Vandermonde Wave Function Ansatz for Improved Variational Monte Carlo

Alberto Acevedo
Dept of Mathematics
University of Arizona
Tucson, AZ, USA
albertoacevedo@math.arizona.edu

Michael Curry
Dept of Computer Science
University of Maryland
College Park, MD, USA
curry@cs.umd.edu

Shantanu H. Joshi
Depts of Neurology and Bioengineering
University of California, Los Angeles
Los Angeles, CA, USA
s.joshi@g.ucla.edu

Brett Leroux
Dept of Mathematics
University of California, Davis
Davis, CA, USA
beleroux@ucdavis.edu

Nicholas Malaya
AMD Research
Advanced Micro Devices, Inc
Austin, TX, USA
nicholas.malaya@amd.com

***Abstract*—Solutions to the Schrödinger equation can be used to predict the electronic structure of molecules and materials and therefore infer their complex physical and chemical properties. Variational Quantum Monte Carlo (VMC) is a technique that can be used to solve the weak form of the Schrödinger equation. Applying VMC to systems with N electrons involves evaluating the determinant of an N by N matrix. The evaluation of this determinant scales as $O(N^3)$ and is the main computational cost in the VMC process. In this work, we investigate an alternative VMC technique based on the Vandermonde determinant. The Vandermonde determinant is a product of pairwise differences and so evaluating it scales as $O(N^2)$. Therefore, this approach reduces the computational cost by a factor of N.**

The Vandermonde determinant was implemented in PyTorch and the performance was assessed in approximating the ground state energy of various quantum systems against existing techniques. The performance is evaluated in a variety of systems, starting with the one-dimensional particle in a box, and then considering more complicated atomic systems with multiple particles. The Vandermonde determinant was also implemented in PauliNet, a deep-learning architecture for VMC. The new method is shown to be computationally efficient, and results in a speed-up as large as 5X. In these cases, the new ansatz obtains a reasonable approximation for wavefunctions of atomic systems, but does not reach the accuracy of the Hartree-Fock method that relies on the Slater determinant. It is observed that while the use of neural networks in VMC can result in highly accurate solutions, further work is necessary to determine an appropriate balance between computational time and accuracy.

***Index Terms*—Schrödinger equation, Vandermonde ansatz, Variational Monte Carlo,**

I. Introduction

Variational Monte Carlo (VMC) is an established technique for computing the ground-state energy of quantum systems. It is one algorithm provided by major HPC applications for quantum Monte Carlo such as QMCPACK [1]. Recently, a number of papers [2]–[4] have recognized that techniques from deep learning can be adopted naturally into the VMC setting. In VMC, the wavefunction of a quantum system is approximated as a parametric function, while modern neural networks make excellent function approximators.

When applying VMC to many-electron atomic and molecular systems, it is typical to use the Born-Oppenheimer approximation and assume the nuclei have fixed coordinates relative to the electrons; then a number of single-electron orbitals are modeled, with separate terms to account for the interactions of electrons.

A particularly important property to enforce is antisymmetry – under exchanges of same-spin electrons, the sign of the wavefunction must undergo flipping. The typical way to enforce antisymmetry is to arrange the single-electron wavefunctions in an aggregate Slater matrix, whose determinant is computed with a computational cost of $O(N^3)$, this is seen in section III subsection B. This technique forms the basis of the neural architectures described in [2], [4], and asymptotically dominates the computational costs.

The Vandermonde determinant has been proposed as an alternative construction for enforcing antisymmetry [5]. Since the Vandermonde determinant can be written as a product of pairwise differences, it can be evaluated using only $O(N^2)$ operations. In this paper, we investigate the use of the Vandermonde determinant as an alternative means of enforcing antisymmetry when solving Schrödinger's equation and also implement it under the framework of the PauliNet architecture [4].

II. Background

A. Variational Monte Carlo

Variational Monte Carlo (VMC) is a technique that can be used to estimate the ground state of a quantum system. We assume that we have access to the Hamiltonian H that characterizes the system. We would like to approximate the ground state wavefunction $|\psi\rangle$, also written ψ when working in the dual space, which is the solution of the time-independent Schrödinger equation

$$H\,|\psi\rangle = E\,|\psi\rangle$$

978-0-7381-1052-3/20/$31.00 ©2020 IEEE
DOI 10.1109/DLS51937.2020.00010

for the lowest-energy state of the system. The variational principle of quantum mechanics states that all wavefunctions ϕ other than the true ground state wavefunction Ψ_0 yield higher expectation values for the energy: $\mathbb{E}_{\Psi_0}[E] \leq \mathbb{E}_{\phi}[E]$. Variational Monte Carlo uses this principle to estimate the ground state wavefunction and ground state energy. The procedure, roughly speaking, consists of the following steps (as described in [2]):

1) Start with some initial variational ansatz $|\psi_\alpha(x)\rangle$ from some class of functions parameterized by α.
2) Using Markov-chain Monte Carlo (MCMC) techniques [6], [7], sample from the probability distribution implied by this wavefunction: $\pi(x) \propto ||\psi_\alpha(x)\rangle|^2$.
3) Minimize the expected value of the energy of the system

$$\mathbb{E}[E] = \frac{\langle\psi_\alpha|H|\psi_\alpha\rangle}{\langle\psi_\alpha|\psi_\alpha\rangle} = \frac{\int \psi_\alpha^*(x) H \psi_\alpha(x) dx}{\int \psi_\alpha^*(x)\psi_\alpha(x)dx}.$$

Using the samples $x_i \sim \pi(x)$, we can approximate this expected value by computing

$$\frac{1}{N}\sum_{i=1}^{N} \psi_\alpha^*(x) H \psi_\alpha(x)$$

and approximate the variance by computing

$$\frac{1}{N}\sum_{i=1}^{N} (\psi_\alpha^*(x) H \psi_\alpha(x) - \mathbb{E}[E])^2$$

and optimize the parameters α of the ansatz to minimize some combination of these losses.

B. Deep neural network ansatz

PauliNet [4] is a neural network architecture for solving N-body electron systems (i.e., molecular and atomic systems). It uses deep learning techniques to approximate electronic wavefunctions. Although it uses neural networks, it is also physically inspired and borrows ideas from traditional techniques. In particular, it effectively incorporates the Hartree-Fock method which is a baseline analytic approximate solution to a simplified version of the appropriate Schrödinger equation. Below, we provide a high-level overview of the architecture of PauliNet, and the procedure for training it to minimize the ground-state energy. Then, we give more detailed explanations of specific important properties of its architecture.

PauliNet uses a Slater-Jastrow-backflow ansatz. Ansatzes of this general form have long been popular for variational Monte Carlo, and are widely used including in QMCPACK [1]. PauliNet follows this basic structure but replaces various parts of it with neural networks. It constructs the ansatz in the following form:

$$\begin{aligned} \Psi_\theta(r) &= e^{\gamma(r)+J_\theta(r)} \sum_p c_p \det(\tilde{\varphi}^{\uparrow}_{\mu_p}(r)) \det(\tilde{\varphi}^{\downarrow}_{\mu_p}(r)) \\ \tilde{\varphi}_\mu(r) &= \varphi_\mu(r_i) f_{\mu,\theta}(r) \end{aligned} \tag{1}$$

This expression is described in more detail below.

- The μ denotes an index when there are multiple copies of certain basis sets or other functions.
- $\varphi_\mu(r)$ are Hartree-Fock functions, a well-known approximate solution derived by making simplifying assumptions to the physical problem until it becomes analytically tractable; they are essentially treated as inputs. They are modified to enforce nuclear cusp constraints, and they naturally enforce decay of the wavefunction to zero as electrons stray far from the nucleus.
- The terms J_θ and $f_{\mu,\theta}$ are learnable neural networks; the $f_{\mu,\theta}$ are the backflow factors which model electron-electron interactions and scale the basis functions; the J_θ is the Jastrow factor which models global correlations between electrons but does not affect antisymmetry.
- The γ are fixed, physical cusp constraints: as electrons approach each other the wavefunction must obey certain physical behavior.
- The rescaled φ are combined using Slater determinants, split according to spin up ($\uparrow$) or spin down ($\downarrow$) electrons (antisymmetry only applies to electrons with the same spin).
- There may be multiple copies of each Slater determinant, which are indexed by p and combined in a weighted sum. We focus on the single-determinant case for most of our work, however.
- Implicit in the implementation of Equation (1) is the use of SchNet [8] to transform electron coordinate inputs into a 128-dimensional vector for input into f_θ and J_θ.

The architecture combines physically motivated inputs and architectural choices with flexible transformations represented by general function approximators.

III. Vandermonde and Slater Ansatz on Simple Systems

A. Getting started

As a testbed for comparing different ansatzes, we implemented our own variational Monte Carlo pipeline. This uses a simple Metropolis-Hastings sampler, and the standard "score function" stochastic gradient descent estimator, also used in PauliNet:

$$\begin{aligned} \mathcal{L}(\theta) &= \mathbb{E}_{\boldsymbol{r}\sim|\Psi_\theta(\boldsymbol{r})|^2}\left[E_{\text{loc}}[\Psi_\theta](\boldsymbol{r})\right] \\ \nabla_\theta \mathcal{L}(\theta) &= 2\mathbb{E}_{r\sim|\Psi_\theta|^2}\left[(E_{\text{loc}}[\Psi_\theta](\boldsymbol{r}) - \mathcal{L}(\theta))\nabla_\theta \log|\Psi_\theta|\right] \end{aligned}$$

We experiment with a number of simple physical systems but primarily focus on systems where antisymmetry of the wave function is a concern.

B. Antisymmetry: Slater and Vandermonde determinants

Due to the Pauli exclusion principle, it is standard to require that the wavefunction Ψ for a system of fermions obey antisymmetry. The traditional way to enforce this is to employ a Slater determinant constructed from the components of a Hartree product [9, section 14.5] as presented below.

$$\Psi_{HP}(r_1, r_2, ..., r_N) = \phi_1(r_1)\phi_2(r_2)...\phi_N(r_N).$$

The ϕ_i are hydrogenic functions. And the Slater determinant is

$$\Psi_{Slater} = \begin{vmatrix} \phi_1(r_1) & \phi_2(r_1) & \cdots & \phi_N(r_1) \\ \phi_1(r_1) & \phi_2(r_2) & \cdots & \phi_N(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(r_N) & \phi_2(r_N) & \cdots & \phi_N(r_N) \end{vmatrix}$$

However, computing a determinant has a complexity of $O(N^3)$, which is costly.

An alternate method of enforcing antisymmetry is the Vandermonde determinant. This has been proposed, for instance, in [5]. In brief, the Vandermonde determinant takes as input a map $(\phi(r_1), \phi(r_2), \cdots, \phi(r_N))$ and takes the determinant of the following matrix:

$$\det_V = \begin{vmatrix} 1 & \phi(r_1)^1 & \phi(r_1)^2 & \cdots & \phi(r_1)^{N-1} \\ 1 & \phi(r_2)^1 & \phi(r_2)^2 & \cdots & \varphi(r_2)^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \phi(r_N)^1 & \phi(r_N)^2 & \cdots & \phi(r_N)^{N-1} \end{vmatrix}$$

This is a determinant, and assuming the vector of φ is permutation equivariant (as it is in our applications), exchanging two inputs will exchange rows of the matrix, swapping signs of the determinant, so that antisymmetry is enforced.

The Vandermonde determinant has an advantageous property: one does not need to write out the full matrix and actually take its determinant; instead, it is possible to use the following expression, requiring only $O(N^2)$ operations:

$$\det_V = \prod_{i<j} (\varphi(r_i) - \varphi(r_j)).$$

Thus the Vandermonde determinant is a potentially efficient technique for enforcing antisymmetry with a lower computational cost than the widely-used Slater determinant.

C. Deficiencies of the Vandermonde ansatz

When trading the Slater ansatz for the Vandermonde ansatz we constrained ourselves to the physical information carried by a single ϕ_i in the array used to compute the Slater determinant. To overcome this loss of physical information, the Vandermonde determinant can be multiplied by an appropriate symmetric function in order to add more physical information to the ansatz such as the boundary conditions imposed by the system. See, for example, the ansatz for the two fermions in a box system in Equation (4).

D. Two fermions in a box

For our experiments, we considered the following physical systems:

- To verify our solutions, we considered the toy problem of two fermions in a box – a slight variation on a textbook problem in quantum mechanics.
- We also considered n-electron systems: Helium (2 electrons) and Lithium (three electrons).

We will return to these models in the experimental results section, for now we will develop the two fermions in a box system as a base for what is to come. The two fermions in a box system consists of two fermions (particles where the Pauli exclusion principle applies) in a box (finite volume infinite potential well) that the fermions are restricted to. The system is relatively simple and analytically tractable – in fact, for the most naturally defined ansatzes the local energy does not depend on the particle configuration, making variational Monte Carlo totally unnecessary. This model is nevertheless a good initial testing ground for VMC. As a concrete example, we consider the Schrödinger equation for the 1-D box case,

$$-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2}\right)\Psi(x_1, x_2) = E\Psi(x_1, x_2).$$

The time independent Schrödinger in this case amounts to the simple eigenvalue problem corresponding to the Laplacian. The Laplacian indeed has a discrete spectrum over a bounded domain. For a 1-dimensional box of length L and the particles bounded to the domain [0, L], the eigenvalue/eigenvector pairs are the energies $E_{n_1,n_2} = \frac{\hbar^2\pi^2}{2L^2}(n_1^2 + n_2^2)$ with the corresponding states $\Psi_{n_1,n_2}(x_1, x_2) = \sqrt{\frac{2}{L}}\sin(\frac{n_1\pi x_1}{L})\sqrt{\frac{2}{L}}\sin(\frac{n_2\pi x_1}{L})$. Let us now calculate the corresponding local energy:

$$E_{loc} = \frac{1}{\Psi_{n_1,n_2}(x_1, x_2)}\{-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_1^2}\right)\Psi_{n_1,n_2}(x_1, x_2)\}$$

This local energy is independent of configuration, therefore structurally for an ansatz will yield trivial results: VMC would be completely superfluous as samples from the wavefunction would serve no purpose. The same issues arise even after implementing antisymmetry via a Slater determinant. In order to test VMC, in Section V-A, we devise a deliberately deficient ansatz, using both Slater and Vandermonde determinants. Below, we denote parameters by α_1, α_2, etc. After working through this toy model we will consider the Helium atom, Lithium, Beryllium and then finally Lithium hydride using the conventional Slater determinant techniques as well as our Vandermonde approach.

IV. Vandermonde Ansatz in PauliNet

A. Vandermonde determinants

The main motivation for replacing the Slater determinants with Vandermonde determinants in PauliNet is in reducing the computational cost of evaluating the local energy. In PauliNet, the cost of evaluating the wavefunction scales as $O(N^3)$ where N is the number of electrons. This is because the cost of computing the Slater determinants is N^3 and the determinant calculations dominate the cost of evaluating the wavefunction. Due to the kinetic energy, evaluating the local energy scales as N times the complexity of evaluating the wavefunction. Therefore, the computational cost of PauliNet scales asymptotically as $O(N^4)$.

In contrast, evaluating the Vandermonde determinants scales as $O(N^2)$. Therefore, replacing the Slater determinants in PauliNet with Vandermonde determinants reduces the computational cost from $O(N^4)$ to $O(N^3)$.

In the next sections, we outline our implementation of the Vandermonde determinant in PauliNet. We compare performance on various atomic systems including Beryllium and Lithium hydride. Finally we provide a visual comparison of the wavefunctions obtained by using PauliNet with the Vandermonde determinant and PauliNet with the Slater determinant.

B. General approach

There are two Slater determinants in the PauliNet ansatz. Instead of taking a Slater determinant, our modification to PauliNet first takes a product of the $\tilde{\varphi_\mu}$ in eq. (1) and then takes the Vandermonde determinant of the product. That is, letting $\tilde{\varphi} = \prod_{\mu=1}^{N} \tilde{\varphi}_\mu(r)$, we set

$$\det_V := \prod_{i<j} (\tilde{\varphi}(r_i) - \tilde{\varphi}(r_j)),$$

and replace every Slater determinant in PauliNet with the corresponding Vandermonde determinant $\det_V$.

C. Enforcement of cusp and other conditions

In order for antisymmetry to be enforced, the $\tilde{\varphi}$ must be permutation equivariant. The same is true for the Vandermonde determinant: this motivates our use of the products $\prod_{\mu=1}^{N} \tilde{\varphi}_\mu(r)$, which incorporate information from the whole basis while ensuring permutation equivariance – exchanging two electrons will exchange only the positions of their respective $\tilde{\varphi}$ without changing values.

While this construction can ensure that antisymmetry is enforced, it has a drawback compared to the Slater determinant. If any electron violates the boundary conditions, its $\tilde{\varphi}$ goes to zero: then every term in the Slater determinant will go to zero, while the same is not true of the Vandermonde determinant. After observing and debugging unexpected samples of configurations with infinite or zero energy, we realized that it was necessary to have a separate term outside the Vandermonde determinants enforcing these properties.

Since the underlying Hartree-Fock functions $\varphi)\mu$ enforce these conditions, we simply select a single φ_μ and multiply the entire wavefunction by $\prod_{i=1}^{N} |\varphi_\mu(r_i)|$ – thus ensuring that if any electron violates cusp or boundary conditions, the entire wavefunction still goes to zero. This term, which falls outside of the determinants, is a symmetric function as required.

V. Experimental Results

A. VMC for simple systems

For two fermions in a box we must make sure that the wave function satisfies antisymmetric conditions. Recall that the Pauli exclusion principle tells us that two particles cannot occupy the same state. To enforce this we must find a function, $\Psi_{excited}(x_1)$ for particle one, that has the same geometry as the second excited state and use this in conjunction with the first excited state wave function,$\Psi_{ground}(x_2)$, to create a total ground state ansatz for our two particle system, $\Psi_{total} = \Psi_{excited}(x_1)\Psi_{ground}(x_2)$. A possible pair of ansatz are $\Psi_{excited}(x) = -(1-x^{2\alpha_1})x$ and $\Psi_{ground}(x) = (1-x^{2\alpha_2})$. Finally, implementing a Slater determinant on the appropriate Slater matrix afforded by the above ansatz we may achieve antisymmetry.

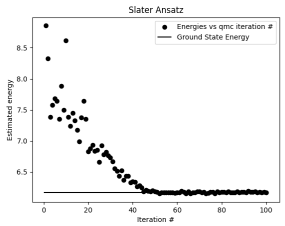

Fig. 1: Two particles in a box. Gradient descent for the Slater Ansatz. Using the ansatz in Equation (2) we estimate the ground state energy.

Figure 1 shows the training results for the following ansatz, which is a 2-dimensional Slater determinant of the Slater matrix constructed from the ansatz $\Psi_{excited}$ and Ψ_{ground} discussed above:

$$\begin{aligned}\Psi(x_1, r_2) = (1 - x_1^{2\alpha_1})(1 - x_2^{2\alpha_2})x_2 \\ - (1 - x_2^{2\alpha_1})(1 - x_1^{2\alpha_2})x_1\end{aligned} \tag{2}$$

Figure 2 shows the training results for another Slater-based ansatz, with slightly different basis:

$$\Psi(x_1, x_2) = (1 - x_1^{2\alpha_1})e^{-\alpha_2 x_2^2} - (1 - x_2^{2\alpha_1})e^{-\alpha_2 x_1^2} \tag{3}$$

as well as the results for a basic Vandermonde-based ansatz:

$$\begin{aligned}\Psi(x_1, x_2) = (1 - x_1^2)^{\alpha_1}(1 - x_2^2)^{\alpha_1} \\ \times (\sin(\alpha_2 x_1) - \sin(\alpha_2 x_2)).\end{aligned} \tag{4}$$

Both Slater ansatz cases achieve convergence to the true

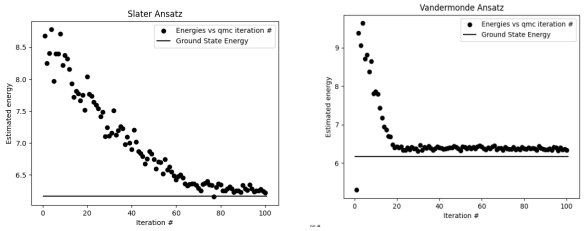

Fig. 2: Convergence plots for the two-fermions-in-a-box system near the ground state energy using the Slater ansatz from Equation (3) (left) and the Vandermonde ansatz from Equation (4) (right).

ground state energy as seen in plots on Figure 1 and Figure 2. Despite not fully converging to the ground truth, the Vandermonde ansatz allows for speed up in convergence. Figure 2 shows on the right a relatively rapid convergence in vicinity of the true ground state energy, on the other hand, the Gaussian-type Slater ansatz on the left of Figure 2 converges about 20 iterations after our Vandermonde type ansatz, Equation (4), has already converged. It is surprising nonetheless, that the ansatz used in the leftmost figure of Figure 2 converges. The Gaussian terms do not satisfy the necessary boundary

conditions, but the variational parameters optimize to alleviate this issue. Despite the Vandermonde ansatz's ability to yield the fastest converging approximate energy it is the weakest of the ansatzes, in this two particle in a box case, due to the relatively significant discrepancy between the attained approximate ground state energy and the true ground state energy.

We now consider a few atomic systems. Due to the Pauli exclusion principle we will need to force antisymmetry via a physically inspired ansatz.

The traditional Helium Slater ansatz is

$$\Psi(r_1, r_2) = e^{-\alpha_1 r_1} e^{-\alpha_2 r_2} - e^{-\alpha_1 r_2} e^{-\alpha_2 r_1}. \tag{5}$$

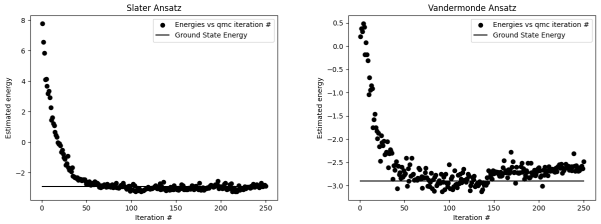

Fig. 3: Convergence plots for the Helium system using the Slater ansatz from Equation (5) (left) and the Vandermonde ansatz from Equation (6) (right) The times of realization are 11 min and 8 min, respectively. The Vandermonde type ansatz yields a significant speed up in computation time.

This ansatz uses *Hydrogenic* type wavefunctions, $\phi(r) = e^{-\alpha r}$, as a basis for the construction of the Slater determinant. Indeed, if the two electrons orbiting the nucleus of the Helium atom did not interact with one another, each electron would simply behave as if it were part of a hydrogen atom.

In Figure 3 we see the convergence of our estimated local energy to the true ground state energy for both the Slater and Vandermonde type ansatz. The corresponding Vandermonde ansatz is of the form presented below.

$$\Psi(r_1, r_2) = e^{-\alpha_1 r_1} e^{-\alpha_1 r_2} (e^{-\alpha_2 r_1} - e^{-\alpha_2 r_2}). \tag{6}$$

The component $e^{-\alpha_1 r_1} e^{-\alpha_1 r_2}$ is symmetric with respect to r_1 and r_2 while the term $(e^{-\alpha_2 r_1} - e^{-\alpha_2 r_2})$ yields the antisymmetry desired.

As a final test of the efficacy of the Vandermonde ansatz we perform gradient decent on the parameterized approximate energies afforded by both a Vandermonde and Slater determinant ansatz. Lithium Vandermonde is as follows

$$\Psi(r_1, r_2, r_3) = e^{-\alpha_1 r_1} e^{-\alpha_1 r_2} e^{-\alpha_1 r_3} \times \sum_{l=1}^{2} \left(\prod_{i<j} e^{-\alpha_{l+1} r_i} - e^{-\alpha_{l+1} r_j} \right). \tag{7}$$

The Slater determinant ansatz in this case is

$$\Psi(r_1, r_2, r_3) = \begin{vmatrix} e^{-\alpha_1 r_1} & e^{-\alpha_1 r_2} & e^{-\alpha_1 r_3} \\ e^{-\alpha_2 r_1} & e^{-\alpha_2 r_2} & e^{-\alpha_2 r_3} \\ e^{-\alpha_3 r_1} & e^{-\alpha_3 r_2} & e^{-\alpha_3 r_3} \end{vmatrix}. \tag{8}$$

Speedup due to the Vandermonde determinant			
	Two Fermions in a box	Helium	Lithium
Vandermonde versus Slater speedup	2.5	1.38	5.26

TABLE I: Comparison of the convergence speedup attained from the usage of a standard Salter determinant ansatz and the Vandermonde ansatz (Values are approximated from figures 2, 3 and 4). The Vandermonde type ansatz converges faster than the Slater ansatz.

The results can be seen in Figure 4. This result shows that we achieve faster convergence for the Vandermonde ansatz. We note that with more iterations we may see the Slater determinant ansatz further converge to the true ground state but this was not tested in our experiments. It also remains to be tested whether the variability around the true ground state energy for the Vandermonde ansatz is less than that of the Slater ansatz.

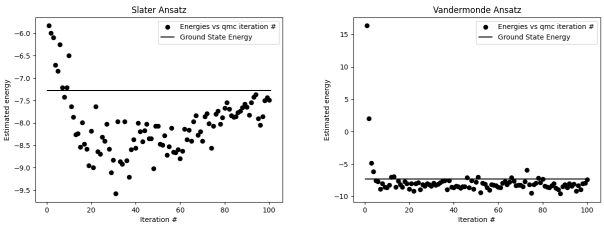

Fig. 4: Convergence plots for the Lithium (3-electron) system near the ground state energy using the Slater ansatz from Equation (8) (left) and the Vandermonde ansatz from Equation (7) (right)

In this section we have so far analyzed a variety of ansatzes for three different models that obey the Pauli exclusion principle. TABLE 1 shows a summary of the speed up for all of the models studied with our software. We compute the values presented in the table by taking the time elapsed prior to convergence for with use of the Slater ansatz and dividing this by the time elapsed prior to convergence with the use of the Vandermonde ansatz. The value 2.5 in the second cell of the second row says that the Salter ansatz requires 2.5 times the time required by the Vandermonde ansatz in order for the corresponding gradient decent to converge. One major challenge was tuning the hyperparameters of the VMC pipeline (i.e., learning rate, proposal distribution standard deviation, momentum, wavefunction input restrictions, number of walkers and number of steps for the Monte Carlo). Exploitation of known physical properties and symmetries or in some instances trial and error was needed in order to find the correct initialization necessary to attain a convergent gradient decent. To study more complicated systems such as Lithium hydride and Beryllium, we employed the PauliNet software.

B. VMC for atoms and molecules with PauliNet

We used the same architectural designs as the original PauliNet paper (see Table 2 in [4] for the full set of parameters), with a few exceptions. We only used a single determinant for both the Slater and Vandermonde cases. All network sizes and dimensionalities of feature representations remained the same, and we used the same basis functions. We found it necessary to lower the maximum learning rate to 0.005 from the default 0.01. We ran for 600 or 800 iterations, less than for the results collected originally in [4].

1) Results and Comparisons: With the Vandermonde determinants, convergence was achieved but to an energy relatively far from the correct answer – the Slater determinants reached much closer. To visualize the wavefunction during training, we made plots of the wavefunction sliced along a single coordinate of a single electron.

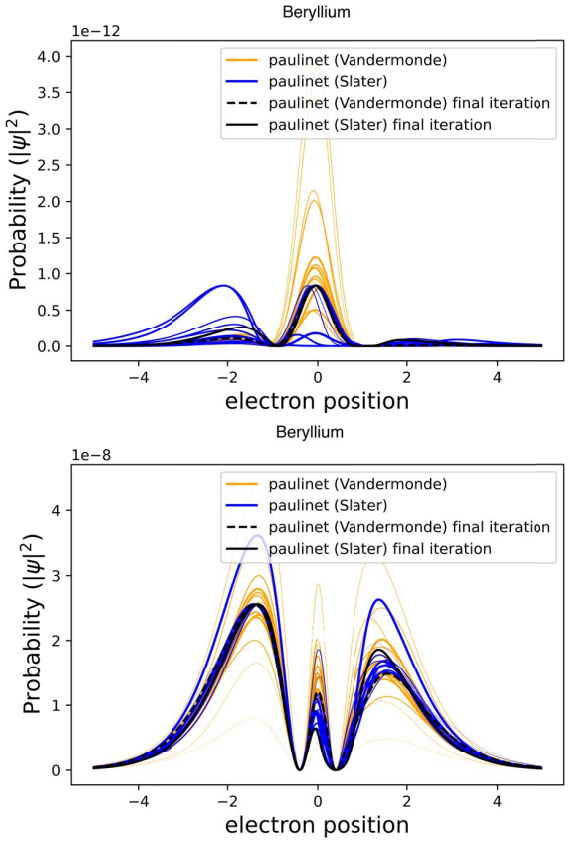

Fig. 5: Two slices of the PauliNet wavefunction for Beryllium. Two different slices of the wave function are plotted every 50 steps over the course of 800 total training steps. Thinner lines are earlier time steps, thicker are later. The ground state energy estimation at the final iteration is -13.2703 for PauliNet with the Vandermonde determinant, and -14.6479 for PauliNet with the Slater determinant. The true ground state energy is -14.6674.

Figure 7 shows the evolution of the wavefunction for LiH, a system with four electrons (one for hydrogen and three for lithium). Figure 5 shows the same evolution for Beryllium, a single-atom system with four electrons.

In both of these systems, we find that the Slater determinant reaches a lower energy – much closer to the true values in both cases. The energies reached are listed in the captions of Figures 5 and 7.

We also make plots showing how the wavefunction changes after every 50 training iterations over the course of 800 total iterations. In particular, we train PauliNet using both the Slater determinant and the Vandermonde determinant, and every 50 training iterations, we estimate $||\Psi_x|^2 - |\Psi_{x-1}|^2|$ by evaluating Ψ_x and Ψ_{x-1} at 3000 random points, where Ψ_x is the wavefunction after $50x$ training iterations. This is a measure of how much the wavefunction changes during training. The results are shown in Figure 6. We find that PauliNet with the Slater determinant continues to train, while the Vandermonde determinant trains quickly at first but appears to get "stuck" in a local minimum far from the true ground state energy. The relative accuracy of various methods is most readily

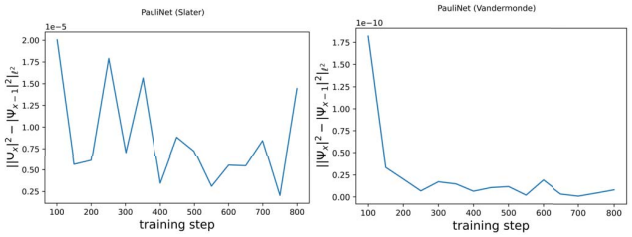

Fig. 6: Norm of difference in wavefunction outputs evaluated at 3000 random points, every 50 training iterations.

understood by comparing the estimation of the ground-state energy. In particular, we compared our estimation of the ground state energy to the estimation obtained by the Hartree-Fock method and the original PauliNet. PauliNet uses the Hartree-Fock method as a baseline solution and as a starting point for the variational Monte Carlo method. It first performs a Hartree-Fock calculation to produce one electron orbitals $\varphi_\mu(r_i)$ (see Equation (1)) which are then used as inputs to PauliNet. After performing variational Monte Carlo, PauliNet significantly surpasses the accuracy of Hartree-Fock. The authors of PauliNet demonstrate this in [4] by comparing the accuracy of their solution after 7000 training steps to the Hartree-Fock method and another deep learning based VMC method.

After 7000 training steps, our modification with the Vandermonde determinant in general does not exceed Hartree-Fock accuracy. Others who used similar constructions [3] also reported poor results. In order to present our method as a viable low-cost alternative to PauliNet, we would at least want to surpass the accuracy of Hartree-Fock for most systems. We believe that there may be other modifications that one could make in order to achieve this level of accuracy while still maintaining the N^3 asymptotic complexity.

VI. Conclusion

A possible limitation of our adaptations to PauliNet might be the use of the Hartree-Fock-derived basis functions as

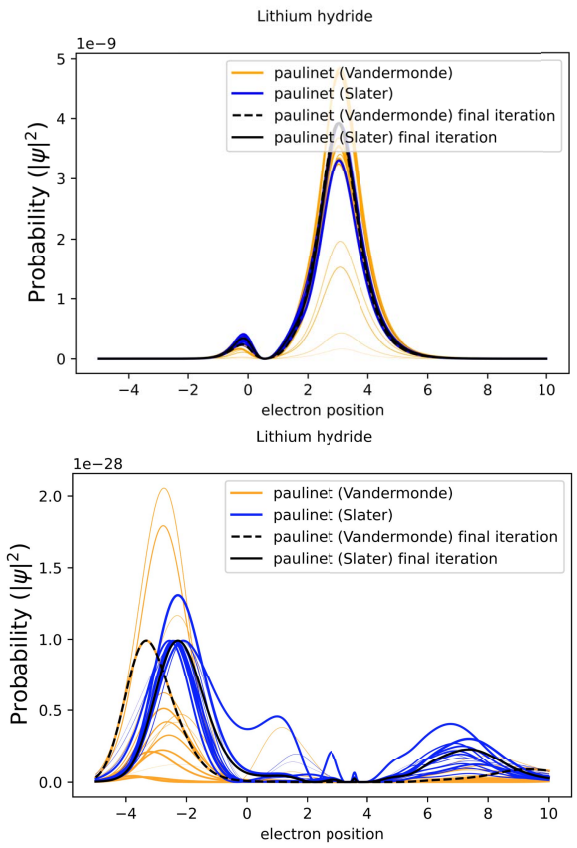

Fig. 7: Two slices of the PauliNet wavefunction for Lithium hydride. Two different slices of the wave function are plotted every 50 steps over the course of 800 total training steps. Thinner lines are earlier time steps, thicker are later. The ground state energy estimation at the final iteration is -7.1686 for PauliNet with the Vandermonde determinant, and -8.0619 for PauliNet with the Slater determinant. The true ground state energy is -8.07.

predetermined inputs to the ansatz. The core Hartree-Fock ansatz itself uses a Slater determinant, so the parameters of the basis functions might be particularly well-tuned for use in conjunction with Slater determinants. The flexible neural networks of the rest of the PauliNet architecture are still capable of learning, but when using the Vandermonde determinant they might have to overcome this limitation.

As such, one possible avenue for improvement might be replacing the basis functions φ with ones somehow better suited to the Vandermonde determinant. The authors of PauliNet claim that part of the success of their method is the fact that it begins with the physically plausible Hartree-Fock functions. It might be possible to initialize by solving some analogue to the Hartree-Fock wavefunction that uses a Vandermonde determinant instead of a Slater determinant, potentially yielding better performance.

Variational Monte Carlo is a widely-used algorithm for approximating the ground state energy of quantum systems. While more complex variations of Monte Carlo approaches can give better performance, VMC is widely used and implemented in many popular high-performance codes including QMCPACK [1]. Variational Monte Carlo relies on the use of a flexible parametric ansatz to represent the wavefunction.

Recently, several authors including [2], [4] have proposed to use deep-learning-based ansatzes in the variational Monte Carlo pipeline. These techniques show promise – using the relatively simple VMC algorithm, they are able to achieve performance commensurate with more complex algorithms. In this work, we focused on the PauliNet architecture from [4].

For many-electron systems, most common ansatzes (including both PauliNet and QMCPACK) use a construction known as the Slater determinant to enforce antisymmetry. For an N-electron system, this requires taking the determinant of an $N \times N$ matrix, an $O(N^3)$ operation which in some cases is the asymptotically dominant term for computational cost. There is however another possible construction known as the Vandermonde determinant. The authors of [5] show that the Vandermonde determinant can function as a universal approximator for antisymmetric functions, and [3] uses a Vandermonde-type construction, suggesting that in principle, it could work as well as the Slater determinant as part of a flexible deep-learning-based ansatz. Yet it is potentially only an $O(N^2)$ operation.

To investigate the feasibility of replacing the Slater determinant with the Vandermonde determinant, we have followed two approaches:

- We implemented our own variational Monte Carlo pipeline on which we tested simple Slater and Vandermonde based ansatzes on simple physical systems.
- We proposed the use of the Vandermonde determinant in place of the Slater determinant in the PauliNet architecture.

In general, we found that the Vandermonde determinant does not work as well as the Slater determinant. In our own implementation, its performance is tolerable, suggesting that as systems scale up and the asymptotic difference in evaluation cost becomes relevant, the trade-off in computational cost could be reasonable. Within PauliNet, the use of the Vandermonde determinant results in a wavefunction that, while somewhat viable, performs significantly worse than PauliNet with Slater determinants.

With regards to PauliNet, a major challenge during development was ensuring the appropriate enforcement of physical boundary conditions – this falls out naturally when using the Slater determinant but required extra effort to guarantee using the Vandermonde determinant. It is possible that different techniques for enforcing these conditions might give better performance than the ones we used. Additionally, it is possible that the input Hartree-Fock basis functions are not well-adapted for the Vandermonde determinant, and that deriving new basis functions could also significantly enhance performance.

VII. Acknowledgements

AMD, the AMD Arrow logo, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product

names used in this publication are for identification purposes only and may be trademarks of their respective companies.

REFERENCES

[1] J. Kim, A. D. Baczewski, T. D. Beaudet, A. Benali, M. C. Bennett, M. A. Berrill, N. S. Blunt, E. J. L. Borda, M. Casula, D. M. Ceperley *et al.*, "QMCPACK: an open source ab initio quantum Monte Carlo package for the electronic structure of atoms, molecules and solids," *Journal of Physics: Condensed Matter*, vol. 30, no. 19, p. 195901, 2018.

[2] D. Pfau, J. S. Spencer, A. G. d. G. Matthews, and W. M. C. Foulkes, "Ab-initio solution of the many-electron Schrödinger equation with deep neural networks," *arXiv preprint arXiv:1909.02487*, 2019.

[3] J. Han, L. Zhang, and E. Weinan, "Solving many-electron Schrödinger equation using deep neural networks," *Journal of Computational Physics*, vol. 399, p. 108929, 2019.

[4] J. Hermann, Z. Schätzle, and F. Noé, "Deep-neural-network solution of the electronic Schrödinger equation," *Nature Chemistry*, vol. 12, no. 10, pp. 891–897, Sep. 2020. [Online]. Available: https://doi.org/10.1038/s41557-020-0544-y

[5] J. Han, Y. Li, L. Lin, J. Lu, J. Zhang, and L. Zhang, "Universal approximation of symmetric and anti-symmetric functions," *CoRR*, vol. abs/1912.01765, 2019. [Online]. Available: http://arxiv.org/abs/1912.01765

[6] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. CRC press, 2011.

[7] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 04 1970. [Online]. Available: https://doi.org/10.1093/biomet/57.1.97

[8] K. 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," *Nature communications*, vol. 10, no. 1, pp. 1–10, 2019.

[9] M. Hjorth-Jensen, "Computational physics," *Lecture notes*, 2011.