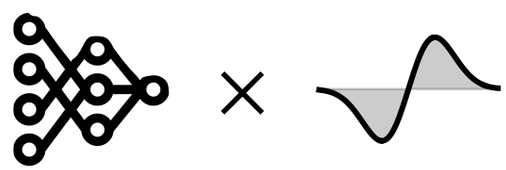
$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) := \det \varphi_{k,\boldsymbol{\theta}}(\mathbf{r}_i), \qquad k = 1,\ldots,N$$

2. Generalized Slater determinant

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) := \det f_{ki,\boldsymbol{\theta}}(\mathbf{r}), \qquad \mathbf{f}_{\boldsymbol{\theta}}(\mathscr{P}_{ij}\mathbf{r}) = \mathscr{P}_{ij}\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{r})$$

3. Real-space baseline/envelope

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) := \det f_{ki,\boldsymbol{\theta}}(\mathbf{r})\phi_k(\mathbf{r}_i)$$



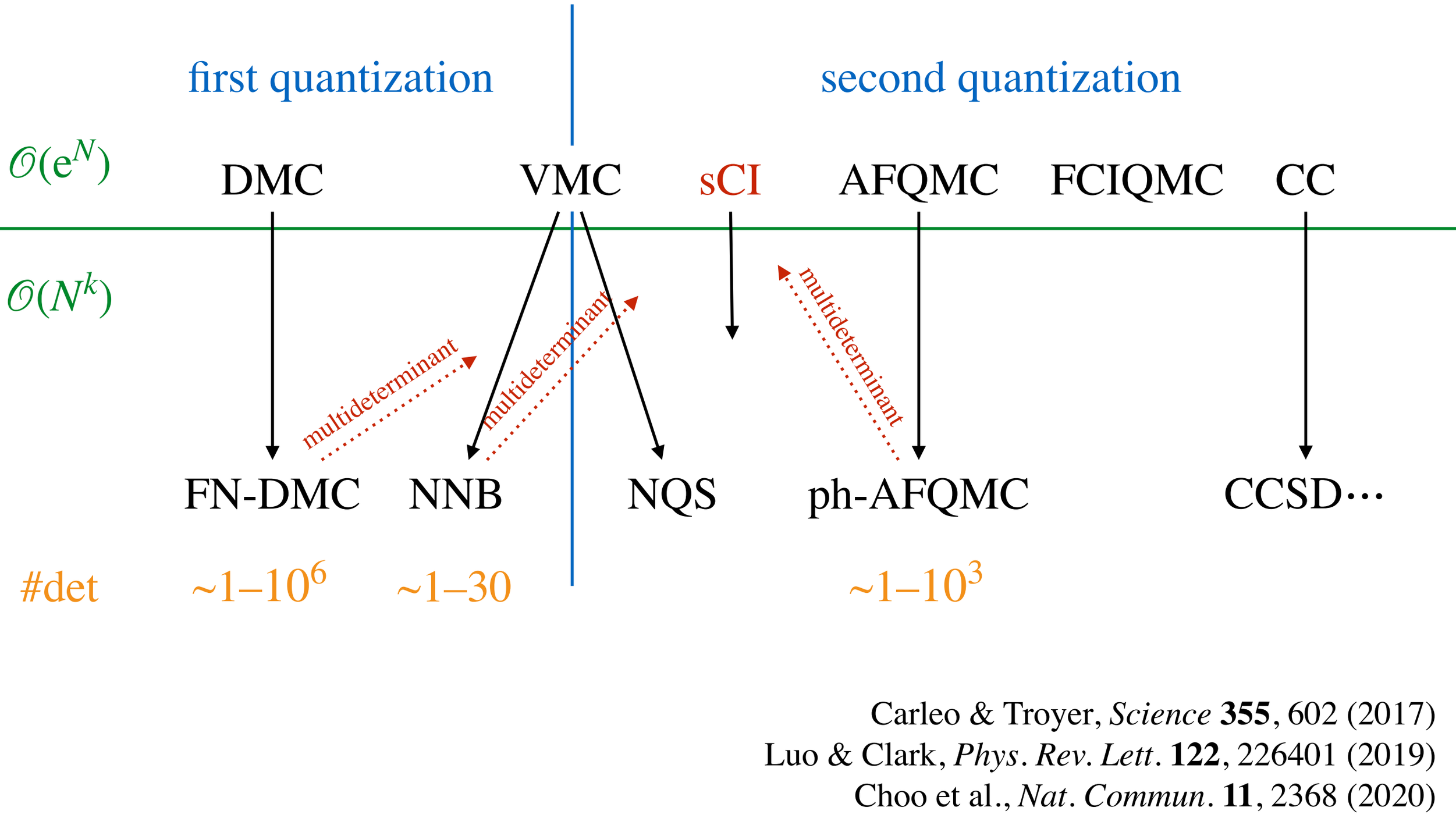
4. Multiple Slater determinants

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) := \sum_p c_p \det f^{(p)}_{ki,\boldsymbol{\theta}}(\mathbf{r})\phi^{(p)}_k(\mathbf{r}_i)$$

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

Han et al., *J. Comp. Phys.* **399**, 108929 (2019)
JH, Schätzle & Noé, *Nat. Chem.* **12**, 891 (2020)

Intermezzo: Determinant expansion curse

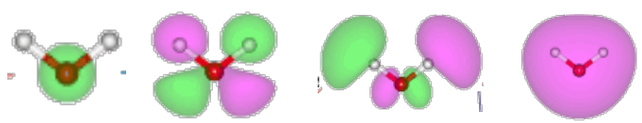


Carleo & Troyer, *Science* **355**, 602 (2017)
Luo & Clark, *Phys. Rev. Lett.* **122**, 226401 (2019)
Choo et al., *Nat. Commun.* **11**, 2368 (2020)

PauliNet: Physics-inspired neural-network ansatz

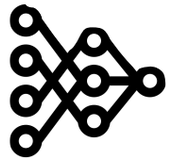
$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) := \sum_p \det f^{(p)}_{ki,\boldsymbol{\theta}}(\mathbf{r})\phi^{(p)}_k(\mathbf{r}_i) := \mathrm{e}^{J_{\boldsymbol{\theta}}(\mathbf{r})+\gamma(\mathbf{r})} \sum_p c_p \det f'_{\kappa_{pk}i,\boldsymbol{\theta}}(\mathbf{r})\varphi_{\kappa_{pk}}(\mathbf{r}_i)$$

Molecular orbitals



$$\phi_k(\mathbf{r}) := \varphi_k(\mathbf{r})$$

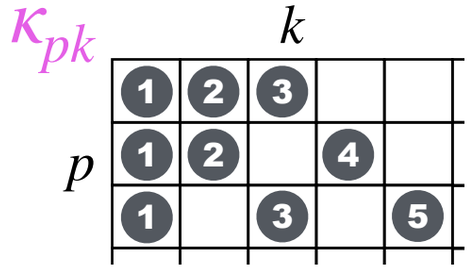
Neural-network Jastrow and generalized backflow



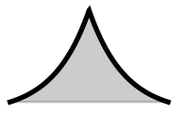
$$f_{ki,\boldsymbol{\theta}}(\mathbf{r}) := \mathrm{e}^{J_{\boldsymbol{\theta}}(\mathbf{r})+\gamma(\mathbf{r})} f'_{ki,\boldsymbol{\theta}}(\mathbf{r})$$

Electron configurations

$$J^{(p)} := J, \quad f'^{(p)}_{ki} := f'_{\kappa_{pk}i}, \quad \phi^{(p)}_k := \phi_{\kappa_{pk}}$$



Cusp corrections



n–e in $\varphi_k(\mathbf{r})$, in via $\gamma(\mathbf{r})$

JH, Schätzle & Noé, *Nat. Chem.* **12**, 891 (2020)
Ceperley, *J. Stat. Phys.* **63**, 1237 (1991)

SchNet for molecules

- Instance of a graph neural network
- Original SchNet is a force field

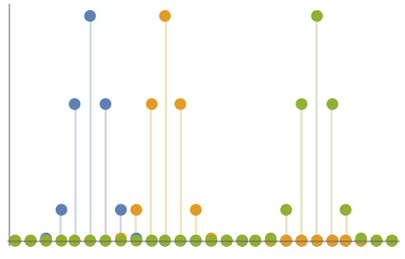
$$E\left(\{Z_i, \mathbf{R}_i\}\right)$$

$$\mathbf{x}_i^{(0)} := \mathbf{X}_{\boldsymbol{\theta}, Z_i}$$

$$\mathbf{z}_i^{(n)} := \sum_j \mathbf{w}_{\boldsymbol{\theta}}^{(n)}\left(\mathbf{e}(|\mathbf{R}_i - \mathbf{R}_j|)\right) \odot \mathbf{h}_{\boldsymbol{\theta}}^{(n)}\left(\mathbf{x}_j^{(n)}\right)$$

$$\mathbf{x}_i^{(n+1)} := \mathbf{x}_i^{(n)} + \mathbf{g}_{\boldsymbol{\theta}}^{(n)}\left(\mathbf{z}_i^{(n)}\right)$$

$$E := \sum_i \varepsilon_{\boldsymbol{\theta}}\left(\mathbf{x}_i^{(N)}\right)$$



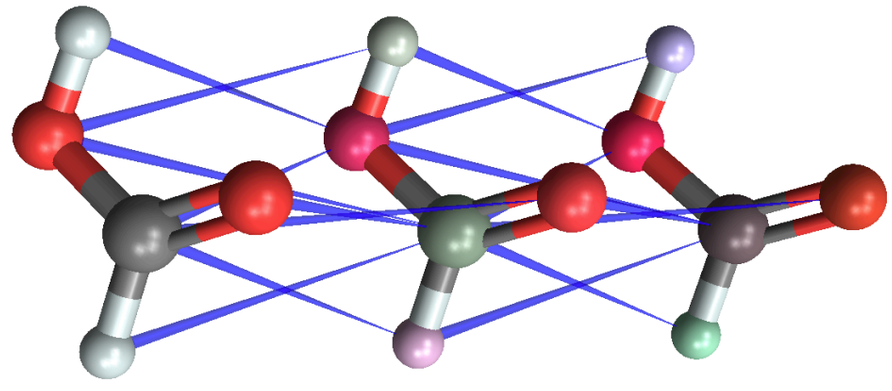
Graph neural networks

initialize $\mathbf{x}_i^{(0)}$

$$\mathbf{z}_i^{(n)} := \sum_j \mathbf{p}_{\boldsymbol{\theta}}^{(n)}(\mathbf{x}_i^{(n)}, \mathbf{x}_j^{(n)}, \mathbf{e}_{ij})$$

$$\mathbf{x}_i^{(n+1)} := \mathbf{q}_{\boldsymbol{\theta}}^{(n)}\left(\mathbf{x}_i^{(n)}, \mathbf{z}_i^{(n)}\right)$$

$$\mathcal{O}(N_{\text{edge}}^2)$$



Schütt et al. *J. Chem. Phys.* **148**, 241722 (2018)

SchNet for electrons in molecules

- Nodes represent electrons
- Separate same-spin, opposite-spin, and nuclear messages
- Single SchNet instance for both the Jastrow factor $J_{\boldsymbol{\theta}}(\mathbf{r})$ and backflow $\mathbf{f}_{\boldsymbol{\theta}}(\mathbf{r})$

$$\mathbf{x}_i^{(0)} := \mathbf{X}_{\boldsymbol{\theta},s_i}$$

$$\mathbf{z}_i^{(n,\pm)} := \sum\nolimits_{j\neq i}^{\pm} \mathbf{w}_{\boldsymbol{\theta}}^{(n,\pm)}\left(\mathbf{e}(|\mathbf{r}_i-\mathbf{r}_j|)\right) \odot \mathbf{h}_{\boldsymbol{\theta}}^{(n)}\left(\mathbf{x}_j^{(n)}\right)$$

$$\mathbf{z}_i^{(n,\mathrm{n})} := \sum\nolimits_{J} \mathbf{w}_{\boldsymbol{\theta}}^{(n,\mathrm{n})}\left(\mathbf{e}(|\mathbf{r}_i-\mathbf{R}_J|)\right) \odot \mathbf{Y}_{\boldsymbol{\theta},J}$$

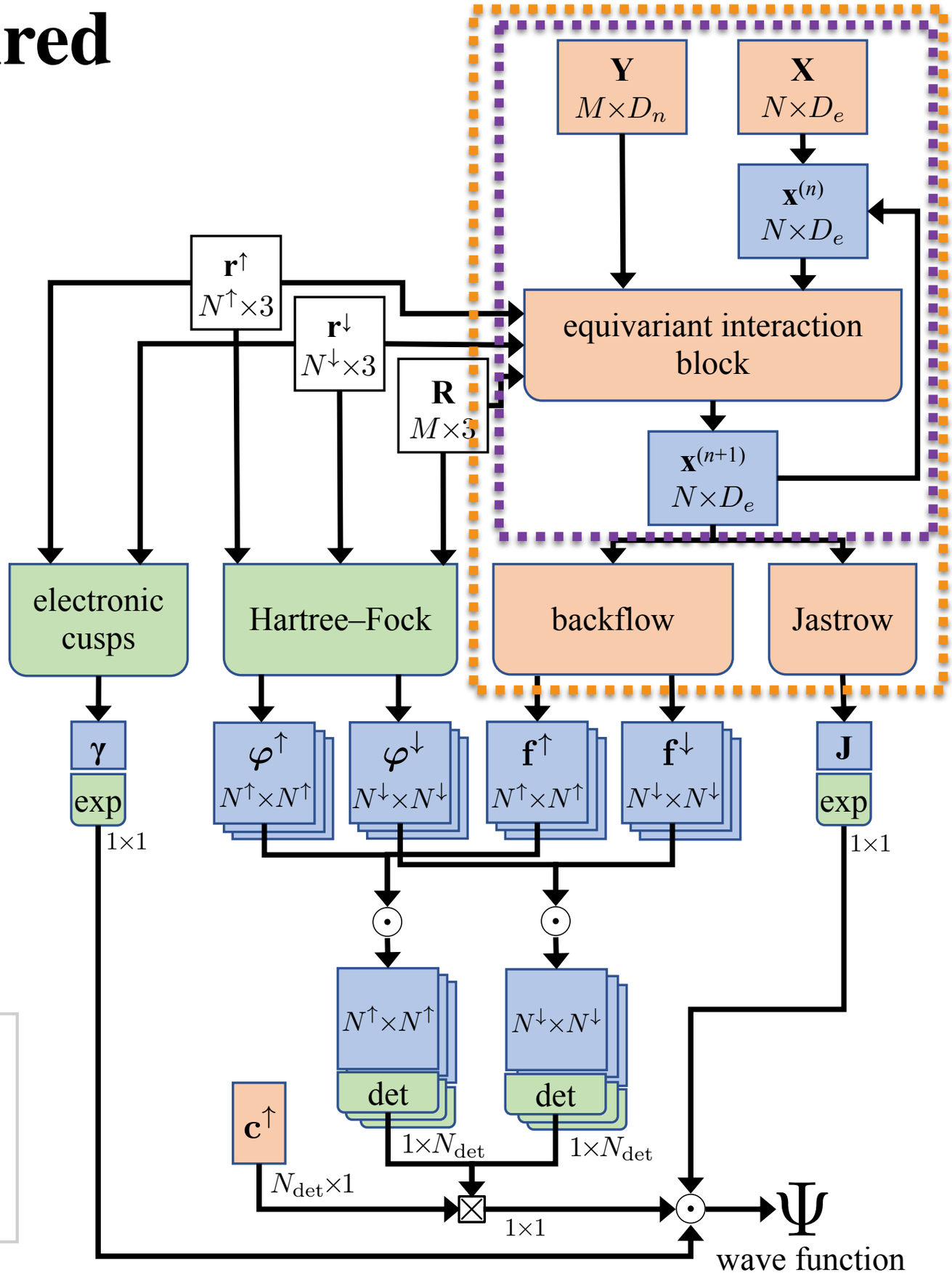
$$\mathbf{x}_i^{(n+1)} := \mathbf{x}_i^{(n)} + \sum\nolimits_{\pm} \mathbf{g}_{\boldsymbol{\theta}}^{(n,\pm)}\left(\mathbf{z}_i^{(n,\pm)}\right) + \mathbf{g}_{\boldsymbol{\theta}}^{(n,\mathrm{n})}\left(\mathbf{z}_i^{(n,\mathrm{n})}\right)$$

$$J := \eta_{\boldsymbol{\theta}}\left(\sum\nolimits_i \mathbf{x}_i^{(N)}\right)$$

$$\mathbf{f}_i := \boldsymbol{\kappa}_{\boldsymbol{\theta}}\left(\mathbf{x}_i^{(N)}\right)$$

Graph neural network

PauliNet: Physics-inspired neural-network ansatz

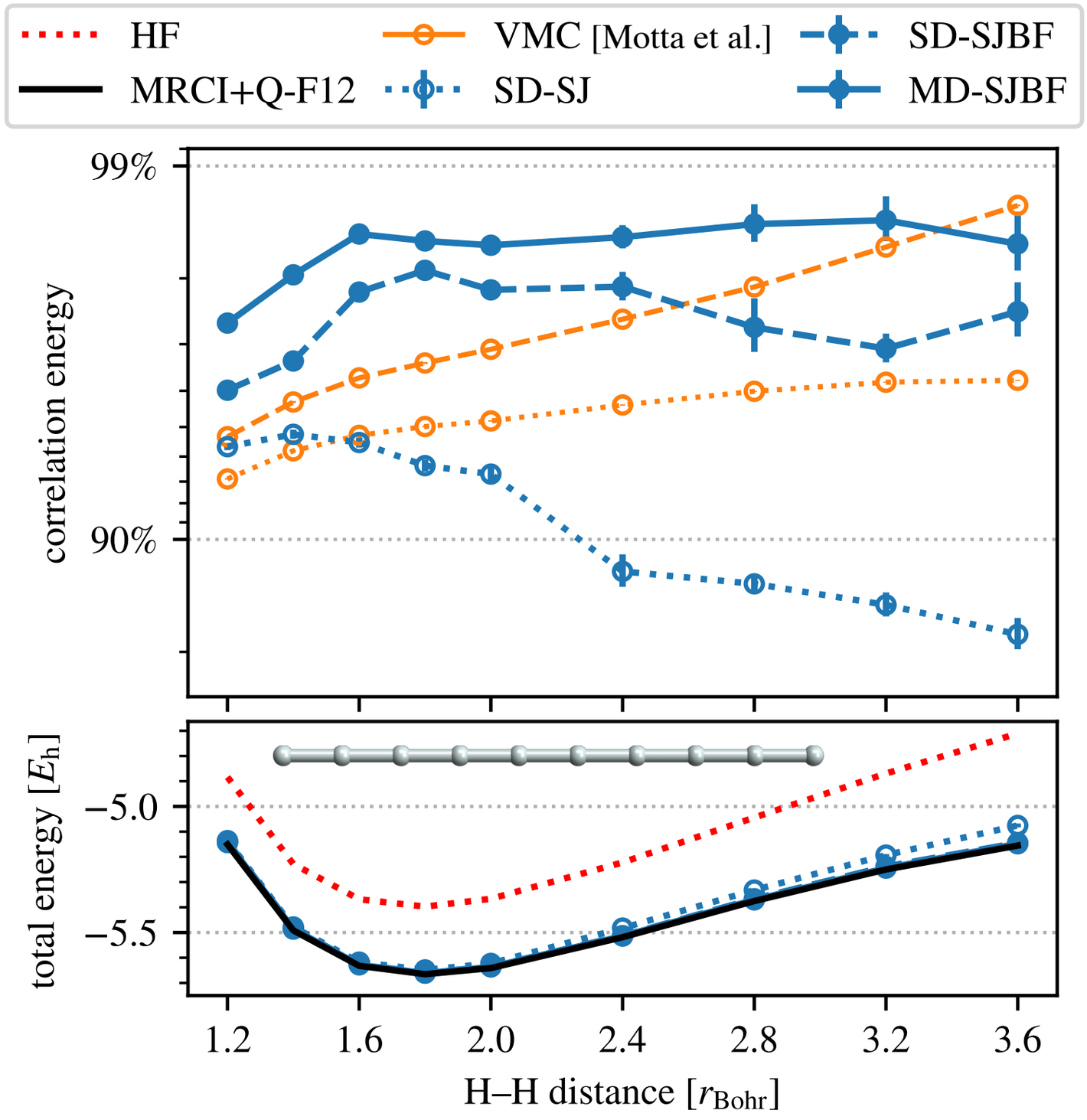


Deep learning

Graph neural network

N number of electrons $N^{\uparrow}$ number of spin-up electrons $N^{\downarrow}$ number of spin-down electrons M number of nuclei D_e embedding dimension, electrons D_n embedding dimension, nuclei	neural network (trainable function)	trainable array	fixed function	input array	hidden array

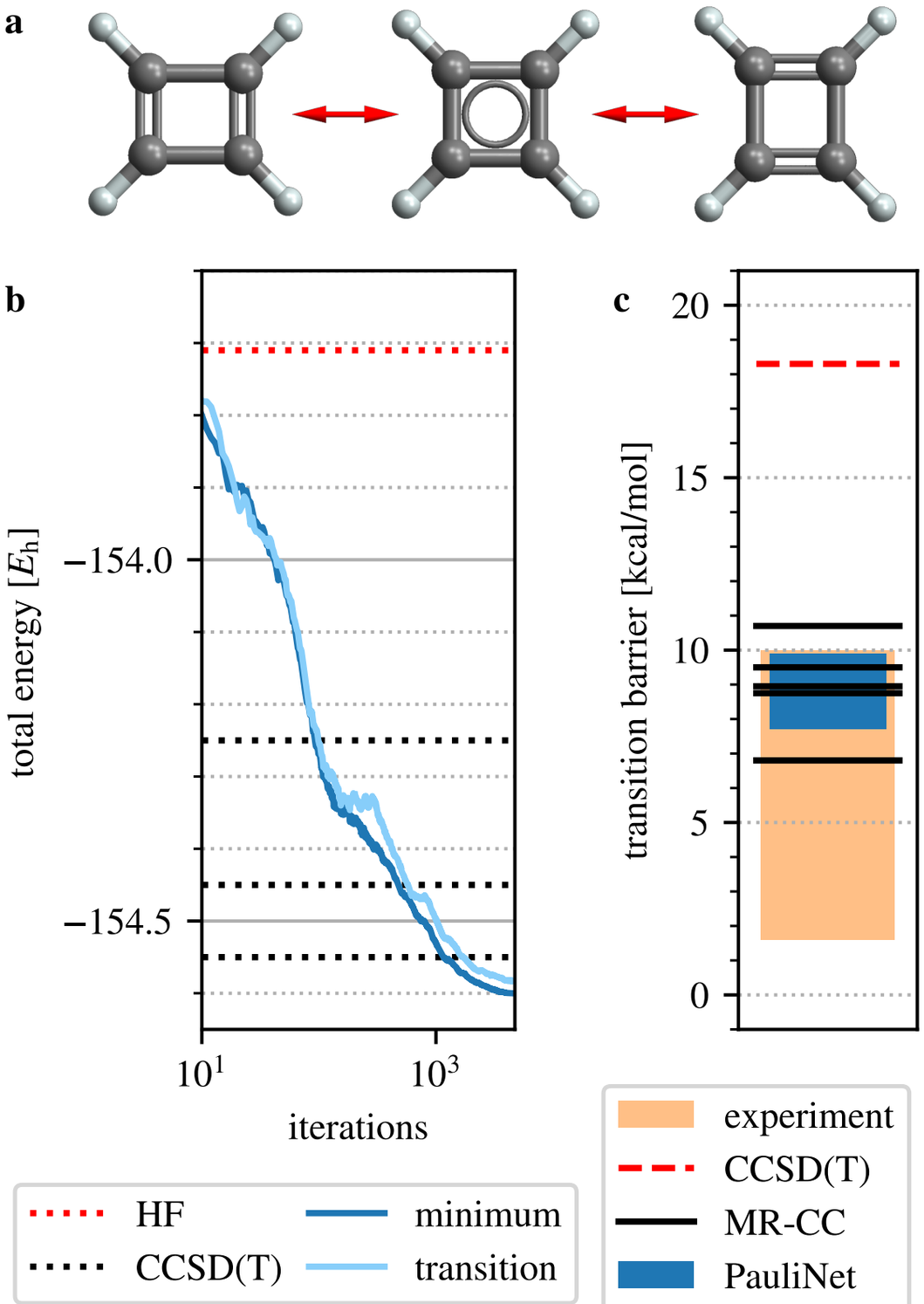
Capturing strong correlation



- Dissociation of linear H_{10} as a test case of strong correlation[1]
- PauliNet with 16 determinants captures 98–99% of the correlation energy along the dissociation

Motta et al., *Phys. Rev. X* **7**, 031059 (2017)

Targeting realistic quantum chemistry



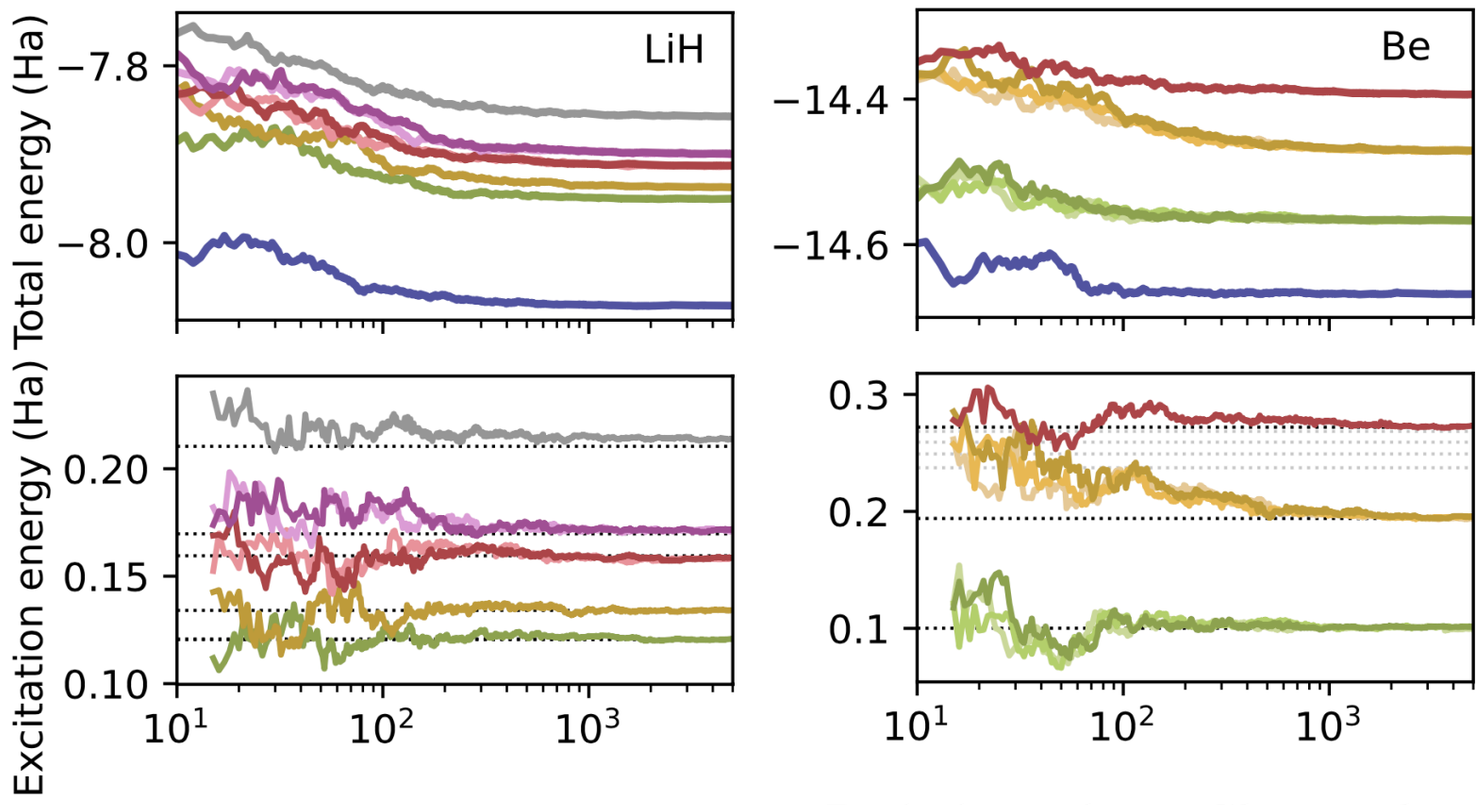
- Barrier of the automerization of cyclobutadiene (28 electrons) ~~still not fully resolved~~
- Strong multireferential character—CCSD(T) ~~overestimates two-fold~~
- Experiment: 1.6–10 kcal/mol, MR-CC 7–11 kcal/mol
- PauliNet with 10 determinants: 8–10 kcal/mol
- 3 days on GTX 1080 Ti GPU

Lyakh et al., *Chem. Rev.* **112**, 182 (2012)
Monino et al., `arXiv:2204.05098` (2022)

Excited states with deep variational QMC

$$\min_{\boldsymbol{\theta}} \left[\sum_i E[\psi_{\boldsymbol{\theta},i}] + \alpha \sum_{j>i} \frac{1}{1 - |\langle \psi_{\boldsymbol{\theta},i} | \psi_{\boldsymbol{\theta},j} \rangle|} \right]$$

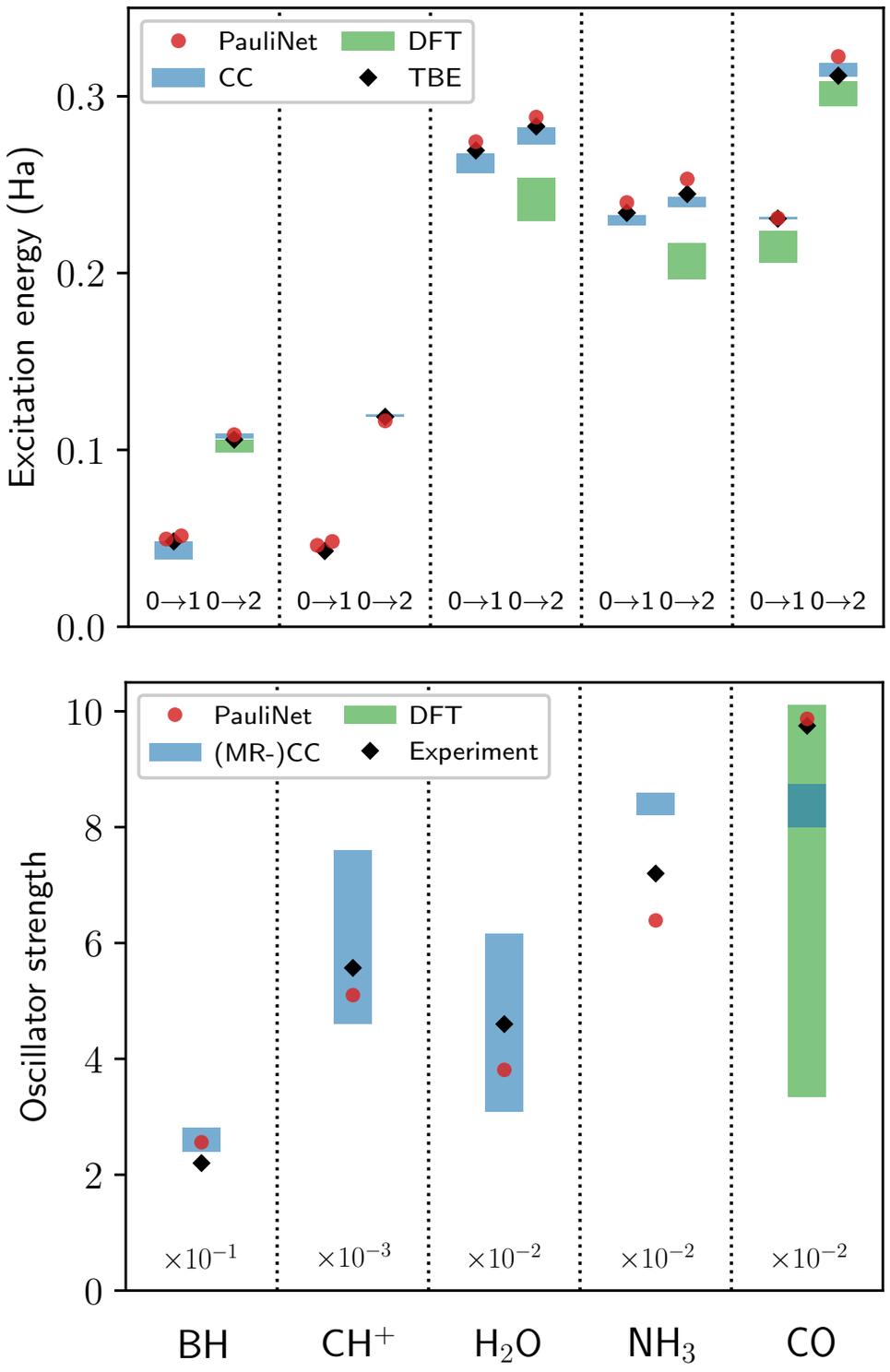
- Extra term in the loss function ensures orthogonality between different eigenstates



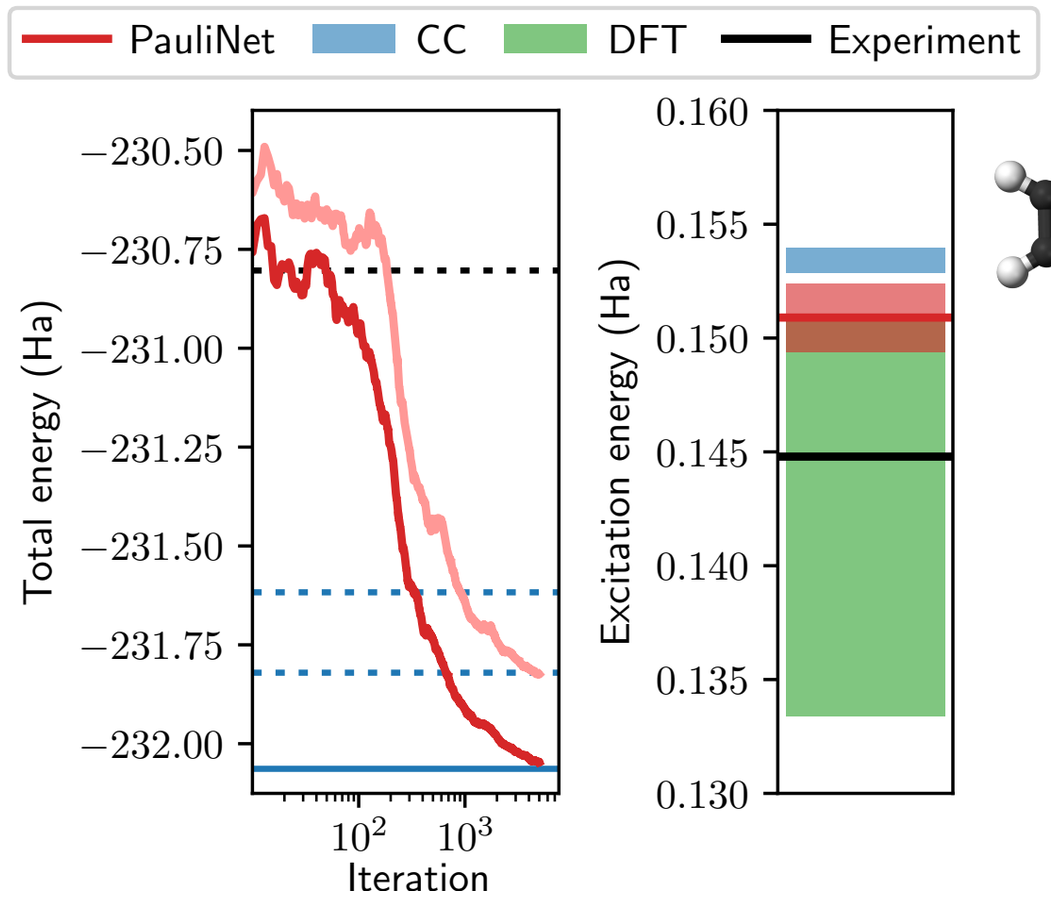
Pathak et al., *J. Chem. Phys.* **154**, 034101 (2021)

Entwistle, Schätzle, Erdman, **JH** & Noé, arXiv:2203.09472 (2022)

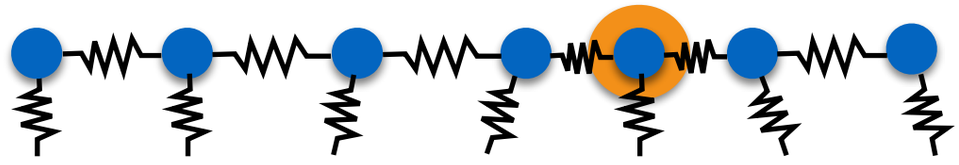
PauliNet for excited states



- Accurate excitation energies and intensities of low-lying states in molecules as large as benzene



Exciton–phonon coupling



- Model for excitations in α-helix
- Excitons in second quantization: $n_i = 0,1\,, i = 1,\ldots, N$
- Phonons in first quantization, $\mathfrak{R} = (R_i, \ldots, R_N)$
- Wave function $\psi(\mathbf{n}, \mathfrak{R})$
- Empirical Hamiltonian

$$\hat{H} := \hat{H}^{\text{ex}} + \hat{H}^{\text{ph}} + \hat{H}^{\text{ex-ph}}$$

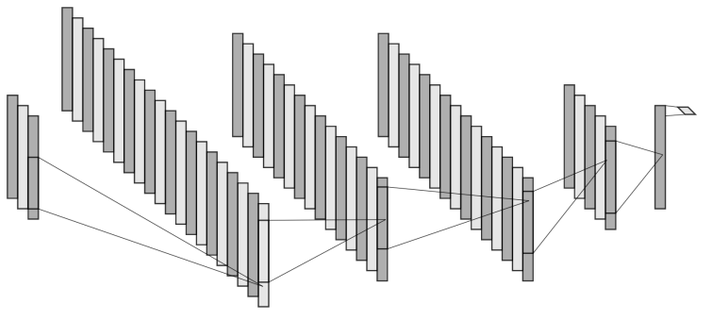
$$\hat{H}^{\text{ex}} = \sum_i \left(\alpha \hat{b}_i^\dagger \hat{b}_i + \beta(\hat{b}_i^\dagger \hat{b}_{i+1} + \hat{b}_i \hat{b}_{i+1}^\dagger)\right)$$

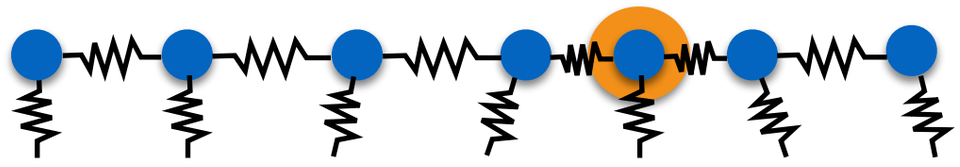
$$\hat{H}^{\text{ph}} = \sum_i \left(\frac{\Delta_i}{2m} + \frac{1}{2} m\nu^2 R_i^2 + \frac{1}{2}\mu\omega^2 (R_i - R_{i+1})^2\right)$$

$$\hat{H}^{\text{ex-ph}} = \sum_i \sigma \hat{b}_i^\dagger \hat{b}_i (R_{i+1} - R_{i-1})$$

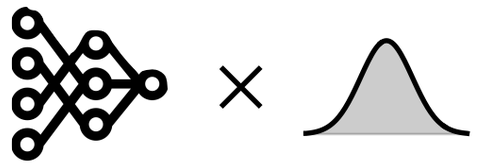
- Local energy: Laplacian of ψ w.r.t. $\mathfrak{R}$, ψ for nearest-neighbor hops

Exciton–phonon neural-network ansatz





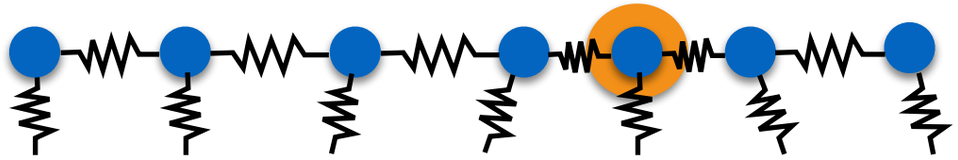
$$\psi_{\boldsymbol{\theta}}(\mathbf{n}, \mathfrak{R}) := \mathrm{ConvNet}\left(\left(n_i, \frac{R_i}{w}, \frac{i}{N} \right)_{i=1}^{N} \right) \mathrm{e}^{-\sum_i \left(\frac{R_i}{w_{\boldsymbol{\theta}}} \right)^2}$$



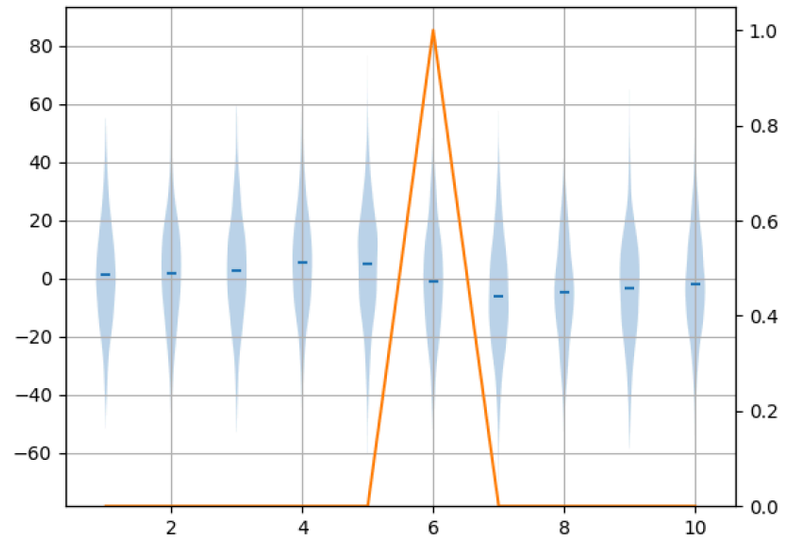
- Product of a neural network and a real-space envelope
- Ordinary 1D convolutional neural network

Davydov solitons

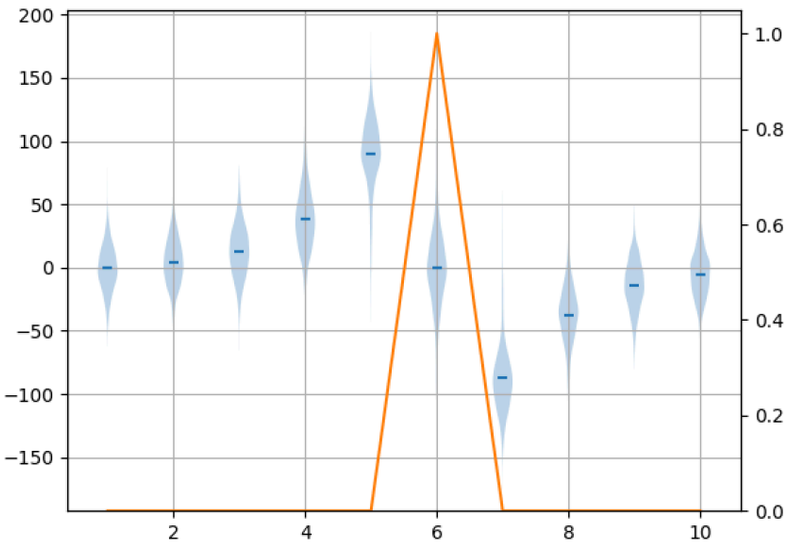
- Self-trapped excitons
- Singly excited Hilbert subspace

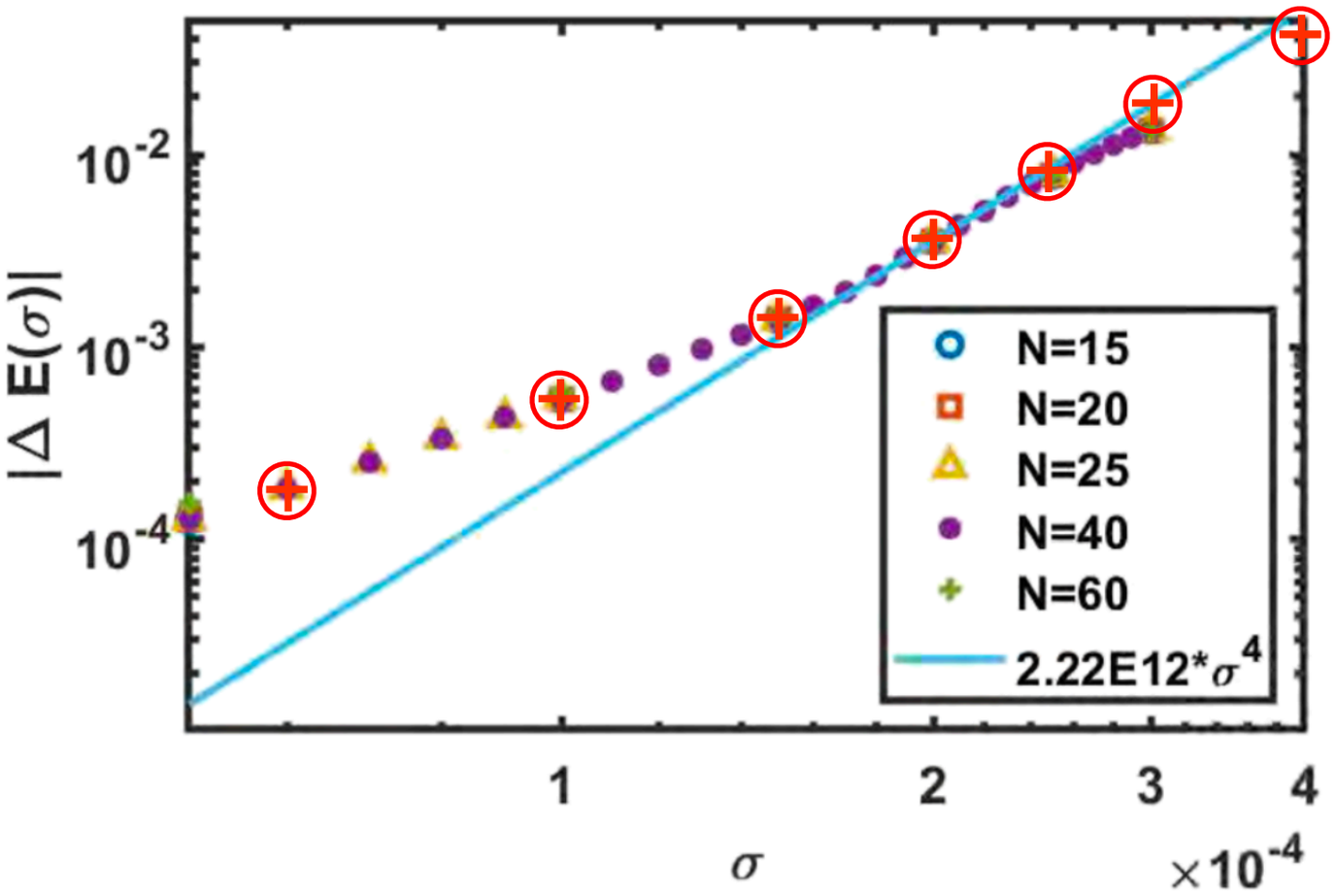


$\sigma = 10^{-4}$



$\sigma = 3 \times 10^{-4}$





Davydov, *J. Theor. Biol.* **38**, 559 (1973)
Geiß et al., *J. Chem. Phys.* **156**, 024109 (2022)

Summary & outlook

- Deep variational QMC is a powerful framework to solve both ab-initio and model Hamiltonians in chemistry with high accuracy
- Active field: periodic systems, transferable wave functions, excited states, pseudopotentials, diffusion QMC, better architectures
- Expensive – scaling up? Benchmarks? Failure modes? Size consistency? Optimization vs. expressivity?
- Aspiration: Accurate black-box method for systems with 100+ electrons and complex electronic structure