Deep-neural-network

Tingyu Zhang[1]

[1]Department of Physics, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

1   The Slater Jastrow Wave Function

The Slater-Jastrow wave function can be written as

$$\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) = \exp\left[J(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})\right] D(\{\mathbf{r}_i\}) \tag{1}$$

where $\{\mathbf{r}_i\}$ and $\{\mathbf{R}_I\}$ denote the electron and nucleus coordinates, respectively. $\exp[J]$ is the Jastrow factor, and $D(\{\mathbf{r}_i\})$ denotes the Slater part, which only depends on the $\{\mathbf{r}_i\}$.

The Jastrow factor can be written as the sum of homogeneous, isotropic electron–electron terms u, isotropic electron–nucleus terms χ centered on the nuclei, isotropic electron–electron-nucleus terms f, also centered on the nuclei and, in periodic systems, plane-wave expansions of electron–electron separation and electron position, p and q. The form is

$$
\begin{aligned}
J\left(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}\right) = {} & \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} u\left(r_{ij}\right) + \sum_{I=1}^{M}\sum_{i=1}^{N}\chi_I\left(r_{iI}\right) + \sum_{I=1}^{M}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} f_I\left(r_{iI}, r_{jI}, r_{ij}\right) \\
& + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N} p\left(\mathbf{r}_{ij}\right) + \sum_{i=1}^{N} q\left(\mathbf{r}_i\right)
\end{aligned}
\tag{2}
$$

where N is the number of electrons and M is the number of nuclei, $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{r}_{iI} = \mathbf{r}_i - \mathbf{r}_I$.

2   The Cusp Conditions

When an electron approaches another electron or a nucleus of charge Z the potential energy contribution to the local energy E_L diverges as Z/r, where r is the distance of two particles. The kinetic energy operator acting on the cusps in the wave function must therefore supply an equal and opposite divergence in the local kinetic energy.

The antiparallel-spin electron–electron cusp condition

Consider the situation where two electrons of opposite spin, i and j, approach one another and the wave functioncis non-zero at the two-particle coalescence point. The cusp condition is

$$\left(\frac{\partial \hat{\Psi}}{\partial r_{ij}}\right)_{r_{ij}=0} = \frac{1}{2}\hat{\Psi}_{r_{ij}=0} \tag{3}$$

where we write the wave function in terms of center-of-mass and difference coordinates of electrons i and j, $\bar{\mathbf{r}}_{ij} = (\mathbf{r}_i + \mathbf{r}_j)/2$ and $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\hat{\Psi}(\bar{\mathbf{r}}_{ij}, r_{ij})$ is the spherical average of $\hat{\Psi}(\bar{\mathbf{r}}_{ij}, \mathbf{r}_{ij})$ about the coalescence point. Neglecting the cuspless p and q terms, the Slater–Jastrow wave function may be written as

$$\Psi\left(\bar{\mathbf{r}}_{ij}, \mathbf{r}_{ij}\right) = \exp\left[J\left(r_i, r_j, r_{ij}\right)\right] D\left(\bar{\mathbf{r}}_{ij}, \mathbf{r}_{ij}\right) \tag{4}$$

to simplify, we assume there is only one nucleus located at the origin.

$$\begin{aligned}
\delta\Psi = &\Psi_{r_{ij}=0} \times \left(\left[\left(\frac{\partial J}{\partial r_i}\right) - \left(\frac{\partial J}{\partial r_j}\right)\right]_{r_{ij}=0} \delta r_i + \left(\frac{\partial J}{\partial r_{ij}}\right)_{r_{ij}=0} r_{ij}\right) \\
&+ \exp\left[J_{r_{ij}=0}\right] \left(\nabla_{ij}D\right)_{r_{ij}=0} \cdot \mathbf{r}_{ij} + \mathcal{O}\left(r_{ij}^2\right)
\end{aligned} \tag{5}$$

where $\delta\mathbf{r}_i$ and $\delta\mathbf{r}_j$ are the changes in $\mathbf{r}_i$ and $\mathbf{r}_j$ and $\delta\mathbf{r}_i = -\delta\mathbf{r}_j$. If the spherical average about the coalescence point is taken then the terms involving δr_i and $\mathbf{r}_{ij}$ vanish to $\mathcal{O}(r_{ij})$, so that

$$\delta\hat{\Psi} = \hat{\Psi}_{r_{ij}=0} \left(\frac{\partial J}{\partial r_{ij}}\right)_{r_{ij}=0} r_{ij} + \mathcal{O}\left(r_{ij}^2\right) \tag{6}$$

Also notice that

$$\begin{aligned}
\delta\hat{\Psi} &= \left(\frac{\partial \hat{\Psi}}{\partial r_{ij}}\right)_{r_{ij}=0} r_{ij} = \frac{1}{2}\hat{\Psi}_{r_{ij}=0} r_{ij} \\
&= \hat{\Psi}_{r_{ij}=0} \left(\frac{\partial J}{\partial r_{ij}}\right)_{r_{ij}=0} r_{ij} + \mathcal{O}\left(r_{ij}^2\right)
\end{aligned} \tag{7}$$

Therefore the antiparallel cusp condition is equivalent to

$$\left(\frac{\partial J}{\partial r_{ij}}\right)_{\substack{r_{ij}=0 \\ r_i=r_j}} = \frac{1}{2} \tag{8}$$

The parallel-spin electron–electron cusp condition

The cusp condition for electrons i and j with parallel spins is

$$\left(\frac{\partial \Psi_{1m}}{\partial r_{ij}}\right)_{r_{ij}=0} = \frac{1}{4}\left(\Psi_{1m}\right)_{r_{ij}=0}. \tag{9}$$

where Ψ_{1m} is the $r_{ij}Y_{1m}$ component of Ψ, and Y_{1m} is the $(1, m)$th spherical harmonic. This cusp condition is equivalent to the requirement that

$$\left(\frac{\partial J}{\partial r_{ij}}\right)_{\substack{r_{ij}=0 \\ r_i=r_j}} = \frac{1}{4} \tag{10}$$

The electron-nucleus cusp condition

The cusp condition for a electron i approaching a nucleus of atomic number Z is

$$\left(\frac{\partial \bar{\Psi}}{\partial r_i}\right)_{r_i=0} = -Z\bar{\Psi}_{r_i=0} \tag{11}$$

where $\bar{\Psi}(r_i)$ is The spherical average of $\Psi(r_i)$ about the nucleus. if the Slater determinant is continuously differentiable at the nucleus then the Jastrow factor must satisfy

$$\left(\frac{\partial J}{\partial r_{ij}}\right)_{\substack{r_i=0 \\ r_{ij}=r_j}} = -Z. \tag{12}$$

If the Slater part of the wave function satisfies the electron–nucleus cusp condition, or if a non-divergent pseudopotential is used, then the Jastrow factor is required to be cuspless at the nuclei.

3 Backflow Transformation

The Slater-Jastrow wave function is written as Eq.(1). After backflow transformation, the wave function becomes

$$\Psi^{\mathrm{BF}}(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) = \exp\left[J(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})\right] D(\{\mathbf{X}\}) \tag{13}$$

where $\mathbf{X} = \{\mathbf{x}_i\}$ is a set of *collective coordinates*,

$$\mathbf{x}_i = \mathbf{r}_i + \boldsymbol{\xi}_i(\mathbf{r}, \mathbf{R}), \tag{14}$$

where $\boldsymbol{\xi}_i$ is the backflow displacement of electron i, which depends on the configuration of the whole system. The electron–electron contribution to the backflow displacement $\boldsymbol{\xi}_i$ is

$$\boldsymbol{\xi}_i^{\mathrm{ee}} = \sum_{j \neq i}^{N} \eta_{ij}\mathbf{r}_{ij} \tag{15}$$

where N is the number of electrons and $\eta_{ij} = \eta(r_{ij})$ is a function of the interparticle distance r_{ij}. The electron–nucleus contribution to the backflow displacement $\boldsymbol{\xi}_i$ is

$$\boldsymbol{\xi}_i^{\mathrm{en}} = \sum_{I}^{M} \mu_{iI}\mathbf{r}_{iI} \tag{16}$$

3

where $\mu_{iI} = \mu(r_{iI})$ and M is the number of nuclei.

Also, there is an electron-electron-nucleus term to describe two-electron backflow displacements in the presence of a nearby nucleus,

$$\boldsymbol{\xi}_i^{\text{een}} = \sum_{j \neq i}^{N} \sum_{I}^{M} \left(\Phi_i^{jI} \mathbf{r}_{ij} + \Theta_i^{jI} \mathbf{r}_{iI} \right) \tag{17}$$

where $\Phi_i^{jI} = \Phi^I(r_{iI}, r_{jI}, r_{ij})$ and $\Theta_i^{jI} = \Theta^I(r_{iI}, r_{jI}, r_{ij})$. Since the vector $\Phi_i^{jI} \mathbf{r}_{ij} + \Theta_i^{jI} \mathbf{r}_{iI}$ is able to span the plane defined by $\mathbf{r}_i$, $\mathbf{r}_j$ and $\mathbf{r}_I$, there is no need to introduce a component along the direction of $\mathbf{r}_{jI}$. The total backflow displacement is the sum of these three components,

$$\boldsymbol{\xi}_i = \boldsymbol{\xi}_i^{\text{ee}} + \boldsymbol{\xi}_i^{\text{en}} + \boldsymbol{\xi}_i^{\text{een}} \tag{18}$$

4   DNN electronic wavefunction ansatz

In this work, the PauliNet, a deep learning QMC approach is developed by replacing existing ad hoc functional forms used in the standard Jastrow factor and backflow transformation with more powerful deep neural network (DNN) representations. Using this approach, for systems with a few tens of electrons, we can obtain high accuracy with orders of magnitude fewer determinants compared to traditional QMC methods. (Fig.1)
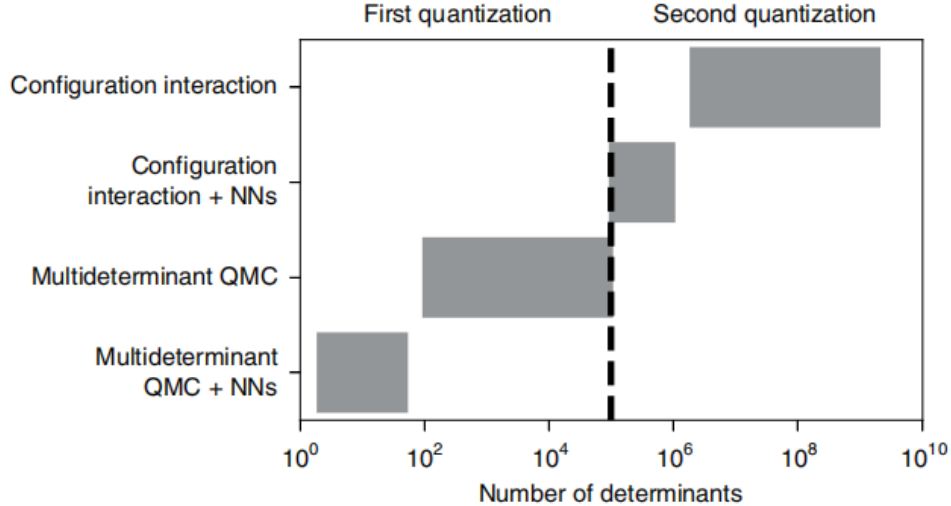

Fig. 1: Combinatorial explosion of the number of Slater determinants in quantum chemistry.This figure shows typical numbers of Slater determinants used by high-accuracy quantum chemistry methods in state-of-the-art calculations on atomic systems with at most a few tens of electrons.

At the core of the deep-learning approach to the electronic Schrödinger equation is a wavefunction ansatz, named PauliNet, which incorporates both the well-established essential physics of electronic wavefunctions, which contains 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. Our proposed trial wavefunction is of the multideterminant Slater–Jastrow–backflow type:

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

$$\tilde{\varphi}_{\mu i}(\mathbf{r}) = \varphi_\mu(\mathbf{r}_i) f_{\boldsymbol{\theta},\mu i}(\mathbf{r}) \tag{19}$$

where $\mathbf{r} = (\mathbf{r}_1 \cdots \mathbf{r}_N)$ and both the Jastrow factor J and backflow function $\mathbf{f}$ are represented by DNNs with trainable parameters $\boldsymbol{\theta}$.(Fig.2)

The electronic cusp conditions are enforced by $\gamma(\mathbf{r})$ to satisfy the cusp condition (Eq.3 and Eq.9),

$$\gamma(\mathbf{r}) := \sum_{i<j} -\frac{c_{ij}}{1 + |\mathbf{r}_i - \mathbf{r}_j|} \tag{20}$$

where c_{ij} is $\frac{1}{2}$ for electrons with opposite spins and $\frac{1}{4}$ for electrons with the same spin. Besides, the nuclear cusp conditions are built in $\varphi_\mu(\mathbf{r}_i)$. To preserve the cusp conditions built into φ_μ and γ, the Jastrow factor and backflow DNNs must be cuspless,

$$\nabla_{\mathbf{r}_i} J(\mathbf{r})|_{\mathbf{r}_i=\{\mathbf{r}_k,\mathbf{R}_I\}} = 0, \quad \nabla_{\mathbf{r}_i} f_{\mu i}(\mathbf{r})|_{\mathbf{r}_i=\{\mathbf{r}_k,\mathbf{R}_t\}} = 0 \tag{21}$$

PauliNet uses an adapted form of one such DNN architecture, called SchNet, which 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, $\chi_{\boldsymbol{\theta}}$.

$$\mathbf{x}_i^{(n+1)} := \mathbf{x}_i^{(n)} + \chi_{\boldsymbol{\theta}}^{(n)} \left(\left\{ \mathbf{x}_j^{(n)}, \{|\mathbf{r}_j - \mathbf{r}_k|\} \right\} \right). \tag{22}$$

In PauliNet we implement the following interation rule

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

$$\mathbf{z}_i^{(n,\mathrm{n})} := \sum_I \mathbf{w}_{\boldsymbol{\theta}}^{(n,\mathrm{n})} \left(\mathbf{e}\left(|\mathbf{r}_i - \mathbf{R}_I|\right) \right) \odot \mathbf{Y}_{\boldsymbol{\theta},I} \tag{23}$$

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

where '$\odot$' denotes element-wise multiplication and $\mathbf{w}_{\boldsymbol{\theta}}$, $\mathbf{h}_{\boldsymbol{\theta}}$, $\mathbf{g}_{\boldsymbol{\theta}}$ are trainable functions represented by ordinary fully connected DNNs; and $\mathbf{e}$ is a radial basis function that featurizes the interatomic distances. The messages $z_i^{(n)}$ received by the electron feature vectors at each iteration are split into three channels, corresponding to same-spin electrons (+), opposite-spin

5

electrons ($-$) and the nuclei (n). Besides, each channel has a separate receiving function g , again increasing fexibility without substantially increasing the number of parameters.
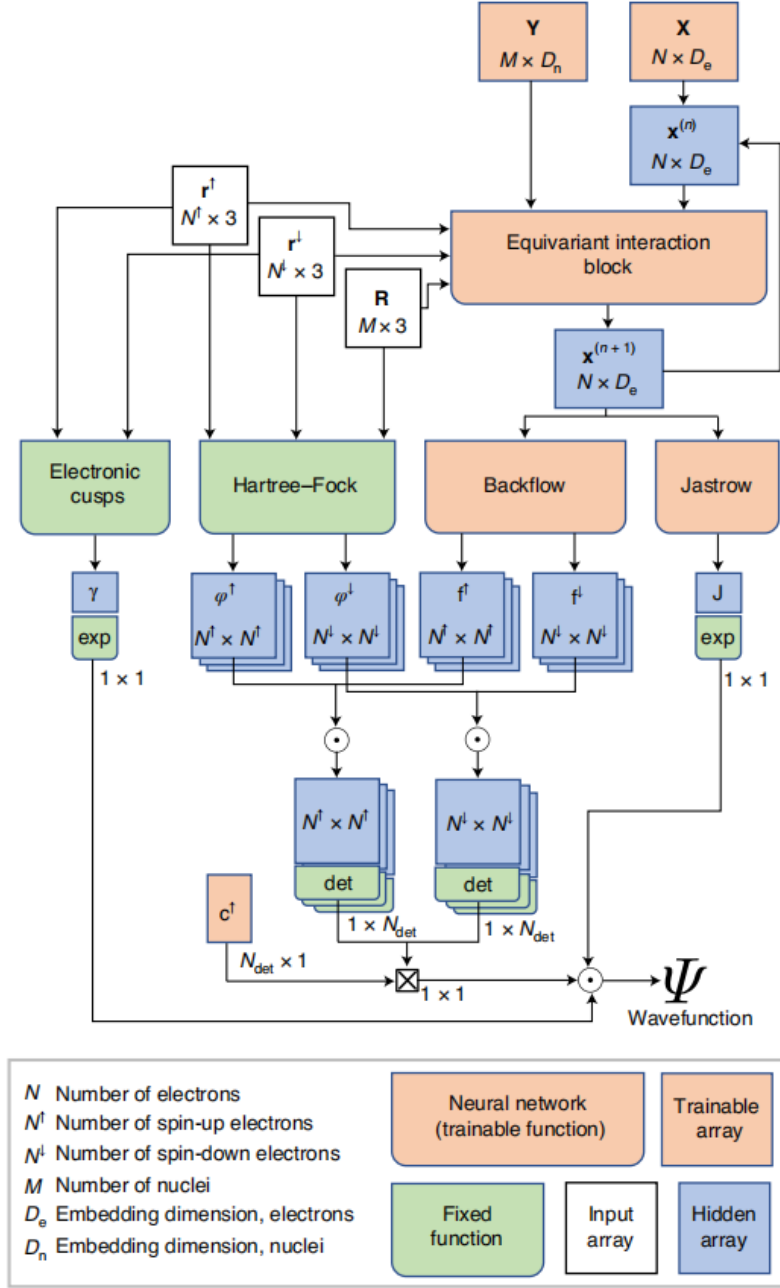

Fig. 2: Architecture of the newly developed PauliNet wavefunction ansatz. The information flows from the input electron and nuclear coordinates, $\mathbf{r}$ and $\mathbf{R}$, to the output wavefunction value, Ψ.

We expand the distances in a basis of Gaussians, which are modified to forces all the

features and their derivatives to zero at zero distance,

$$
e_k(r) := r^2 \mathrm{e}^{-r-(r-\mu_k)^2/\sigma_k^2}
$$
$$
\mu_k := r_\mathrm{c} q_k^2, \quad \sigma_k := \tfrac{1}{7}\left(1+r_\mathrm{c} q_k\right)
$$
(24)

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

After a fixed number of iterations, the final electron representations, $x_i^{(L)}$, 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:

$$
J := \eta_\theta\left(\sum_i \mathbf{x}_i^{(L)}\right), \quad \mathbf{f}_i := \boldsymbol{\kappa}_\theta\left(\mathbf{x}_i^{(L)}\right)
$$
(25)

Since the feature vectors $\mathbf{x}_i^{(n)}$ are equivariant with respect to electron exchange at each iteration, we can easily prove that the Jastrow factor J is invariant while the backflow vectors $\mathbf{f}_i$ are equivariant with respect to exchanges of electrons.

$$
J\left(\mathcal{P}_{ij}\mathbf{r}\right) = J(\mathbf{r}), \quad \mathcal{P}_{ij}f_{\mu i}(\mathbf{r}) = f_{\mu j}\left(\mathcal{P}_{ij}\mathbf{r}\right),
$$
(26)

where $\mathcal{P}_{ij}$ is the exchange of same-spin electrons i and j. Therefore the PauliNet wave function is constructed to be antisymmetric,

$$
\psi_\theta\left(\ldots,\mathbf{r}_i,\ldots,\mathbf{r}_j,\ldots\right) = -\psi_\theta\left(\ldots,\mathbf{r}_j,\ldots,\mathbf{r}_i,\ldots\right).
$$
(27)

5 Approaching exact solutions with few determinants

We train PauliNet via the variational principle, minimizing the total electronic energy (variational QMC).

$$
E_0 = \min_\psi E[\psi] \leq \min_\theta E\left[\psi_\theta\right],
$$
(28)

where

$$
E[\psi] = \int \mathrm{d}\mathbf{r}\psi^\dagger(\mathbf{r})\hat{H}\psi(\mathbf{r})
$$
(29)

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 $|\psi(\mathbf{r})|^2$,

$$
E[\psi] = \mathbb{E}_{\mathbf{r}\sim|\psi|^2}[E_\mathrm{loc}[\psi](\mathbf{r})]
$$
(30)

To optimize the parameters $\boldsymbol{\theta}$ in the Jastrow and backflow neural networks, we use the weighted Adam algorithm:

Require: α: Stepsize

Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0: Initial parameter vector

$\quad m_0 \leftarrow 0$ (Initialize 1$^{\text{st}}$ moment vector)

$\quad v_0 \leftarrow 0$ (Initialize 2$^{\text{nd}}$ moment vector)

$\quad t \leftarrow 0$ (Initialize timestep)

$\quad$ **while** θ_t not converged **do**

$\quad\quad t \leftarrow t + 1$

$\quad\quad g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep)

$\quad\quad m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_1$ (Update biased first moment estimate)

$\quad\quad v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_1^2$ (Update biased second raw moment estimate)

$\quad\quad \hat{m}_t \leftarrow m_t/(1 - \beta_1^t)$ (Compute bias-corrected first moment estimat)

$\quad\quad \hat{v}_t \leftarrow v_t/(1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)

$\quad\quad \theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t/(\sqrt{\hat{v}_t} + \varepsilon)$ (Update parameters)

$\quad$ **end while**

$\quad$ **return** θ_t (Resulting parameters)

In the algorithm above, g_t^2 indicates the elementwise square $g_t \odot g_t$, and β_1^t and β_2^t denote β_1 and β_2 to the power of t. Besides, we use the total energy directly as the loss function,

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{r} \sim |\psi_\theta|^2} \left[E_{\text{loc}} \left[\psi_\theta \right] (\mathbf{r}) \right]. \tag{31}$$

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. Since

$$
\begin{aligned}
\frac{\partial}{\partial \theta} \mathbb{E}_{\mathbf{r} \sim p(\mathbf{r};\theta)}[f(\mathbf{r};\theta)] &= \frac{\partial}{\partial \theta} \int d^3 r\, p(\mathbf{r};\theta) f(\mathbf{r};\theta) \\
&= \int d^3 r \left[\frac{\partial p(\mathbf{r};\theta)}{\partial \theta} f(\mathbf{r};\theta) + p(\mathbf{r};\theta) \frac{\partial f(\mathbf{r};\theta)}{\partial \theta} \right] \\
&= \int d^3 r \left[p(\mathbf{r};\theta) \frac{\partial \ln p(\mathbf{r};\theta)}{\partial \theta} f(\mathbf{r};\theta) + p(\mathbf{r};\theta) \frac{\partial f(\mathbf{r};\theta)}{\partial \theta} \right] \\
&= \mathbb{E}_{\mathbf{r} \sim p(\mathbf{r};\theta)} \left[\left(\frac{\partial}{\partial \theta} \ln p(\mathbf{r};\theta) \right) f(\mathbf{r};\theta) + \frac{\partial}{\partial \theta} f(\mathbf{r};\theta) \right],
\end{aligned}
\tag{32}
$$

Therefore we can write the gradient of the loss function w.r.t. θ:

$$\nabla_\theta \mathcal{L}(\boldsymbol{\theta}) = 2\mathbb{E}_{\mathbf{r} \sim |\psi_\theta|^2}[E_{\text{loc}}[\psi_\theta](\mathbf{r}) \nabla_\theta \ln |\psi_\theta|] - 2\mathcal{L}(\boldsymbol{\theta}) \mathbb{E}_{\mathbf{r} \sim |\psi_\theta|^2}[\ln |\psi_\theta|]. \tag{33}$$

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).

6   Performance of PauliNet on some small moleculars

They first investigate 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 reaches the correlation energy between 97% and 99.9% with one to two orders of magnitude less determinants than standard variational ansatzes (Fig.3).
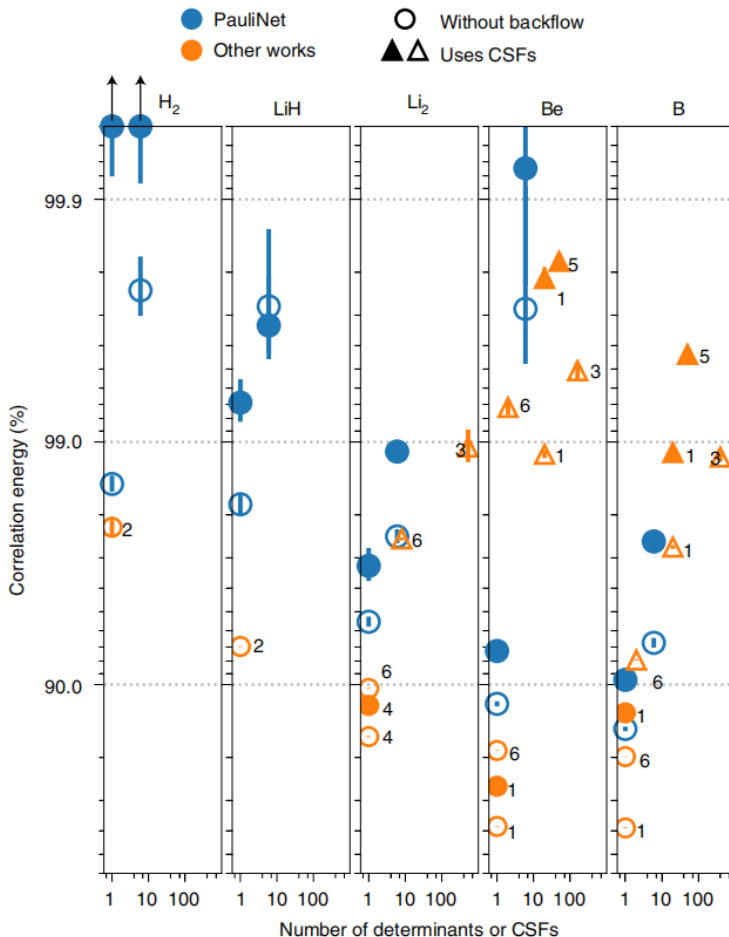

Fig. 3: Performance of PauliNet with one and six determinants on atoms and diatomic molecules. Four variants of PauliNet are shown, single and multideterminant as well as with and without backflow. The reference results are taken from (1) Brown et al., (2) Casalegno et al., (3) Morales et al., (4) Lopez Ríos et al., (5) Seth et al. and (6) Toulouse and Umrigar.

9

Four variants with a single or multiple Slater determinant(s) (SD or MD), with or without backflow (BF) are shown in Fig.4, which shows that both the backflow and the use of a few determinants is crucial for reaching high accuracy.
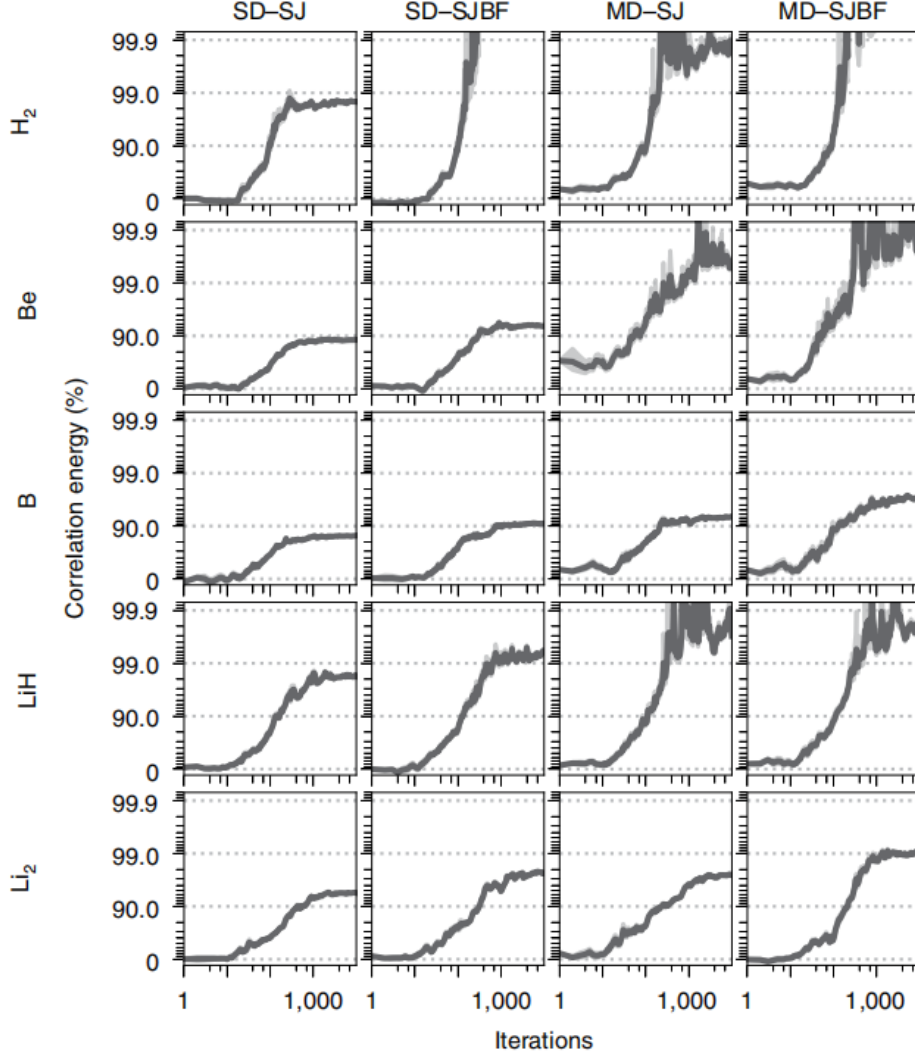

Fig. 4: Roles of backflow and multiple determinants in training of PauliNet ansatz.

For linear hydrogen chain H_{10} which exhibits strong correlation, they recover $98.41(8)\%$ and $98.4(3)\%$ of the correlation energy in the equilibrium and stretched geometries, respectively, using 16 determinants (Fig.5). 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 $82(2)\%$), which indicates that the backflow plays a much larger role than multiple determinants in this strong correlation system, highlighting the

10

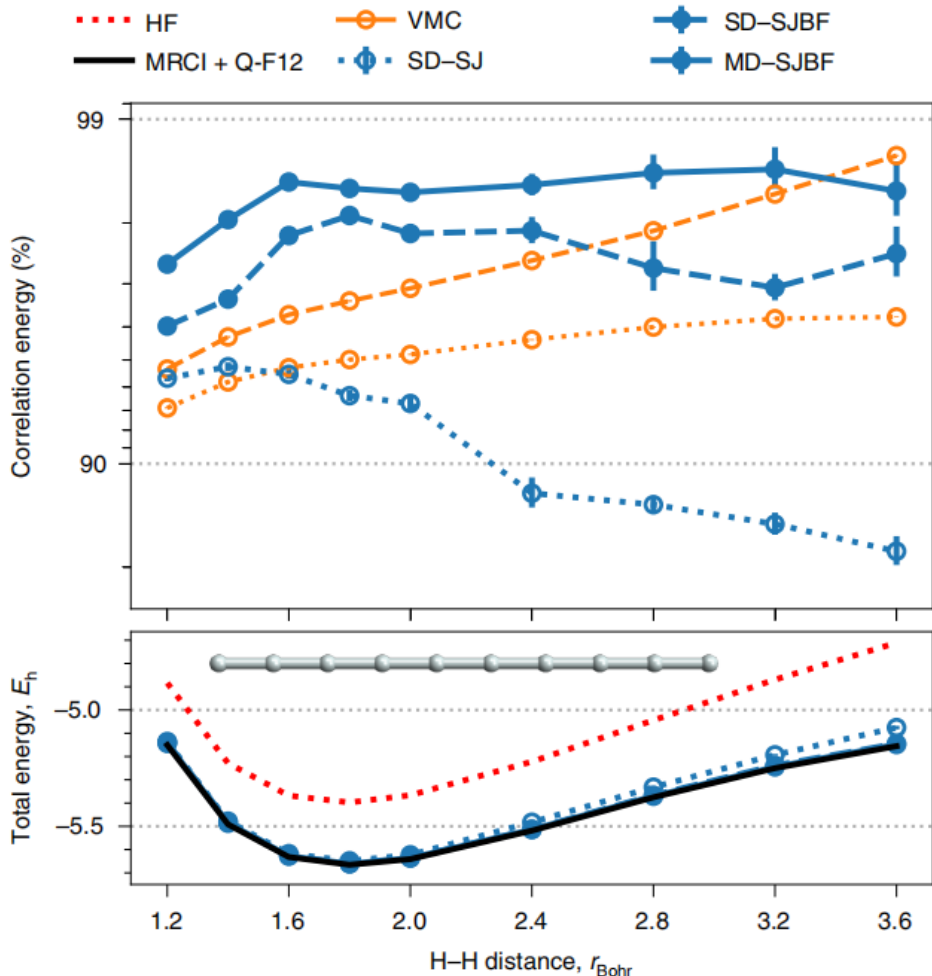central role of the trainable backflow in PauliNet.

Fig. 5: The correlation energy recoverd in linear hydrogen chain H_{10}. PauliNet results (blue) with single determinant or 16 determinants (MD) and with or without backflow are shown. The correlation energy is calculated with respect to multireference configuration-interaction (MRCI) results. E_h, Hartree energy; r_{Bohr}, Bohr radius.

7 Straightforward generalization to larger molecules

The automerization of cyclobutadiene (环丁二烯) (Fig.6a, 28 electrons) is a chemical process that has received considerable attention from both experiment and theory. The experimental estimates of the energy barrier range between 1.6 and 10 kcal·mol^{-1}. However, 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

11

from various flavours of the multireference coupled-cluster (MR-CC) theory and fall between 7 and 11 $kcal \cdot mol^{-1}$, without a decisive answer as to which of the variants is closer to the ground truth.
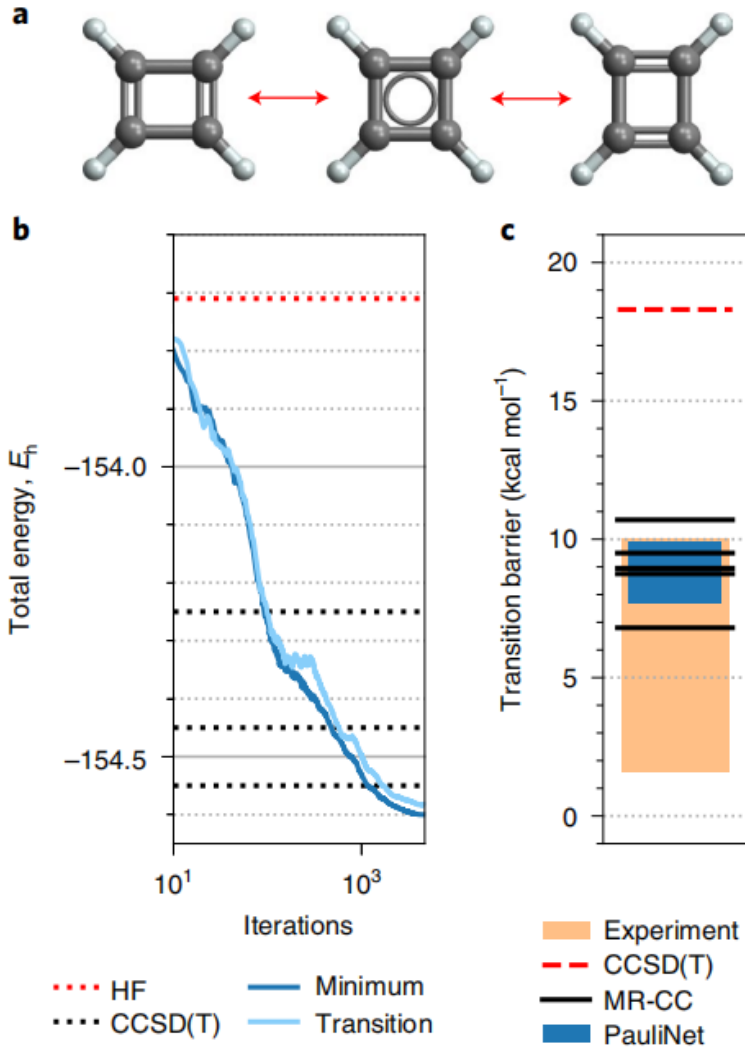

Fig. 6: Calculation of the transition barrier of cyclobutadiene automerization. **a**, Cyclobutadiene automerization. The transition state has a highly multireferential character. **b**, Convergence of the total energy of the energy minimum and transition state with training. **c**, Energy barrier obtained by sampling the trained PauliNet wavefunctions, compared with results obtained by other approaches.

The PauliNet ansatz with ten determinants and the same hyperparameters as used for the much smaller systems is able to give all-electron variational energies for the energy minimum and transition states of cyclobutadiene, and thus for the energy barrier (Fig.6b).

12