DTU Library

Neural network ansatz for periodic wave functions and the homogeneous electron gas

Wilson, Max; Moroni, Saverio; Holzmann, Markus; Gao, Nicholas; Wudarski, Filip; Vegge, Tejs; Bhowmik, Arghya

Neural network ansatz for periodic wave functions and the homogeneous electron gas

Max Wilson [1,*] Saverio Moroni [2] Markus Holzmann,[3] Nicholas Gao [4] Filip Wudarski [5,6] Tejs Vegge [7] and Arghya Bhowmik [7,†]

[1]*Technical University of Denmark, Department of Applied Mathematics and Computer Science, 2800 Kgs. Lyngby, Denmark*
[2]*Consiglio Nazionale delle Ricerche, Istituto Officina dei Materiali and Scuola Internazionale Superiore di Studi Avanzati, 34136 Trieste, Italy*
[3]*Univ. Grenoble Alpes, CNRS, LPMMC, 38000 Grenoble, France*
[4]*Technical University of Munich, 85748 Munich, Germany*
[5]*Quantum Artificial Intelligence Lab. (QuAIL), Exploration Technology Directorate, National Aeronautics and Space Administration Ames Research Center, Moffett Field, California 94035, USA*
[6]*Universities Space Research Association, Research Institute for Advanced Computer Science, Mountain View, California 94043, USA*
[7]*Technical University of Denmark, Department of Energy Conversion and Storage, 2800 Kgs. Lyngby, Denmark*

We design a neural network *Ansatz* for variationally finding the ground-state wave function of the homogeneous electron gas, a fundamental model in the physics of extended systems of interacting fermions. We study the spin-polarized and paramagnetic phases with 7, 14, and 19 electrons over a broad range of densities from $r_s = 1$ to $r_s = 100$, obtaining similar or higher accuracy compared to a state-of-the-art iterative backflow baseline even in the challenging regime of very strong correlation. Our work extends previous applications of neural network *Ansätze* to molecular systems with methods for handling periodic boundary conditions, and makes two notable changes to improve performance: splitting the pairwise streams by spin alignment and generating backflow coordinates for the orbitals from the network. We illustrate the advantage of our high-quality wave functions in computing the reduced single-particle density matrix. This contribution establishes neural network models as flexible and high-precision *Ansätze* for periodic electronic systems, an important step towards applications to crystalline solids.

I. INTRODUCTION

Electronic structure theory forms the backbone of *ab initio* calculations of systems in quantum chemistry and condensed matter physics. Method development of approximate solutions to the many-electron Schrödinger equation is a central request of theory, pushed further with algorithmic improvements and the continuing growth of computational resources [1]. Approximate solutions to the electronic Schrödinger equation together with advanced high-throughput, automated methods have broad applications ranging from pharmaceutical design to new materials for clean energy technologies such as batteries, and machine learning methods are promising to address the various challenges involved.

In this paper, we develop a periodic neural network *Ansatz*, dubbed wave function *Ansatz* (but periodic) (WAP)-net and outlined in Fig. 1, to describe the ground state of the homogeneous electron gas (HEG) over a broad density regime. The model is composed of permutation equivariant layers [2], products of Hartree-Fock plane-wave orbitals appropriate to a homogeneous system and prefactors [3], periodic input features [4], and backflow coordinates [5]. Additionally, improvements are made to previous related networks by splitting the pairwise interaction data by spin alignment and including orbital backflow coordinates. We compare the obtained wave function *Ansatz* (but periodic) (WAP)-net energies to those from iterative backflow (IB) network wave functions [6] and show that our method provides the most accurate explicit wave functions in almost all cases and often competes with the energies obtained from fixed-node diffusion Monte Carlo (DMC) calculations [7] performed on the top of our best IB wave function for the HEG.

The HEG, also known as jellium or the uniform electron gas [8], is a simplified model allowing us to capture intricate properties of solid systems such as the quantum nature of delocalized electrons without requiring the introduction of a specific atomic lattice [7,9–12]. It has played a prominent role in the development of theoretical approaches such as diagrammatic perturbation theory [7,13,14], quantum Monte Carlo (QMC) [15,16], and, more recently, in the exportation of quantum chemistry methods to extended systems [17–19]. In all of these approaches, the HEG has been used as a first milestone on the way towards more accurate description of electronic correlations in materials and is a natural first step to developing novel neural network *Ansätze* for solids.

Methods involving neural networks have already demonstrated state-of-the-art performance in problems spanning fields including physics [20], chemistry [2], games [21], and autonomous systems [22], driven by rapid evaluation on GPUs, availability of data, and design innovations. First
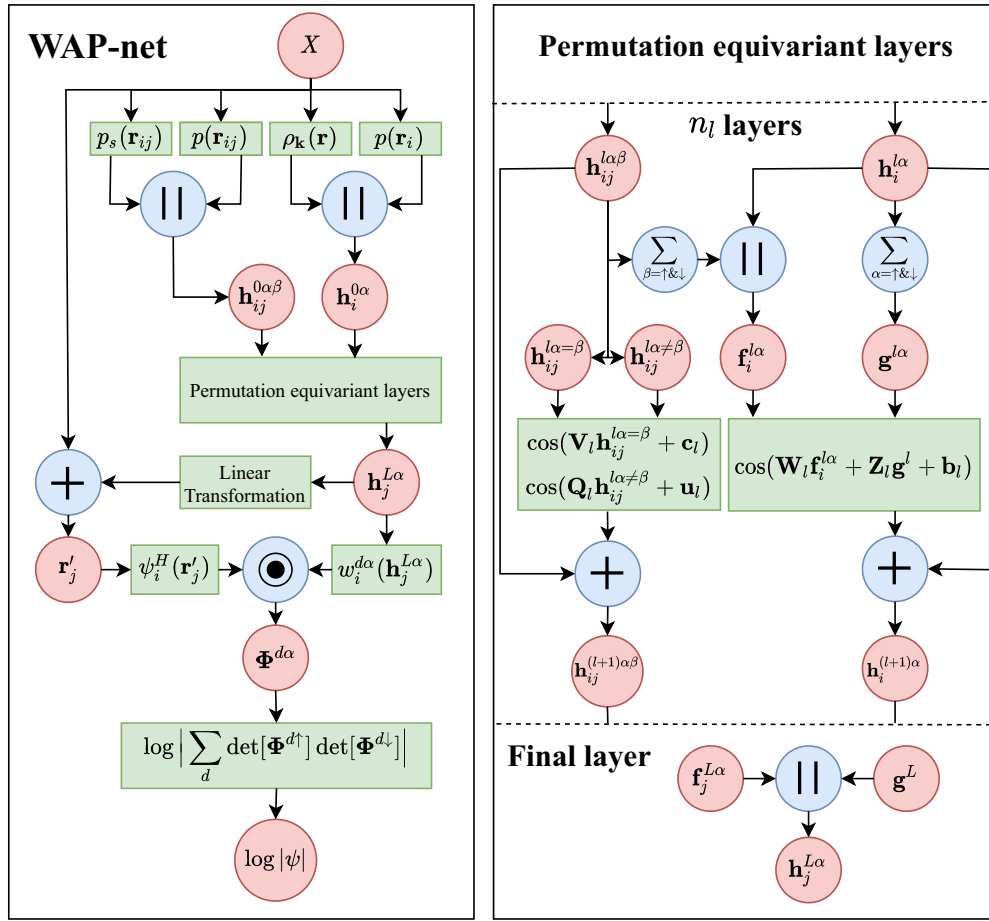
*amawi@dtu.dk
†arbh@dtu.dk

FIG. 1. Diagram showing key functions (green), variables (red), and operations (blue) in WAP-net. All functions and operations are described in Sec. III (*Ansatz*). Roughly, system configurations (X) are embedded by periodic functions [$p_s(\cdot)$, $p(\cdot)$, $\rho_{\mathbf{k}}(\cdot)$], split into single and pairwise streams, and passed through equivariant layers. The resulting variables are used to generate orbitals in Slater determinants that compose the log-amplitude of the *Ansatz*. In the operations, || represents a concatenation, $+$ the addition, and $\bullet$ the product. In an abuse of notation the "&" symbol in the sums indicates the sums are taken over both ↑ and ↓ separately. The diagram is representative but not exact, it does not contain all operations, for an exact description see Sec. III.

established as an *Ansatz* for quantum many-body systems in 2017 [23], subsequent methods introduced other ideas from the rapidly developing field of neural networks. Recent attempts have included physics-based structure into the *Ansätze*, where more sophisticated techniques in network design and optimization have even advanced this subfield to the state-of-the-art performance on molecular systems [2–4,24–26].

Recent diagrammatic Monte Carlo calculations [27] have benchmarked values of the renormalization factor, Z, characterizing the Fermi liquid behavior of the HEG, against previous quantum Monte Carlo results [28]. Here we use WAP-net to calculate the reduced single-particle density matrix and estimate the possible fixed-node bias in Ref. [28].

The remainder of this paper is organized as follows. Section II highlights relevant literature and its relationship to this research, Section III gives a complete theoretical description of the method, including hyperparameters and setup, Section IV outlines the results, first on the ground-state energy (Sec. IV A) needed to judge the performance of WAP-net (Sec. IV B). Then we apply our wave functions to calculate the reduced single-particle density matrix (Sec. IV C). Finally, the paper is summarized and next steps are proposed in Sec. V.

Recently, multiple groups have made grand strides in the application of neural networks to quantum Monte Carlo, including self-attention networks [29] and extended systems [30], further establishing the importance, flexibility, and potential of these methods.

II. RELATED WORK

Regarding fermionic systems, the accuracy of variational QMC methods are, in general, ultimately limited by the underlying form of the trial wave function (the *Ansatz*). In standard implementations, antisymmetry is guaranteed by one or several Slater determinants constructed from single-particle orbitals that are correlated via an explicit pairwise bosonic (permutation invariant) potential, frequently called the Slater-Jastrow wave function. Backflow wave functions further correlate the particle positions entering the orbitals of the Slater determinants [31], or more general permutation equivarant forms [5,32,33], from which IB networks [6,34] currently provide the most accurate systematic improvement in homogeneous systems [35].

Research into neural network *Ansätze* has covered Boltzmann machines and feed-forward neural networks over

a wide range of systems [23,36–38], including more recent Slater determinant *Ansätze* [39,40]. Other more recent examples incorporate physics-based structure [3], or large and deep networks [2,24,26] on molecular systems. Associated implementations such as the PESNet attempt to curtail the learning process of a potential energy surface [25] and the neural canonical transformation for the computation of the electron effective mass [41]. Finally, there is work on incorporating periodic functions to neural networks and/or applying these methods to periodic data/systems [4,42,43] and equivariances to periodic data [44,45].

In this paper, we focus on a QMC neural network *Ansatz* method applied to the HEG model. Other recent work tackles the problem of designing a neural network *Ansatz* for bosonic systems on a torus [4], resulting in some similarities in function design such as the application of periodic embedding functions to the electron coordinates. Quantum chemistry benchmarks on the HEG [17–19] are limited to the high-density region and small number of electrons. Since neural networks are highly parameterizable and can model arbitrary smooth functions [46], one can expect that this computational model in conjugation with QMC techniques will be flexible enough to approach high accuracy uniformly in density, similar or better than IB *Ansatz*. Additionally, it may potentially provide a more favorable scaling (for a given accuracy) with electron number than other quantum chemistry methods.

III. METHODS

A. Hamiltonian

In the following we consider a system of N_e electrons in three spatial dimensions and seek the ground-state wave function $\psi(X)$ of the time-independent Schrödinger equation

$$H\psi(X) = E\psi(X), \quad (1)$$

where $X = (\mathbf{r}_1, \ldots, \mathbf{r}_{N_e})$ denotes a full configuration of the electron coordinates $\mathbf{r}_i = (r_{i,x}, r_{i,y}, r_{i,z})$. In atomic units, the Hamiltonian

$$H = -\frac{1}{2}\nabla_X^2 + V(X) \quad (2)$$

contains the kinetic energy operator ∇_X^2 of each electron where $\nabla_X = (\nabla_{\mathbf{r}_1}, \ldots, \nabla_{\mathbf{r}_{N_e}})$ and

$$\nabla_X^2 = \sum_{i=1}^{N_e}\left(\frac{\partial^2}{\partial r_{i,x}^2} + \frac{\partial^2}{\partial r_{i,y}^2} + \frac{\partial^2}{\partial r_{i,z}^2}\right). \quad (3)$$

The potential energy is $V(X)$, a function of pairwise interactions $v(\mathbf{r})$ depending on the distance between two particles $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$,

$$V(X) = \sum_{i<j} v(\mathbf{r}_{ij}). \quad (4)$$

In order to describe the bulk of a extended systems of electron density $\rho_e = N_e/L^3$, with electrons in a cubic box of side L, we want to impose periodic boundary conditions, e.g., $\psi(\ldots, r_{i,x} + L, \ldots) = \psi(\ldots, r_{i,x,\ldots})$. Since the Coulomb interaction is long ranged, some care is needed for proper setup and computation of the interaction potential. Here, we use the standard Ewald summation [1,47] in direct and reciprocal

space for its computation given charges q_i and q_j, for electrons in atomic units $q_i = 1$,

$$v(\mathbf{r}) = q_i q_j \left[\sum_{\mathbf{T}} \frac{\text{erfc}[\kappa||\mathbf{r} - \mathbf{T}||]}{||\mathbf{r} - \mathbf{T}||} - \frac{1}{\kappa^2 L^3} \right.$$
$$\left. + \frac{4\pi}{L^3} \sum_{\mathbf{k}}' \frac{\exp(-\mathbf{k}^2/(4\kappa^2))}{\mathbf{k}^2} \exp(-i\mathbf{k} \cdot \mathbf{r}) \right], \quad (5)$$

where the first summation is over all the lattice vectors to image cells, $\mathbf{T} = (i_x, i_y, i_z)L$ with integers i_α, whereas the second summation is over all reciprocal lattice vectors $\mathbf{k} = 2\pi(i_x, i_y, i_z)/L$, and the prime indicates the omission of $\mathbf{k} = (0, 0, 0)$ term. The convergence of both summations is determined by the hyperparameter κ, which should be set to minimize the size of the sets required to make the sums converge.

For convenience, we also add the interaction energy of each electron with its own images, in our case for a system of N_e electrons it is $N_e v_M$, to the total interaction energy $V(X)$ where

$$v_M = \lim_{\mathbf{r}_{ij} \to 0}\left[v(\mathbf{r}_{ij}) - \frac{q_i q_j}{||\mathbf{r}_{ij}||}\right] \quad (6)$$

is the Madelung constant, which is independent of the configuration X, and depends only on L. Although its contribution to the total energy is negligible in the limit $N_e \to \infty$, it is a standard term, which is in general systematically included to accelerate the extrapolation to the thermodynamic limit.

For the homogeneous electron gas, electron density is usually expressed in terms of the Wigner-Seitz parameter $r_s = a/a_B = (4\pi\rho_e a_B^3/3)^{-1/3}$, the ratio between the Bohr radius ($a_B = 1$ in atomic units) and the mean interparticle distance $a = (4\pi\rho_e/3)^{-1/3}$. Energies are given in Rydbergs. In the high-density limit, $r_s \to 0$, kinetic energy dominates and the system approaches the ideal Fermi gas whereas at low densities kinetic energy quantum effects become less important and the electrons will eventually form a Wigner crystal to minimize the classical potential energy [12].

B. Variational Monte Carlo

Given that the Hamiltonian is a lower-bounded operator, the variational principle states that upper-bound estimates of the ground-state energy can be obtained by minimizing the expectation value of the energy with respect to the parameters θ of an *Ansatz* $\psi(X; \theta)$ [48]. The high-dimensional integral required to compute the expectations can, in general, be performed with Monte Carlo integration. In this work, M samples are drawn from $p(X; \theta) \propto |\psi(X; \theta)|^2$ in a Markov chain process via the Metropolis Hastings algorithm [49]. For reviews of QMC methods for solids see Refs. [10,50,51].

The estimate of the energy is computed

$$\mathbb{E}_{X \sim p(X;\theta)}[E_L(X; \theta)] \approx \frac{1}{M}\sum_{i=1}^{M} E_L(X_i; \theta), \quad (7)$$

where the expectation is taken over the distribution $p(X; \theta)$ and this explicit notation is dropped from now on. $E_L(X; \theta)$,

the local energy, is

$$E_L(X;\theta) = -\tfrac{1}{2}\big[\nabla_X^2 \log \psi(X;\theta)$$
$$+ (\nabla_X \log \psi(X;\theta))^2\big] + V(X). \quad (8)$$

The derivatives of the expectation of the energy with respect to the *Ansatz* parameters θ can be framed as the vector of derivatives ∇_θ of a loss function $\mathcal{L}(\theta)$ and are computed as

$$\nabla_\theta \mathcal{L}(\theta) = \mathbb{E}_X[(E_L(X;\theta) - \mathbb{E}_X[E_L(X;\theta)])\nabla_\theta \log \psi(X;\theta)]. \quad (9)$$

C. Iterative backflow *Ansatz*

For fermions, the trial wave function, $\psi(X;\theta) = D(X;\theta)e^{-U(X;\theta)}$, is conveniently split into a generalized Jastrow factor, $U(X;\theta)$, symmetric with respect to any permutation in the particle labels i, and a manifestly antisymmetric from, $D(X;\theta)$, usually a Slater determinant composed of N_e orbitals, $\psi_k(\cdot)$, a set of plane waves for homogeneous systems, $\psi_{\mathbf{k}}(\mathbf{r}) = \exp(i\mathbf{k} \cdot \mathbf{r})$.

For a spin-independent Hamiltonian, we can further label the orbitals according to the spin polarization, $\uparrow/\downarrow$, such that the Slater determinant factorizes, e.g., $D(X) = [\det_{ij} \psi_{\mathbf{k}_j}(\mathbf{r}_i^\uparrow)] \times [\det_{ij} \psi_{\mathbf{k}_j}(\mathbf{r}_i^\downarrow)]$. Typically, the wave vectors used in the respective Slater determinant correspond to those of the ground-state wave function of the ideal Fermi gas.

A standard two-body Jastrow function writes $U_0(X) = \sum_{i<j} u_\theta^0(\mathbf{r}_{ij})$, and backflow wave functions are obtained by the use of backflow coordinates $Q_0(X) = (\mathbf{q}_1^0, \dots, \mathbf{q}_{N_e}^0)$ with $\mathbf{q}_i^0 = \mathbf{r}_i + \sum_{j \neq i} \mathbf{r}_{ij}\eta_\theta^0(\mathbf{r}_{ij})$ as arguments in the orbitals of the Slater determinant

$$\Big[\det_{ij} \psi_{\mathbf{k}_j}(\mathbf{q}_i^{0\uparrow})\Big] \times \Big[\det_{ij} \psi_{\mathbf{k}_j}(\mathbf{q}_i^{0\downarrow})\Big] \exp\Big[-\sum_{i<j} u_\theta^0(\mathbf{r}_{ij})\Big]. \quad (10)$$

Higher-order many-body correlations can then be systematically constructed by an iterative procedure $U_n(Q_{n-1}) = \sum_{i<j} u_\theta^n(\mathbf{q}_{ij}^{n-1})$, and, analogously, new sets of backflow coordinates, $Q_n(X)$, from $\mathbf{q}_i^n = \mathbf{q}_i^{n-1} + \sum_{j \neq i} \mathbf{q}_{ij}^{n-1}\eta_\theta^n(\mathbf{q}_{ij}^{n-1})$ yielding the general form of IB wave functions [6]

$$\Big[\det_{ij} \psi_{\mathbf{k}_j}(\mathbf{q}_i^{n\uparrow})\Big] \times \Big[\det_{ij} \psi_{\mathbf{k}_j}(\mathbf{q}_i^{n\downarrow})\Big] \exp\Big[-\sum_{n' \leqslant n} U_{n'}(Q_{n-1})\Big]. \quad (11)$$

This structure not only captures many-body correlations in a compact way, but also allows for an efficient calculation of the wave function, its gradient and local energy, of order N_e^3 independent of the number n of iterations.

Each iteration defines a new permutation equivariant structure, introducing a new set of parameters θ in the respective functions, $u_\theta^n(\cdot)$ and $\eta_\theta^n(\cdot)$, possibly separated into spinlike and spin-unlike parts, by use of a basis set expansion. Details of the implementation used here are given in Ref. [35].

D. WAP-net *Ansatz*

The *Ansatz* developed here, dubbed WAP-net and represented in Fig. 1, is a progression on other implementations

[2,3,24] designed for molecular systems and is related to other periodic networks developed for bosonic systems [4]. Key differences with previous work for molecular systems are embedding coordinates with periodic functions as in Ref. [4], and other changes to the *Ansatz* include splitting pairwise features by spin alignment and feeding backflow coordinates to the plane wave Hartree-Fock orbitals. In our implementation of the WAP-net we rescale the lengths so that the box side becomes $L = 1$. The complete model is described below.

The *Ansatz* comprises linear layers, second-order differentiable nonlinear activations, and Slater determinants. Before the Slater determinant, nonlinear layers of functions operate on the electron coordinates ($\mathbf{r}_j$) and their displacements ($\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$), to maintain indexes of intermediate variables such that they correspond to a coordinate or displacement from the inputs. Intermediate variables corresponding to coordinates and displacements flow through parts of the network called the single and pairwise streams, respectively. The permutation equivariance of these functions ensures that exchange of same spin electron coordinates results in the same permutation of the intermediate variables, resulting in the exchange of rows (or columns) of the Slater determinant, and flipping the amplitude sign, fulfilling the antisymmetry required for fermionic systems.

Coordinates are first embedded by periodic functions to satisfy the periodicity of the system, including continuity of the function and derivatives across the boundary,

$$p(\mathbf{r}) = [\cos(2\pi\mathbf{r}), \sin(2\pi\mathbf{r}), \dots,$$
$$\dots, \cos(2n_p\pi\mathbf{r}), \sin(2n_p\pi\mathbf{r})], \quad (12)$$

where n_p is the number of functions and the functions are applied elementwise. In the case of the pairwise streams a distancelike feature, $||p_s(\mathbf{r}_{ij})||$, is computed and concatenated

$$p_s(\mathbf{r}_{ij}) = \tfrac{1}{2}\big[\sin(\pi r_{ij}^x), \sin(\pi r_{ij}^y), \sin(\pi r_{ij}^z)\big]. \quad (13)$$

For the single streams, these periodic input features are concatenated with density fluctuations

$$\rho_{\mathbf{k}} = \Big(\sum_j \cos(\mathbf{k} \cdot \mathbf{r}_j,), \sum_j \sin(\mathbf{k} \cdot \mathbf{r}_j)\Big) \quad (14)$$

given a set $\mathbb{K}$ of n_k reciprocal lattice vectors $\mathbf{k}$. Overall, the variables in the zeroth layer are

$$\mathbf{h}_i^{0\alpha} = [\{\rho_{\mathbf{k}} : \mathbf{k} \in \mathbb{K}\}, p(\mathbf{r}_i)], \quad (15)$$

$$\mathbf{h}_{ij}^{0\alpha\beta} = [p(\mathbf{r}_{ij}), ||p_s(\mathbf{r}_{ij})||]. \quad (16)$$

Permutation equivariant functions then operate on data from both streams to compute single-stream variables

$$\mathbf{f}_i^{l\alpha} = \Big(\mathbf{h}^{l\alpha}, \frac{1}{n_\uparrow} \sum_{j \text{ if } \beta \neq \downarrow} \mathbf{h}_{ij}^{l\alpha\beta}, \frac{1}{n_\downarrow} \sum_{j \text{ if } \beta \neq \uparrow} \mathbf{h}_{ij}^{l\alpha\beta}\Big) \quad (17)$$

$$\mathbf{g}^l = \Big(\frac{1}{n_\uparrow} \sum_{i \text{ if } \alpha \neq \downarrow} \mathbf{h}_i^{l\alpha}, \frac{1}{n_\downarrow} \sum_{i \text{ if } \alpha \neq \uparrow} \mathbf{h}_i^{l\alpha}\Big). \quad (18)$$

Updates on the single and pairwise streams at layer l are computed as

$$\mathbf{h}_i^{(l+1)\alpha} = \cos(\mathbf{W}_l\mathbf{f}_i^{l\alpha} + \mathbf{Z}_l\mathbf{g}^l + \mathbf{b}_l) + \mathbf{h}_i^{l\alpha}, \quad (19)$$
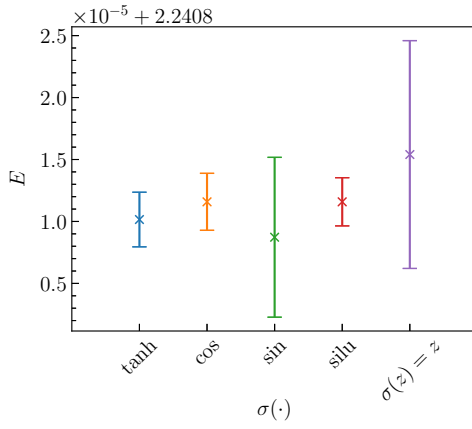
FIG. 2. Comparison of different activation functions in the back-flow layer, Eq. (23), for $rs = 1$ and $N = 14$. Discussion on SILU can be found in Ref. [52]. Mean and error bars represent the statistics over a small ensemble of models trained with different seeds.

$$\mathbf{h}_{ij}^{(l+1)\alpha=\beta} = \cos\left(\mathbf{V}_l \mathbf{h}_{ij}^{l\alpha=\beta} + \mathbf{c}_l\right) + \mathbf{h}_{ij}^{l\alpha=\beta}, \tag{20}$$

$$\mathbf{h}_{ij}^{(l+1)\alpha\neq\beta} = \cos\left(\mathbf{Q}_l \mathbf{h}_{ij}^{l\alpha\neq\beta} + \mathbf{u}_l\right) + \mathbf{h}_{ij}^{l\alpha\neq\beta}, \tag{21}$$

for weights $\mathbf{W}_l$, $\mathbf{Z}_l$, $\mathbf{V}_l$, and $\mathbf{Q}_l$, and biases $\mathbf{b}_l$, $\mathbf{c}_l$, $\mathbf{u}_l$. Residual connections are added to all layers where $\dim(\mathbf{h}^l) = \dim(\mathbf{h}^{(l-1)})$. This implementation differs from previous similar works [2,24,26] as the pairwise stream layers are split into two spin-alignment-dependent blocks: $\alpha = \beta$ and $\alpha \neq \beta$ corresponding to when the displacements are computed between electrons of the same and different spins, respectively. This method is motivated as correlations change depending on spin alignment.

There are n_l of these nonlinear parameterised layers. The last layer is half the size of the previous layer and the outputs $\mathbf{f}_i^{L\alpha}$ and $\mathbf{g}^L$ are concatenated

$$\mathbf{h}_i^{L\alpha} = \mathbf{f}_i^{L\alpha} \,\|\, \mathbf{g}^L, \tag{22}$$

then split into spin-dependent data blocks. These variables are first used to construct backflow coordinates as inputs to the Hartree-Fock orbitals

$$\mathbf{r}_j^{\prime\alpha} = \mathbf{r}_j^{\alpha} + \tanh\left(\mathbf{W}_L^{d\alpha} \mathbf{h}_j^{L\alpha} + \mathbf{b}_y^{d\alpha}\right), \tag{23}$$

where d indexes the Slater determinant where the backflow coordinates will be used, and a tanh activation was used to limit the outputs between -1 and 1. This gives good performance. Other activations were tested but the results showed no improvement and in the case of no activation function, the computed energies had higher variance and final energy, see Fig. 2. Second, $\mathbf{h}_j^{L\alpha}$ are mapped to scalar orbital prefactors via a linear transformation $w_i^{d\alpha}(\cdot)$ giving Slater determinant orbitals

$$\phi_{ij}^{d\alpha}(X) = w_i^{d\alpha}\left(\mathbf{h}_j^{L\alpha}\right)\psi_i^H\left(\mathbf{r}_j^{\prime\alpha}\right), \tag{24}$$

where $\psi_i^H(\mathbf{r}_j)$ are the Hartree-Fock plane wave orbitals

$$\psi_i^H(\mathbf{r}_j) = \begin{cases} \sin(\mathbf{k}_i\mathbf{r}_j) & \text{if } i \text{ is even,} \\ \cos(\mathbf{k}_i\mathbf{r}_j) & \text{if } i \text{ is odd,} \end{cases} \tag{25}$$

TABLE I. Hyperparameters used to define the model.

Symbol	Description	Value
N/A	Single stream hidden units	128
N/A	Pairwise stream hidden units	32
n_l	Number of layers	3
n_p	Number of periodic functions	5
n_d	Number of determinants	$\{1, 8\}$
n_k	Number of k-points used in the input	19

where the set of $\mathbf{k}$ are ordered such that $\mathbf{k}_{i+1}$ is opposite to $\mathbf{k}_i$, i.e., $-\mathbf{k}_{i+1} = \mathbf{k}_i$, for even i and $\mathbf{k}_1 = (0, 0, 0)$. We take the shortest N_α k points to form the Hartree-Fock orbitals for the N_α electrons of spin α.

The Slater determinant of a spin becomes

$$\det[\mathbf{\Phi}^{d\alpha}] = \begin{vmatrix} \phi_{11}^{d\alpha}(X) & \dots & \phi_{N_\alpha 1}^{d\alpha}(X) \\ \vdots & & \vdots \\ \phi_{1N_\alpha}^{d\alpha}(X) & \dots & \phi_{N_\alpha N_\alpha}^{d\alpha}(X) \end{vmatrix}. \tag{26}$$

for d defining the index.

The full wave function is written

$$\log|\psi(X)| = \log\left|\sum_d \det[\mathbf{\Phi}^{d\uparrow}]\det[\mathbf{\Phi}^{d\downarrow}]\right|. \tag{27}$$

We have taken the log-absolute value of the amplitude in the implementation. While in general one can write $\log\psi = \log|\psi| + i\phi$, with the periodic boundary conditions adopted in this work the wave function is real. Therefore we can ignore the phase ϕ, which is a constant wherever the modulus $|\psi|$ is nonzero.

The complete set of *Ansatz* hyperparameters is given in Table I; these hyperparameters were chosen based on intuition from previous research [2,24,26] and sweeps over options (including for example alternate activations and hidden layer size). We tested different models and found best performance with the design outlined in this work, Figure 3 portrays how the different features (splitting the pairwise stream by spin alignment and backflow coordinates) affected performance for a small demonstration example. Some early design choices behaved better than others, notably, the inclusion of a Jastrow factor $U(X)$ did not systematically improve the performance, similar to atomic and molecular systems [2]. This suggests that the model is able to efficiently describe the electron-electron cusp conditions [10] at low r_s and the strong curvature of the correlation hole [10] at large r_s. The model and algorithm were implemented in JAX [53] and run on RTX 3090 GPUs.

E. Optimization

Before parameters are optimized via variational Monte Carlo (VMC), the *Ansatz* is first pretrained on Hartree-Fock orbitals using methods described in other Refs. [2,24]. At a high level, *Ansatz* orbitals in the Slater determinants are fit to Hartree-Fock orbitals in a supervised way: A loss is computed as the squared difference between them. This helps the *Ansatz* to start closer to the ground state, to avoid being stranded in
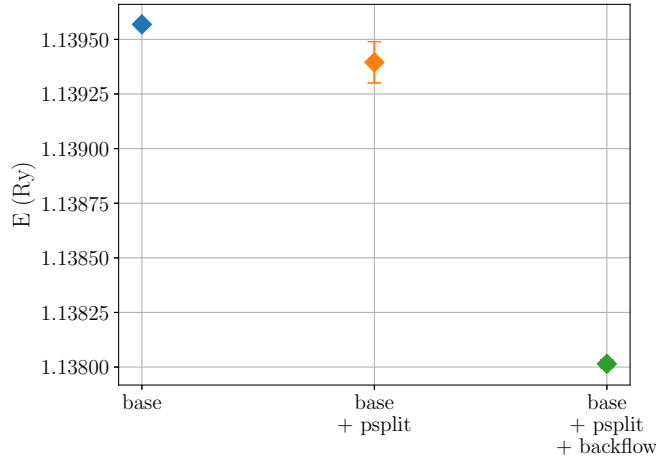
FIG. 3. Plot showing the change in performance for a small model on a paramagnetic ($N_e = 14$) system over 3×10^4 training iterations. psplit indicates the method of splitting the pairwise streams by spin alignment [Eqs. (20)–(21)] and backflow the adjustment to the Hartree-Fock orbital inputs [Eq. (23)]. The error bars (only visible for base+psplit) are average over the two runs for the standard error on the mean of the energy calculation, computed with 1000 batches of 512 walkers.

local minima, and not to diverge during early training due to large gradients.

Additionally, during VMC optimization, parameter gradients computed from Eq. (9) are transformed to approximate natural gradients, using a Kronecker factored approximate curvature (KFAC) [54–56] approximation to the Fisher information matrix,

$$\delta_l = \mathbb{E}_{\mathbf{a}_l}\big[\mathbf{a}_l\mathbf{a}_l^T\big]\nabla_\theta L(\theta)\mathbb{E}_{\mathbf{s}_l}\big[\mathbf{s}_l\mathbf{s}_l^T\big], \qquad (28)$$

where $\mathbf{a}$ and $\mathbf{s}$ are the activations and sensitivities [55], respectively, and l is the layer index. Though correct, this equation ignores a complete description of the algorithm used to compute the approximate natural gradients, which requires damping, constraining of the norm of the transformed gradients, and smoothing approximations to the covariance matrices. Particularly, damping and the covariance matrices are computed via methods developed for convolutional layers [54]. Norm constraint and other details of the method are described in the Appendix of Ref. [24], more background on these optimization routines can be found in Refs. [54,55], and a list of optimization hyperparameters can be found in Table II.

Finally, similar to stochastic reconfiguration [57], the optimal step size in a natural gradient descent scales roughly

TABLE II. Table of KFAC hyperparameters.

Description	Value	Decay	Floor
Learning rate	1×10^{-3}	1×10^{-4}	1×10^{-4}
Damping	1×10^{-4}	1×10^{-2}	1×10^{-6}
Norm constraint	1×10^{-4}	1×10^{-4}	1×10^{-6}
Number of walkers	2048	N/A	N/A
Number of iterations	1×10^5	N/A	N/A

with the inverse of the energy. In order to capture variations between $r_s = 1$ and $r_s = 100$, the gradients are heuristically scaled depending on the density parameter of the system

$$\delta_l' = r_s^{1+r_s/100}\delta_l. \qquad (29)$$

This helps to amplify the gradients when the variance of the energy is small, particularly for systems with large r_s. This heuristic is essential to achieving state-of-the-art accuracy for these systems for both spin-polarized ($N_\uparrow = N_e$) and paramagnetic ($N_\uparrow = N_\downarrow = N_e/2$) systems.

IV. RESULTS

A. Ground-state energy

In order to illustrate the accuracy of WAP-net, we have computed VMC and fixed-node DMC energies based on the iterative backflow (IB) network wave function. Our IB results of Table III outperform the corresponding best single-determinant BF-VMC and BF-DMC values previously published ($N = 19$, $r_s = 1$: Table 5 in the Supplemental Material of Ref. [19], $N = 14$, $r_s = \{1, 2, 5, 10, 20, 50\}$ in Ref. [58]). Comparing to extrapolated transcorrelated FCIQMC [58] available for $N = 14$ at high densities, $r_s \leqslant 5$, we estimate the residual fixed-node error to be around 0.7 mRy or below in this region, becoming more accurate with decreasing density. For the spin-polarized system, IB-DMC values are expected to be in general closer to the exact ground-state values, e.g. around 0.2 mRy for $N = 19$, $r_s = 1$ [19]. Therefore, our baseline IB results provide a challenging benchmark for WAP-net over the full density region of $r_s = 1$ to $r_s = 100$ for the spin-polarized and paramagnetic system.

During the design phase of WAP-net many different model features and implementations were tested. Representative results we include here from those tests are shown in Fig. 3 highlighting the hierarchy of performance observed while developing the model. Base defines a model with periodic input features and permutation equivariant layers. Psplit adds split pairwise layers [Eqs. (20) and (21)], and backflow indicates channeling backflow coordinates from the permutation equivariant layers output to the Hartree-Fock orbitals in a backflowlike way, showing in this small example a clear improvement in performance as these features are added.

The following results refer to the WAP-net with model and KFAC hyperparameters stated in Tables I and II, respectively, for example the number of training iterations was 1×10^5 for each experiment. For the comparative methods (WAP-net with $n_d = 1$ and the IB baseline, which was run with a single Slater determinant, second and fourth column in Table III, respectively, counting the density parameter r_s as the first column), we find that our approach improves upon previous VMC results in all cases $r_s \in \{1, 2, 5, 10, 20, 50, 100\}$ for systems $N_e \in \{7, 14\}$, and most $r_s \in \{1, 2, 5, 10, 20\}$ for the spin-polarized $N_e = 19$ case. Additionally, in all cases, other than $r_s = 2$, for the spin-polarized $N_e = 7$ systems, WAP-net improved upon the IB wave function with DMC. In some cases, $r_s \in \{1, 2\}$ for $N_e = 14$ and $r_s = 1$ for $N_e = 19$,

TABLE III. Results obtained by the simulations described in this paper. The reported values are the energies per particle in Rydbergs and the titles spin-polarised ($N_\uparrow = N_e$) and paramagnetic ($N_\uparrow = N_\downarrow = N_e/2$) describe the systems. Errors are the standard error on the mean represented as bracketed numbers with the same precision as the last digit, for example 2.13(2) is equivalent to 2.13 ± 0.02, where 0.02 is the standard error on the mean of 2.13. For WAP-net, the energies and errors are computed from 1000 batches of 2048 walkers.

r_s	WAP-Net ($n_d = 1$)	WAP-Net ($n_d = 8$)	$E_{\text{VMC}}^{\text{IB}}$	$E_{\text{DMC}}^{\text{IB}}$
		$N_e = 7$ (spin-polarized)		
1.0	2.240785(7)	2.240715(7)	2.24084(2)	2.24080(1)
2.0	0.221766(5)	0.221700(2)	0.221803(3)	0.221758(4)
5.0	−0.132825(2)	−0.132845(1)	−0.132773(1)	−0.132811(1)
10.0	−0.1062797(5)	−0.1062872(5)	−0.1062376(6)	−0.106276(1)
20.0	−0.0645378(2)	−0.0645379(3)	−0.0644823(3)	−0.064533(4)
50.0	−0.02932178(3)	−0.02932161(4)	−0.0292957(1)	−0.0293208(2)
100.0	−0.01549309(2)	−0.01549304(1)	−0.01547628(7)	−0.0154925(1)
		$N_e = 14$ (paramagnetic)		
1.0	1.137912(9)	1.13793(1)	1.13832(1)	1.13795(1)
2.0	−0.016666(5)	−0.016662(5)	−0.016408(5)	−0.01665(1)
5.0	−0.159669(1)	−0.159672(1)	−0.159524(1)	−0.159684(3)
10.0	−0.1104049(6)	−0.1104076(6)	−0.1103229(7)	−0.110436(1)
20.0	−0.0648883(3)	−0.0648868(2)	−0.0648444(2)	−0.0649157(5)
50.0	−0.02924445(6)	−0.02924422(7)	−0.02922920(2)	−0.0292573(2)
100.0	−0.01546048(3)	−0.01545996(3)	−0.01545525(6)	−0.01546859(5)
		$N_e = 19$ (spin-polarized)		
1.0	2.092495(5)	2.092483(5)	2.092544(7)	2.092450(1)
2.0	0.192608(1)	0.192606(2)	0.192637(3)	0.192605(3)
5.0	−0.1345001(7)	−0.1345021(5)	−0.134480(1)	−0.134508(1)
10.0	−0.1057197(2)	−0.1057207(2)	−0.1057090(6)	−0.1057304(6)
20.0	−0.0640219(1)	−0.0640164(2)	−0.0640200(2)	−0.0640333(2)
50.0	−0.02912995(5)	−0.02913142(4)	−0.02913147(5)	−0.02913839(8)
100.0	−0.01543521(2)	−0.01542724(3)	−0.01543798(2)	−0.01544285(2)

respectively, WAP-net improved upon the IB wave function with DMC.

We additionally ran networks with $n_d = 8$ and found no consistent improvement over the single determinant case. The results were marginally improved in the cases $N_e = 7$ $r_s \in \{1, 2, 5, 10, 20\}$, $N_e = 14$ $r_s \in \{5, 10\}$, and $N_e = 19$ $r_s \in \{1, 2, 5, 10, 50\}$.

In order to understand the scaling of WAP-net with respect to system size we computed the average time taken for one iteration of each of the subroutines (sampling, local energy, and KFAC), as seen in Fig. 4. Each set of timing data was taken with one, four, and eight GPUs. Polynomials of form $f(N_e) = aN_e^p + c$ were fit to the data where a and c were adjustable parameters and $p \in \{2, 3, 4\}$. Lines were fit to each of the data sets and the best fit (as measured by the minimum squared residual) was chosen. The cost of the subroutines scales as expected: $O(N_e^3)$ for sampling, which is dominated by the determinant computation [known to scale as $O(N_e^3)$]; $O(N_e^4)$ for the local energy computation [which in some cases was close to $O(N_e^3)$], and can be understood in terms of the scaling as N_e evaluations of the determinant; and finally $O(N_e^2)$ in the case of KFAC. There is a small discontinuity in

the local energy data around $N_e = 32$ for four and eight GPUs. This might be explained by reaching some threshold of data in the GPUs, for example if operations need to be queued as the number of variables increase.

B. Discussion of performance

Whereas our network achieved an overall excellent accuracy compared to the IB baseline, the quality deteriorates for larger r_s and larger systems. This trend might be somewhat expected. The network hyperparameters are kept the same across all systems meaning the setup chosen would need to be extended: Adding more or larger hidden layers, more walkers (to approximate the natural gradients better), or training for more iterations to tackle more complex systems to the same relative performance as the less complex systems (fewer electrons and smaller r_s).

Having that in mind, the energy gain with respect to IB-VMC calculations is comparable to those from IB-DMC for $N = 7$ and $N = 19$ providing a rough demonstration of size consistency of the WAP-net *Ansatz*. However, further
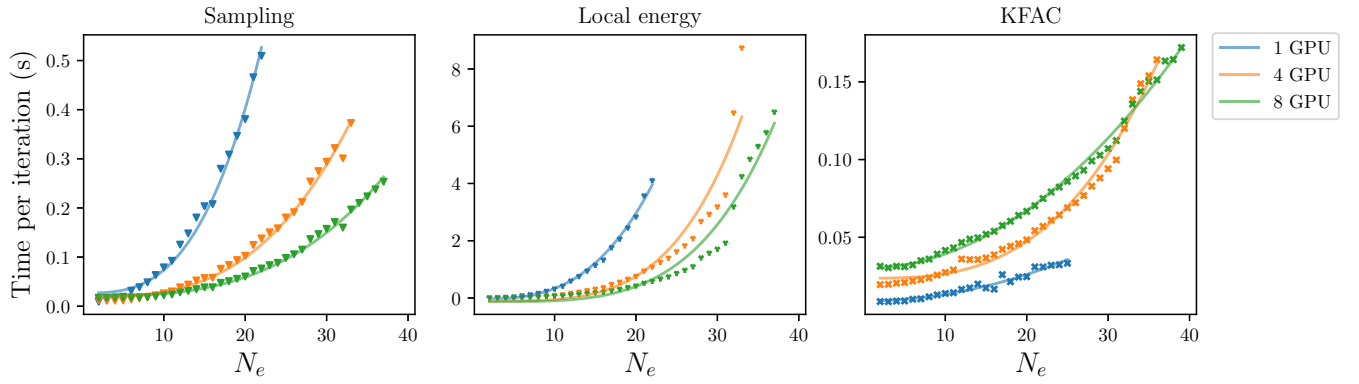
FIG. 4. Plots containing timing data for subroutines (Sampling, left, Local energy, middle, and KFAC, right) for spin-polarized systems as a function of the number of electrons N_e. Each point represents the average of 1000 executions of the subroutine and each plot contains three lines representing the subroutines run with one (blue), four (orange), and eight (green) GPUs. Fit lines are the best fit (the smallest mean squared residual) from a selection of three polynomials of the form $f(N_e) = aN_e^p + c$, where a and c are parameters and $p \in \{2, 3, 4\}$. In most cases, the Sampling routine scales as $O(N_e^3)$, the Local energy as $O(N_e^4)$, and KFAC as $O(N_e^2)$. The only cases where this did not hold were for the local energy subroutine in the spin-polarized system with one GPU, finding a cubic fit, however, the difference between the mean residuals of the fit between the cubic and quartic cases was small (within 2% mean residual).

tests on larger systems are needed to fully confirm size consistency in practice.

In contrast to previous networks on atomic and molecular systems [2,3], we do not observe a consistent improvement increasing the number of Slater determinants in the wave function. This may simply be a consequence of the translational symmetry when considering extended systems and working with closed shell situations, as multideterminant wave functions do not considerably improve the ground-state energies of noble gas atoms, e.g., Ne [2,59].

Our results indicate quadratic scaling of the KFAC subroutine. We expected linear scaling in the number of electrons as adding additional electrons only increases the number of layers for which a KFAC step is performed (computation of the covariances and their inversion). It is possible to see this trend in the one-GPU case and observe that adding additional GPUs adds overhead for moving data between them, potentially explaining the difference between expected and observed scaling. Nevertheless, this subroutine is relatively negligible (in terms of the absolute time) for a complete iteration of training.

In testing and in other works [2,24,26], KFAC shows clear advantages in optimization over more standard techniques such as Adam [3,60]. However, the algorithm requires careful balancing of the hyperparameters (learning rate, damping, and norm constraint) to achieve optimal performance, which consumes time in development and tweaking of model design. Alternate methods such as conjugate gradient, which approximate the Fisher information matrix directly [61] and has been demonstrated in other works [25,62], may improve results or require less heuristics.

C. Single-particle density matrix

So far, our discussion was naturally focused on the calculation of the energy expectation value, intrinsically connected to all variational approaches. Parametrizations of the HEG correlation energy from QMC calculations, provide the input of practically all Kohn-Sham density functionals (DFT)

based on the local density approximation (LDA). Although our WAP-net and IB results provide the basis to reduce the fixed-node error of previous calculations based on Slater-Jastrow or simple backflow wave function, current DFT functionals are more likely affected by corrections to LDA than by these comparably small changes in the HEG energies. On the other hand, improvements of the underlying wave function may be more relevant to correlation functions, which encode direct physical insight.

One of the physical observables where electron-electron interactions are directly visible is the electronic momentum distribution. Whereas electrons only occupy momentum states up to the Fermi surface in an ideal Fermi gas, electronic correlations also involve the occupation of states above the Fermi surface. This leads to a reduction of the size of the jump of the momentum distribution at the Fermi surface, the so-called renormalization factor Z. Values of Z for the HEG are relevant both qualitatively, as the defining feature of the Fermi liquid paradigm, and quantitatively, to explain experimental measurements of the momentum distribution in solid Na [63] and Li [64], corrected by band structure [65] and core electron effects [66].

In the following we will focus on the reduced single-particle density matrix

$$g_1(r) = \mathbb{E}_{X \sim p(X)} \left[\frac{\psi(\mathbf{r}_1 + \mathbf{r}, \mathbf{r}_2, \dots \mathbf{r}_N)}{\psi(\mathbf{r}_1, \mathbf{r}_2, \dots \mathbf{r}_N)} \right], \quad (30)$$

where $p(X)$ is proportional to $|\psi(X)|^2$. From $g_1(r)$ the momentum distribution can then be obtained by Fourier transform.

Having an explicitly parameterized wave function $\psi(X)$, it is straightforward to compute $g_1(r)$ within VMC [67]. The DMC calculation, instead, is more involved. Within DMC the trial wave function, $\psi(X)$, is stochastically projected to its fixed-node ground state, $\psi_0(X)$, resulting in lower energy values. However, the configurations X are now sampled according to the weight $p_{\text{DMC}} \sim |\psi_0(X)\psi(X)|$. For a general operator A, for which ψ_0 is not an eigenfunction, the

calculation of the expectation value $\langle \psi_0 | A | \psi_0 \rangle$ based on the mixed distribution p_{DMC} typically introduces a large variance [68]. Alternatively, the mixed estimator bias can be avoided by reptation Monte Carlo (RMC) methods [69]; however, the application of RMC to off-diagonal properties such as $g_1(r)$ is rather elaborated [28]. Instead, we follow here the common DMC practice to assume ψ sufficiently close to ψ_0, so that

$$\langle \psi_0 | A | \psi_0 \rangle \approx \langle \psi | A | \psi \rangle + [\langle \psi_0 | - \langle \psi |] A | \psi \rangle$$
$$+ \langle \psi | A [| \psi_0 \rangle - | \psi \rangle], \qquad (31)$$

leading to the so-called extrapolated estimator [70]

$$A_{ext} \equiv 2 \mathbb{E}_{X \sim p_{DMC}(X)}[A] - \mathbb{E}_{X \sim p(X)}[A]. \qquad (32)$$

Although the extrapolated estimator can reduce the bias and eventually recover the exact expectation value of local operators (see Ref. [71] for an application to long-range properties of the pair correlation function), this is not guaranteed for nonlocal observables such as $g_1(r)$: key physical properties such as the presence of a condensate for Bose systems [72] or a finite Z for Fermi systems [73] are hard coded in the analytic form of standard wave functions, and they cannot be modified by either the mixed or the extrapolated DMC estimators. In practice, the Fermi liquid nature of the HEG is guessed in advance and maintained through the IB optimization and subsequent DMC projection, and the study of competing phases typically requires generalizations of the wave function specific to the targeted state of the system.

The WAP-net *Ansatz*, on the contrary, is essentially agnostic, because it addresses individually each of the N_α^2 orbitals of Eq. (24) through the many-body prefactors $w_i^{d\alpha}(\mathbf{h}_j^{L\alpha})$. This has the potential to alter the Fermi liquid nature expressed by filling the N_a lowest-energy plane waves $\psi_i^H(\mathbf{r}_j^\alpha)$.

We now present our results for $g_1(r)$ obtained with SJ and IB wave functions, and then discuss the insight we gain using the WAP-net *Ansatz*. Here, SJ denotes a wave function based on a Slater determinant and Jastrow function using only bare electron coordinates.

Figure 5 illustrates the reduced single-particle density matrix and the resulting momentum distribution for the unpolarized $N = 14$ electron system at three different densities, $r_s = 1$, 5, and 100 using the extrapolated estimator with IB wave functions. Increasing r_s (decreasing density), $g_1(r)$ progressively departs from the ideal gas curve. In the momentum distribution, shown in the inset, these deviations correspond to a reduction of the jump, Z, at the Fermi wave vector.

Differences between $g_1(r)$ obtained from optimized SJ, IB, and WAP-net wave functions relative to those of extrapolated IB-DMC are shown in Fig. 6. Under the plausible assumption that the IB-DMC values are more accurate, the SJ-VMC and SJ-DMC results indicate that the DMC extrapolated estimates are not uniformly better than VMC across the range of density and distance explored. This may cast doubt on the full reliability of the correction from the IB-VMC to the IB-DMC results.

The WAP-net *Ansatz* offers a completely independent route to the calculation of $g_1(r)$. Figure 6 shows that the WAP-net values of $g_1(r)$ are much closer to IB-VMC than IB-DMC. While this may seem disappointing at first, there are good reasons to prefer the WAP-net to the IB-DMC results. The
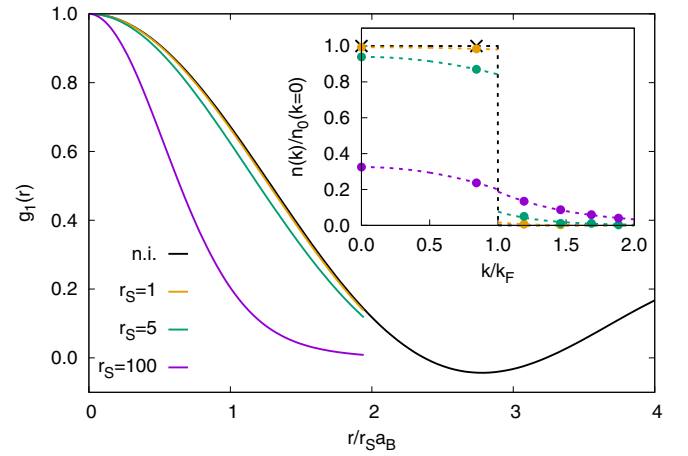

FIG. 5. The reduced single-particle density matrix $g_1(r)$ of the paramagnetic HEG. The colored lines are extrapolated estimates calculated with IB wave functions for $r_s = 1$, 5, and 100 at $N_e = 14$. The black line is the analytic result for the noninteracting Fermi gas at the same system size. The inset shows the corresponding momentum distributions (dotted lines are a guide to the eye).

energies listed in Table III imply that the WAP-net wave function ψ is nearly as good or better, depending on r_s, than the fixed-node IB-DMC projected eigenstate ψ_0. The WAP-net estimate of $g_1(r)$ is better than IB-DMC because it uses an approximation to the true ground state of similar quality and it does not make the further assumption of Eq. (32).

Having better control on $g_1(r)$, we will now conclude on assessing the accuracy of existing QMC of Z for the HEG [28], addressing possible sources of discrepancy with experiment [63,64], as well as with recent diagrammatic Monte Carlo calculations [27]. The momentum distribution as well as the single-particle density matrix shown in Fig. 5 suffer from strong finite-size effects, and thermodynamic limit extrapolation is beyond the scope of the present article. Nevertheless, noting that leading-order finite-size corrections are linear in $g_1(r)$ [28], we can study the systematic error due to quality of
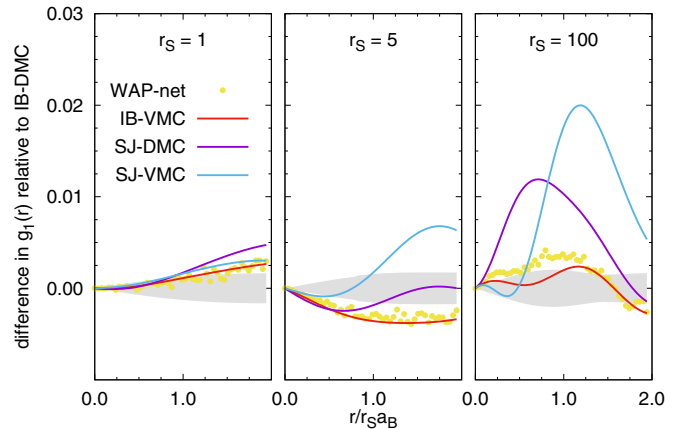

FIG. 6. The difference between various calculations of $g_1(r)$ and the IB extrapolated estimate IB-DMC of Fig. 5. The shaded area indicates the statistical uncertainty of IB-DMC. The WAP-net result is very close to the IB variational estimate IB-VMC.

the wave function, which roughly decouples from that of the system size.

From Fig. 6 we see that IB-VMC agrees well with WAP-net at all three densities whereas changes between SJ-VMC and WAP-net increase with increasing r_s. The corresponding variations in the momentum distribution would indicate a lowering of Z compared to optimized SJ-VMC of ~ 0.03 for $r_s = 5$ and ~ 0.04 for $r_s = 100$, whereas at $r_s = 1$ differences are below our resolution. Calculations done in Ref. [28] are based on analytical Slater-Jastrow and (noniterated) backflow wave functions. The quality of BF-RMC in Ref. [28] is therefore expected to lie between our optimized SJ-VMC and IB-VMC/WAP-net ones. Thus, the systematic error due to the quality of the wave function used in Ref. [28] is less than the error quoted, whereas Z may lower by at most 3% at $r_s = 5$. Although small, such a shift may be relevant for improving comparison with experiments [63,64], but would also place them slightly below those of diagrammatic Monte Carlo calculations [27].

V. CONCLUSION

In this work we have presented a neural network *Ansatz*, WAP-net, for approximating the many-electron ground state of the homogeneous electron gas, extending previous work on atoms and molecules to periodic electronic systems. Using the WAP-net *Ansatz* with variational Monte Carlo methods, we have obtained accuracies comparable to or beyond our best diffusion Monte Carlo calculations based on iterative backflow wave functions for spin-polarized and paramagnetic systems over a broad density region. The comparison of results for polarized systems from $N_e = 7$ to $N_e = 19$ shows that the *Ansatz* is size consistent (at least in the single determinant case, $n_d = 1$). Larger systems are within reach, such that thermodynamic limit values can be properly addressed based

on finite-size extrapolations [71]. This method can be extended to complex wave functions with more general twisted boundary conditions [74] to accelerate the thermodynamic limit extrapolation.

Applying the WAP-net *Ansatz* to the homogeneous electron gas, we focused on the description of intrinsic electron correlations, not captured by independent particle approaches, e.g., the reduced single-particle density matrix where we discussed possible bias of previous calculations. In order to describe the electronic structure of real materials, the periodic lattice potential due to the ionic crystal must be further included in the Hamiltonian. We expect WAP-net to still provide an accurate description to describe the electronic structure in materials, a detailed study to benchmark its performance for solids will be addressed in a future work.

Note added. Recently, two related papers studying the electron gas with similar a neural network approach appeared, [75] focusing on the transition from the Fermi liquid to the Wigner crystal, and [30] towards more realistic systems.

[1] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, 2020).

[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, Phys. Rev. Res. **2**, 033429 (2020).

[3] J. Hermann, Z. Schätzle, and F. Noé, Deep-neural-network solution of the electronic schrödinger equation, Nature Chem. **12**, 891 (2020).

[4] G. Pescia, J. Han, A. Lovato, J. Lu, and G. Carleo, Neural-network quantum states for periodic systems in continuous space, Phys. Rev. Res. **4**, 023138 (2022).

[5] M. Holzmann and S. Moroni, Orbital-dependent backflow wave functions for real-space quantum monte carlo, Phys. Rev. B **99**, 085121 (2019).

[6] M. Taddei, M. Ruggeri, S. Moroni, and M. Holzmann, Iterative backflow renormalization procedure for many-body ground-state wave functions of strongly interacting normal fermi liquids, Phys. Rev. B **91**, 115106 (2015).

[7] R. M. Martin, L. Reining, and D. M. Ceperley, *Interacting Electrons* (Cambridge University Press, Cambridge, 2016).

[8] G. D. Mahan, Homogeneous electron gas, in *Many-Particle Physics* (Springer, Berlin, 2000), pp. 295–374.

[9] K. W. Jacobsen, J. K. Norskov, and M. J. Puska, Interatomic interactions in the effective-medium theory, Phys. Rev. B **35**, 7423 (1987).

[10] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, Quantum monte carlo simulations of solids, Rev. Mod. Phys. **73**, 33 (2001).

[11] W. A. Harrison, *Solid State Theory* (Courier Corporation, Mineola, 1980).

[12] G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, 2005).

[13] L. Hedin, New method for calculating the one-particle green's function with application to the electron-gas problem, Phys. Rev. **139**, A796 (1965).

[14] K. Chen and K. Haule, A combined variational and diagrammatic quantum monte carlo approach to the many-electron problem, Nature Commun. **10**, 3725 (2019).

[15] D. Ceperley, Ground state of the fermion one-component plasma: A monte carlo study in two and three dimensions, Phys. Rev. B **18**, 3126 (1978).

[16] D. M. Ceperley and B. J. Alder, Ground State of the Electron Gas by a Stochastic Method, Phys. Rev. Lett. **45**, 566 (1980).

[17] J. J. Shepherd, G. H. Booth, and A. Alavi, Investigation of the full configuration interaction quantum monte carlo method using homogeneous electron gas models, J. Chem. Phys. **136**, 244101 (2012).

[18] J. J. Shepherd and A. Grüneis, Many-Body Quantum Chemistry for the Electron Gas: Convergent Perturbative Theories, Phys. Rev. Lett. **110**, 226401 (2013).

[19] M. Ruggeri, P. L. Ríos, and A. Alavi, Correlation energies of the high-density spin-polarized electron gas to mev accuracy, Phys. Rev. B **98**, 161105(R) (2018).

[20] M. Raissi, P. Perdikaris, and G. E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. **378**, 686 (2019).

[21] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, Mastering the game of go without human knowledge, Nature (London) **550**, 354 (2017).

[22] B. Kiumarsi, K. G. Vamvoudakis, H. Modares, and F. L. Lewis, Optimal and autonomous control using reinforcement learning: A survey, IEEE Trans. Neural Netw. Learning Syst. **29**, 2042 (2017).

[23] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[24] M. Wilson, N. Gao, F. Wudarski, E. Rieffel, and N. M. Tubman, arXiv:2103.12570 [physics, physics:quant-ph].

[25] N. Gao and S. Günnemann, Ab-initio potential energy surfaces by pairing gnns with neural wave functions, in *International Conference on Learning Representations* (ICLR, Appleton, 2022).

[26] J. S. Spencer, D. Pfau, A. Botev, and W. M. C. Foulkes, arXiv:2011.07125.

[27] K. Haule and K. Chen, Single-particle excitations in the uniform electron gas by diagrammatic monte carlo, Sci. Rep. **12**, 2294 (2022).

[28] M. Holzmann, B. Bernu, C. Pierleoni, J. McMinis, D. M. Ceperley, V. Olevano, and L. Delle Site, Momentum Distribution of the Homogeneous Electron Gas, Phys. Rev. Lett. **107**, 110402 (2011).

[29] I. von Glehn, J. S. Spencer, and D. Pfau, arXiv:2211.13672.

[30] X. Li, Z. Li, and J. Chen, *Ab initio* calculation of real solids via neural network ansatz, Nat Commun **13**, 7895 (2022).

[31] K. E. Schmidt, M. A. Lee, M. H. Kalos, and G. V. Chester, Structure of the Ground State of a Fermion Fluid, Phys. Rev. Lett. **47**, 807 (1981).

[32] M. Holzmann, B. Bernu, and D. M. Ceperley, Many-body wavefunctions for normal liquid ^{3}He, Phys. Rev. B **74**, 104510 (2006).

[33] T. Bouabça, B. Braïda, and M. Caffarel, Multi-jastrow trial wavefunctions for electronic structure calculations with quantum monte carlo, J. Chem. Phys. **133**, 044111 (2010).

[34] M. Ruggeri, S. Moroni, and M. Holzmann, Nonlinear Network Description for Many-Body Quantum Systems in Continuous Space, Phys. Rev. Lett. **120**, 205302 (2018).

[35] M. Holzmann and S. Moroni, Itinerant-Electron Magnetism: The Importance of Many-Body Correlations, Phys. Rev. Lett. **124**, 206404 (2020).

[36] J. Kessler, F. Calcavecchia, and T. D. Kühne, Artificial neural networks as trial wave functions for quantum monte carlo, Adv. Theory Simul. **4**, 2000269 (2021).

[37] J. Han, L. Zhang, and W. E, Solving many-electron Schrödinger equation using deep neural networks, J. Comput. Phys. **399**, 108929 (2019).

[38] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for ab-initio electronic structure, Nature Commun. **11**, 2368 (2020).

[39] A. Acevedo, M. Curry, S. H. Joshi, B. Leroux, and N. Malaya, Vandermonde Wave Function *Ansatz* for Improved Variational Monte Carlo, in *2020 IEEE/ACM Fourth Workshop on Deep Learning on Supercomputers* (IEEE, USA, 2020), pp. 40–47.

[40] H. Li, Q. Zhai, and J. Z. Y. Chen, Neural-network-based multi-state solver for a static Schrödinger equation, Phys. Rev. A **103**, 032405 (2021).

[41] H. Xie, L. Zhang, and L. Wang, *Ab-initio* study of interacting fermions at finite temperature with neural canonical transformation, arXiv:2105.08644.

[42] L. Ziyin, T. Hartwig, and M. Ueda, in *Advances in Neural Information Processing Systems*, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc., 2020), Vol. 33, pp. 1583–1594.

[43] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, Implicit neural representations with periodic activation functions, in *Advances in Neural Information Processing Systems* edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc., 2020), Vol. 33, pp. 7462–7473.

[44] 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. **148**, 241722 (2018).

[45] T. Xie, X. Fu, O.-E. Ganea, R. Barzilay, and T. Jaakkola, arXiv:2110.06197.

[46] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, Neural Networks **2**, 359 (1989).

[47] P. P. Ewald, Die berechnung optischer und elektrostatischer gitterpotentiale, Annalen der Physik **369**, 253 (1921).

[48] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Courier Corporation, Mineola, 2012).

[49] W. K. Hastings, *Monte Carlo Sampling Methods using Markov Chains and Their Applications* (Oxford University Press, Oxford, 1970).

[50] M. Suzuki, *Quantum Monte Carlo Methods in Condensed Matter Physics* (World Scientific, Singapore, 1993).

[51] W. A. Lester Jr, L. Mitas, and B. Hammond, Quantum monte carlo for atoms, molecules and solids, Chem. Phys. Lett. **478**, 1 (2009).

[52] S. Elfwing, E. Uchibe, and K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, Neural Networks **107**, 3 (2018).

[53] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python+NumPy programs, 2018.

[54] R. Grosse and J. Martens, A kronecker-factored approximate fisher matrix for convolution layers, in *International Conference on Machine Learning* (PMLR, Brookline, 2016), pp. 573–582.

[55] J. Martens and R. Grosse, Optimizing neural networks with kronecker-factored approximate curvature, in *International Conference on Machine Learning* (PMLR, Brookline, 2015) pp. 2408–2417.

[56] J. Ba, R. Grosse, and J. Martens, Distributed second-order optimization using kronecker-factored approximations, *International Conference on Learning Representations* (ICLR, 2017).

[57] S. Sorella, Generalized lanczos algorithm for variational quantum monte carlo, Phys. Rev. B **64**, 024512 (2001).

[58] K. Liao, T. Schraivogel, H. Luo, D. Kats, and A. Alavi, Towards efficient and accurate ab initio solutions to periodic systems via transcorrelation and coupled cluster theory, Phys. Rev. Res. **3**, 033072 (2021).

[59] M. D. Brown, J. R. Trail, P. López Ríos, and R. J. Needs, Energies of the first row atoms from quantum monte carlo, J. Chem. Phys. **126**, 224110 (2007).

[60] M. Scherbela, R. Reisenhofer, L. Gerard, P. Marquetand, and P. Grohs, Nat. Comput. Sci. **2**, 331 (2022).

[61] E. Neuscamman, C. J. Umrigar, and G. K.-L. Chan, Optimizing large parameter sets in variational quantum Monte Carlo, Phys. Rev. B **85**, 045103 (2012).

[62] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, in *3rd International Conference for Learning Representations* (ICLR, Appleton, 2014).

[63] S. Huotari, J. A. Soininen, T. Pylkkänen, K. Hämäläinen, A. Issolah, A. Titov, J. McMinis, J. Kim, K. Esler, D. M. Ceperley, M. Holzmann, and V. Olevano, Momentum Distribution and Renormalization Factor in Sodium and the Electron Gas, Phys. Rev. Lett. **105**, 086403 (2010).

[64] N. Hiraoka, Y. Yang, T. Hagiya, A. Niozu, K. Matsuda, S. Huotari, M. Holzmann, and D. M. Ceperley, Direct observation of the momentum distribution and renormalization factor in lithium, Phys. Rev. B **101**, 165124 (2020).

[65] V. Olevano, A. Titov, M. Ladisa, K. Hämäläinen, S. Huotari, and M. Holzmann, Momentum distribution and compton profile by the ab initio gw approximation, Phys. Rev. B **86**, 195123 (2012).

[66] Y. Yang, N. Hiraoka, K. Matsuda, M. Holzmann, and D. M. Ceperley, Quantum monte carlo compton profiles of solid and liquid lithium, Phys. Rev. B **101**, 165125 (2020).

[67] W. L. McMillan, Ground state of liquid he⁴, Phys. Rev. **138**, A442 (1965).

[68] K. S. Liu, M. H. Kalos, and G. V. Chester, Quantum hard spheres in a channel, Phys. Rev. A **10**, 303 (1974).

[69] S. Baroni and S. Moroni, Reptation Quantum Monte Carlo: A Method for Unbiased Ground-State Averages and Imaginary-Time Correlations, Phys. Rev. Lett. **82**, 4745 (1999).

[70] D. Ceperley and M. Kalos, *Quantum Many-Body Problems in Monte Carlo Methods in Statistical Physics*, edited by K. Binder (Springer-Verlag, Berlin, 1979).

[71] M. Holzmann, R. C. Clay III, M. A. Morales, N. M. Tubman, D. M. Ceperley, and C. Pierleoni, Theory of finite size effects for electronic quantum monte carlo calculations of liquids and solids, Phys. Rev. B **94**, 035126 (2016).

[72] L. Reatto and G. V. Chester, Phonons and the properties of a bose system, Phys. Rev. **155**, 88 (1967).

[73] S. Moroni, G. Senatore, and S. Fantoni, Momentum distribution of liquid helium, Phys. Rev. B **55**, 1040 (1997).

[74] C. Lin, F. H. Zong, and D. M. Ceperley, Twist-averaged boundary conditions in continuum quantum monte carlo algorithms, Phys. Rev. E **64**, 016702 (2001).

[75] G. Cassella, H. Sutterud, S. Azadi, N. D. Drummond, D. Pfau, J. S. Spencer, and W. M. C. Foulkes, Discovering Quantum Phase Transitions with Fermionic Neural Networks, Phys. Rev. Lett. **130**, 036401 (2023).