Convergence to the fixed-node limit in deep variational Monte Carlo

Z. Schätzle
FU Berlin
Department of Mathematics and Computer Science
Arnimallee 6, 14195 Berlin, Germany
`zeno.schaetzle@fu-berlin.de`

J. Hermann
FU Berlin
Department of Mathematics and Computer Science
Arnimallee 6, 14195 Berlin, Germany
TU Berlin
Machine Learning Group
Marchstr. 23, 10587 Berlin, Germany
`jan.hermann@fu-berlin.de`

F. Noé
FU Berlin
Department of Mathematics and Computer Science
Arnimallee 6, 14195 Berlin, Germany
FU Berlin
Department of Physics
Arnimallee 14, 14195 Berlin, Germany
Rice University
Department of Chemistry
Houston, TX 77005, USA
`frank.noe@fu-berlin.de`

Abstract

Variational quantum Monte Carlo (variational QMC) is an ab-initio method for solving the electronic Schrödinger equation that is exact in principle, but limited by the functional form used to represent the electronic wavefunction in practice. The recent deep-neural-network ansatzes PauliNet and FermiNet increased the accuracy of variational QMC dramatically, but little is understood about the convergence behavior of such ansatzes. Here, we show that a deep neural network can overcome the limitations of a small one-electron basis set and approach the complete-basis-set limit, and that an ansatz with a deep Jastrow factor can systematically converge to the fixed-node limit. Benchmarks on H_2O show that the optimized deep ansatz recovers by half an order of magnitude more fixed-node correlation energy compared to previous variational QMC results.

1 Introduction

The fundamental problem in quantum chemistry is to solve the electronic Schrödinger equation as accurately as possible at a manageable cost. Variational quantum Monte Carlo (variational QMC) is

Third Workshop on Machine Learning and the Physical Sciences (NeurIPS 2020), Vancouver, Canada.

an ab-initio method based on the stochastic evaluation of expectation values that scales favorably with the system size and provides explicit access to the electronic wavefunction [1]. Although exact in principle, the accuracy of variational QMC (VMC) strongly depends on the expressiveness of the trial wavefunction ansatz. Recently, deep QMC has been introduced, which involves a new class of ansatzes based on deep neural networks (DNNs) [2, 3] and elevates the accuracy of VMC within quantum chemistry to state-of-the-art levels or beyond.

Currently, there is little understanding why such DNN ansatzes work well and how their individual components contribute to the approximation of the ground-state wavefunction and energy. Such understanding, however, is necessary for scaling the accuracy of deep QMC for small molecules to larger systems. Here, we develop a hierarchy of approximation methods based on the traditional QMC methodology that enables us to disentangle the contributions of individual components of the ansatz to the resulting accuracy, and thus identify which of them are amenable to improvement.

An electronic wavefunction ansatz, $\psi_{\boldsymbol{\theta}}(\mathbf{r})$, is a parametric function that is antisymmetric with respect to exchange of the input coordinates $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ of electrons with same spin, and whose parameters are optimized via the variational principle of quantum mechanics. A general form for VMC approaches is the product of an antisymmetric part, usually composed of Slater determinants, and a symmetric Jastrow factor,

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = \underbrace{\mathrm{e}^{J(\mathbf{r})}}_{\text{symmetric}} \underbrace{\textstyle\sum_p c_p \det[\mathbf{A}_p^{\uparrow}(\mathbf{r})] \det[\mathbf{A}_p^{\downarrow}(\mathbf{r})]}_{\text{antisymmetric}} \tag{1}$$

In traditional ansatzes, the Slater determinants are formed from one-electron functions, called molecular orbitals, $A_{ij} = \varphi_i(\mathbf{r}_j)$, which are optimized and fixed prior to the actual VMC simulation in a Hartree–Fock (HF) calculation. This also fixes the *nodal surface* of the ansatz—the hyperplane in the space of $\mathbf{r}$, on which the wavefunction changes sign. The orbitals are formed as a linear combination of basis functions, which allows arbitrarily accurate orbitals only in the limit of infinitely many basis function—the complete-basis-set (CBS) limit. This makes accurate calculations costly in practice. The Jastrow factor, J, introduces correlation between electrons and traditionally involves power series expansions in one, two, and three-body terms [4], which does not provide sufficient expressiveness to reach high accuracy. VMC is therefore usually followed by a much more costly fixed-node diffusion QMC (FN-DMC), a projector method based on evolving electronic configurations according to the imaginary time Schrödinger equation, which overcomes any deficiencies of the Jastrow factor, but cannot improve the nodal surface of the antisymmetric part of the ansatz [5].

Related work The ability of neural networks to represent antisymmetric (wave) functions has been explored theoretically [6, 7]. The treatment of antisymmetry can be avoided in the second-quantized formalism at the cost of the finite-basis-set error [8]. The formulation of quantum mechanics based on the imaginary-time Schrödinger equation offers yet another entry point for the use of neural networks to represent quantum states [9, 10]. On the applied side, deep QMC can be used for many-particle quantum systems other than electrons [11]. The unsupervised learning based on the variational principle is not the only way how machine learning can represent electronic structure [12, 13].

Contributions In this work we investigate how neural networks in deep QMC compensate for incomplete basis sets and reach high accuracy with a deep Jastrow factor without the need for diffusion QMC. In particular, this paper contains the following contributions:

1. We demonstrate that DNNs can correct the single-particle orbitals of a HF calculation in a small basis and obtain energies equivalent to the complete-basis-set limit.
2. We show that for two nodeless systems, H_2 and He, a deep Jastrow can achieve results within five significant digits of the exact energy.
3. We conduct an extensive hyperparameter search for the deep Jastrow and find that its expressiveness can be systematically increased to converge to the fixed-node limit.
4. We combine both the orbital correction and deep Jastrow and obtain state-of-the-art VMC results for the water molecule.

2 Methods

Our starting point is PauliNet [2], a DNN ansatz that follows directly eq. 1,

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = \mathrm{e}^{\gamma(\mathbf{r})+J_{\boldsymbol{\theta}}(\mathbf{r})} \sum_p c_p \det[\tilde{\varphi}^{\uparrow}_{\boldsymbol{\theta},\mu_p i}(\mathbf{r})] \det[\tilde{\varphi}^{\downarrow}_{\boldsymbol{\theta},\mu_p i}(\mathbf{r})] \tag{2}$$

$$\tilde{\varphi}_{\boldsymbol{\theta},\mu i}(\mathbf{r}) = \varphi_\mu(\mathbf{r}_i) f^{\otimes}_{\boldsymbol{\theta},\mu i}(\mathbf{r}) + f^{\oplus}_{\boldsymbol{\theta},\mu i}(\mathbf{r}) \tag{3}$$

Here, γ is a fixed part of the Jastrow factor that ensures correct asymptotics [14], J is a deep trainable Jastrow factor, and both f form a deep backflow that turns the one-electron orbitals into general many-electron functions. Jastrow factor and backflow transformation are obtained from a joint latent space many-body representation, encoded in the final node features of a convolutional neural network acting on the molecular graph of electrons and nuclei,

$$\mathbf{x}_i^{(n+1)} := \mathbf{x}_i^{(n)} + \boldsymbol{\chi}_{\boldsymbol{\theta}}^{(n,\pm)}\big(\{\mathbf{x}_j^{(n)}, \{|\mathbf{r}_i - \mathbf{r}_j|\}\}\big) + \boldsymbol{\chi}_{\boldsymbol{\theta}}^{(n,\mathrm{n})}\big(\{\mathbf{Y}_{\boldsymbol{\theta},I}, \{|\mathbf{r}_i - \mathbf{R}_I|\}\}\big) \tag{4}$$

The neural network is a modification of SchNet [15], where electronic embeddings $\mathbf{x}_i^{(n)}$ are iteratively updated within the interaction phase, while the stationary nuclei are represented by trainable arrays $\mathbf{Y}_{\boldsymbol{\theta},I}$. Initializing same-spin electrons with identical embeddings and applying permutation invariant convolutions $\boldsymbol{\chi}_{\boldsymbol{\theta}}$ the many-body representation is equivariant under the exchange of spin-up and spin-down electrons respectively. We form three restricted variants of the single-determinant form of PauliNet that enables us to study the individual components separately.

Deep orbital correction Version with no Jastrow factor and a restricted version of the deep backflow from (3), where f take as an input only the coordinate of the i-th electron, leaves the orbitals as one-electron functions, but modifies their shape. This is an alternative to the traditional approach, in which the linear basis-set expansion of $\varphi_\mu(\mathbf{r}_i)$ is improved by increasing the basis-set size, which converges only slowly to the complete-basis-set limit.

Deep Jastrow factor PauliNet without the backflow. The deep Jastrow factor $J_{\boldsymbol{\theta}}$ is obtained from the latent space many-body representation,

$$J_{\boldsymbol{\theta}}(\mathbf{r}) := \eta_{\boldsymbol{\theta}}\big(\textstyle\sum_i \mathbf{x}_i^{(L)}(\mathbf{r})\big) \tag{5}$$

The full specification of the Jastrow factor involves the configuration of the graph neural network fixing the dimension of the node features, node messages, distance kernels, and distance features, depths of all the message-passing subnetworks, and the number of message-passing iterations, as well as the depth of the Jastrow network $\eta_{\boldsymbol{\theta}}$. These together form a set of hyperparameters that affect the expressiveness of the deep Jastrow.

Mean-field Jastrow factor A restricted version of the deep Jastrow, which is separable in the individual electrons, and hence does not introduce electron correlation,

$$J_{\boldsymbol{\theta}}(\mathbf{r}) := \textstyle\sum_i \zeta_{\boldsymbol{\theta}}(\mathbf{r}_i) \tag{6}$$

The mean-field Jastrow is able to optimize the one-electron density of the ansatz.

Training procedure The PauliNet ansatz is optimized by applying the variational principle of quantum mechanics and following the standard VMC approach of approximating the energy expectation value $\int d\mathbf{r}\psi(\mathbf{r})H\psi(\mathbf{r})$ by Monte Carlo integration. The training then consists of an alternating scheme of stochastically sampling the probability density associated with the square of the trial wavefunction using a Langevin Monte Carlo approach and updating the model parameters to minimize the energy expectation value estimated from minibatches of electronic configurations,

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{r}\sim\psi^2_{\theta'}}\Big[\frac{H\psi_\theta(\mathbf{r})}{\psi_\theta(\mathbf{r})}\Big] \tag{7}$$

3 Results

Correcting finite-basis-set error of HF orbitals with a DNN The deep orbital correction can be employed to improve the orbitals of an imprecise HF@6-31G baseline. For the test systems H_2, He, and Be, energies within 0.1 mH of the extrapolated complete-basis-set limit were achieved (Fig. 1A), demonstrating that a real-space orbital correction can compensate for the finite-basis-set error.

Exact solutions with deep Jastrow for two-electron systems The wavefunctions of closed-shell two-electron systems H_2 and He have no nodal surface, providing a pure test of any Jastrow factor. Our deep Jastrow factor recovered 99.97(3)% and 99.98(2)% of the total correlation energy, respectively. On the dissociation curve of H_2, the deep Jastrow outperforms full configuration interaction (FCI) with the cc-pV5Z basis set for all distances (Fig. 1B).

Systematic approach to the fixed-node limit with deep Jastrow The performance of the Jastrow factor is evaluated for LiH, a four-electron systems that exhibits more intricate interactions, while

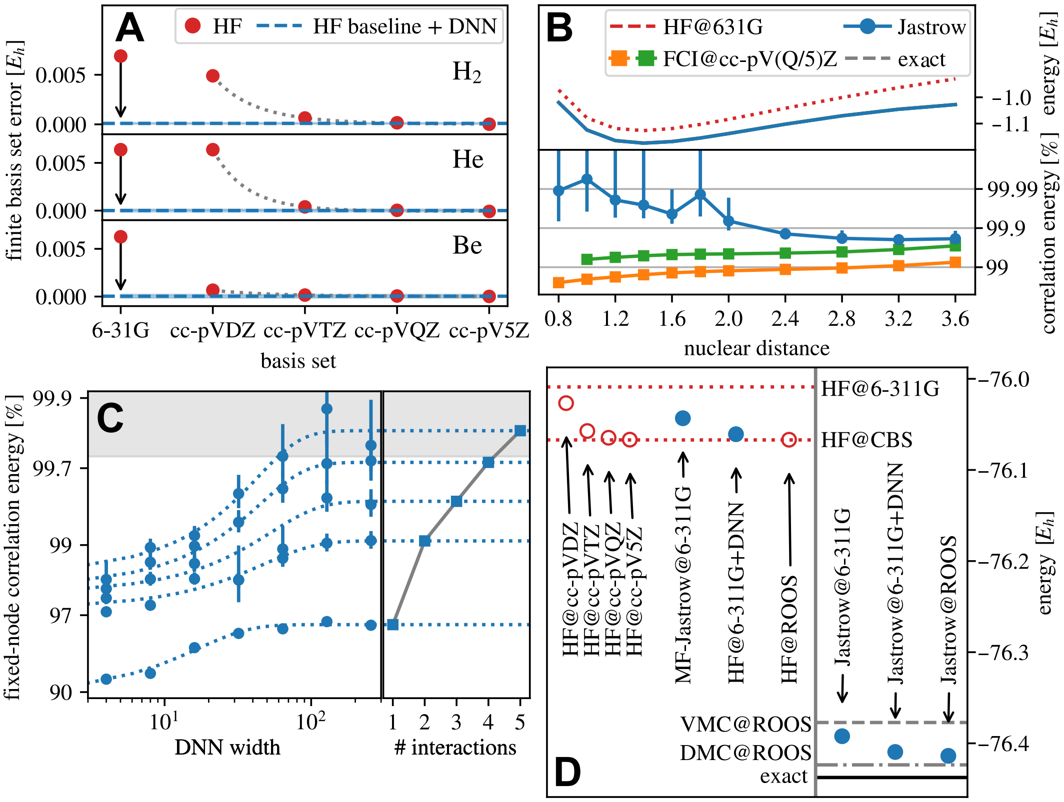

Figure 1: **(A) Removing the basis-set error in a HF calculation.** The baseline HF calculation with the small 6-31G basis in combination with the deep orbital correction reaches directly the complete-basis-set limit, which traditionally requires increasingly larger basis sets from the cc-pVnZ series. **(B) Dissociation curve of the H_2 molecule.** Total energy and correlation energy with HF, deep Jastrow and FCI are compared, the latter being indistinguishable from the exact energy in the upper plot. Exact energy was taken from [16] and FCI results were obtained with PySCF [17] in the cc-pVQZ basis (orange) and cc-pV5Z basis (green). **(C) Approaching FN-DMC accuracy with deep Jastrow on LiH molecule.** Increasingly expressive deep Jastrow factors are trained to approach the FN-DMC energy of a single-determinant (HF@TZP [18]) trial wavefunction [19]. The most expressive ansatzes give results within the sampling error of the FN-DMC energy (shaded region at the top). The dependence on the width of the distance kernel (DNN width) and the number of interactions L (# interactions) in the Jastrow graph-neural-network is shown. **(D) Benchmarking deep QMC on H_2O.** All three restricted variants of PauliNet introduced in Methods are compared: the deep orbital correction (6-311G+DNN), the full deep Jastrow, and the mean-field (MF) Jastrow. ROOS denotes the Roos-aug-DZ-ANO basis set. VMC and DMC references are taken from [20].

Table 1: **Benchmarking single-determinant (SD) Slater–Jastrow (SJ) ansatzes on H_2O.**

reference	HF	VMC (SD-SJ)	DMC (SD-SJ)	corr. energy [%]	basis set
PauliNet	-76.009	-76.3923(7)	–	91.2(2)*	6-311G
PauliNet	-76.0612	-76.4096(7)	–	96.0(2)*	6-311G+DNN
PauliNet	-76.0672	**-76.4139**(5)	–	**97.2**(1)*	Roos-aug-DZ-ANO
Ref. [20]	-76.0672	-76.3773(2)	-76.42376(5)	87.01(6)	Roos-aug-DZ-ANO
Ref. [21]	-76.063	-76.327(5)	-76.4219(1)	73.3(1)	Roos-aug-DZ-ANO
Ref. [22]	-76.0587	-76.327(1)	-76.42102(4)	73.5(3)	6–311++G(2d,2p)

*The fixed-node correlation energy is computed with respect to the reference FN-DMC energy of [20].

still computationally lightweight, such that the hyperparameter space of the deep ansatz can be explored exhaustively. Since the single-determinant Slater–Jastrow ansatz possesses an inherent fixed-node error, we measure the performance of the deep Jastrow factor with respect to the fixed-node limit estimated from FN-DMC calculations with an identical Slater determinant. We performed a comprehensive hyperparameter scan that revealed a systematic convergence to the fixed-node limit with increasing network size (Fig. 1C), indicating that the deep Jastrow factor can be extended towards completeness.

Application of different levels of theory to the water molecule To investigate how deep orbital correction and deep Jastrow behave for more complex systems, we evaluated the performance of these two restricted variants of PauliNet on the water molecule (Fig. 1D). Starting within the mean-field theory we corrected the HF@6-311g determinant with the deep orbital correction and resolved 90% of the finite-basis-set error. We then estimated how much of the finite-basis-set error amounts to the fixed-node error, by applying the mean-field Jastrow, which recovered only about half of the finite-basis-set error. This suggests that when approaching the complete-basis-set limit, the nodal surface of the HF solution is altered significantly. Next, we benchmark the deep Jastrow factor with a HF determinant in the Roos-aug-DZ-ANO basis [23], a common basis set that gives HF energies at the complete-basis-set limit. We compare to VMC and DMC results from the literature, achieving 97.2(1)% of the fixed-node correlation energy and surpassing the accuracy of previous VMC calculations by half an order of magnitude (Table 1). We computed the energies of the deep Jastrow with a HF@6-311g determinant with and without the orbital correction, demonstrating that both methods can be combined to reach high accuracy even with a minimal baseline. We recovered 96.0(2)% of the fixed-node correlation energy with respect to the Roos-aug-DZ-ANO basis. The results further show that the differences between the energies of the mean-field ansatzes match the differences of the corresponding Slater–Jastrow trial wavefunctions.

Conclusions We have demonstrated that there are no fundamental limitations to the investigated components of the PauliNet ansatz. With both the deep orbital correction and deep Jastrow, close to exact energies for the respective level of theory can be obtained, highlighting the expressiveness of deep QMC. A single ansatz without any problem-specific modifications can compete with state-of-the-art VMC energies on a variety of systems and can be extended systematically to improve the accuracy without introducing new components to the architecture. The analysis for the water molecule shows that the same instances of deep orbital correction and deep Jastrow factor evaluated on the small systems can be successfully applied to a larger test case and there is evidence that PauliNet can achieve state-of-the-art energies for molecules with up to a few dozen of electrons [2]. Though the accuracy of a fixed architecture for systems with increasing size is yet to be investigated in detail, a systematically extensible expressiveness in combination with the generally favourable scaling of the VMC method (N^4) with respect to the number of electrons might constitute the basis for a scalable QMC algorithm. Albeit the results emphasize the potential of the deep QMC approach, the major benefit of deep QMC over FN-DMC calculations remains the capability of altering the nodal surface. The presented analysis paves the way for future investigations on how the full PauliNet ansatz and other deep QMC approaches improve the nodal surface and overcomes the fixed-node limitations.

Broader Impact

Quantum chemistry is essential for the understanding of fundamental properties of molecules and materials. Deep learning methods to solve the quantum-mechanical equations, such as deep QMC, can provide a deeper understanding of our world at the nanoscale and provide useful insight and training data upon which chemicals or materials with desirable properties can be developed or optimized.

As most machine-learning methods, the convergence of deep QMC methods to good approximations of the electronic Schrödinger equation depend on various choices and hyperparameters. For larger molecules, for which no ground truth is known with other highly accurate methods, wrong conclusions could be made when convergence is not reached. Since the deep QMC methods discussed here are variational, better solutions will always reveal themselves with lower variational energies. This property should be exploited to test the robustness of predictions before drawing conclusions that affect real-world decisions.

References

[1] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, "Quantum Monte Carlo simulations of solids," *Rev. Mod. Phys.*, vol. 73, pp. 33–83, Jan. 2001.

[2] J. Hermann, Z. Schätzle, and F. Noé, "Deep-neural-network solution of the electronic Schrödinger equation," *Nature Chemistry*, vol. 12, pp. 891–897, Oct. 2020.

[3] 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," *Phys. Rev. Research*, vol. 2, p. 033429, Sept. 2020.

[4] N. D. Drummond, M. D. Towler, and R. J. Needs, "Jastrow correlation factor for atoms, molecules, and solids," *Phys. Rev. B*, vol. 70, p. 235119, Dec. 2004.

[5] 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*, vol. 22, p. 023201, Dec. 2009.

[6] J. Han, Y. Li, L. Lin, J. Lu, J. Zhang, and L. Zhang, "Universal approximation of symmetric and anti-symmetric functions." Preprint at http://arxiv.org/abs/1912.01765, Dec. 2019.

[7] M. Hutter, "On Representing (Anti)Symmetric Functions." July 2020.

[8] K. Choo, A. Mezzacapo, and G. Carleo, "Fermionic neural-network states for ab-initio electronic structure," *Nat. Commun.*, vol. 11, p. 2368, May 2020.

[9] A. Barr, W. Gispen, and A. Lamacraft, "Quantum Ground States from Reinforcement Learning," in *Proceedings of Machine Learning Research*, vol. 107, pp. 635–653, July 2020.

[10] J. Han, J. Lu, and M. Zhou, "Solving high-dimensional eigenvalue problems using deep neural networks: A diffusion Monte Carlo like approach." Preprint at http://arxiv.org/abs/2002.02600, July 2020.

[11] C. Adams, G. Carleo, A. Lovato, and N. Rocco, "Variational Monte Carlo calculations of A $\leq$ 4 nuclei with an artificial neural-network correlator ansatz." Preprint at http://arxiv.org/abs/2007.14282, July 2020.

[12] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, "Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions," *Nat. Commun.*, vol. 10, p. 5024, Dec. 2019.

[13] M. Gastegger, A. McSloy, M. Luya, K. T. Schütt, and R. J. Maurer, "A deep neural network for molecular wave functions in quasi-atomic minimal basis representation," *J. Chem. Phys.*, vol. 153, p. 044123, July 2020.

[14] T. Kato, "On the eigenfunctions of many-particle systems in quantum mechanics," *Communications on Pure and Applied Mathematics*, vol. 10, no. 2, pp. 151–177, 1957.

[15] 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.*, vol. 148, p. 241722, Mar. 2018.

[16] W. Kołos and L. Wolniewicz, "Potential-Energy Curves for the X 1Σg+, b3Σu+, and C 1Πu States of the Hydrogen Molecule," *J. Chem. Phys.*, vol. 43, pp. 2429–2441, Oct. 1965.

[17] Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, "Recent developments in the PySCF program package," *J. Chem. Phys.*, vol. 153, p. 024109, July 2020.

[18] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus, "New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community," *J. Chem. Inf. Model.*, vol. 59, pp. 4814–4820, Nov. 2019.

[19] 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.*, vol. 118, pp. 7193–7201, Apr. 2003.

[20] I. G. Gurtubay and R. J. Needs, "Dissociation energy of the water dimer from quantum Monte Carlo calculations," *J. Chem. Phys.*, vol. 127, p. 124306, Sept. 2007.

[21] N. A. Benedek, I. K. Snook, M. D. Towler, and R. J. Needs, "Quantum Monte Carlo calculations of the dissociation energy of the water dimer," *J. Chem. Phys.*, vol. 125, p. 104302, Sept. 2006.

[22] I. G. Gurtubay, N. D. Drummond, M. D. Towler, and R. J. Needs, "Quantum Monte Carlo calculations of the dissociation energies of three-electron hemibonded radical cationic dimers," *J. Chem. Phys.*, vol. 124, p. 024318, Jan. 2006.

[23] P.-O. Widmark, P.-Å. Malmqvist, and B. O. Roos, "Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions," *Theoret. Chim. Acta*, vol. 77, pp. 291–306, Sept. 1990.