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, Dept. of Mathematics and Computer Science, Arnimallee 6, 141 95 Berlin, Germany
[2]TU Berlin, Machine Learning Group, Marchstr. 23, 105 87 Berlin, Germany
[3]FU Berlin, Dept. of Physics, Arnimallee 14, 141 95 Berlin, Germany
[4]Rice University, Dept. of Chemistry, Houston, TX 77005, USA
{jan.hermann,frank.noe}@fu-berlin.de

Abstract

The electronic Schrödinger equation describes fundamental properties of molecules and materials, but cannot be solved exactly for larger systems than the hydrogen atom. Quantum Monte Carlo is a high-accuracy approximation method, in principle only limited by the flexibility of the used wave-function ansatz. Here we develop the deep-learning real-space wave-function ansatz PauliNet. PauliNet has the Hartree–Fock solution built in as a baseline, incorporates the physics of valid wave functions, and is trained using variational quantum Monte Carlo (VMC). PauliNet achieves higher accuracy than comparable state-of-the-art VMC for atoms, diatomic molecules, and a strongly-correlated hydrogen chain.

1 Introduction

A cornerstone of predicting the physical and chemical properties of a given molecule or material defined by Hamiltonian operator, $\hat{H}$ is the solution of the corresponding electronic Schrödinger equation for the unknown electronic wave function, $\psi(\mathbf{r}_1, \ldots, \mathbf{r}_N)$, and electronic energy, E,

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

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

Here, $(\mathbf{r}_1, \ldots, \mathbf{r}_N) \equiv \mathbf{r} \in \mathbb{R}^{3N}$ are the spatial coordinates of N electrons, and $(Z_I, \mathbf{R}_I)$ are the charges and coordinates of M atomic nuclei. In addition to spatial coordinates, each electrons has a $s_i \in \{\uparrow, \downarrow\}$, and the solution of Eq. (1-2) must obey antisymmetry with respect to exchanging equal-spin electrons. The antisymmetry constraint makes solving the electronic Schrödinger equation to a fixed accuracy level NP-hard, i.e. it scales exponentially with N [Troyer and Wiese, 2005]. An exact solution of Eq. (1-2) is only available for $N = M = 1$, the hydrogen atom, but quantum chemistry has developed numerous approximation methods with varying degrees of accuracy and computational efficiency [Piela, 2014].

Many quantum chemistry methods build on the *variational principle*, which states that the solution of (1) with the lowest energy, ψ_0, E_0, can be obtained by minimizing the energy expectation value over all antisymmetric wave functions. Thus, minimizing over a *trial* subset of all wave functions, called an *ansatz*, parametrized by $\boldsymbol{\theta}$, gives an upper bound to the true ground-state energy E_0:

$$E_0 = \min_\psi E[\psi] \leq \min_{\boldsymbol{\theta}} E[\psi_{\boldsymbol{\theta}}], \qquad E[\psi] = \int \mathrm{d}\mathbf{r}\, \psi(\mathbf{r})\hat{H}\psi(\mathbf{r}) \tag{3}$$
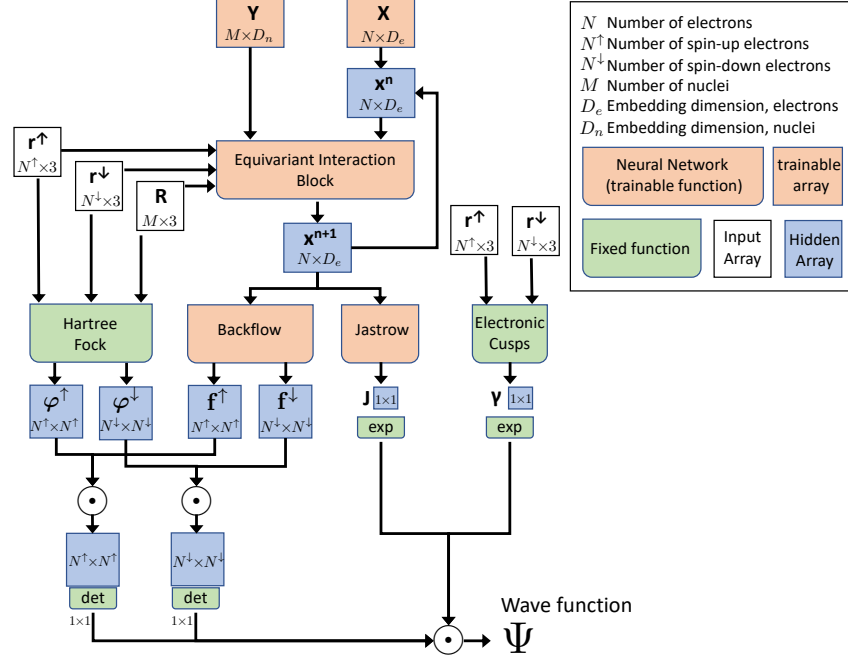
Figure 1: **Schematic of the PauliNet neural network structure and information flow.**

In variational quantum Monte Carlo (VMC) [Foulkes et al., 2001, Needs et al., 2010, Austin et al., 2012], the energy integral is written as an expected value of the *local energy*, $E_{\text{loc}}[\psi](\mathbf{r}) = \hat{H}\psi(\mathbf{r})/\psi(\mathbf{r})$, over the probability distribution $|\psi^2(\mathbf{r})|$,

$$E[\psi] = \mathbb{E}_{\mathbf{r}\sim|\psi|^2}\big[E_{\text{loc}}[\psi](\mathbf{r})\big] \tag{4}$$

The primary motivation to improve existing variational wave-function ansatzes is that although there are several extensions of VMC that increase its accuracy, such as the diffusion quantum Monte Carlo (DMC), improving the actual ansatz is the only way to improve the *nodal surface* of a wave function, which is the hyperplane of electronic coordinates on which the wave function changes sign, and whose accuracy sets fundamental limitation to the accuracy of all quantum Monte Carlo (QMC) techniques. The existing wave-function ansatzes are partially motivated by the physics of electrons, partially by trial-and-error experience. Here, we demonstrate that ansatzes that incorporate the right physics while using DNNs in place of the existing ad-hoc functional forms can outperform state-of-the-art variational ansatzes of comparable types.

2 Method

2.1 Neural wave function PauliNet incorporating the physics of electrons

Our proposed PauliNet wave function is of the Slater–Jastrow–backflow type, where both the Jastrow factor J and the backflow $\mathbf{f}$ are represented by DNNs with trainable parameters $\boldsymbol{\theta}$ (Figure 1),

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = \det[\varphi_\mu(\mathbf{r}_i^\uparrow)f_{\boldsymbol{\theta},\mu i}(\mathbf{r})]\det[\varphi_\mu(\mathbf{r}_i^\downarrow)f_{\boldsymbol{\theta},\mu i}(\mathbf{r})]\mathrm{e}^{\gamma(\mathbf{r})+J_{\boldsymbol{\theta}}(\mathbf{r})} \tag{5}$$

$J_{\boldsymbol{\theta}}$ and $f_{\boldsymbol{\theta}}(\mathbf{r})$ are invariant and equivariant DNNs, respectively, with respect to exchanging same-spin electrons, and φ_μ and γ are fixed functions. This ansatz ensures basic physics of electronic wave functions as follows.

Antisymmetry We fix the first $N_\uparrow$ ($N_\downarrow$) electrons to be spin-up (spin-down), The wave function then must be antisymmetric in the first $N_\uparrow$ and in the last $N_\downarrow = N - N_\uparrow$ coordinates.

$$\psi\big(\mathbf{r}_1^\uparrow,\dots,\mathbf{r}_{N_\uparrow}^\uparrow,\mathbf{r}_{N_\uparrow+1}^\downarrow,\dots,\mathbf{r}_N^\downarrow\big) \equiv \psi(\mathbf{r}^\uparrow,\mathbf{r}^\downarrow) \equiv \psi(\mathbf{r}) \tag{6}$$

$$\psi(\dots,\mathbf{r}_i,\dots,\mathbf{r}_j,\dots) = -\psi(\dots,\mathbf{r}_j,\dots,\mathbf{r}_i,\dots), \tag{7}$$

As common in quantum chemistry, our ansatz enforces antisymmetry among same-spin electrons via the matrix determinants in Eq. (5).

Hartree–Fock baseline   To ensure a good starting point for the optimization problem, we build the standard Hartree–Fock (HF) wave function as a baseline into PauliNet. The HF wave function uses one-electron functions, $\varphi_\mu(\mathbf{r})$, called molecular orbitals, and achieved antisymmetry by the so-called Slater determinant.

$$\psi_{\mathrm{HF}}(\mathbf{r}) := \det[\varphi_\mu(\mathbf{r}_i^\uparrow)] \det[\varphi_\mu(\mathbf{r}_i^\downarrow)] \tag{8}$$

HF is straightforward, computationally efficient, captures much of the basic physics of atoms and molecules, and makes qualitative reasonable predictions. We take the molecular orbitals directly from the HF method and keep them fixed in our ansatz (5).

Beyond Hartree–Fock with trainable Jastrow factor and backflow   The main limitation to accuracy in HF is that it does not depend on electron-electron distances, and therefore captures no electron-electron correlation besides the antisymmetric constraint. We combine two established strategies to improve the HF baseline. First, the HF wave function is multiplied by a nonnegative totally symmetric function $e^{J_\theta(\mathbf{r})}$, with the so-called Jastrow factor J [Drummond et al., 2004]. This modification can build complex electron correlations into the wave function, but cannot modify the nodal surface inherited from HF. Second, we can generalize the molecular orbitals to all-electron functions as long as we retain equivariance (and thus the determinant antisymmetry) with respect to exchanging same-spin electrons,

$$\varphi_\mu(\mathbf{r}_i) \to \varphi_{\mu i}(\mathbf{r}), \qquad \mathcal{P}\varphi_{\mu i}(\mathbf{r}) = \varphi_{\mu i}(\mathcal{P}\mathbf{r}) \tag{9}$$

where $\mathcal{P}$ is the exchange operator for same-spin electrons. This approach is called the backflow, and unlike the Jastrow factor it can modify the nodal surface of the HF wave function [López Ríos et al., 2006]. We achieve backflow by multiplying the orbitals φ_μ with trainable DNN functions $f_{\theta,\mu}$.

Cusp conditions   Any ground-state electronic wave function obeys exact asymptotic behavior defined by the *cusp conditions* as electrons approach each other and the nuclei [Kato, 1957],

$$\frac{1}{\psi_0}\frac{\partial\psi_0}{\partial|\mathbf{r}_i - \mathbf{R}_I|}\bigg|_{\mathbf{r}_i=\mathbf{R}_I} = -Z_I, \qquad \frac{1}{\psi_0}\frac{\partial\psi_0}{\partial|\mathbf{r}_i - \mathbf{r}_j|}\bigg|_{\mathbf{r}_i=\mathbf{r}_j} = \begin{cases} \frac{1}{4} & s_i = s_j \\ \frac{1}{2} & s_i \neq s_j \end{cases} \tag{10}$$

PauliNet ensures the nuclear cusp conditions via the molecular orbitals $\varphi_\mu(\mathbf{r}_i)$ as described in Ma et al. [2005]. The electronic cusp conditions are enforced by a fixed function $\gamma(\mathbf{r})$. To preserve the cusp conditions built into φ_μ and γ, the Jastrow factor and backflow DNNs must be cuspless, which we ensured by construction of the DNN architecture (see below).

$$\nabla_{\mathbf{r}_i} J(\mathbf{r})\big|_{\mathbf{r}_i=\{\mathbf{r}_k,\mathbf{R}_a\}} = 0, \qquad \nabla_{\mathbf{r}_i} f_{\mu i}(\mathbf{r})\big|_{\mathbf{r}_i=\{\mathbf{r}_k,\mathbf{R}_a\}} = 0 \tag{11}$$

Neural network architecture   The requirements of invariance and equivariance with respect to permutation of particles, and the fact that particle interactions are a function of their distances are closely related with the aim of constructing DNNs that learn potential energy functions. Therefore, we adapt the graph-convolutional neural network SchNet [Schütt et al., 2018], as follows: (i) The iterated feature vectors $\mathbf{x}_i^{(n)}$ of the graph represent electrons, not atoms as in SchNet, and we only distinguish two initial embeddings: spin-up and spin-down. (ii) The graph messages received by the electron feature vectors at each iteration are split into three channels: same-spin electrons, opposite-spin electrons, and nuclei. (iii) Each nucleus is represented by a trainable embedding $\mathbf{Y}_{\theta,I}$, shared across all iterations of the graph network. (iv) Two new trainable functions η_θ and κ_θ of the final electron feature vectors return the values of the Jastrow factor and the backflow, respectively. As feature vectors $\mathbf{x}_i^{(n)}$ are equivariant with respect to exchange at each iteration, so are the backflow vectors $\mathbf{f}_i$. (v) The distance features $\mathbf{e}$ are constructed to be cuspless to maintain the cusp behavior encoded in φ_μ and γ.

Training   To generate samples of $\mathbf{r}$ from the wave-function distribution $|\psi^2(\mathbf{r})|$, needed to evaluate the expected value in (4), we use a standard Langevin Monte Carlo approach [Umrigar et al., 1993]. To optimize the parameters θ in the Jastrow and backflow neural networks, we use Adam [Kingma and Ba, 2014] with the variational energy (4) as the loss function. The stochastic loss gradient over a batch of samples is computed with a formula that exploits the Hermitian nature of the Hamiltonian in (3) [Ceperley et al., 1977]. This formulation only requires second derivatives of the wave function (for the Laplace operator), whereas direct differentiation would require third derivatives.

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{r}\sim|\psi'^2|}\big[E_{\mathrm{loc}}(\mathbf{r};\boldsymbol{\theta})\big], \qquad \nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}) = 2\mathbb{E}_{\mathbf{r}\sim|\psi'^2|}\big[\big(E_{\mathrm{loc}}(\mathbf{r};\boldsymbol{\theta}) - \mathcal{L}(\boldsymbol{\theta})\big)\nabla_{\boldsymbol{\theta}}\ln|\psi_{\boldsymbol{\theta}}|\big] \tag{12}$$

Table 1: **Ground-state energies of five test systems obtained by four different methods.**

system	VMC		DMC		DeepWF		PauliNet	
			E_0^a/E_h (% E_{corr})					
H_2					-1.1738	98.4%	$-1.174\,37(6)$	99.7%
LiH	-8.0635^f	91.5%	-8.0703^f	99.7%	-7.8732	[b]	$-8.0690(3)$	98.1%
Be	-14.6311^d	61.6%	-14.6572^d	89.2%	-14.6141	43.6%	$-14.6546(7)$	86.5%
B	-24.6056^d	60.0%	-24.6398^d	88.3%	-24.2124	[b]	$-24.634(2)$	83.5%
H_{10}^e					-5.5685	63.8%	$-5.655(2)$	96.0%

[a]Reference exact ground-state and HF energies are taken from Kołos and Wolniewicz [1965], Cencek and Rychlewski [2000], Brown et al. [2007], Motta et al. [2017], Casalegno et al. [2003]. [b]For LiH and B, DeepWF does not reach the accuracy of the HF method.
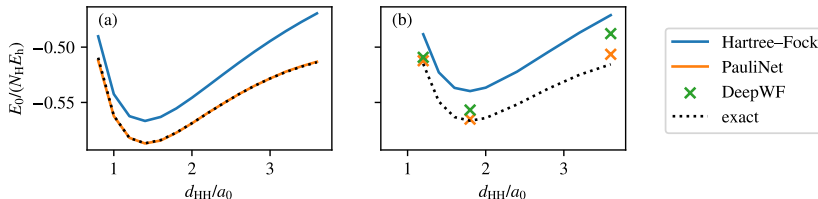
Figure 2: **Electronic energy as a function of distance between atoms in H_2 (a) and H_{10} (b).**

3   Results and Conclusions

In this work, we test only the Jastrow part of our architecture and turn the backflow part off, which enables us to directly compare against fixed-node DMC calculations. We use the same set of atoms and small molecules that were used to test DeepWF in [Han et al., 2018] (Table 1).

PauliNet outperforms state-of-the-art Jastrow factors   A common way to compare the accuracy of *ab-initio* methods in quantum chemistry is via the fraction of *correlation energy* captured, defined as the difference between the exact ground-state energy and that of the HF method. Whereas DeepWF does not even reach the accuracy of the HF method for LiH and B, our PauliNet gives variational energies for atoms and diatomics that are better than those from state-of-the-art VMC calculations, and are only outperformed by the fixed-node limits of DMC calculations by several percents of the correlation energy. It took several hundreds of iterations to converge the energy to the reported values, corresponding to tens of minutes on a single GTX1080Ti GPU.

PauliNet captures strong correlation   Unlike the atoms and diatomics, the linear hydrogen chain H_{10} exhibits *strong correlation*, which describes a situation where the single-determinant description of the HF method is qualitatively insufficient [Motta et al., 2017]. This is a hard test for the ansatz, because the Jastrow factor needs to be a complicated many-body function under such circumstances [Goetz and Neuscamman, 2017]. Already H_2 exhibits strong correlation when the two atoms are dissociated, but our Jastrow factor is able to cover the transition from dynamic to strong correlation smoothly (Figure 2a). For H_{10}, we recover 96% (80%) of the correlation energy in the equilibrium (stretched) geometry (Figure 2b). The supposedly harder problem of learning the Jastrow factor for a strongly-correlated system leads to a longer optimization time, here on the order of hours.

Conclusions   DNN representations of real-space electronic wave functions can outperform state-of-the-art variational ansatzes for simple and strongly-correlated systems, as a result of the increased flexibility in the functional form. In parallel, [Pfau et al., 2019] developed a deep-learning architecture that exhibits higher accuracy, but also much greater computational costs. Combining the physical constraints built into PauliNet with the improved optimizer of [Pfau et al., 2019] has the potential to combine their advantages. In future work, we plan to explore numerical performance of the neural backflow in PauliNet, which in principle removes any inherent limitations of the flexibility of our ansatz. We hope that introducing DNNs to quantum Monte Carlo opens the possibility to exploit the striking advances in deep learning for scientific discovery.

References

B. M. Austin, D. Y. Zubarev, and W. A. Lester. Quantum Monte Carlo and Related Approaches. *Chem. Rev.*, 112(1):263–288, Jan. 2012. doi: 10.1021/cr2001564.

M. D. Brown, J. R. Trail, P. López Ríos, and R. J. Needs. Energies of the first row atoms from quantum Monte Carlo. *J. Chem. Phys.*, 126(22):224110, June 2007. doi: 10.1063/1.2743972.

M. Casalegno, M. Mella, and A. M. Rappe. Computing accurate forces in quantum Monte Carlo using Pulay's corrections and energy minimization. *J. Chem. Phys.*, 118(16):7193, 2003. doi: 10.1063/1.1562605.

W. Cencek and J. Rychlewski. Benchmark calculations for Heq2 and LiH molecules using explicitly correlated Gaussian functions. *Chem. Phys. Lett.*, 320:549–552, 2000.

D. Ceperley, G. V. Chester, and M. H. Kalos. Monte Carlo simulation of a many-fermion study. *Phys. Rev. B*, 16(7):3081–3099, Oct. 1977. doi: 10.1103/PhysRevB.16.3081.

N. D. Drummond, M. D. Towler, and R. J. Needs. Jastrow correlation factor for atoms, molecules, and solids. *Phys. Rev. B*, 70(23):235119, Dec. 2004. doi: 10.1103/PhysRevB.70.235119.

W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.*, 73(1):33–83, Jan. 2001. doi: 10.1103/RevModPhys.73.33.

B. V. D. Goetz and E. Neuscamman. Suppressing Ionic Terms with Number-Counting Jastrow Factors in Real Space. *J. Chem. Theory Comput.*, 13(5):2035–2042, May 2017. doi: 10.1021/acs.jctc.7b00158.

J. Han, L. Zhang, and W. E. Solving Many-Electron Schrödinger Equation Using Deep Neural Networks. *arXiv*, 1807.07014, July 2018. URL http://arxiv.org/abs/1807.07014.

T. Kato. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.*, 10(2):151–177, Jan. 1957. doi: 10.1002/cpa.3160100201.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv*, 1412.6980, Dec. 2014. URL http://arxiv.org/abs/1412.6980.

W. Kołos and L. Wolniewicz. Potential-Energy Curves for the X 1Σg+, b3Σu+, and C 1Πu States of the Hydrogen Molecule. *J. Chem. Phys.*, 43(7):2429–2441, Oct. 1965. doi: 10.1063/1.1697142.

P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs. Inhomogeneous backflow transformations in quantum Monte Carlo calculations. *Phys. Rev. E*, 74(6), Dec. 2006. doi: 10.1103/PhysRevE.74.066701.

A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs. Scheme for adding electron–nucleus cusps to Gaussian orbitals. *J. Chem. Phys.*, 122(22):224322, June 2005. doi: 10.1063/1.1940588.

M. Motta, D. M. Ceperley, G. K.-L. Chan, J. A. Gomez, E. Gull, S. Guo, C. A. Jiménez-Hoyos, T. N. Lan, J. Li, F. Ma, A. J. Millis, N. V. Prokof'ev, U. Ray, G. E. Scuseria, S. Sorella, E. M. Stoudenmire, Q. Sun, I. S. Tupitsyn, S. R. White, D. Zgid, S. Zhang, and Simons Collaboration on the Many-Electron Problem. Towards the Solution of the Many-Electron Problem in Real Materials: Equation of State of the Hydrogen Chain with State-of-the-Art Many-Body Methods. *Phys. Rev. X*, 7(3):031059, Sept. 2017. doi: 10.1103/PhysRevX.7.031059.

R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos. Continuum variational and diffusion quantum Monte Carlo calculations. *J. Phys. Condens. Matter*, 22(2):023201, Jan. 2010. doi: 10.1088/0953-8984/22/2/023201.

D. Pfau, J. S. Spencer, A. G. d. G. Matthews, and W. M. C. Foulkes. Ab-Initio Solution of the Many-Electron Schrödinger Equation with Deep Neural Networks. *arXiv*, 1909.02487, Sept. 2019. URL http://arxiv.org/abs/1909.02487.

L. Piela. *Ideas of Quantum Chemistry*. Elsevier, 2nd edition, 2014. ISBN 978-0-444-52227-6.

K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.*, 148(24):241722, June 2018. doi: 10.1063/1.5019779.

M. Troyer and U.-J. Wiese. Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations. *Phys. Rev. Lett.*, 94(17):170201, May 2005. doi: 10.1103/PhysRevLett.94.170201.

C. J. Umrigar, M. P. Nightingale, and K. J. Runge. A diffusion Monte Carlo algorithm with very small time-step errors. *J. Chem. Phys.*, 99(4):2865–2890, Aug. 1993. doi: 10.1063/1.465195.