

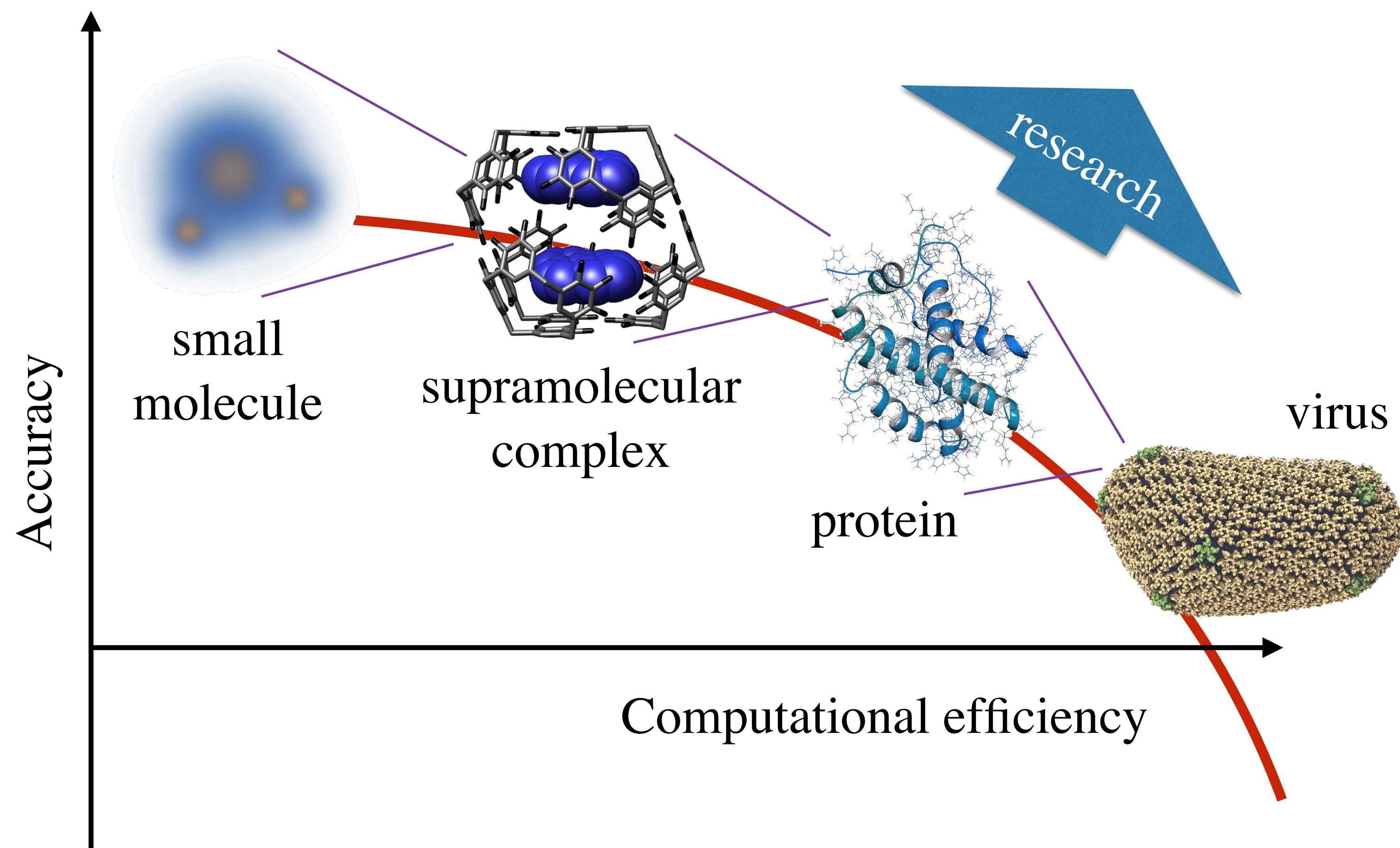
Solving the electronic Schrödinger equation with deep learning

Jan Hermann

FU Berlin, Department of Mathematics, Berlin, Germany

ACS Fall Meeting, 26 August 2021

Computational chemistry: Story of accuracy vs. efficiency



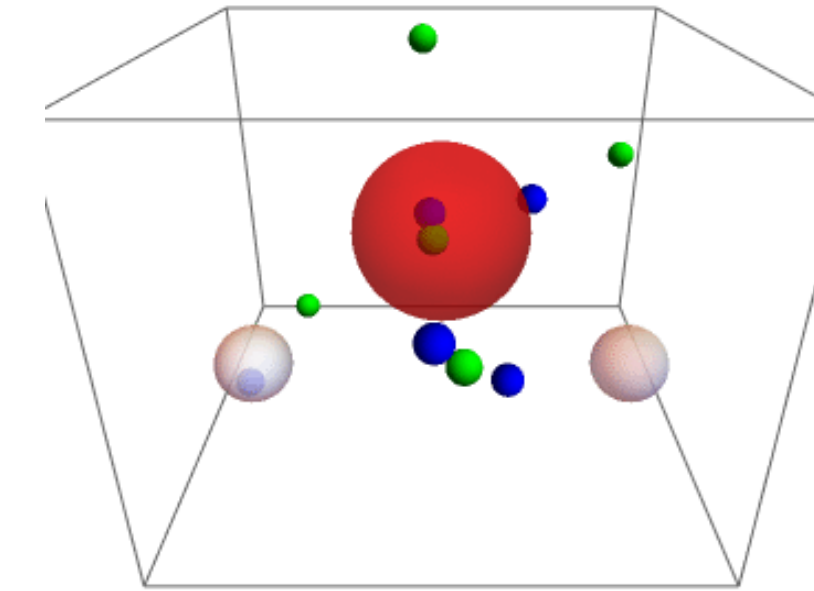
- The electronic Schrödinger equation holds the keys to predicting molecular properties, but impossible to solve exactly in practice
- Plethora of ab-initio methods and empirical models that trade accuracy for efficiency

Electronic Schrödinger equation

$$\hat{H}\psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = E\psi(\mathbf{r}_1, \dots, \mathbf{r}_N)$$
$$\hat{H} := \sum_i \left(-\frac{1}{2} \nabla_{\mathbf{r}_i}^2 - \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \right) + \sum_{i < j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

- Differential equation for eigenstates ψ_n and eigenvalues (energies) E_n , including the ground (lowest-energy) state ψ_0, E_0
- Electron coordinates: $(\mathbf{r}_1, \dots, \mathbf{r}_N) \equiv \mathbf{r}$
- Hamiltonian operator \hat{H} specified by charges Z_I and coordinates \mathbf{R}_I of M atomic nuclei
- Analytically solvable only for $M = N = 1$, but many approximate numerical methods in quantum chemistry

Electronic wave functions and antisymmetry



- Hamiltonian has no spin – electrons have spin coordinates, but those can be considered fixed

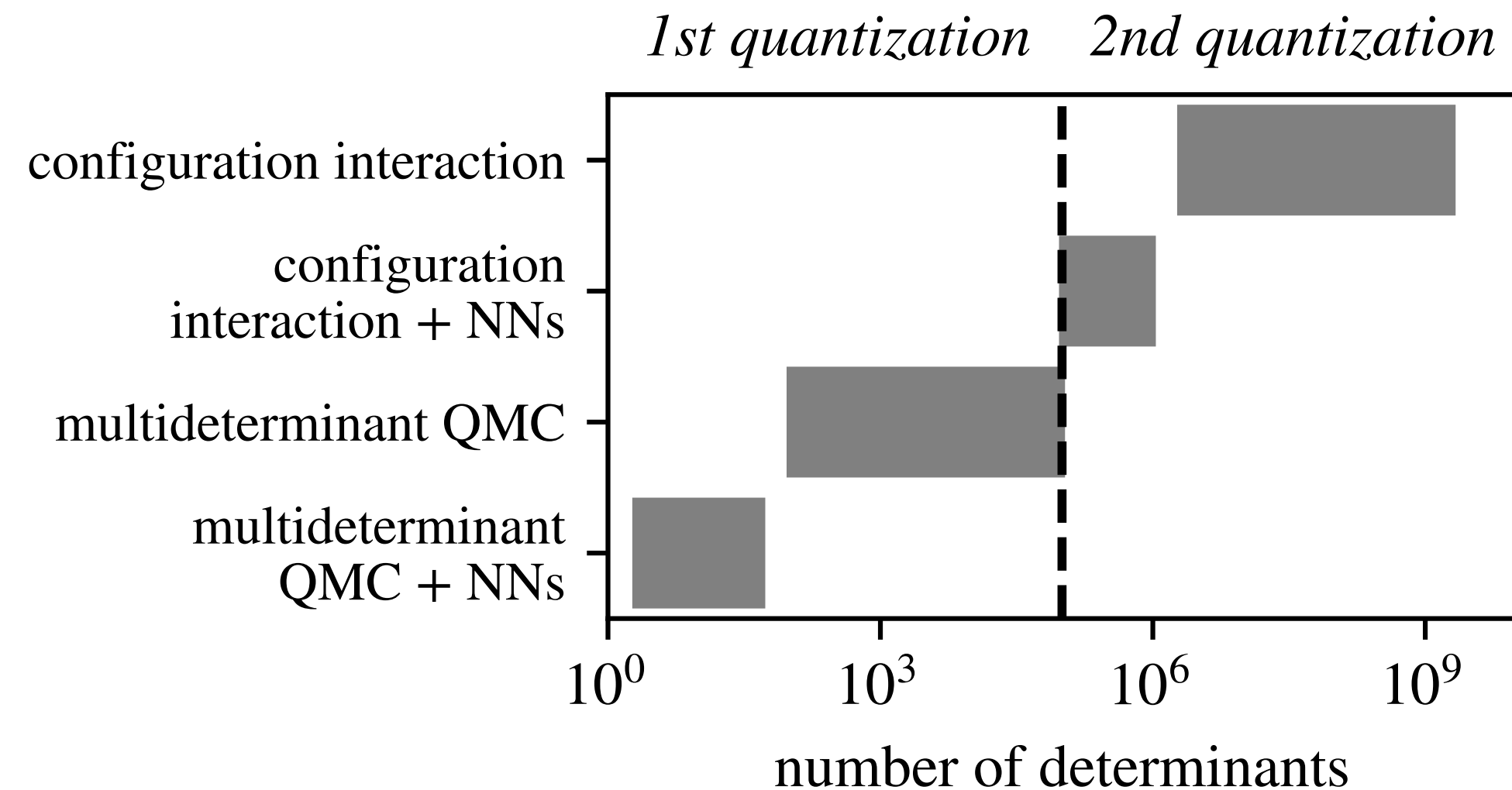
$$\psi(\mathbf{r}_1^\uparrow, \dots, \mathbf{r}_{N_{\text{up}}}^\uparrow, \mathbf{r}_{N_{\text{up}}+1}^\downarrow, \dots, \mathbf{r}_N^\downarrow), \quad s_i = \begin{cases} \uparrow & i \leq N_{\text{up}} \\ \downarrow & i > N_{\text{up}} \end{cases}$$

- Electrons are fermions – wave function must be *antisymmetric* w.r.t. exchange of same-spin electrons ($s_i = s_j$)

$$\psi(\dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots) = -\psi(\dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots)$$

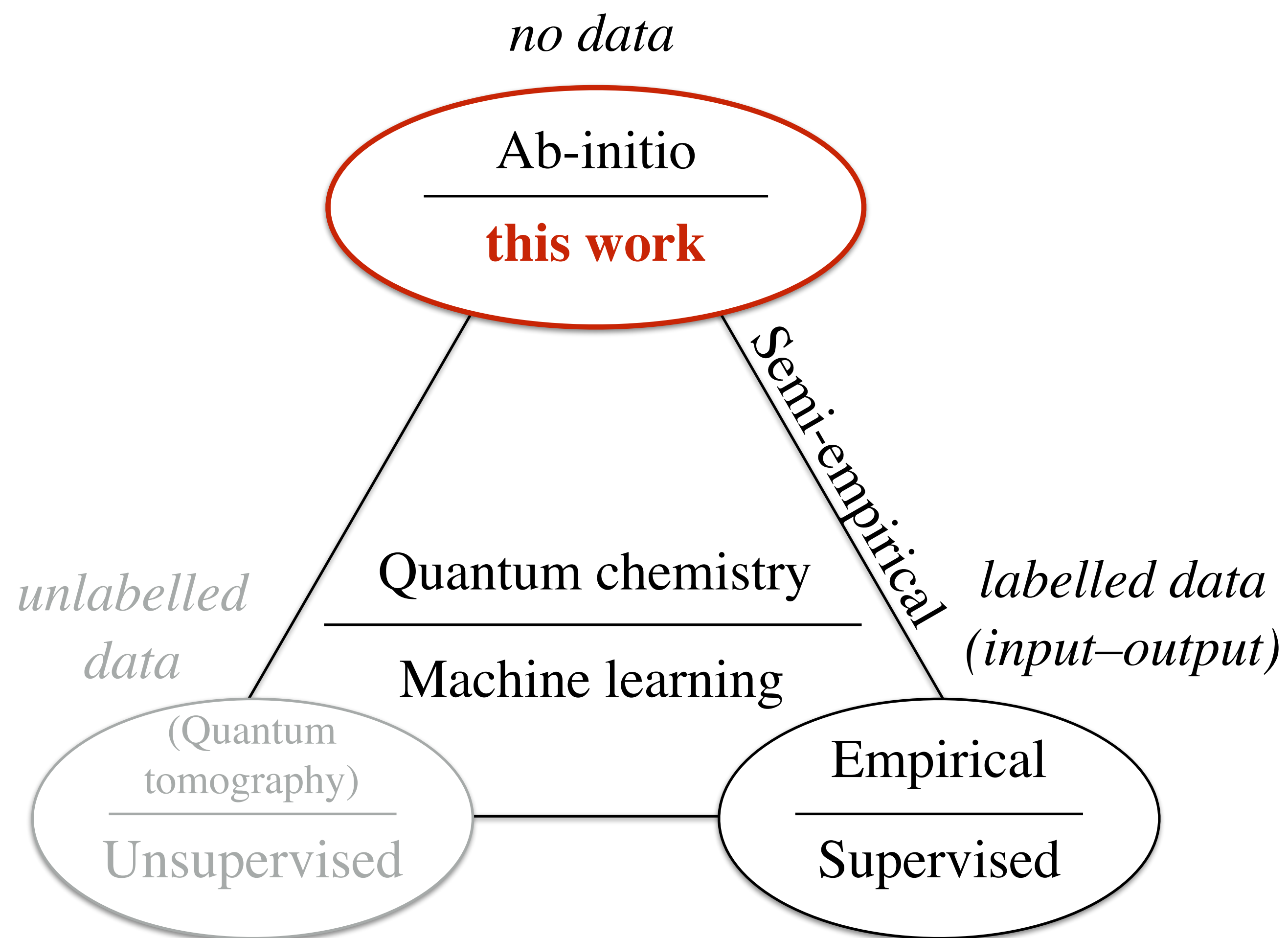
- $\psi(\mathbf{r}) = 0$ defines a $(3N - 1)$ -dimensional *nodal surface* on which the wave function changes sign
- The topology of a nodal surface determines the electronic eigenstate
- If known, efficient solution of the Schrödinger equation with diffusion QMC

Multideterminant expansion curse



- Natural basis for antisymmetric functions—matrix determinant
- NP-hardness of the many-body electronic problem manifest in the rapidly growing determinant expansions in standard methods
- This prevents application of high-level accurate methods to larger systems
- Escape the exponential curse with machine learning?

Machine learning for quantum chemistry



Variational quantum Monte Carlo

$$E_0 = \min_{\psi} E[\psi] \leq \min_{\theta} E[\psi_{\theta}]$$

$$E[\psi] := \int d\mathbf{x} \psi(\mathbf{x}) \hat{H} \psi(\mathbf{x})$$

- Standard quantum chemistry: $E[\psi]$ can be evaluated analytically
- Quantum Monte Carlo (QMC): quantum *expectation* value expressed as statistical *expected* value of local energy

$$E[\psi] = \mathbb{E}_{\mathbf{x} \sim |\psi|^2} [E_{\text{loc}}[\psi](\mathbf{x})]$$

$$E_{\text{loc}}[\psi] = \hat{H}\psi/\psi$$

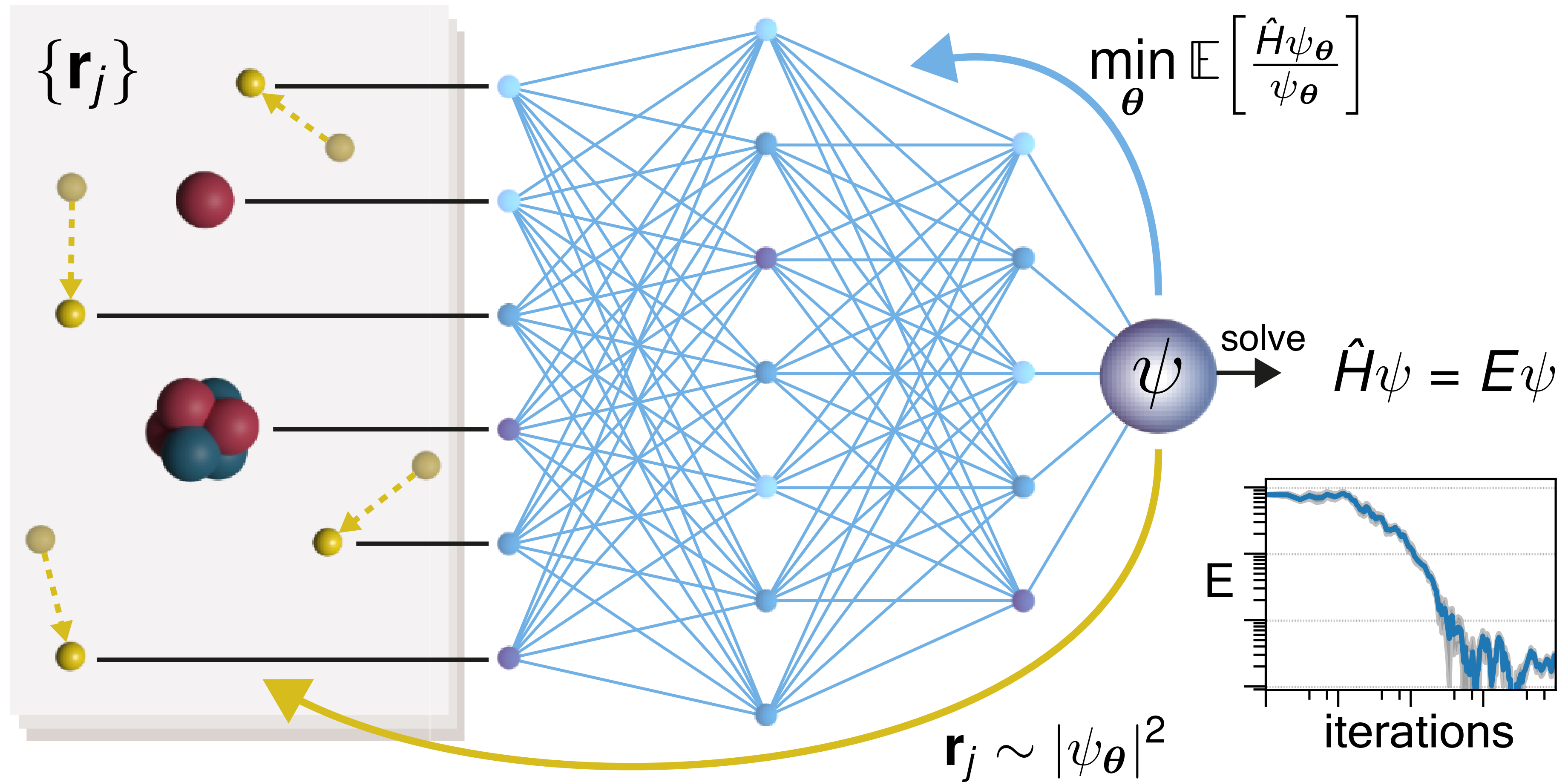
- Traditional ansatz of the Slater–Jastrow–backflow type¹

$$\psi_{\text{SJBF}}(\mathbf{r}) := e^{J(\{\mathbf{r}_i^{\uparrow}\}, \{\mathbf{r}_i^{\downarrow}\})} \sum_p c_p \det[\varphi_{\mu_p}(\mathbf{r}_i^{\uparrow})] \det[\varphi_{\mu_p}(\mathbf{r}_i^{\downarrow})]$$

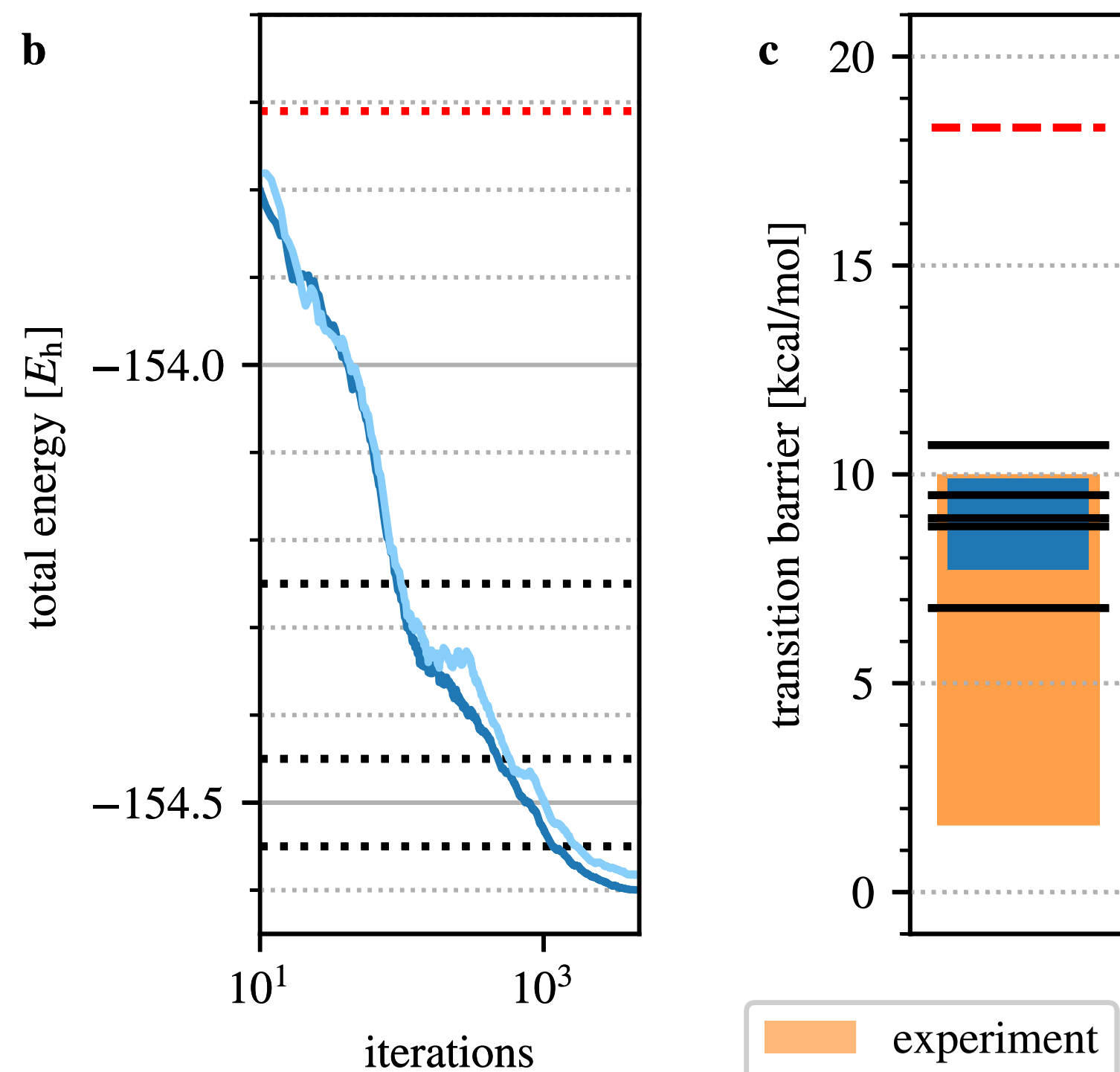
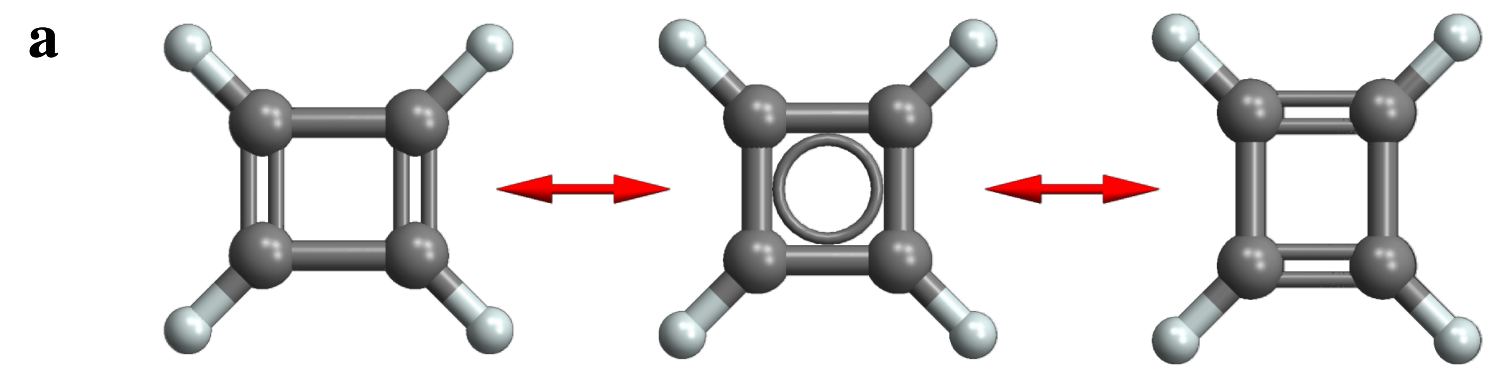
$$\mathbf{r}_i \rightarrow \mathbf{r}_i + \mathbf{q}_i(\mathbf{r})$$

- For larger systems multideterminant Slater–Jastrow ansatzes are used—rapid scaling with system size
- Accuracy of variational QMC limited only by the parametric flexibility of the ansatz – use neural networks²

Deep neural networks for the electronic Schrödinger equation



Targeting state-of-the-art quantum chemistry



..... HF — minimum
..... CCSD(T) — transition

■ experiment
- - - CCSD(T)
— MR-CC
■ PauliNet

- Barrier of the automerization of cyclobutadiene (28 electrons) still not fully resolved¹
- Strong multireferential character—CCSD(T) overestimates two-fold
- Experiment: 1.6–10 kcal/mol, multireference coupled clusters 7–11 kcal/mol
- PauliNet with 10 determinants: 8–10 kcal/mol
- 3 days on GTX 1080 Ti GPU, comparable to multireference methods, but better scaling

¹Lyakh et al., *Chem. Rev.* **112**, 182 (2012)

PauliNet: Deep-neural-network trial wave function

- Slater–Jastrow–backflow ansatz with Jastrow and backflow represented by neural networks

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

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

- Slater determinants of fixed molecular orbitals $\varphi_{\mu}(\mathbf{r}_i)$ from (multi-reference) Hartree–Fock as a baseline
- Generalized backflow¹ for improving the nodal surface
- Built-in n–e and e–e cusp conditions
- Jastrow factor $J_{\theta}(\mathbf{r})$ invariant and backflow equivariant $\mathbf{f}_{\theta}(\mathbf{r})$ w.r.t. electron permutation from adapted cusp-less SchNet² to retain antisymmetry and cusp conditions

Hermann, Schätzle & Noé, *Nat. Chem.* **12**, 891 (2020)

- $\varphi_{\mu}(\mathbf{r}_i)$: HF orbitals
- $J_{\theta}(\mathbf{r})$: Jastrow-factor NN
- $\mathbf{f}_{\theta}(\mathbf{r})$: backflow NN
- $\gamma(\mathbf{r})$: e–e cusp conditions

¹Ceperley, *J. Stat. Phys.* **63**, 1237 (1991)

²Schütt et al. *J. Chem. Phys.* **148**, 241722 (2018)

Variational optimization protocol

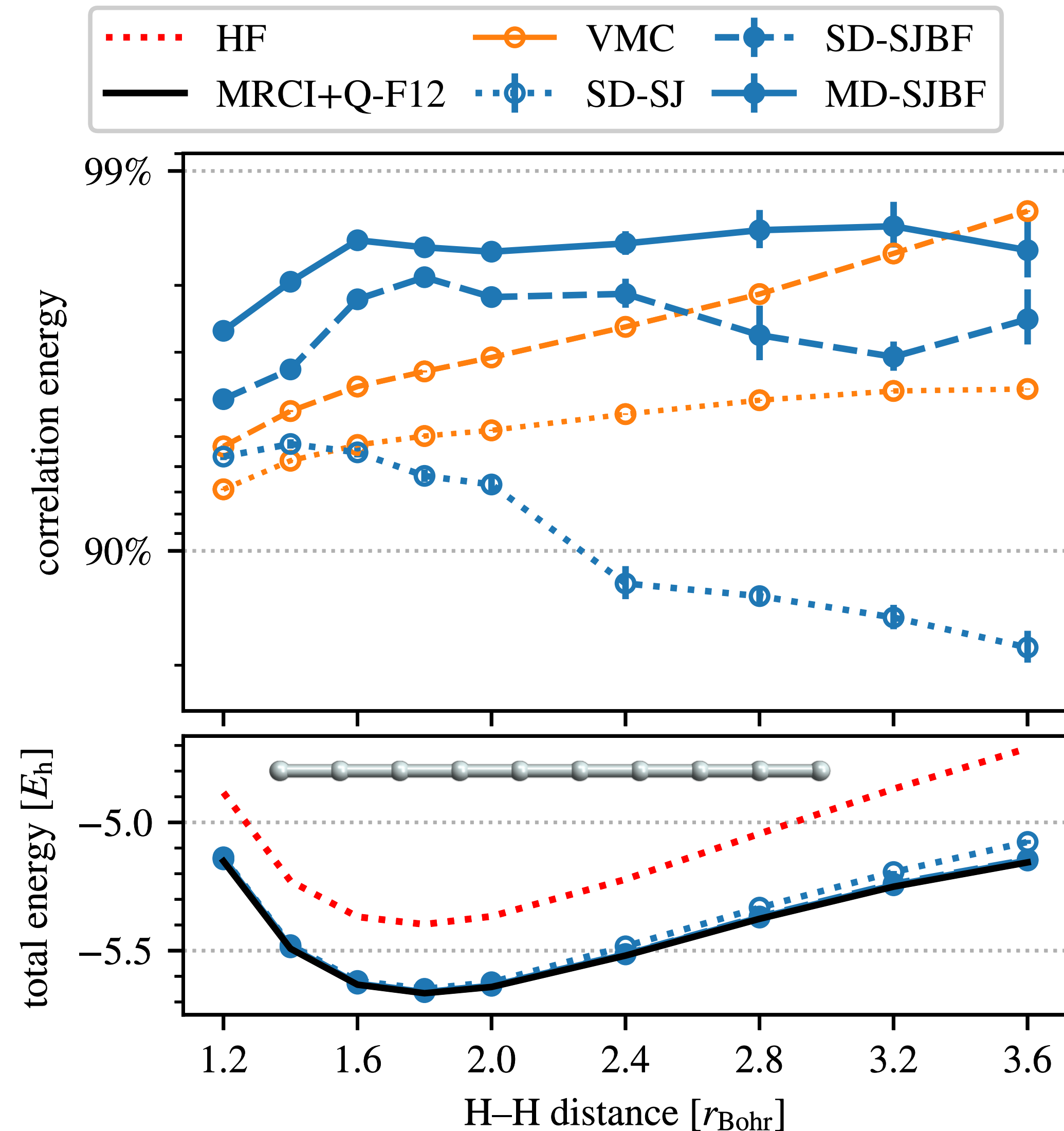
- Energy expectation value directly as a loss function
- Requires calculation of second derivatives (kinetic energy) done by automatic differentiation
- Hermiticity of Hamiltonian enables efficient evaluation of the loss gradient

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{r} \sim |\psi_{\boldsymbol{\theta}}|^2} [E_{\text{loc}}[\psi_{\boldsymbol{\theta}}](\mathbf{r})] \quad (\equiv E[\psi_{\boldsymbol{\theta}}])$$

$$\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) = 2 \mathbb{E}_{\mathbf{r} \sim |\psi_{\boldsymbol{\theta}}|^2} [(E_{\text{loc}}[\psi_{\boldsymbol{\theta}}](\mathbf{r}) - E[\psi_{\boldsymbol{\theta}}]) \nabla_{\boldsymbol{\theta}} \ln |\psi_{\boldsymbol{\theta}}|]$$

- Alternating periods of sampling (Markov-chain Monte Carlo) and optimization (Adam)

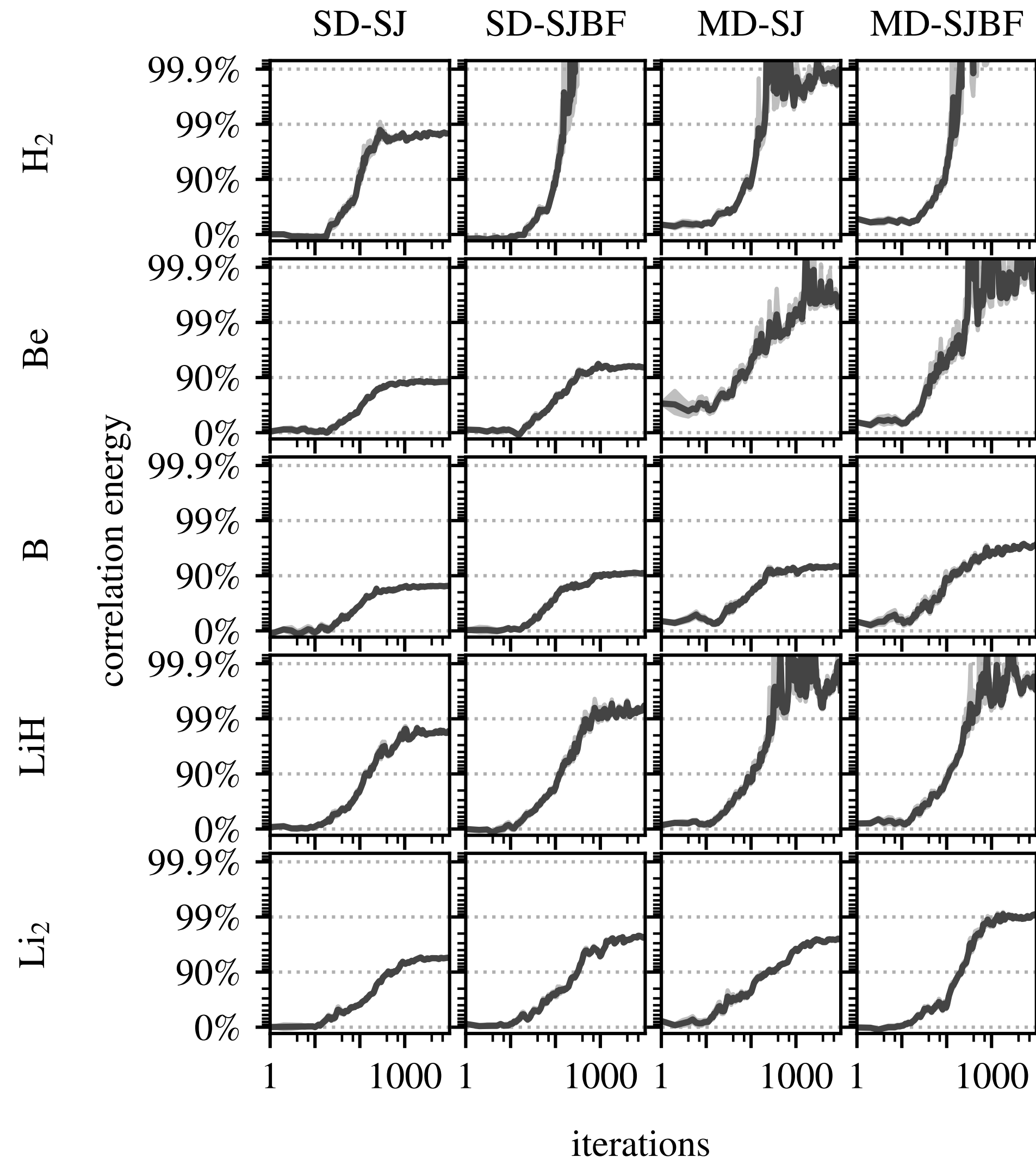
Capturing strong correlation



- Dissociation of linear H_{10} as a test case of *strong correlation*¹ – baseline HF gives unphysical dissociation energy
- PauliNet with 16 determinants captures 98–99% of the correlation energy along the dissociation
- PauliNet surpasses a traditional VMC approach specifically adapted for H_{10}

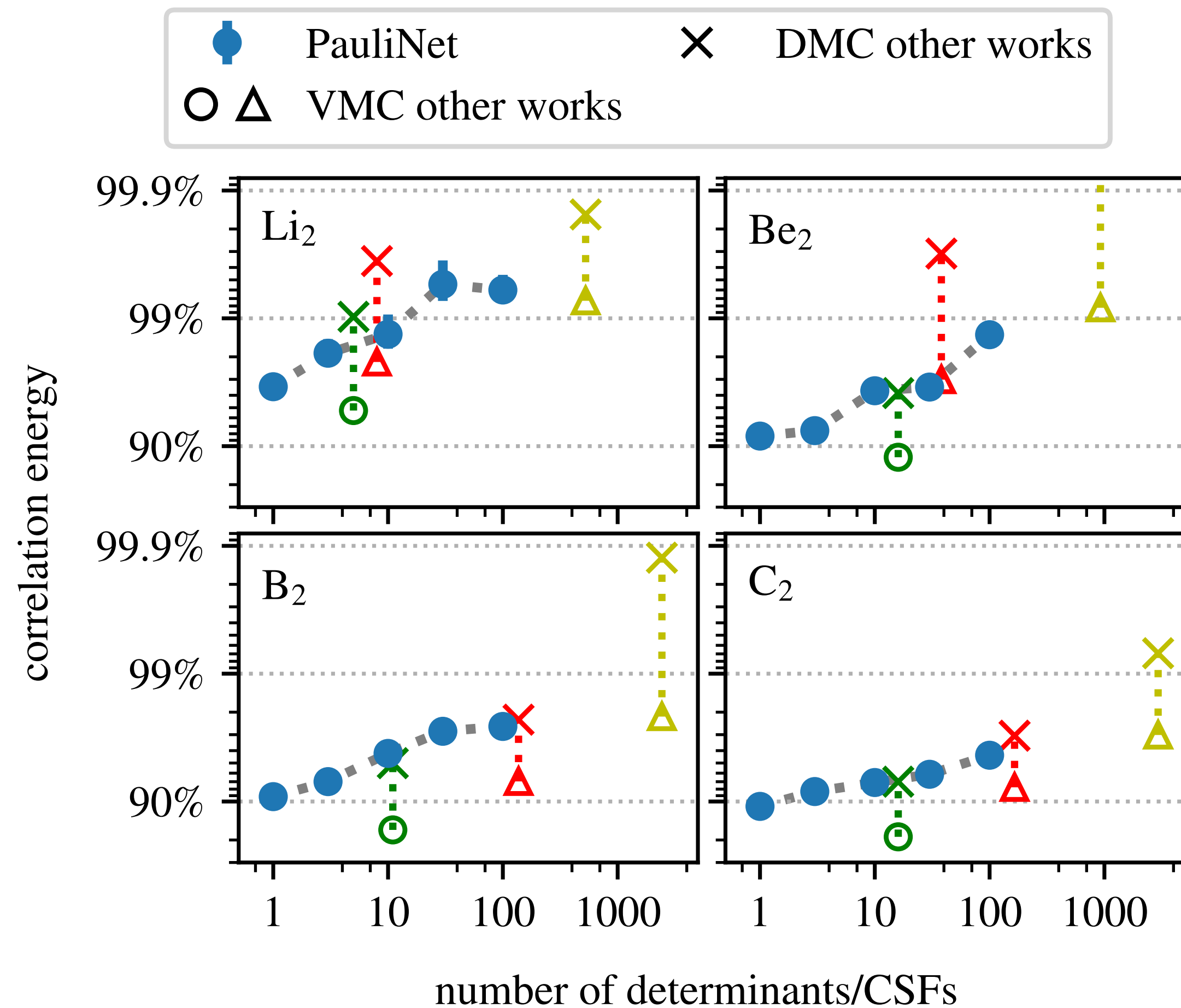
¹Motta et al., *Phys. Rev. X* 7, 031059 (2017)

Approaching exact solution with few determinants



- PauliNet recovers 97% to 99.9% of the electron correlation energy after tens of minutes to a few hours on a single GTX 1080 Ti GPU
- Correlation energy increases monotonously during training
- Only 6 determinants are necessary to substantially reduce the correlation energy from a single-determinant ansatz

Scaling with determinants and system size



- Substantial improvement over existing VMC results and comparable to DMC results in the regime of a few number of determinants

Summary & outlook

- Deep QMC opens a new path to highly accurate methods for electronically complex systems with favorable scaling with system size
- Reuse all standard techniques: diffusion QMC, pseudopotentials
- Improvements in neural network architecture
- Will the performance gain translate straightforwardly to solids?
- How quickly do we lose accuracy as we go to larger systems?

<https://github.com/deepqmc/deepqmc>

```
from deepqmc import Molecule, train
from deepqmc.wf import PauliNet

mol = Molecule.from_name('LiH')
net = PauliNet.from_hf(mol).cuda()
train(net)
```