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

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

(Dated: 13 October 2020)

Variational quantum Monte Carlo (QMC) is an ab-initio method for solving the electronic Schrödinger equation that is exact in principle, but limited by the flexibility of the available ansatzes in practice. The recently introduced deep QMC approach, specifically two deep-neural-network ansatzes PauliNet and FermiNet, allows variational QMC to reach the accuracy of diffusion QMC, but little is understood about the convergence behavior of such ansatzes. Here, we analyze how deep variational QMC approaches the fixed-node limit with increasing network size. First, we demonstrate that a deep neural network can overcome the limitations of a small basis set and reach the mean-field complete-basis-set limit. Moving to electron correlation, we then perform an extensive hyperparameter scan of a deep Jastrow factor for LiH and H_4 and find that variational energies at the fixed-node limit can be obtained with a sufficiently large network. Finally, we benchmark mean-field and many-body ansatzes on H_2O, increasing the fraction of recovered fixed-node correlation energy by half an order of magnitude compared to previous VMC results. This analysis helps understanding the superb performance of deep variational ansatzes, and will guide future improvements of the neural network architectures in deep QMC.

I. 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 (VMC) is a variational ab-initio method based on the stochastic evaluation of expectation values that scales favorably with the system size and provides explicit access to the wavefunction[1]. Although exact in principle, the method strongly depends on the quality of the trial wavefunction, which determines both efficiency and accuracy of the computation and typically constitutes the limiting factor of VMC calculations.

Recently, deep QMC has been introduced. Deep QMC involves a new class of ansatzes that complement traditional trial wavefunctions with the expressiveness of deep neural networks (DNNs). This ab-initio approach is orthogonal to the supervised learning of electronic structure that requires external datasets[2,3]. The use of neural-network trial wavefunctions has been pioneered for spin lattice systems[4] and later generalized to molecules in second quantization[5]. The first application to molecules in real space was a proof-of-principle effort, but did not reach the accuracy close to traditional VMC[6]. The DNN architectures PauliNet and FermiNet advanced the real space deep QMC approach[7,8], increasing the accuracy to state-of-the-art levels and beyond. Application of the deep QMC to many-particle quantum systems other than electrons is also possible[9].

Currently, there is little understanding of why these DNN wavefunctions work well and how their individual components contribute to the approximation of the ground-state wavefunction and energy. Examining the expressive power of the components and measuring their performance with respect to traditional approaches is essential to establish neural-network trial wavefunctions as a standard technique in VMC and to guide further development. Here, we identify a hierarchy of model ansatzes based on the traditional VMC methodology (Fig. 1) that enables us to distinguish the effects of improving single-particle orbitals and adding correlation in the symmetric part of the wavefunction ansatz.

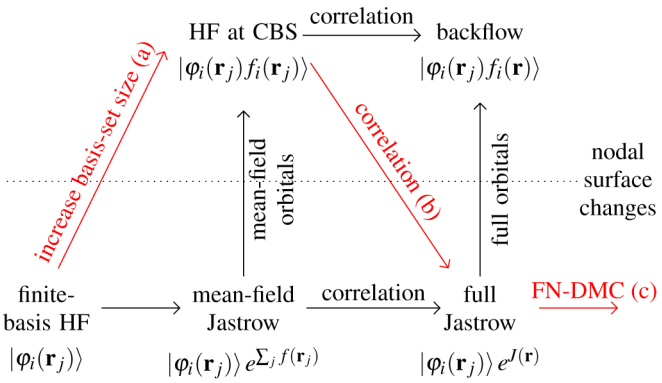

FIG. 1. **Hierarchy of single-determinant ansatzes in QMC.** The starting point of a finite-basis Hartree–Fock (HF) calculation can be extended by a "mean-field" Jastrow factor to improve the one-electron density of the ansatz. From that point, the ansatz can be improved in one of two directions—by modifying the orbitals (bottom–top) or introducing electron correlation (left–right). The red pathway shows a standard approach in traditional QMC and is the path we pursue in our analysis.

The trial wavefunctions in QMC are typically constructed by combining a symmetric Jastrow factor with an antisymmetric part that implements the Pauli exclusion principle for fermions by specifying the nodal surface of the ansatz—the hyperplane in the space of electron coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$, on which the wavefunction changes sign.

Expressing the antisymmetric part as a linear combination of Slater determinants gives rise to the ansatz of the Slater–Jastrow–backflow-type that comprises most VMC ansatzes, including the deep variants PauliNet and FermiNet:

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

The ability of neural networks to represent antisymmetric (wave) functions has been explored theoretically[10,11].

Traditionally, Slater determinants are antisymmetrized product states constructed from single-particle molecular orbitals, which are expressed in a single-electron basis set consisting of basis functions ϕ_k.

$$A_{ij} = \varphi_i(\mathbf{r}_j) = \sum_k c_{ik} \phi_k(\mathbf{r}_j) \tag{2}$$

Employing such basis sets transforms the problem of searching over infinitely many functions into a problem of searching coefficients in a system of equations, which can be solved by means of linear algebra by applying for instance the Hartree–Fock (HF), the multi-configurational self-consistent field (MCSCF), or the full configuration interaction (FCI) method. The projection comes at the cost of introducing the finite-basis-set error, which completely vanishes only in the limit of infinitely many basis function—the complete-basis-set (CBS) limit (Fig. 1a). Finite-basis-set errors are inherent to a second-quantized representation, which nevertheless provides an alternative platform to introduce deep learning to quantum chemistry[12].

The real-space formulation of VMC allows to introduce explicit electron correlation efficiently by modelling many-body interactions with a Jastrow factor (Fig. 1b). The Jastrow factor is a symmetric function of the electron coordinates that traditionally involves an expansion in one, two, and three-body terms[13]. Although strongly improving the ansatz, traditional Jastrow factors do not have sufficient expressiveness to reach high accuracy and an initial VMC calculation is typically followed by a computationally demanding fixed-node diffusion QMC (FN-DMC) simulation (Fig. 1c) that eventually projects out the exact solution for the given nodal surface—the fixed-node limit[14]. Diffusion QMC is based on the imaginary-time Schrödinger equation and offers yet another entry point for the use of neural networks to represent quantum states[15,16].

The nodal surface of the trial wavefunctions can be improved by increasing the number of determinants or applying the backflow technique, transforming single-particle orbitals to many-body orbitals under consideration of the symmetry constraints. These are key concepts to efficiently reach very high accuracy with VMC and integral features of deep QMC.

In this paper we take a closer look at how neural networks compensate errors arising from finite basis sets and demonstrate converge to the fixed-node limit within the VMC framework by systematically increasing the expressiveness of a deep Jastrow factor. For the sake of disentangling the individual contributions to the overall accuracy, we conduct our analysis with the Slater–Jastrow-type trial wavefunctions with an antisymmetric part consisting of a single Slater determinant, that is, with ansatzes possessing a mean-field nodal surface. We compare neural-network variants with traditional functional forms, as well as with DMC results. In particular we investigate the PauliNet, a recently proposed neural-network trial wavefunction[7]. PauliNet combines ideas from conventional trial wavefunctions, such as a symmetric Jastrow factor, a generalized backflow transformation, multi-determinant expansions, quantum chemistry baselines, and an explicit implementation of physical constraints of ground-state wavefunctions. Since PauliNet is a powerful instance of the general ansatz in (1), we can obtain traditional types of QMC ansatzes at different levels of the theory by deactivating certain trainable parts of PauliNet. This poses well-defined subproblems, which we use to investigate the performance of the individual components of the PauliNet ansatz.

The remaining paper is organized as follows. In Section II we review the general PauliNet ansatz and show how different levels of the model hierarchy (Fig.1) can be obtained. In Section III we use these instances of PauliNet to investigate several subproblems within the deep QMC approach. First we demonstrate that DNNs can be employed to correct the single-particle orbitals of a HF calculation in a small basis and obtain energies close to the CBS limit. Next we benchmark the deep Jastrow factor. We start by applying it to two node-less test systems, H_2 and He, where results within five significant digits of the exact energy are achieved. Next we conduct an extensive hyperparameter search for two systems with four electrons, LiH and the H_4 rectangle, revealing that the expressiveness of the ansatz can be systematically increased to converge to the fixed-node limit imposed by the employed antisymmetric ansatz. Finally we combine both the orbital correction and deep Jastrow factor in an analysis of the water molecule. Section IV discusses the results.

II. THEORY AND METHODS

A. PauliNet

The central object of our investigation is PauliNet, a neural-network trial wavefunction of the form in (1). PauliNet extends the traditional multi-determinant Slater–Jastrow–backflow-type trial wavefunctions, retaining physically motivated structural features while replacing ad-hoc parameterizations with highly expressive DNNs.

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = \mathrm{e}^{\gamma(\mathbf{r}) + J_{\boldsymbol{\theta}}(\mathbf{r})} \sum_p c_p \det[\tilde{\varphi}^{\uparrow}_{\boldsymbol{\theta},\mu_{pi}}(\mathbf{r})] \det[\tilde{\varphi}^{\downarrow}_{\boldsymbol{\theta},\mu_{pi}}(\mathbf{r})] \tag{3}$$

$$\tilde{\varphi}_{\boldsymbol{\theta},\mu i}(\mathbf{r}) = \varphi_\mu(\mathbf{r}_i) f^{\otimes}_{\boldsymbol{\theta},\mu i}(\mathbf{r}) + f^{\oplus}_{\boldsymbol{\theta},\mu i}(\mathbf{r}) \tag{4}$$

The model is initialized with orbitals φ_μ from a standard quantum-chemistry calculation, such as HF or MCSCF. This baseline is augmented with a Jastrow factor J_θ and the generalized backflow transformation f_θ, which are obtained from a joint equivariant many-body representation encoded by a message-passing neural network. The analytically known electron–nucleus and electron–electron cusp conditions[17] are enforced within the single-particle orbitals and as a fixed part γ of the Jastrow factor, respectively. The PauliNet ansatz is

trained in the VMC framework. This involves stochastically sampling minibatches of electron configurations and optimizing a loss function such as the variational energy. For more details of the training methodology see Ref. 7. Numerical calculations have been carried out with the DeepQMC Python package[18], with training hyperparameters as reported in Table A1. Next we show how to obtain the ansatzes of Fig. 1 from the general PauliNet architecture and introduce the respective optimization problems to be solved.

B. Deep orbital correction

The simplest way to approach the quantum many-body problem is by considering a mean-field theory. The HF method gives the optimal mean-field solution up to the finite-basis-set error of the employed basis set. A mean-field variant of the PauliNet architecture can be used to account for inaccuracies in the HF baseline, by introducing a real-space correction to the single-particle orbitals.

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = \det[\tilde{\varphi}^{\uparrow}_{\boldsymbol{\theta},\mu}(\mathbf{r}_i)]\det[\tilde{\varphi}^{\downarrow}_{\boldsymbol{\theta},\mu}(\mathbf{r}_i)] \tag{5}$$

$$\tilde{\varphi}_{\boldsymbol{\theta},\mu}(\mathbf{r}_i) = \varphi_{\mu}(\mathbf{r}_i) f^{\otimes}_{\boldsymbol{\theta},\mu}(\mathbf{r}_i) + f^{\oplus}_{\boldsymbol{\theta},\mu}(\mathbf{r}_i) \tag{6}$$

The functions $f^{\otimes}_{\boldsymbol{\theta}}$ and $f^{\oplus}_{\boldsymbol{\theta}}$ are DNNs that generate a multiplicative and an additive correction to the HF orbitals φ_{μ}, respectively. This approach is a special case of the generalized backflow transformation in (4), in which the backflow correction depends on the position of the i-th electron only. If Gaussian-type orbitals are used, it is common to correct the missing nuclear cusp at the coalescence points within the orbitals. We employ the cusp correction of Ma *et al.* [19] and construct the DNNs to be cuspless. Though the DNN could in principle approximate the orbitals from scratch, providing a baseline that ensures the correct asymptotics and offers a good initial guess reduces the training cost and makes the training process more robust. In the mean-field theory the energy at the CBS limit constitutes a benchmark for the best possible solution to the optimization problem.

C. Deep Jastrow factor

The Slater–Jastrow-type ansatz goes beyond the mean-field theory by introducing explicit electronic correlation. The symmetric Jastrow factor however cannot alter the nodal surface and the Slater–Jastrow-type ansatz hence is a many-body ansatz possessing a mean-field nodal surface.

$$\psi_{\boldsymbol{\theta}}(\mathbf{r}) = \mathrm{e}^{\gamma(\mathbf{r})+J_{\boldsymbol{\theta}}(\mathbf{r})}\det[\varphi^{\uparrow}_{\mu}(\mathbf{r}_i)]\det[\varphi^{\downarrow}_{\mu}(\mathbf{r}_i)] \tag{7}$$

The deep Jastrow factor $J_{\boldsymbol{\theta}}$ is obtained from the graph-neural-network many-body representation, which is at the core of the PauliNet architecture.

$$J_{\boldsymbol{\theta}}(\mathbf{r}) := \eta_{\boldsymbol{\theta}}\big(\textstyle\sum_i \mathbf{x}^{(L)}_i(\mathbf{r})\big) \tag{8}$$

The many-body embeddings $\mathbf{x}^{(L)}_i$ serve as input to a fully-connected neural network η_{θ}. The embeddings are the internal representation of the electrons within a message-passing neural network. The architecture is a modification of SchNet[20], a graph-convolutional neural network that was designed to meet the symmetry constraints resulting from the indistinguishability of quantum particles. It involves multiple smaller components, such as trainable arrays and fully-connected neural networks, whose full specification gives rise to a collection of hyperparameters. A subset of hyperparameters that we investigate in this paper is listed in Table A2. A comprehensive description of the architecture can be found in the appendix.

Benchmarking Jastrow factors comes with the difficulty of distinguishing errors arising from the nodal surface from those present due to a lack of expressiveness in the Jastrow factor. We evaluate the performance of the deep Jastrow factor with respect to FN-DMC results, as the FN-DMC algorithm gives the exact ground state of the Schrödinger equation under the fixed-node constraint.

D. Mean-field Jastrow factor

We furthermore implement a mean-field Jastrow factor, which constitutes another point in the space of ansatz classes (Fig. 1).

$$J_{\boldsymbol{\theta}}(\mathbf{r}) := \textstyle\sum_i \eta_{\boldsymbol{\theta}}\big(\mathbf{x}^{(L)}_i(\mathbf{r}_i)\big) \tag{9}$$

The mean-field Jastrow factor can optimize the one-electron density of the ansatz without modifying the nodal surface or introducing correlation. This equips us with an intermediate step in approaching the finite-basis-set limit, that can be used to relate the finite-basis-set error to the fixed-node error of the HF wavefunction.

III. RESULTS

A. Correct finite-basis-set errors of HF orbitals with a DNN

Our first experiment aims at employing the PauliNet ansatz to remove the finite-basis-set error from a HF calculation. We start from a HF baseline obtained in the small 6-31G basis set. Instead of introducing more basis functions, the PauliNet ansatz follows the alternative approach of correcting the orbitals directly in real space. We trained the mean-field variant of the PauliNet ansatz (Sec. II B) on H_2, He and Be. For all three test systems we obtained energies within 0.1mH of the extrapolated CBS limit (Fig. 2). This is a proof of principle that the conceptually simple approach of fitting the orbitals in real space can be applied in practice.

B. Exact solutions for two electron systems

Next we turn to modelling electron correlation with the deep Jastrow factor (Sec. II C). We start by evaluating the deep

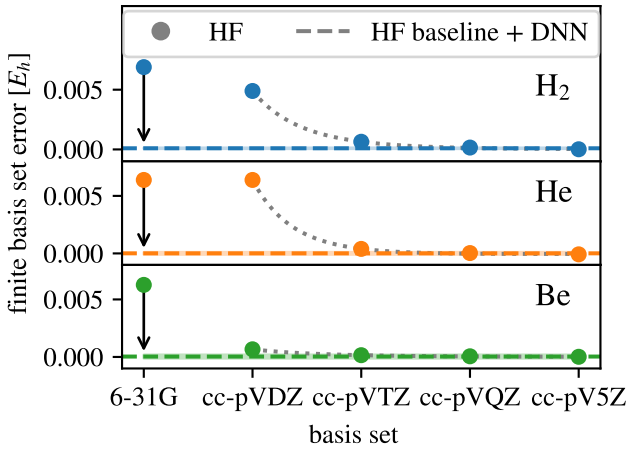

FIG. 2. **Removing the basis-set error of a HF calculation**. The error with respect to the estimated CBS limit of HF calculations with increasingly large basis sets, as well as the result of employing a DNN to correct the single-particle orbitals of a HF calculation in a small basis (6-31G) are shown. The statistical error of the Monte Carlo integration is smaller than the line width. The DNN is capable to correct deficiencies arising from the finite basis, producing energies close to the CBS limit.

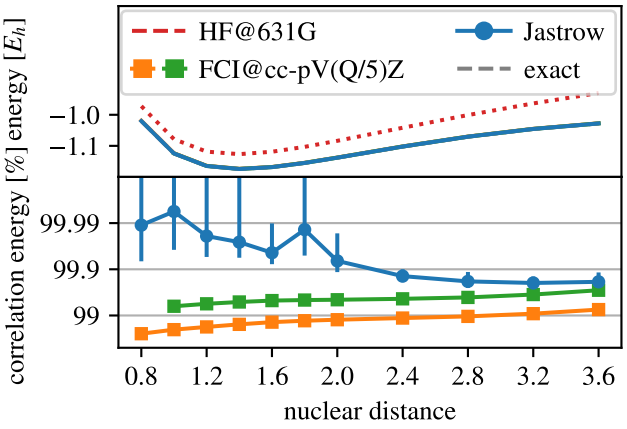

FIG. 3. **Dissociation curve of the hydrogen molecule.** Upper panel shows total energy. Exact results cannot be distinguished from FCI and the deep Jastrow factor. Lower panel shows the percentage of correlation energy recovered. Exact results were taken from Kołos and Wolniewicz [21] and FCI results were obtained with PySCF[23] in the cc-pVQZ basis (orange) and pp-pV5Z basis (green). Deep QMC surpasses the FCI accuracy for the entire dissociation curve.

Jastrow factor for H_2 and He, two-electron closed-shell systems for which the ground state is node-less (the antisymmetry comes from the spin part of the wavefunction only), such that the Jastrow factor is, in principle, sufficient to reproduce exact results. This yields a pure test for the expressiveness of the deep Jastrow factor. The recovered many-body correlation is measured by the correlation energy.

$$\eta = \frac{E_{\mathrm{VMC}} - E_{\mathrm{HF}}}{E_{\mathrm{exact}} - E_{\mathrm{HF}}} \tag{10}$$

TABLE I. **Results for two-electron node-less systems.**

system	deep Jastrow factor	exact energy	η [%]
H_2 ($d = 1.4$)	-1.17446(1)	-1.1744748[21]	99.97(3)
He	2.90372(1)	-2.9037247[22]	99.98(2)

For both systems we obtain energies matching five significant digits of the exact references (Table I). We evaluate the ansatz for various distances on the dissociation curve of H_2 (Fig. 3). Deep QMC outperforms FCI even with the large cc-pV5Z basis set, reducing the error in correlation energy by one to two orders of magnitude at compressed geometries and being still more accurate for stretched geometries, where the system exhibits strong correlation and the restricted HF baseline gives qualitatively wrong results. The results point out the difficulty of modelling dynamic correlation in Slater determinant space when applying purely second-quantized approaches and showcase the advantages of explicitly encoding many-body effects.

C. Systematically approach the fixed-node limit

The complexity of modelling correlation increases steeply with the number of particles. We evaluated the performance of the deep Jastrow factor for the hydrogen rectangle H_4 and LiH. While these four-electron systems exhibit more intricate interactions, they are computationally lightweight, such that the hyperparameter space of the deep ansatzes can be explored exhaustively. Modelling multiple same-spin electrons, the spacial wavefunction is no longer node-less and the single-determinant Slater–Jastrow ansatz possesses a fixed-node error. Instead of comparing to exact energies, we therefore measure the performance of the Jastrow factor with respect to the fixed-node limit estimated from FN-DMC calculations, reporting the fixed-node correlation energy.

$$\eta_{\mathrm{FN}} = \frac{E_{\mathrm{VMC}} - E_{\mathrm{HF}}}{E_{\mathrm{DMC}} - E_{\mathrm{HF}}} \tag{11}$$

As the fixed-node correlation energy is defined for ansatzes with an identical nodal surface, the nodes of the FN-DMC benchmark have to be reconstructed. For the mean-field nodal surface this implies starting from a HF computation with the same basis set. For H_4 we performed a comprehensive hyperparameter scan for our deep ansatz. The experiment revealed that by increasing the total number of trainable parameters, the fixed-node limit can be reached (Fig. A1). From these results several important hyperparameters for scaling the architecture can be identified, such as the depth of the DNN η_θ for obtaining the Jastrow factor from the many-body representation, the number of message passes L and the width of the convolutional kernel. Furthermore the experiment showed that the energy behaves smoothly with respect to changes in the hyperparameters and there are no strong mutual dependencies of hyperparameters.

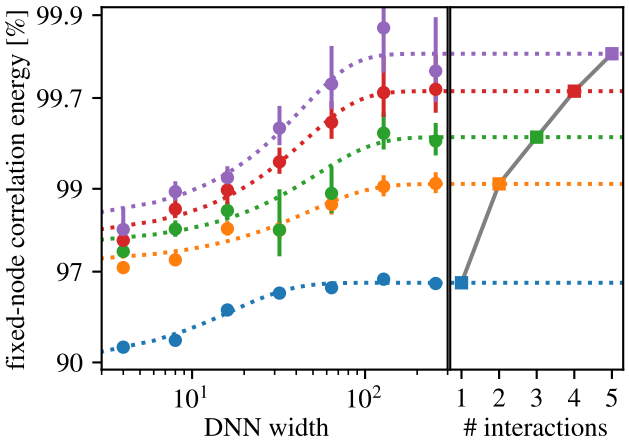

FIG. 4. **Approaching FN-DMC accuracy with deep Jastrow factor.** For LiH increasingly expressive Jastrow factors are trained to approach the FN-DMC energy of a single-determinant trial wavefunction[24]. The orbitals of the antisymmetric part are expressed in the TZP basis[25]. The most accurate trial wavefunctions give results within the sampling error of FN-DMC energy, indicated by the shaded region at the upper end of the graph. The dependence on the width of the convolutional kernel (DNN width) and the number of message passes (# iterations) in the Jastrow graph-neural-network is shown.

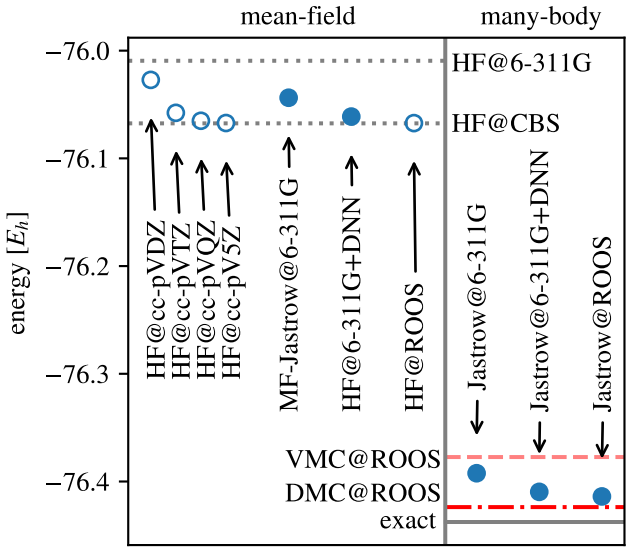

FIG. 5. **Deep Jastrow factor and orbital correction for H_2O.** All three restricted variants of PauliNet introduced in Methods are compared: the deep orbital correction (6-311G+DNN), the full deep Jastrow factor, and the mean-field (MF) Jastrow factor. ROOS denotes the Roos-aug-DZ-ANO basis set. VMC and DMC references are taken from Gurtubay and Needs [26]. HF@CBS and exact energy are taken from Rosenberg and Shavitt [27]. The Monte Carlo sampling error is smaller than the size of the markers.

The results were used to perform a thorough investigation of the convergence behavior on LiH, varying a subset of hyperparameters and fixing the remaining hyperparameters at suitable values. We show a systematic convergence to the fixed-node limit with an increasing size of the convolutional kernel as well as the number of message passes (Fig. 4). This is an indication that the deep Jastrow factor can be extended towards completeness in a computationally feasible way. The remaining fluctuations of the fixed-node correlation energy are caused by the stochasticity of the training and the sampling error of evaluating the energy of the wavefunction.

D. Application of different levels of theory to H_2O

The results for the small test systems showed that DNNs can be used to converge to the CBS limit within the mean-field theory, and that by adding correlation with a deep Jastrow factor the fixed-node limit can be approached. To investigate how these ansatzes behave for more complex systems, we evaluated the performance of the respective instances of PauliNet on the water molecule (Fig. 5).

We again start with the mean-field theory and consider the finite-basis-set error. We corrected the HF orbitals in the small 6-311G basis set with the deep orbital correction (Sec. II B), which recovered 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 factor (Sec. II D), which can recover about half of the finite-basis-set error only. This suggests that upon approaching the finite-basis-set limit the nodal surface is altered significantly.

Next, we investigate single-determinant ansatzes with the full Jastrow factor. We benchmark the deep Jastrow factor with a HF determinant in the Roos augmented double-zeta basis (Roos-aug-DZ-ANO)[28], a basis set that is frequently used for calculations on H_2O and gives HF energies at the CBS 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 II).

In order to study how finite-basis-set errors manifest in the mean-field nodal surface of both the HF and the many-body ansatzes, we furthermore computed the energies of the deep Jastrow factor with a HF determinant in a 6-311G basis with and without the orbital correction. This demonstrates that both the orbital correction and the deep Jastrow factor can be combined to achieve energies close to the fixed-node limit of the best mean-field nodal surface, even starting from a minimal baseline. We recover 96.0(2)% of the fixed-node correlation energy with respect to the Roos-aug-DZ-ANO basis, which is expected to be close to the CBS limit. The results furthermore show that finite-basis-set errors of the HF calculations transfers directly to the many-body regime. The differences between the energies of the mean-field ansatzes match the differences of the respective Slater–Jastrow trial wavefunctions and errors due to finite-basis-set effects are not altered by the many-body correlation.

TABLE II. **Benchmarking single-determinant (SD) Slater–Jastrow (SJ) ansatzes on H_2O.**

reference	HF	VMC (SD-SJ)	DMC (SD-SJ)	η_{FN} [%]	basis set
PauliNet	−76.009	−76.3923(7)	–	91.2(2)[a]	6-311G
PauliNet	−76.0612	−76.4096(7)	–	96.0(2)[a]	6-311G+DNN
PauliNet	−76.0672	**−76.4139**(5)	–	**97.2**(1)[a]	Roos-aug-DZ-ANO
Gurtubay and Needs [26]	−76.0672	−76.3773(2)	−76.42376(5)	87.01(6)	Roos-aug-DZ-ANO
Benedek *et al.* [29]	−76.063	−76.327(5)	−76.4219(1)	73.3(1)	Roos-aug-DZ-ANO
Gurtubay *et al.* [30]	−76.0587	−76.327(1)	−76.42102(4)	73.5(3)	6–311++G(2d,2p)

[a] The fixed-node correlation energy is computed with respect to the reference FN-DMC energy of Gurtubay and Needs [26].

IV. DISCUSSION

Overall we have demonstrated that there are no fundamental limitations in the investigated components of the PauliNet ansatz. For both the deep orbital correction and deep Jastrow factor, close to exact energies for the respective level of theory can be obtained. This highlights the generality and 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 trial wavefunction architecture. Though the results emphasize the potential of the deep QMC approach, the major benefit of deep QMC over FN-DMC calculations remains that deep QMC can in principle go beyond the fixed-node approximation by faithfully representing the nodal surface upon using multiple determinants and introducing many-body correlation at the level of the orbitals with the generalized backflow transformation. The presented analysis paves the way for future investigations on how the full PauliNet ansatz improves the nodes and overcomes the fixed-node limitations.

DATA AVAILABILITY

The data that support the findings of this study are openly available at `http://doi.org/10.6084/m9.figshare.13077158.v1`.

ACKNOWLEDGMENTS

We gratefully acknowledge funding and support from the European Research Commission (ERC CoG 772230), the Berlin Mathematics Research Center MATH+ (Projects EF1-2, AA2-22) and the German Ministry for Education and Research (Berlin Institute for the Foundations of Learning and Data BIFOLD). J. H. would like to thank K.-R. Müller for support and acknowledge funding from TU Berlin.

[1] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).
[2] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, Nat. Commun. **10**, 5024 (2019).
[3] M. Gastegger, A. McSloy, M. Luya, K. T. Schütt, and R. J. Maurer, J. Chem. Phys. **153**, 044123 (2020).
[4] G. Carleo and M. Troyer, Science **355**, 602 (2017).
[5] K. Choo, A. Mezzacapo, and G. Carleo, Nature Communications **11**, 2368 (2020).
[6] J. Han, L. Zhang, and W. E, Journal of Computational Physics **399**, 108929 (2019).
[7] J. Hermann, Z. Schätzle, and F. Noé, Nature Chemistry **12**, 891 (2020).
[8] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, Phys. Rev. Research **2**, 033429 (2020).
[9] 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," (2020), preprint at http://arxiv.org/abs/2007.14282.
[10] J. Han, Y. Li, L. Lin, J. Lu, J. Zhang, and L. Zhang, "Universal approximation of symmetric and anti-symmetric functions," (2019), preprint at http://arxiv.org/abs/1912.01765.
[11] M. Hutter, "On Representing (Anti)Symmetric Functions," (2020).
[12] K. Choo, A. Mezzacapo, and G. Carleo, Nat. Commun. **11**, 2368 (2020).
[13] N. D. Drummond, M. D. Towler, and R. J. Needs, Phys. Rev. B **70**, 235119 (2004).
[14] R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos, J. Phys.: Condens. Matter **22**, 023201 (2009).
[15] A. Barr, W. Gispen, and A. Lamacraft, in *Proceedings of Machine Learning Research*, Vol. 107 (2020) pp. 635–653.
[16] J. Han, J. Lu, and M. Zhou, "Solving high-dimensional eigenvalue problems using deep neural networks: A diffusion Monte Carlo like approach," (2020), preprint at http://arxiv.org/abs/2002.02600.
[17] T. Kato, Communications on Pure and Applied Mathematics **10**, 151 (1957).
[18] J. Hermann, Z. Schätzle, and F. Noé, "Deepqmc/deepqmc: DeepQMC 0.2.0," Zenodo (2020), doi:10.5281/zenodo.3992004.
[19] A. Ma, M. D. Towler, N. D. Drummond, and R. J. Needs, J. Chem. Phys. **122**, 224322 (2005).
[20] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, J. Chem. Phys. **148**, 241722 (2018).
[21] W. Kołos and L. Wolniewicz, J. Chem. Phys. **43**, 2429 (1965).
[22] T. Kinoshita, Phys. Rev. **115**, 366 (1959).
[23] 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, J. Chem. Phys. **153**, 024109 (2020).
[24] M. Casalegno, M. Mella, and A. M. Rappe, J. Chem. Phys. **118**, 7193 (2003).
[25] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus, J. Chem. Inf. Model. **59**, 4814 (2019).
[26] I. G. Gurtubay and R. J. Needs, J. Chem. Phys. **127**, 124306 (2007).
[27] B. J. Rosenberg and I. Shavitt, The Journal of Chemical Physics **63**, 2162 (1975).
[28] P.-O. Widmark, P.-Å. Malmqvist, and B. O. Roos, Theoret. Chim. Acta **77**, 291 (1990).

[29] N. A. Benedek, I. K. Snook, M. D. Towler, and R. J. Needs, J. Chem. Phys. **125**, 104302 (2006).
[30] I. G. Gurtubay, N. D. Drummond, M. D. Towler, and R. J. Needs, J. Chem. Phys. **124**, 024318 (2006).
[31] K. Gasperich, M. Deible, and K. D. Jordan, J. Chem. Phys. **147**, 074106 (2017).

APPENDIX

Message-passing neural-network architecture

At the level of quantum chemistry a molecule can be represented as a complete graph, where nodes correspond to electrons and nuclei and the distances of each pair of particles is assigned to the edge between their respective nodes. In the message-passing neural network each node is associated with an embedding, that is successively updated to gradually encode many-body correlations. At every iteration a filter-generating function $\mathbf{w}_\theta$ creates messages based on an expansion of the distance stored as edge feature of the interacting entities. The messages are exchanged throughout the graph and the electronic embeddings are updated based on the incoming information.

$$\begin{aligned}
\mathbf{x}_i^{(0)} &:= \mathbf{X}_{\boldsymbol{\theta},s_i} \\
\mathbf{z}_i^{(n,\pm)} &:= \textstyle\sum_{j\neq i}^{\pm} \mathbf{w}_{\boldsymbol{\theta}}^{(n,\pm)}\left(\mathbf{e}(|\mathbf{r}_i-\mathbf{r}_j|)\right) \odot \mathbf{h}_{\boldsymbol{\theta}}^{(n)}\left(\mathbf{x}_j^{(n)}\right) \\
\mathbf{z}_i^{(n,\mathrm{n})} &:= \textstyle\sum_I \mathbf{w}_{\boldsymbol{\theta}}^{(n,\mathrm{n})}\left(\mathbf{e}(|\mathbf{r}_i-\mathbf{R}_I|)\right) \odot \mathbf{Y}_{\boldsymbol{\theta},I} \\
\mathbf{x}_i^{(n+1)} &:= \mathbf{x}_i^{(n)} + \textstyle\sum_{\pm} \mathbf{g}_{\boldsymbol{\theta}}^{(n,\pm)}\left(\mathbf{z}_i^{(n,\pm)}\right) + \mathbf{g}_{\boldsymbol{\theta}}^{(n,\mathrm{n})}\left(\mathbf{z}_i^{(n,\mathrm{n})}\right)
\end{aligned} \tag{12}$$

The functions $\mathbf{w}_\theta$, $\mathbf{h}_\theta$ and $\mathbf{g}_\theta$ are represented by fully-connected neural networks and the embeddings $\mathbf{X}_{\boldsymbol{\theta},s_i}$ and $\mathbf{Y}_{\boldsymbol{\theta},I}$ are trainable arrays. All fully-connected neural networks are chosen with a logarithmic scaling of the width from input to output dimension. The initial electronic embeddings $\mathbf{X}_{\boldsymbol{\theta},s_i}$ are shared among same-spin electrons. The distance expansion $\mathbf{e}$ is designed to make the architecture cusp less, as to retain the cusp explicitly implemented by the cusp corrections. The exchange of messages is repeated L times. The continuous-filter convolutions over the molecular graph make the electronic embeddings equivariant with respect to the exchange of same-spin electrons.

$$\mathscr{P}_{ij}\mathbf{x}_i(\mathbf{r}) = \mathbf{x}_j(\mathscr{P}_{ij}\mathbf{r}) \tag{13}$$

This facilitates to construct the architecture as to intrinsically comply with the symmetry requirements of fermionic wavefunctions, without imposing further requirements on the parameters of the model. In order to retain the antisymmetry of the wavefunction the Jastrow factor must be permutation invariant with respect to the exchange of two same-spin electrons. Contracting the the permutation equivariant many-body embeddings $\mathbf{x}_i^{(L)}$ after the message-passing phase guaranties the final Jastrow factor to be symmetric.

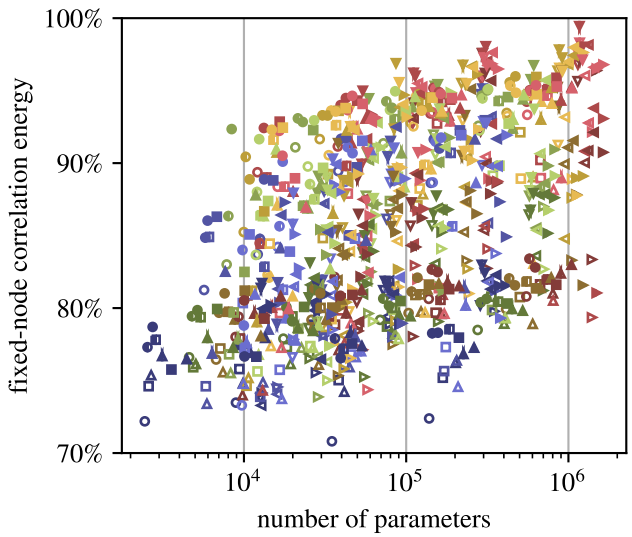

FIG. A1. **Hyperparameter scan H4 rectangle.** The figure maps different combinations of hyperparameters in the space of accuracy with respect to the number of total trainable parameters. The parameters varied in the scan are listed in Table A2. The fixed-node correlation energy is obtained with respect to FN-DMC results from[31], where a single-determinant trial wavefunction with a basis consisting of s functions from the cc-pV5Z basis set and the p and d functions from the cc-pVTZ basis set was used. Increasing the number of trainable parameters the fixed-node limit is approached.

TABLE A1. **Hyperparameters used in numerical calculations.**

Hyperparameter	Value
One-electron basis	6-31G
Dimension of $\mathbf{e}$	16
Dimension of $\mathbf{x}_i$	128
Number of interaction layers L	4
Number of layers in $\eta_{\boldsymbol{\theta}}$	3
Number of layers in $\mathbf{w}_{\boldsymbol{\theta}}$	1
Number of layers in $\mathbf{h}_{\boldsymbol{\theta}}$	2
Number of layers in $\mathbf{g}_{\boldsymbol{\theta}}$	2
Batch size	2000
Number of walkers	2000
Number of training steps	H_4: 5000 H_2: 10000 He: 10000 Be: 10000 LiH: 10000 H_2O: 15000
Optimizer	AdamW
Learning rate scheduler	CyclicLR
Minimum/maximum learning rate	0.0001/0.01
Clipping window q	5
Epoch size	100
Number of decorrelation sampling steps	4
Target acceptance	57%

TABLE A2. **Hyperparameters of the deep Jastrow factor.**

component	type	hyperparameter
$\mathbf{X}_{\boldsymbol{\theta},s_i}$ – electronic embedding	trainable array	embedding dimension
$\mathbf{Y}_{\boldsymbol{\theta},I}$ – nuclear embeddings	trainable array	kernel dimension
$\mathbf{e}$ – distance expansion	fixed function	distance features
$\mathbf{w}_\theta$ – filter generating function	DNN	(distance features $\rightarrow$ kernel dimension), depth
$\mathbf{h}_\theta$ – transformation embedding to kernel space	DNN	(embedding dimension $\rightarrow$ kernel dimension), depth
$\mathbf{g}_\theta$ – transformation kernel to embedding space	DNN	(kernel dimension $\rightarrow$ embedding dimension), depth
$\eta_{\boldsymbol{\theta}}$ – Jastrow network	DNN	(embedding dimension $\rightarrow$ 1), depth
full architecture	—	number of interactions

TABLE A3. **Geometries of test systems.**

Molecule	Atom	position [Å]
LiH	Li	(0.000, 0.000, 0.000)
	H	(1.595, 0.000, 0.000)
H_4	H	(0.000, -0.635, 0.000)
	H	0.000, 0.635, 0.000)
	H	(2.400, -0.635, 0.000)
	H	(2.400, 0.635, 0.000)
H_2O	O	(0.00000, 0.00000, 0.00000)
	H	(0.75695, 0.58588, 0.00000)
	H	(-0.75695, 0.58588, 0.00000)