nature chemistry

ARTICLES

https://doi.org/10.1038/s41557-020-0544-y

Deep-neural-network solution of the electronic Schrödinger equation

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

The electronic Schrödinger equation can only be solved analytically for the hydrogen atom, and the numerically exact full configuration-interaction method is exponentially expensive in the number of electrons. Quantum Monte Carlo methods are a possible way out: they scale well for large molecules, they can be parallelized and their accuracy has, as yet, been only limited by the flexibility of the wavefunction ansatz used. Here we propose PauliNet, a deep-learning wavefunction ansatz that achieves nearly exact solutions of the electronic Schrödinger equation for molecules with up to 30 electrons. PauliNet has a multireference Hartree–Fock solution built in as a baseline, incorporates the physics of valid wavefunctions and is trained using variational quantum Monte Carlo. PauliNet outperforms previous state-of-the-art variational ansatzes for atoms, diatomic molecules and a strongly correlated linear H_{10}, and matches the accuracy of highly specialized quantum chemistry methods on the transition-state energy of cyclobutadiene, while being computationally efficient.

A solution of the time-independent electronic Schrödinger equation of a given atomic system provides, in principle, full access to its chemical properties. This equation can be solved analytically only for an isolated hydrogen atom, but solid-state physics and quantum chemistry have been remarkably successful in developing numerical approximation methods[1]. For small molecules containing up to a few tens of electrons, methods based on the configuration-interaction and the closely related coupled-cluster approaches or the multideterminant quantum Monte Carlo (QMC) can reach impressive accuracy of up to six significant digits in the total electronic energy[2].

Unfortunately, the computational cost of such high-accuracy methods increases with a high power of the number of electrons, N, making these methods impractical for most relevant molecules or materials. Computationally less demanding methods, such as density functional theory (DFT), can scale to larger molecules, but at the price of limited accuracy. Fundamentally, high-accuracy methods scale unfavourably with N because the dimension of the solution space of the Schrödinger equation for a many-body problem scales exponentially. A distinct but related problem that appears in some approaches is the so-called sign problem originating from the Pauli exclusion principle[3]. The trade-off between accuracy and computational cost is apparent when considering that most quantum chemistry methods represent electronic wavefunctions by linear combinations of Slater determinants. A Slater matrix is constructed by selecting N out of $M > N$ molecular orbitals, and assigning N electrons to them, resulting in a combinatorial growth of all possible matrices with system size.

The Slater determinants have different roles in different quantum chemistry methods (Fig. 1). In the configuration-interaction and coupled-cluster approaches, the electronic problem is solved entirely in the basis of the determinants (second quantization), and as such their number in typical applications is the largest, as the determinant expansion must recover all many-body interactions that are missing in individual determinants[4]. Stochastic methods that sample over vast determinant spaces have been developed[5,6], but the underlying scaling trap persists nevertheless. Recent work by Choo et al. suggests that the number of required determinants can be reduced with the use of neural networks[7], but whether this would also reduce the scaling issue has yet to be demonstrated.

Conventional QMC methods[8–10] solve the electronic problem in real space (first quantization), and treat a large portion of the correlation in the electronic motion explicitly, which greatly reduces the number of required determinants. Standard QMC variants are still practical for systems with hundreds of electrons, such as supramolecular complexes[11] and molecular crystals[12]. Even though only a single determinant can be typically used for such large systems, QMC still outperforms other electronic-structure methods applicable to such systems, such as DFT or the random-phase approximation. However, for small systems, where second-quantized approaches are applicable, standard QMC methods need to use at least hundreds of determinants to be competitive, and this number increases rapidly with N. The large number of determinants is necessary to accurately represent the nodal surface of the wavefunction, which is otherwise difficult to improve with standard real-space QMC methods. A key development enabling the progress in the present work is the real-space backflow technique[13]. The idea of the backflow technique is to transform the electrons into pseudoparticles, the position of each of which depends on the positions of all the electrons, and this many-body mixing then leads to an improved nodal surface[14,15]. While the traditional backflow technique does not reach the accuracy of the large determinant expansions and does not generalize well to larger systems, Luo et al. recently showed that representing the backflow with a neural network is a powerful generalization[16].

Machine learning has had a great impact on quantum chemistry, especially in the case of the supervised learning and prediction of electronic energies[17–24], electron densities[25] and molecular orbitals[26]. This approach entirely avoids the solution of the Schrödinger equation, at the price of requiring datasets of preexisting solutions, obtained for instance by DFT or the coupled-cluster method.

By contrast, the direct representation of correlated wavefunctions with neural networks and their unsupervised training via the

[1]Department of Mathematics and Computer Science, FU Berlin, Berlin, Germany. [2]Machine Learning Group, TU Berlin, Berlin, Germany. [3]Department of Physics, FU Berlin, Berlin, Germany. [4]Department of Chemistry, Rice University, Houston, TX, USA. ✉e-mail: jan.hermann@fu-berlin.de; frank.noe@fu-berlin.de

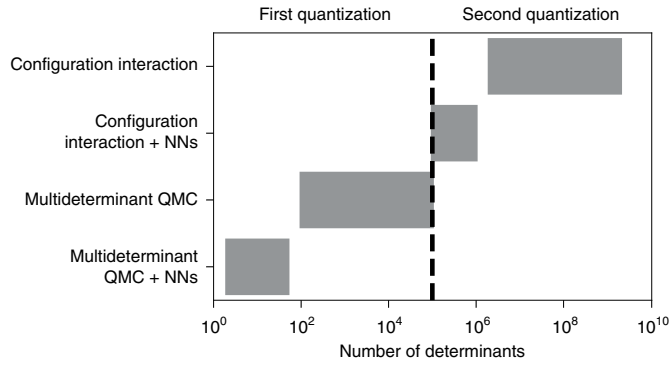

Fig. 1 | Combinatorial explosion of the number of Slater determinants in quantum chemistry. Both configuration-interaction and multideterminant QMC approaches suffer from the rapid scaling of the number of Slater determinants with system size, which in both cases can be reduced with neural networks (NNs). The multideterminant QMC combined with neural networks is the approach developed in this work. The plot shows typical numbers of Slater determinants used by high-accuracy quantum chemistry methods in state-of-the-art calculations on atomic systems with at most a few tens of electrons.

variational principle, first proposed by Carleo and Troyer for discrete spin lattice systems[27], is an ab initio approach that requires no preexisting data and has no fundamental limits to its accuracy. It is motivated by the fact that neural networks are universal function approximators and could therefore provide more efficient means for approximating the exponentially scaling complexity of many-body quantum systems. The initial attempts on lattice systems were later generalized to bosons in real space[28,29] and even electrons in real space[30], but the latter approach does not use a wavefunction ansatz in the form of a Slater determinant, and perhaps for that reason does not reach the accuracy of the baseline Hartree–Fock (HF) method for some systems.

In this work, we develop PauliNet, a deep-learning QMC approach that replaces existing ad hoc functional forms used in the standard Jastrow factor and backflow transformation with more powerful deep neural network (DNN) representations. Besides the sheer gain in expressive power, our neural network architecture is specifically designed to encode the physics of valid wavefunctions and incorporates the multireference HF method as a baseline. These physically motivated choices are essential to obtain a method that not only is highly accurate, but also converges robustly, while maintaining computational efficiency. Using several test systems, we demonstrate that our neural network ansatz substantially outperforms the accuracy of state-of-the-art wavefunction ansatzes using a similar number of determinants. Thanks to the trainable backflow ansatz, high accuracy can be obtained with orders of magnitude fewer determinants compared to traditional QMC methods. Our method has the asymptotic scaling of N^4, and we expect that it will be feasible to apply it to much larger systems than is currently possible with existing high-accuracy methods. We demonstrate this with an accurate calculation of the transition-state energy of the 28-electron cyclobutadiene molecule, which was previously achievable only with highly specialized methods.

The parallel work of Pfau et al. follows the same basic idea as ours, but differs in one important aspect[31]. Their architecture does not encode any physical knowledge about wavefunctions besides the essential antisymmetry, which is compensated by a much larger number of optimized parameters. This difference likely leads to the higher computational cost per iteration. In addition, their architecture is trained substantially longer and as a consequence reaches higher accuracy for some systems.

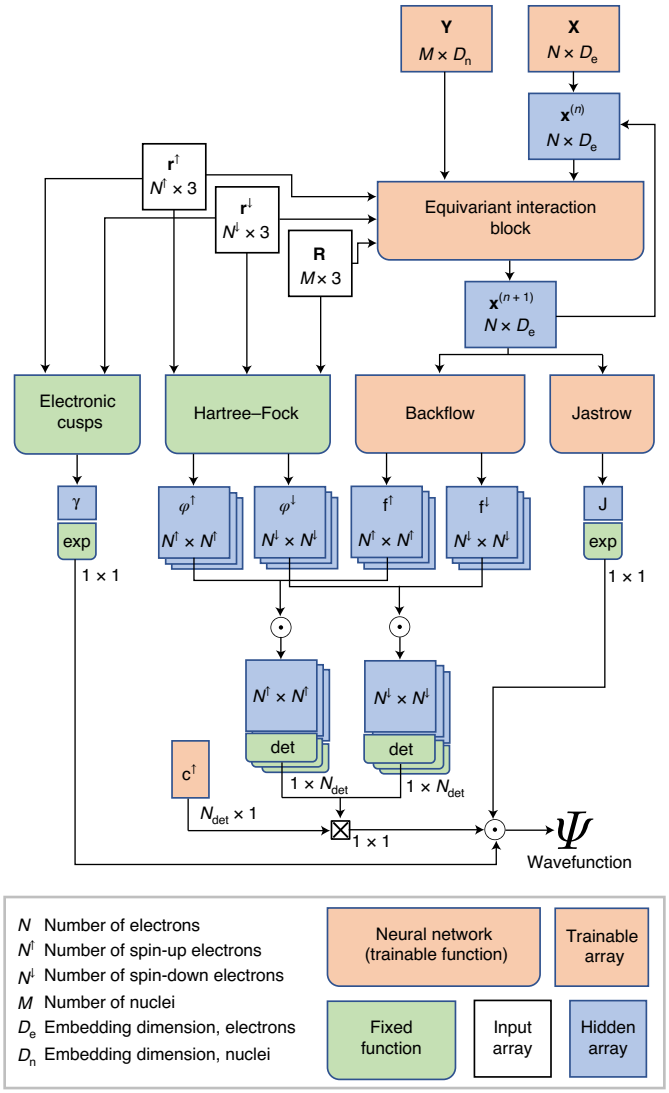

Fig. 2 | Architecture of the newly developed PauliNet wavefunction ansatz. The information flows from the input electron and nuclear coordinates, **r** and **R**, to the output wavefunction value, Ψ. Modelling the wavefunction via Jastrow and backflow functions is common in QMC, but here these functions are learned with DNNs. N_{det}, number of determinants.

Results

DNN electronic wavefunction ansatz. At the core of our deep-learning approach to the electronic Schrödinger equation is a wavefunction ansatz, dubbed PauliNet, which incorporates both the well-established essential physics of electronic wavefunctions—Slater determinants, multideterminant expansion, Jastrow factor, backflow transformation and cusp conditions—as well as DNNs capable of encoding the complex features of the electronic motion in heterogeneous molecular systems. Our proposed trial wavefunction, $\psi_{\boldsymbol{\theta}}(\mathbf{r})$, $\mathbf{r}=(\mathbf{r}_1,\ldots,\mathbf{r}_N)$, is of the multideterminant Slater–Jastrow–backflow type[32], where both the Jastrow factor, J, and the backflow, $\mathbf{f}$, are represented by DNNs with trainable parameters $\boldsymbol{\theta}$ (Fig. 2),

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = e^{\gamma(\mathbf{r})+J_{\boldsymbol{\theta}}(\mathbf{r})} \sum_p c_p \det[\tilde{\varphi}^{\uparrow}_{\boldsymbol{\theta},\mu_p i}(\mathbf{r})] \det[\tilde{\varphi}^{\downarrow}_{\boldsymbol{\theta},\mu_p i}(\mathbf{r})]$$
$$\tilde{\varphi}_{\mu i}(\mathbf{r}) = \varphi_{\mu}(\mathbf{r}_i) f_{\boldsymbol{\theta},\mu i}(\mathbf{r}) \quad (1)$$

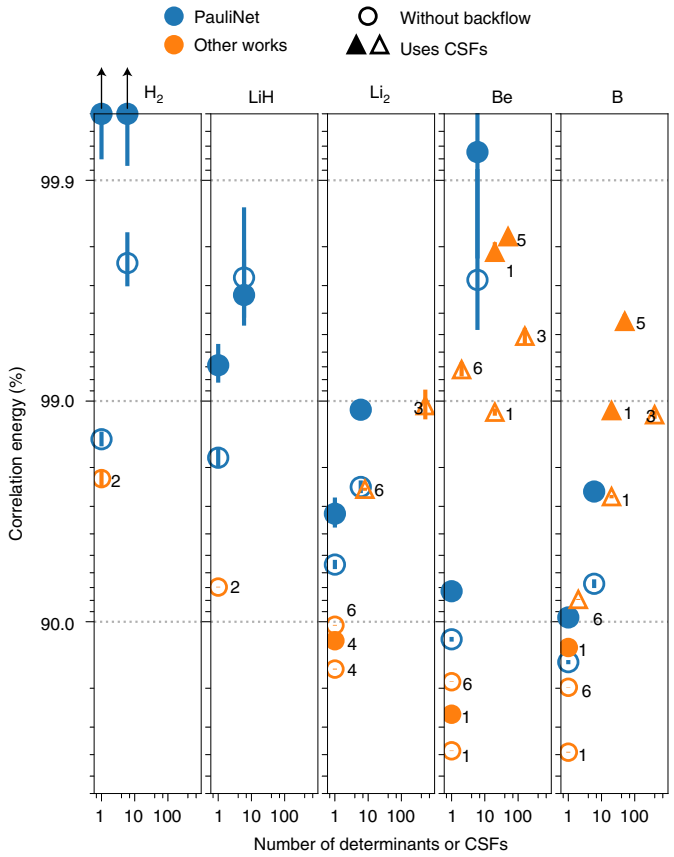

Fig. 3 | Performance of PauliNet with one and six determinants on atoms and diatomic molecules. PauliNet recovers 97% to 99.9% of correlation energy with one to two orders of magnitude less determinants than standard variational ansatzes. Four variants of PauliNet are shown, single- and multideterminant as well as with and without backflow. Data points marked with black arrows are outside the *y*-axis range. The reference results are taken from (1) Brown et al.[32], (2) Casalegno et al.[39], (3) Morales et al.[2], (4) Lopez Ríos et al.[13], (5) Seth et al.[40] and (6) Toulouse and Umrigar[41]. Each CSF corresponds to a few to a few dozen determinants, depending on the system and the particular method. The numerical data can be found in Extended Data Fig. 1.

While the expressiveness of PauliNet is contained in the Jastrow factor and backflow DNNs, the physics is encoded by the determinant form; the one-electron molecular orbitals, φ_μ; and the electronic cusps, γ, in the following way.

Every valid electronic wavefunction must be antisymmetric with respect to the exchange of same-spin electrons,

$$\psi(\dots,\mathbf{r}_i,\dots,\mathbf{r}_j,\dots) = -\psi(\dots,\mathbf{r}_j,\dots,\mathbf{r}_i,\dots) \tag{2}$$

As is common in quantum chemistry, we enforce antisymmetry via matrix determinants, as determinants change sign upon exchanging any two rows or columns.

To ensure a good starting point for the variational optimization problem, we exploit the approximate HF method. Specifically, we use a multireference HF calculation with a small complete active space, and select the most dominant determinants and their orbitals based on the magnitude of their linear coefficient. The HF-optimized one-electron molecular orbitals, $\varphi_\mu(\mathbf{r})$, are then used as an input to PauliNet and are modified during training only by the backflow transformation.

Any ground-state electronic wavefunction obeys exact asymptotic behaviour defined by the cusp conditions as electrons approach each other and the nuclei[33]. We chose to build the cusp conditions directly into the PauliNet functional form as this makes the training more efficient as well as stable by removing divergences from the local electronic energy. We incorporate the nuclear cusps by modifying the molecular orbitals using the technique from Ma et al.[34] and the electronic cusps by the fixed cusp function, $\gamma(\mathbf{r})$. We ensure that the trainable Jastrow factor and backflow DNNs are cuspless, so as to maintain the enforced cusp behaviour (see Methods for details).

Robust deep Jastrow factor and backflow. PauliNet differs from conventional QMC ansatzes by representing the Jastrow factor and backflow functions with specialized DNNs. To retain the antisymmetry of the wavefunction, as enforced by the Slater determinants, the Jastrow factor and backflow DNNs are constructed to be invariant and equivariant, respectively, with respect to the exchange of same-spin electrons, $\mathcal{P}_{ij}$,

$$J(\mathcal{P}_{ij}\mathbf{r}) = J(\mathbf{r}), \quad \mathcal{P}_{ij}f_{\mu i}(\mathbf{r}) = f_{\mu j}(\mathcal{P}_{ij}\mathbf{r}) \tag{3}$$

The Jastrow factor is a nonnegative totally symmetric function, which can encode complex electron correlations into the wavefunction, but cannot modify the nodal surface inherited from the determinant expansion.

We found that attempting to express the standard backflow form of coupled electron coordinates with DNNs leads to a difficult optimization problem. Instead, the PauliNet backflow has the form of multiplying the bare one-electron molecular orbitals with many-electron equivariant functions, $\mathbf{f}$ (equation (1)). In combination with just a few determinants, this presents a powerful representation of the electronic nodal surface.

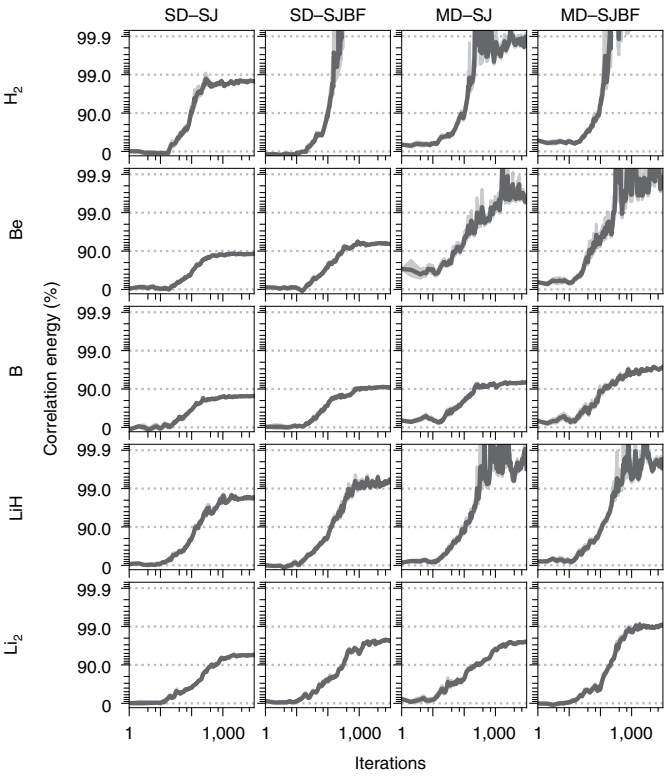

Fig. 4 | Roles of backflow and multiple determinants in training of PauliNet ansatz. The four variants are combinations of a single or multiple Slater determinant(s) (SD or MD), a Jastrow factor (SJ) and an optional backflow (BF). Both the backflow and the use of a few determinants is crucial for reaching high accuracy. Exponential moving average is applied to the energy at each iteration.

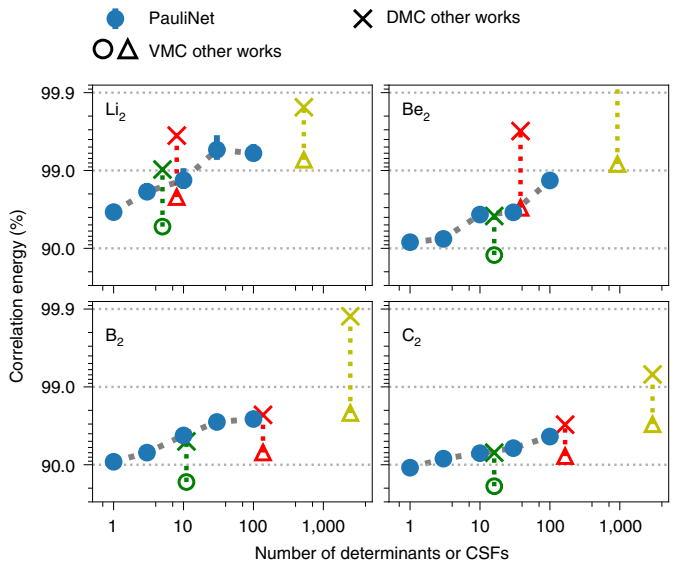

Fig. 5 | Convergence of PauliNet with the number of determinants. The accuracy of PauliNet increases with increasing number of determinants as in standard methods, but reaches the accuracy of DMC already at the variational level. Comparison is shown to VMC (circle/triangle) and corresponding DMC (cross) results taken from Filippi and Umrigar[42] (green), Toulouse and Umrigar[41] (red) and Morales et al.[2] (yellow). Either the number of determinants (circle) or the number of CSFs (triangle) is plotted.

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 to constructing DNNs that learn potential energy functions. PauliNet uses an adapted form of one such DNN architecture, called SchNet[22]. SchNet is a graph DNN that represents each particle with a vector in a high-dimensional abstract feature space, $\mathbf{x}_i$, which is iteratively refined by interactions with other particles through real-space trainable convolutions, χ_θ, which encode the inter-particle distances and are invariant with respect to particle exchange,

$$\mathbf{x}_i^{(n+1)} := \mathbf{x}_i^{(n)} + \chi_\theta^{(n)}\left(\left\{\mathbf{x}_j^{(n)}, \{|\mathbf{r}_j - \mathbf{r}_k|\}\right\}\right) \quad (4)$$

The SchNet architecture and its modifications for PauliNet are described in detail in the Methods. After a fixed number of iterations, the final electron representations, $\mathbf{x}_i^{(L)}$, which now encode complex many-body electron correlations, are used as an input to two trainable functions, η_θ and κ_θ, which return the Jastrow factor and backflow, respectively:

$$J := \eta_\theta\left(\sum_i \mathbf{x}_i^{(L)}\right), \quad \mathbf{f}_i := \kappa_\theta\left(\mathbf{x}_i^{(L)}\right). \quad (5)$$

Since the feature vectors, $\mathbf{x}_i^{(n)}$, are equivariant with respect to electron exchange at each iteration, so are the backflow vectors, $\mathbf{f}_i$. As a result, the Slater determinants in PauliNet produce an antisymmetric wavefunction. Furthermore, J is by construction invariant with respect to exchanges of electrons and therefore a symmetric function that maintains this antisymmetry.

Approaching exact solutions with few determinants. We train PauliNet via the variational principle, minimizing the total electronic energy (variational QMC). The training data are electron configurations that are generated on the fly by sampling the electron distribution, $|\psi|^2$ (see Methods for details). We first investigate the same systems that were used to test DeepWF[30], in particular the hydrogen molecule (H_2), lithium hydride (LiH), beryllium (Be), boron (B) and the linear hydrogen chain H_{10}. For the mono- and diatomic systems, PauliNet recovers between 97% and 99.9% of the electron correlation energy (Fig. 3) after training for tens of minutes to a few hours on a single GTX 1080 Ti graphics processing unit (GPU), in all cases far beyond the accuracy of DeepWF. We compare these results to the standard single-determinant and multideterminant variational Monte Carlo (VMC) methods with and without backflow. In all cases, PauliNet with six determinants is better than all single-determinant and few-determinant ansatzes, and is only surpassed by trial wavefunctions with tens or hundreds of configuration state functions (CSFs), corresponding to hundreds to thousands of determinants.

Figure 4 highlights two crucial aspects of our method. First, the error in the correlation energy decreases monotonously as the training progresses from the initial HF baseline level to the final reported values. The learning curves are not yet fully plateaued in most cases, demonstrating the high expressiveness of our ansatz, and indicating that even higher accuracy could be achieved with more computational resources. Second, we compare our full ansatz to variants using only a single determinant and variants without backflow, and find that both these components are important for refining the nodal surface of the HF baseline and thus reaching high accuracy. The fact that only a few determinants are sufficient to substantially reduce the correlation energy error compared to the single-determinant case indicates that deep learning can be an efficient tool to reduce the large number of determinants sampled

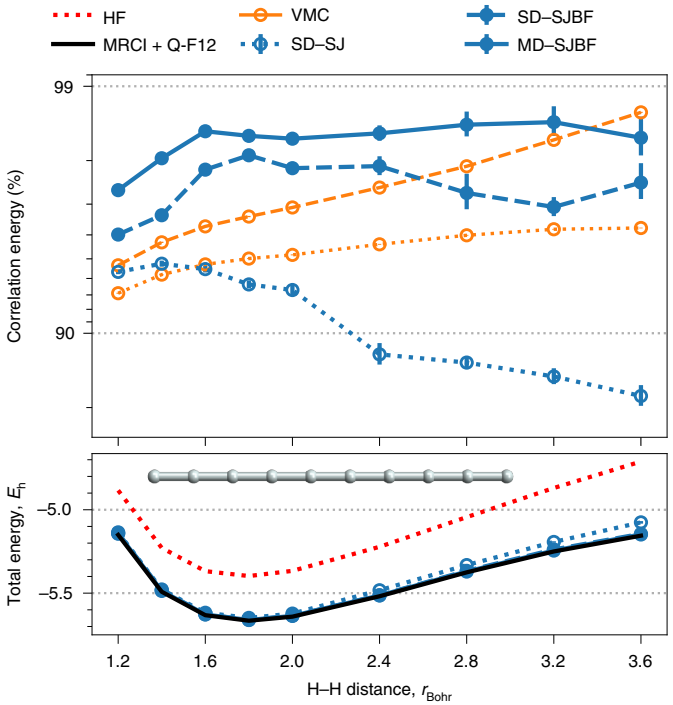

Fig. 6 | PauliNet captures strong correlation in H_{10} along the dissociation curve. PauliNet results (blue) with single determinant or 16 determinants (MD) and with or without backflow are shown. The backflow plays a much larger role than multiple determinants. PauliNet outperforms highly specialized VMC ansatzes (orange) of the single-determinant (dotted) and multideterminant geminal (dashed) form by Motta et al.[35]. The correlation energy is calculated with respect to multireference configuration-interaction (MRCI) results, also by Motta et al. E_h, Hartree energy; r_{Bohr}, Bohr radius.

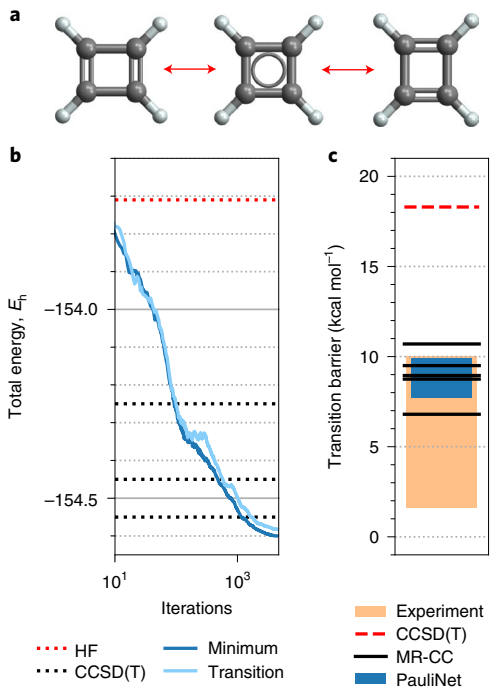

Fig. 7 | Calculating the transition barrier of cyclobutadiene automerization. a, Cyclobutadiene automerization. The transition state has a highly multireferential character. **b**, Convergence of the total energy of the energy minimum and transition state with training. Absolute energies of the energy minimum from HF at the complete basis set limit and from CCSD(T) with the cc-pV*n*Z basis set (n = D, T, Q) are shown. **c**, Energy barrier obtained by sampling the trained PauliNet wavefunctions, compared to references taken from Lyakh et al.[36]. Results from five variants of MR-CC theory are shown (top to bottom): MR-DI-EOMCCSD, RMRCCSD(T), Mk-MRCCSD(T), MRCISD+Q and BW-MRCCSD(T)—note the very small separation between Mk-MRCCSD(T) and MRCISD+Q.

in other VMC approaches that directly operate on determinants of fixed orbitals[5,7]. By having a powerful backflow transformation, each additional determinant substantially increases the flexibility of the ansatz.

We further analyse the scaling of PauliNet with the number of determinants and with system size on a set of four homonuclear diatomic molecules from Li_2 to C_2 (Fig. 5). PauliNet reaches high accuracy quickly with increasing but small numbers of determinants. In the regime of a few to a few dozen determinants, PauliNet surpasses existing variational results and in most cases reaches the accuracy of the corresponding diffusion Monte Carlo (DMC) results. We note that DMC can be implemented in a straightforward manner for our approach and is expected to further increase its accuracy.

Capturing strong correlation. Unlike the atoms and diatomics, the linear hydrogen chain H_{10} exhibits strong correlation, which describes a situation where the single-determinant or even few-determinant description of the HF method is qualitatively insufficient, and the correlation energy constitutes a substantial part of the electronic energy[35]. For H_{10}, we recover 98.41(8)% and 98.4(3)% of the correlation energy in the equilibrium and stretched geometries, respectively, using 16 determinants (Fig. 6). The results are only slightly worse using a single determinant (98.10(9)% and 97.5(4)%), but substantially worse when the trainable backflow is also switched off (93.7(2)% and 82(2)%). Compared to standard VMC ansatzes that were specifically adjusted for this particular application[35], we reach higher accuracy at equilibrium and comparable accuracy at the dissociated limit.

We make three observations based on these results. First, even though the system size and complexity of H_{10} is notably higher compared to the systems in the previous section, we achieve the same level of accuracy while using the same form of the ansatz. Second, there is no decrease in accuracy in the stretched geometry due to the increased strongly correlated character. Third, adding multiple determinants to the ansatz recovers only a fraction of the correlation energy compared to the backflow transformation, highlighting the central role of the trainable backflow in PauliNet.

Straightforward generalization to larger molecules. The previous two sections demonstrated the performance of PauliNet on relatively small benchmark systems, for which essentially exact results are already available from well-established methods. In this section, we show that the same PauliNet ansatz scales in a straightforward manner to larger molecules with complex electronic structures, for which only highly specialized derivatives of standard quantum chemistry methods could deliver satisfactory results as of yet. For this purpose, we chose the automerization of cyclobutadiene (Fig. 7a, 28 electrons), a chemical process that has received considerable attention from both experiment and theory[36]. The experimental estimates of the energy barrier range between 1.6 and 10 kcal mol^{-1}, while the standard coupled-cluster method with up to perturbative triple excitations (CCSD(T)) predicts 18 kcal mol^{-1}, a twofold overestimation. The best computational estimates are available from various flavours of the multireference coupled-cluster (MR-CC) theory and fall between 7 and 11 kcal mol^{-1}, without a decisive answer as to which of the variants is closer to the ground truth.

Using the PauliNet ansatz with ten determinants and the same hyperparameters as used for the much smaller systems, we obtain all-electron variational energies for the energy minimum and transition states of cyclobutadiene, and thus for the energy barrier. Since the energy barrier is only 0.01% of the total energy, we use a modified optimization protocol to stabilize the training of the neural network with respect to the inherent stochasticity, in which ten independent copies are optimized simultaneously and periodically synchronized such that the five copies with higher energies are discarded, and the rest duplicated.

With this modification, the total energies converge smoothly (Fig. 7b), and by running two independent optimizations with the synchronization period of 250 and 375 iterations, we obtain estimates of the energy barrier of 9.9 ± 0.6 and 7.7 ± 0.6 kcal mol^{-1}, respectively. The range larger than the respective statistical sampling errors suggests a remaining degree of stochasticity in the optimization, but nevertheless both results are well within the range spanned by the MR-CC methods. We note that compared to the MR-CC methods, many of which require use of system- and state-specific wavefunction ansatzes, the PauliNet ansatz is constructed in essentially the same way for all the systems studied in this work. The computational cost of the cyclobutadiene optimization is 50 s per iteration on a single GTX 1080 Ti GPU for each of the synchronized optimizations.

Discussion

We have designed PauliNet, a DNN representation of electronic wavefunctions in real space, and shown that it can outperform state-of-the-art variational quantum chemistry methods that do not use large determinant expansions. By contrast, our approach requires only few determinants, and as a result we anticipate that its computational cost scales asymptotically as N^4 (N^3 for a determinant evaluation, and additional N for the evaluation of the kinetic energy), subject to additional technical details[8]. PauliNet is thus a candidate for a quantum chemistry method that can scale to much larger systems with high accuracy.

Compared to standard functional forms used in QMC, the use of DNNs has several advantages. First, the much higher flexibility of DNNs allows a variational approach to reach or exceed the accuracy of DMC, which aids the calculation of accurate derived electronic properties beyond the electronic energy. Second, besides encoding more-complex many-body correlations between electrons, DNNs have an essentially unlimited flexibility in the spatial degrees of freedom, circumventing the curse of incomplete basis sets of quantum chemistry, which can be removed only with DMC when using standard techniques. Third, the rapid oscillations of the local energy close to heavy nuclei in standard QMC mandate the use of pseudopotentials for heavier elements such as transition metals. The flexibility of DNNs could sidestep this necessity by smoothing out such oscillations.

In classical quantum chemistry methods, strong correlation is usually treated by using large multideterminant expansions, which are computationally demanding and introduce the problem of selecting the proper subset of determinants. Treating strong correlation on the level of Jastrow factors traditionally requires construction of specialized many-body forms[37,38]. By contrast, we show that DNNs are capable of learning strong correlation between electrons without any specialized adaptation. Convergence to high accuracy can be achieved with only a few determinants, changing the problem from searching or sampling over exponentially many determinants to letting a DNN search over exponentially many functions. Although it is unclear whether this is advantageous in a strict mathematical sense, this is precisely the task that DNNs have been demonstrated to be strong at in a variety of real-world applications. Complementary approaches that use VMC in a second-quantized form of the electronic problem have also been proposed[7]. This class of methods has the advantage of eliminating much of the complexity of electronic wavefunctions (such as the antisymmetry or cusp conditions) from the machine-learning part of the problem, but needs to cope with the ubiquitous limitations of single-particle basis sets.

Already a brief comparison of our approach with that of Pfau et al. hints at potential improvements of both architectures[31]. The combination of architectural design and optimization methods used in FermiNet with the built-in physical constraints of PauliNet appears to be a promising venue for computationally affordable, scalable, yet highly accurate black-box methods for quantum chemistry. We hope that the introduction of neural networks into the field of electronic QMC opens the possibility to utilize the striking advances in deep learning from the last decade in a new field.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41557-020-0544-y.

Received: 6 December 2019; Accepted: 5 August 2020;
Published online: 23 September 2020

References

1. Piela, L. *Ideas of Quantum Chemistry* 2nd edn (Elsevier, 2014).
2. Morales, M. A., McMinis, J., Clark, B. K., Kim, J. & Scuseria, G. E. Multideterminant wave functions in quantum Monte Carlo. *J. Chem. Theory Comput.* **8**, 2181–2188 (2012).
3. Troyer, M. & Wiese, U.-J. Computational complexity and fundamental limitations to fermionic quantum Monte Carlo simulations. *Phys. Rev. Lett.* **94**, 170201 (2005).
4. Shavitt, I. & Bartlett, R. J. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory* (Cambridge Univ. Press, 2009).
5. Booth, G. H., Thom, A. J. W. & Alavi, A. Fermion Monte Carlo without fixed nodes: a game of life, death, and annihilation in Slater determinant space. *J. Chem. Phys.* **131**, 054106 (2009).
6. Thom, A. J. W. Stochastic coupled cluster theory. *Phys. Rev. Lett.* **105**, 263004 (2010).
7. Choo, K., Mezzacapo, A. & Carleo, G. Fermionic neural-network states for ab-initio electronic structure. *Nat. Commun.* **11**, 2368 (2020).
8. Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **73**, 33–83 (2001).
9. Needs, R. J., Towler, M. D., Drummond, N. D. & Ríos, P. L. Continuum variational and diffusion quantum Monte Carlo calculations. *J. Phys. Condens. Matter* **22**, 023201 (2010).
10. Austin, B. M., Zubarev, D. Y. & Lester, W. A. Quantum Monte Carlo and related approaches. *Chem. Rev.* **112**, 263–288 (2012).
11. Ambrosetti, A., Alfè, D., DiStasio, R. A. Jr & Tkatchenko, A. Hard numbers for large molecules: toward exact energetics for supramolecular systems. *J. Phys. Chem. Lett.* **5**, 849–855 (2014).
12. Zen, A. et al. Fast and accurate quantum Monte Carlo for molecular crystals. *Proc. Natl Acad. Sci. USA* **115**, 1724–1729 (2018).
13. López Ríos, P., Ma, A., Drummond, N. D., Towler, M. D. & Needs, R. J. Inhomogeneous backflow transformations in quantum Monte Carlo calculations. *Phys. Rev. E* **74**, 066701 (2006).
14. Feynman, R. P. & Cohen, M. Energy spectrum of the excitations in liquid helium. *Phys. Rev.* **102**, 1189–1204 (1956).
15. Schmidt, K. E. & Pandharipande, V. R. New variational wave function for liquid ^{3}He. *Phys. Rev. B* **19**, 2504–2519 (1979).
16. Luo, D. & Clark, B. K. Backflow transformations via neural networks for quantum many-body wave functions. *Phys. Rev. Lett.* **122**, 226401 (2019).
17. Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **98**, 146401 (2007).
18. Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
19. Chmiela, S. et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
20. Bartók, A. P. et al. Machine learning unifies the modeling of materials and molecules. *Sci. Adv.* **3**, e1701816 (2017).
21. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
22. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet — a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).
23. Faber, F. A., Christensen, A. S., Huang, B. & von Lilienfeld, O. A. Alchemical and structural distribution based representation for universal quantum machine learning. *J. Chem. Phys.* **148**, 241717 (2018).
24. Welborn, M., Cheng, L. & Miller, T. F. Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **14**, 4772–4779 (2018).
25. Grisafi, A. et al. Transferable machine-learning model of the electron density. *ACS Cent. Sci.* **5**, 57–64 (2019).
26. Schütt, K. T., Gastegger, M., Tkatchenko, A., Müller, K.-R. & Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
27. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
28. Saito, H. Method to solve quantum few-body problems with artificial neural networks. *J. Phys. Soc. Jpn* **87**, 074002 (2018).
29. Ruggeri, M., Moroni, S. & Holzmann, M. Nonlinear network description for many-body quantum systems in continuous space. *Phys. Rev. Lett.* **120**, 205302 (2018).
30. Han, J., Zhang, L. & E, W. Solving many-electron Schrödinger equation using deep neural networks. *J. Comput. Phys.* **399**, 108929 (2019).
31. Pfau, D., Spencer, J. S., Matthews, A. G. d. G. & Foulkes, W. M. C. Ab-initio solution of the many-electron Schrödinger equation with deep neural networks. Preprint at http://arxiv.org/abs/1909.02487 (2019).
32. Brown, M. D., Trail, J. R., López Ríos, P. & Needs, R. J. Energies of the first row atoms from quantum Monte Carlo. *J. Chem. Phys.* **126**, 224110 (2007).
33. Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **10**, 151–177 (1957).
34. Ma, A., Towler, M. D., Drummond, N. D. & Needs, R. J. Scheme for adding electron-nucleus cusps to Gaussian orbitals. *J. Chem. Phys.* **122**, 224322 (2005).
35. Motta, M. et al. 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**, 031059 (2017).
36. Lyakh, D. I., Musiał, M., Lotrich, V. F. & Bartlett, R. J. Multireference nature of chemistry: the coupled-cluster view. *Chem. Rev.* **112**, 182–243 (2012).

37. Neuscamman, E., Umrigar, C. J. & Chan, G. K.-L. Optimizing large parameter sets in variational quantum Monte Carlo. *Phys. Rev. B* **85**, 045103 (2012).
38. Gasperich, K., Deible, M. & Jordan, K. D. H_4: a model system for assessing the performance of diffusion Monte Carlo calculations using a single Slater determinant trial function. *J. Chem. Phys.* **147**, 074106 (2017).
39. Casalegno, M., Mella, M. & Rappe, A. M. Computing accurate forces in quantum Monte Carlo using Pulay's corrections and energy minimization. *J. Chem. Phys.* **118**, 7193 (2003).
40. Seth, P., Ríos, P. L. & Needs, R. J. Quantum Monte Carlo study of the first-row atoms and ions. *J. Chem. Phys.* **134**, 084105 (2011).
41. Toulouse, J. & Umrigar, C. J. Full optimization of Jastrow–Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules. *J. Chem. Phys.* **128**, 174101 (2008).
42. Filippi, C. & Umrigar, C. J. Multiconfiguration wave functions for quantum Monte Carlo calculations of first-row diatomic molecules. *J. Chem. Phys.* **105**, 213–226 (1996).

Methods

Ansatz optimization. We optimize the PauliNet ansatz individually for each atomic system in an unsupervised fashion using the variational principle for the total electronic energy,

$$E_0 = \min_{\psi} E[\psi] \leq \min_{\boldsymbol{\theta}} E[\psi_{\boldsymbol{\theta}}], \qquad E[\psi] = \int \mathrm{d}\mathbf{r}\, \psi(\mathbf{r}) \hat{H} \psi(\mathbf{r}) \tag{6}$$

Following the standard QMC technique, the energy integral is evaluated as an expected value of the local energy, $E_{\mathrm{loc}}[\psi](\mathbf{r}) = \hat{H}\psi(\mathbf{r})/\psi(\mathbf{r})$, over the probability distribution $|\psi^2(\mathbf{r})|$,

$$E[\psi] = \mathbb{E}_{\mathbf{r}\sim|\psi|^2}[E_{\mathrm{loc}}[\psi](\mathbf{r})] \tag{7}$$

We generate training data for PauliNet on the fly by periodically alternating training on one hand and sampling electron positions with a standard Langevin Monte Carlo approach on the other[43]. Each sampled electron configuration is used only once in an optimization run. We use a simplified version of the method by Umrigar et al.[43], in which the radial step proposal is replaced with clipping the step length such that the step size is always shorter than the distance to the nearest nucleus, so the nucleus can never be 'overshot'. The initial electron positions for the Markov chain are sampled from Gaussian distributions around the nuclei such that the effective atomic Mulliken charges obtained from the HF method are respected.

To optimize the parameters $\boldsymbol{\theta}$ in the Jastrow and backflow neural networks, we use the weighted Adam optimizer[44,45] together with the total energy used directly as the loss function. To calculate the stochastic gradient of the loss function over a batch of samples, we use a gradient formula that takes advantage of the fact that the Hamiltonian operator is Hermitian[46],

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{r}\sim|\psi'^2|}[E_{\mathrm{loc}}[\psi_{\boldsymbol{\theta}}](\mathbf{r})]$$
$$\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}) = 2\mathbb{E}_{\mathbf{r}\sim|\psi'^2|}[(E_{\mathrm{loc}}[\psi_{\boldsymbol{\theta}}](\mathbf{r}) - \mathcal{L}(\boldsymbol{\theta}))\nabla_{\boldsymbol{\theta}}\ln|\psi_{\boldsymbol{\theta}}|] \tag{8}$$

This expression for the gradient requires calculating only second derivatives of the wavefunction (for the Laplace operator), whereas direct differentiation would require third derivatives (derivative of the Laplace operator). We smoothly clip the local energy of each sample by a logarithmically growing clipping function outside the window defined as five times the mean deviance from the median local energy in a given batch. The learning rate is controlled by a cyclic scheduling policy[47].

Cusp conditions. Equation (1) ensures the nuclear cusp conditions via the molecular orbitals $\varphi_\mu(\mathbf{r}_i)$. We achieve this by modifying the molecular orbitals using the technique from Ma et al.[34] with one simplification—we optimize the orbital values at atomic nuclei, $\mathbf{r}_i = \mathbf{R}_I$, via the energy variational principle, rather then fitting them against references values. The electronic cusp conditions are enforced by $\gamma(\mathbf{r})$,

$$\gamma(\mathbf{r}) := \sum_{i<j} -\frac{c_{ij}}{1+|\mathbf{r}_i - \mathbf{r}_j|}, \tag{9}$$

where c_{ij} is either $\frac{1}{2}$ or $\frac{1}{4}$ depending on the spins of the two electrons. To preserve the cusp conditions built into φ_μ and γ, the Jastrow factor and backflow DNNs must be cuspless,

$$\nabla_{\mathbf{r}_i} J(\mathbf{r})|_{\mathbf{r}_i=\{\mathbf{r}_k,\mathbf{R}_I\}} = 0, \quad \nabla_{\mathbf{r}_i} f_{\mu i}(\mathbf{r})\Big|_{\mathbf{r}_i=\{\mathbf{r}_k,\mathbf{R}_I\}} = 0 \tag{10}$$

These conditions are ensured by constructing the DNNs appropriately, as detailed below.

PauliNet extension of SchNet. SchNet is an instance of the class of graph convolutional neural networks, and was designed to model the molecular energy as a function of just the nuclear charges and coordinates[22]. In PauliNet, we use SchNet to represent electrons in molecular environments by implementing the iteration rule in equation (4),

$$\mathbf{z}_i^{(n,\pm)} := \textstyle\sum_{j\neq i}^{\pm} \mathbf{w}_{\boldsymbol{\theta}}^{(n,\pm)}\left(\mathbf{e}(|\mathbf{r}_i - \mathbf{r}_j|)\right) \odot \mathbf{h}_{\boldsymbol{\theta}}^{(n)}\left(\mathbf{x}_j^{(n)}\right)$$
$$\mathbf{z}_i^{(n,\mathrm{n})} := \textstyle\sum_I \mathbf{w}_{\boldsymbol{\theta}}^{(n,\mathrm{n})}\left(\mathbf{e}(|\mathbf{r}_i - \mathbf{R}_I|)\right) \odot \mathbf{Y}_{\boldsymbol{\theta},I}$$
$$\mathbf{x}_i^{(n+1)} := \mathbf{x}_i^{(n)} + \textstyle\sum_{\pm} \mathbf{g}_{\boldsymbol{\theta}}^{(n,\pm)}\left(\mathbf{z}_i^{(n,\pm)}\right) + \mathbf{g}_{\boldsymbol{\theta}}^{(n,\mathrm{n})}\left(\mathbf{z}_i^{(n,\mathrm{n})}\right) \tag{11}$$

where '$\odot$' denotes element-wise multiplication; $\mathbf{w}_{\boldsymbol{\theta}}^{(n)}$, $\mathbf{h}_{\boldsymbol{\theta}}^{(n)}$ and $\mathbf{g}_{\boldsymbol{\theta}}^{(n)}$ are trainable functions represented by ordinary fully connected DNNs; and $\mathbf{e}$ is a radial basis function that featurizes the interatomic distances. The modifications of the original SchNet are as follows.

1. Since the wavefunction is a function of electron coordinates, the iterated feature vectors $\mathbf{x}_i^{(n)}$ represent electrons, not atoms.
2. The messages $\mathbf{z}_{i(n)}$ received by the electron feature vectors at each iteration are split into three channels, corresponding to same-spin electrons (+), opposite-spin electrons (−) and the nuclei (n). This builds more flexibility into the architecture, and is motivated by the fact that electrons and nuclei are particles of an entirely different type.
3. Each channel has a separate receiving function $\mathbf{g}_{\boldsymbol{\theta}}$, again increasing flexibility without substantially increasing the number of parameters.
4. Each nucleus is represented by a trainable embedding $\mathbf{Y}_{\boldsymbol{\theta},I}$, which is shared across all iterations and not iteratively updated. In VMC, the wavefunction is always optimized for a given fixed geometry of the nuclei, so the nuclear embeddings can be assumed to already represent each nucleus with its (fixed) atomic environment, hence the absence of need for their iterative refinement.
5. The distance features $\mathbf{e}$ are constructed to be cuspless, as detailed below. We use a distance featurization inspired by the PhysNet architecture[48], with a modified envelope that forces all the Gaussian features and their derivatives to zero at zero distance,

$$e_k(r) := r^2 \mathrm{e}^{-r-(r-\mu_k)^2/\sigma_k^2}$$
$$\mu_k := r_c q_k^2, \qquad \sigma_k := \tfrac{1}{7}(1 + r_c q_k) \tag{12}$$

where q_k equidistantly spans the interval (0, 1) and r_c is a cutoff parameter.

Computational details. All reported methods were implemented in Pytorch[49]. The linear coefficients of the HF orbitals φ_μ as well as of the determinants in a multideterminant expansion were calculated with PySCF[50] using the 6-311G basis set. The plain fully connected DNNs that represent the trainable functions in our architecture were chosen such that the total number of trainable parameters is around 7×10^4 (see Extended Data Fig. 2).

Data availability

All raw data were generated with the accompanying code and are available in Figshare (https://doi.org/10.6084/m9.figshare.12720569.v2)[51]. Processed data used to generate figures are both included with the code and provided with this paper as source data.

Code availability

All computer code developed in this work is released either in the general DeepQMC package available on Zenodo (https://doi.org/10.5281/zenodo.3960827)[52] and developed on Github (https://github.com/deepqmc/deepqmc), or in the project-specific repository (https://doi.org/10.6084/m9.figshare.12720833.v1)[53], both under the MIT license.

References

43. Umrigar, C. J., Nightingale, M. P. & Runge, K. J. A diffusion Monte Carlo algorithm with very small time-step errors. *J. Chem. Phys.* **99**, 2865–2890 (1993).
44. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *3rd International Conference on Learning Representations* (ICLR, 2015); https://dblp.org/rec/journals/corr/KingmaB14.html
45. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. In *7th International Conference on Learning Representations* (ICLR, 2019); https://openreview.net/forum?id=Bkg6RiCqY7
46. Ceperley, D., Chester, G. V. & Kalos, M. H. Monte Carlo simulation of a many-fermion study. *Phys. Rev. B* **16**, 3081–3099 (1977).
47. Smith, L. N. Cyclical learning rates for training neural networks. In *Winter Conference on Applications of Computer Vision* 464–472 (IEEE, 2017); https://ieeexplore.ieee.org/document/7926641
48. Unke, O. T. & Meuwly, M. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
49. Paszke, A. et al. In *Advances in Neural Information Processing Systems* 8026–8037 (Curran Associates, 2019); http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library
50. Sun, Q. et al. PySCF: the Python-based simulations of chemistry framework. *WIREs Comput. Mol. Sci.* **8**, e1340 (2018).
51. Hermann, J., Schätzle, Z. & Noé, F. *Raw data for "Deep neural network solution of the electronic Schrödinger equation"*. Figshare https://doi.org/10.6084/m9.figshare.12720569.v2 (2020).
52. Hermann, J., Schätzle, Z. & Noé, F. *Deepqmc 0.1.1*. Zenodo https://doi.org/10.5281/zenodo.3960827 (2020).
53. Hermann, J., Schätzle, Z. & Noé, F. *Code for "Deep neural network solution of the electronic Schrödinger equation"*. Figshare https://doi.org/10.6084/m9.figshare.12720833.v1 (2020).

Acknowledgements

We thank C. Clementi (Rice, FU Berlin), J. Eisert (FU Berlin), G. Scuseria (Rice), J. Neugebauer (MPIE) and H. Wu (Tongji) for inspiring discussions. Funding is

acknowledged from the European Commission (ERC CoG 772230 'Scale-Cell'), Deutsche Forschungsgemeinschaft (CRC1114/A04, GRK2433 DAEDALUS/P04) and the MATH+ Berlin Mathematics Research Center (AA1x6, EF1x2). J.H. thanks K.-R. Müller for support and acknowledges funding from TU Berlin.

Author contributions

J.H. and F.N. designed the research. J.H. developed the method with contributions from F.N. and Z.S.; J.H. wrote the computer code with contributions from Z.S.; J.H. and Z.S. carried out the numerical calculations. All authors analysed the data. J.H. and F.N. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41557-020-0544-y.

Supplementary information is available for this paper at https://doi.org/10.1038/s41557-020-0544-y.

Correspondence and requests for materials should be addressed to J.H. or F.N.

Reprints and permissions information is available at www.nature.com/reprints.

		Slater–Jastrow			Slater–Jastrow with backflow		
	ref.	SD*	MD†		SD	MD	
H_2	PauliNet	98.5(1)	99.8(1)	6 D	100.02(3)	99.98(3)	6 D
	Casalegno et al.	97.8(2)	–		–	–	
Be	PauliNet	88.0(3)	99.7(2)	6 D	92.7(4)	99.9(1)	6 D
	Toulouse and Umrigar	81.31(5)	99.28(5)	2 CSF	–	–	
	Brown et al.	61.6(1)	98.88(4)	20 CSF	73.8(1)	99.79(2)	20 CSF
	Morales et al.	–	99.49(4)	160 CSF	–	–	
	Seth et al.	–	–		–	99.82(1)	50 CSF
B	PauliNet	84.7(3)	93.3(3)	6 D	90.5(4)	97.4(2)	6 D
	Toulouse and Umrigar	80.19(4)	92.08(4)	2 CSF	–	–	
	Brown et al.	61.0(2)	97.29(5)	20 CSF	86.92(8)	98.89(3)	20 CSF
	Morales et al.	–	98.85(5)	396 CSF	–	–	
	Seth et al.	–	–		–	99.56(3)	50 CSF
LiH	PauliNet	98.1(2)	99.7(1)	6 D	99.3(1)	99.7(1)	6 D
	Casalegno et al.	93.04(2)	–		–	–	
Li_2	PauliNet	94.5(2)	97.6(2)	6 D	96.8(5)	98.9(1)	6 D
	Toulouse and Umrigar	89.63(4)	97.49(4)	8 CSF	–	–	
	Morales et al.	–	99.0(2)	526 CSF	–	–	
	López Ríos et al.	83.62(8)	–		87.81(8)	–	

*Single-determinant ansatz. †Multi-determinant ansatz. "D" denotes a determinant, "CSF" denotes a configuration state function.

Extended Data Fig. 1 | Variational correlation energy (%) of five test systems obtained with four types of trial wave functions.

Hyperparameter	Value
One-electron basis	6-311G
Dimension of $\mathbf{e}$	32
Dimension of $\mathbf{x}_i$	128
Dimension of $\mathbf{z}_i$	64
Number of interaction layers L	3
Number of layers in η_θ	3
Number of layers in κ_θ	3
Number of layers in $\mathbf{w}_\theta$	2
Number of layers in $\mathbf{h}_\theta$	1
Number of layers in $\mathbf{g}_\theta$	1
Batch size	10 000
Number of walkers	2000
Number of training steps	10 000
Optimizer	AdamW
Minimum/maximum learning rate	0.0001/0.01
Cyclic frequency	500
Clipping window q	5
Resampling frequency	10
Number of decorrelation sampling steps	10
Target acceptance	57%

Extended Data Fig. 2 | Hyperparameters used in numerical calculations.