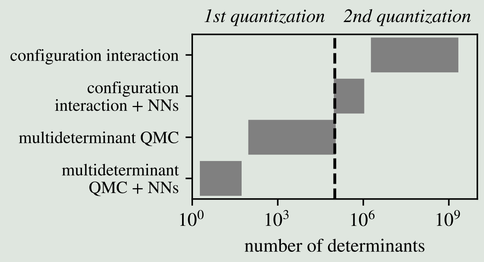
Deep neural network solution of the electronic Schrödinger equation

Jan Hermann[1,2], Zeno Schätzle[1], and Frank Noé[1,3,4]

[1]*FU Berlin, Department of Mathematics and Computer Science, Arnimallee 6, 14195 Berlin, Germany*
[2]*TU Berlin, Machine Learning Group, Marchstr. 23, 10587 Berlin, Germany*
[3]*FU Berlin, Department of Physics, Arnimallee 14, 14195 Berlin, Germany*
[4]*Rice University, Department of Chemistry, Houston, TX 77005, USA*

Background

- Solution of the electronic Schrödinger equation of a given atomic system provides full access to its properties
- Computational cost of highly accurate methods increases rapidly with system size, making them unusable for molecules and materials of practical interest
- Fundamental cause of the unfavorable scaling is the electronic *sign problem*, which is NP-hard (large determinant expansions)



- Neural networks have been shown to tackle the exponential complexity of many-body wave functions efficiently for simpler quantum systems[1]

Variational quantum Monte Carlo

- Electronic Schrödinger equation is a second-order differential eigenvalue equation

$$\hat{H}\psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) = E\psi(\mathbf{r}_1,\ldots,\mathbf{r}_N)$$

$$\hat{H} := \sum_i \left(\frac{1}{2}\nabla^2_{\mathbf{r}_i} + \sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|}\right) + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

- Electrons follow Pauli exclusion principle, implying that valid electronic wave function must be antisymmetric

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

- Schrödinger equation can be reformulated as a minimization problem using variational principle

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

$$E[\psi] \equiv \int \mathrm{d}\mathbf{u}\, \psi(\mathbf{r})\hat{H}\psi(\mathbf{r})$$

- Quantum Monte Carlo (QMC) uses the fact that the square of a wave function is a probability distribution for the electrons and rewrites the energy functional as a statistical expectation value

$$E[\psi] \equiv \mathbb{E}_{\mathbf{r}\sim|\psi|^2}\left[E_{\mathrm{loc}}[\psi](\mathbf{r})\right]$$

- Samples from the wave function are generated with Markov-chain Monte Carlo

Traditional wave function ansatzes

- Hartree–Fock (HF) method of quantum chemistry expresses uncorrelated electronic wave functions as a Slater determinant of molecular orbitals $\varphi_\mu(\mathbf{r}_i)$

$$\psi_{\mathrm{HF}}(\mathbf{r}) = \det[\varphi_\mu(\mathbf{r}_i^\uparrow)]\det[\varphi_\mu(\mathbf{r}_i^\downarrow)]$$

- Traditional QMC uses totally symmetric non-negative Jastrow factors to build correlations between electrons into uncorrelated wave functions, but these do not improve the *nodal surface*

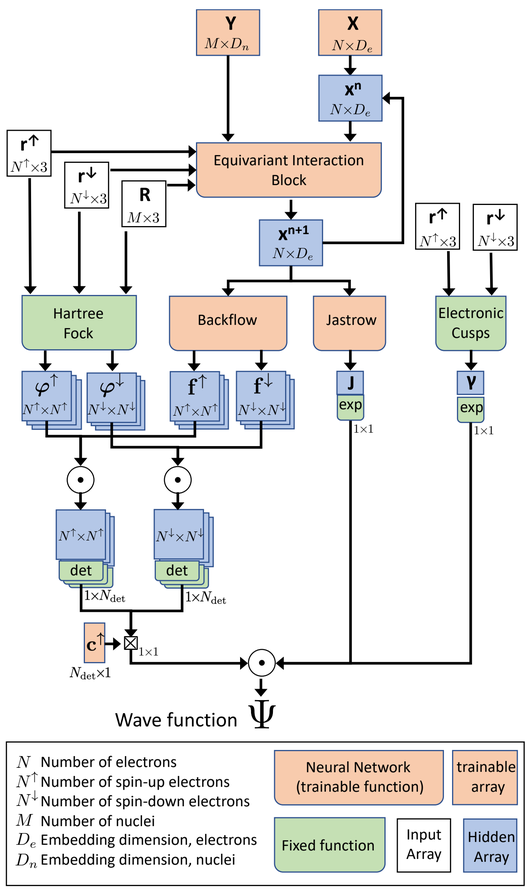
$$\psi(\mathbf{r}) := \psi_{\mathrm{HF}}(\mathbf{r})\mathrm{e}^{J(\mathbf{r})}$$

- Backflow transformation replaces one-electron coordinates with coupled coordinates of pseudoelectrons, which improves the nodal surface, but does not generalize well to larger systems

$$\mathbf{r}_i \mapsto \mathbf{r}_i + \mathbf{b}_i(\mathbf{r})$$

Summary & Outlook

- We develop PauliNet, a neural network architecture designed for representing electronic wave functions in molecules that has several physical constraints built in
- Variational quantum Monte Carlo provides an efficient framework to train such a deep neural network
- The flexibility of deep neural networks enables reaching much higher accuracy in the electronic energy than what is possible with traditional variational ansatzes with the same system-size scaling behavior
- We plan to exploit the superior scaling of our approach to tackle molecular sizes that are inaccessible to traditional methods of quantum chemistry
- We will incorporate ingredients from the parallel work on FermiNet[4]



PauliNet: Deep neural network electronic wave function ansatz

- Our proposed ansatz is of the Slater–Jastrow–backflow type, where both the Jastrow factor J and the backflow $\mathbf{f}$ are deep neural networks

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

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

Antisymmetry Standard technique from quantum chemistry—Linear combination of Slater determinants changes sign upon exchange of any two rows (molecular orbitals)

Hartree–Fock baseline We build in multireference Hartree–Fock solution into PauliNet to ensure a good starting point for the optimization and to increase data efficiency

Asymptotic behavior Electronic eigenstates satisfy *cups conditions* when an electron approaches other electron or a nucleus, which are built into PauliNet through $\boldsymbol{\varphi}(\mathbf{r})$ and $\gamma(\mathbf{r})$

Jastrow factor and backflow from graph convolutional network

- To retain antisymmetry, the Jastrow factor and backflow neural networks must be invariant and equivariant, respectively, with respect to the exchange fo same-spin electrons

$$J(\mathscr{P}_{ij}\mathbf{r}) = J(\mathbf{r}) \qquad \mathscr{P}_{ij} f_{\mu i}(\mathbf{r}) = f_{\mu j}(\mathscr{P}_{ij}\mathbf{r})$$

- PauliNet uses an adapted SchNet architecture,[2] which is a graph neural network with interactions between particles represented as convolutions in real space

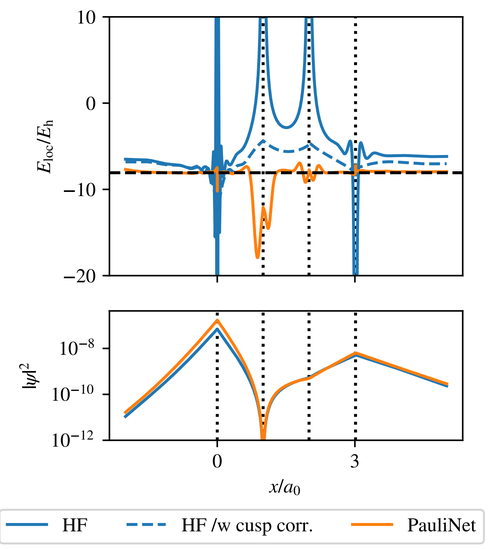
$$\mathbf{x}_i^{(n+1)} := \mathbf{x}_i^{(n)} + \boldsymbol{\chi}_{\boldsymbol{\theta}}^{(n)}\left(\left\{\mathbf{x}_j^{(n)}, \{|\mathbf{r}_j - \mathbf{r}_k|\}\right\}\right)$$

- The final electron many-body representations from SchNet are used as input into vanilla deep neural networks for the Jastrow factor and backflow

$$J := \eta_{\boldsymbol{\theta}}\left(\sum_i \mathbf{x}_i^{(L)}\right) \qquad \mathbf{f}_i := \boldsymbol{\kappa}_{\boldsymbol{\theta}}\left(\mathbf{x}_i^{(L)}\right)$$

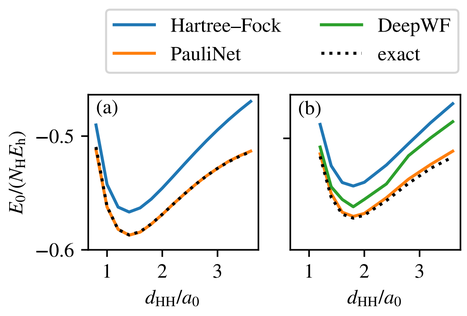
Approaching exact solution with few determinants

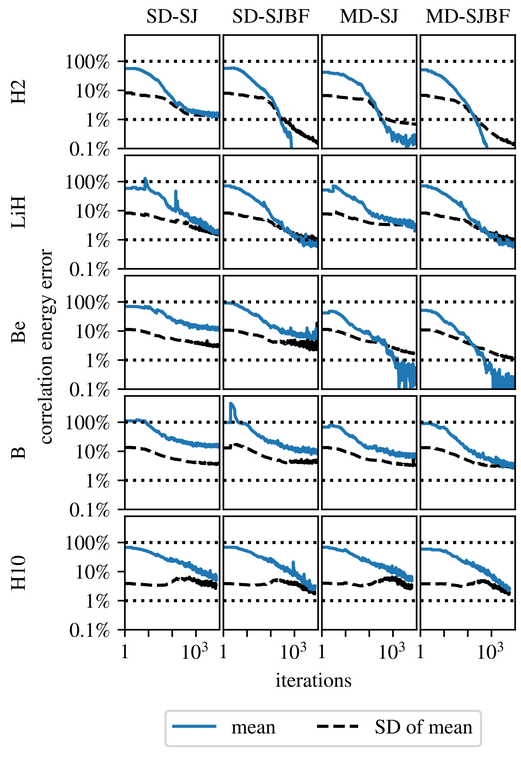
- PauliNet recovers 97% to 99.9% of the electron correlation energy after training for tens of minutes to a few hours on a single GTX 1080 Ti GPU
- Error in the correlation energy decreases almost monotonously as the training progresses from the initial HF baseline level
- Only a few determinants are necessary to substantially reduce the correlation energy from a single-determinant ansatz
- The trained PauliNet generalizes well to regions with low probability



Capturing strong correlation

- Systems that exhibit *strong correlation* are especially hard to describe without large determinant expansions
- PauliNet can smoothly dissociate both the H_2 molecule and the strongly correlated linear hydrogen chain





	Correlation energy					E/E_{h}
System	SD-VMC	SD-DMC	DeepWF[3]	PauliNet		Exact
H_2			98.4 %	99.99 %	**–1.17447**(2)	–1.17447
LiH	91.5 %	99.7 %	–	99.3 %	**–8.070**0(2)	–8.070548
Be	61.6 %	89.2 %	43.6 %	99.94 %	**–14.667**4(3)	–14.66736
B	60.0 %	88.3 %	–	97.3 %	**–24.65**06(11)	–24.65391
H_{10}			63.8 %	98.0 %	**–5.66**02(7)	–5.6655

Manuscript

arXiv:1909.08423

References

(1) Carleo, Troyer, *Science* **355**, 6325 (2017)
(2) Schütt et al., *J. Chem. Phys.* **148**, 241722 (2018)
(3) Han et al., arXiv:1807.07014 (2018)
(4) Pfau et al., arXiv:1909.02487 (Sep 2019)

Acknowledgements

Funding is acknowledged from the European Commission (ERC CoG 772230 "Scale-Cell"), Deutsche Forschungsgemeinschaft (CRC1114/A04, GRK2433 DAEDALUS/P04), the MATH+ Berlin Mathematics research center (AA1x6, EF1x2). J. H. thanks K.-R. Müller for support and acknowledges funding from TU Berlin (Project No. 10032745).