# Message-passing neural quantum states for the homogeneous electron gas

Gabriel Pescia [1,2,*] Jannes Nys [1,2] Jane Kim,[3] Alessandro Lovato [4,5,6] and Giuseppe Carleo [1,2]

[1]*Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[2]*Center for Quantum Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland*
[3]*Department of Physics and Astronomy and Facility for Rare Isotope Beams Michigan State University, East Lansing, Michigan 48824, USA*
[4]*Physics Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[5]*Computational Science Division, Argonne National Laboratory, Argonne, Illinois 60439, USA*
[6]*INFN-TIFPA Trento Institute for Fundamental Physics and Applications, 38123 Trento, Italy*

We introduce a message-passing neural-network (NN)-based wave function *Ansatz* to simulate extended, strongly interacting fermions in continuous space. Symmetry constraints, such as continuous translation symmetries, can be readily embedded in the model. We demonstrate its accuracy by simulating the ground state of the homogeneous electron gas in three spatial dimensions at different densities and system sizes. With orders of magnitude fewer parameters than state-of-the-art NN wave functions, we demonstrate better or comparable ground-state energies. Reducing the parameter complexity allows scaling to $N = 128$ electrons, previously inaccessible to NN wave functions in continuous space, allowing future work on finite-size extrapolations to the thermodynamic limit. We also show the capability of the *Ansatz* to quantitatively represent different phases of matter.

## I. INTRODUCTION

Predicting emergent physical phenomena and system properties from the *ab initio* description of the constituents of the system is notoriously difficult [1,2]. Fermionic systems can exhibit strong correlations among the particles, leading to collective phenomena in the form of exotic phases of matter, e.g., superconductivity and superfluidity [3,4].

In recent years, progress in numerical simulations of strongly correlated systems was triggered by the development of increasingly precise machine learning (ML) approximation techniques. Most notably, artificial neural network (NN) architectures, in combination with variational Monte Carlo (VMC), have shown great promise in representing the ground states of quantum spin systems, especially in >1 spatial dimension [5–12]. Due to the universal approximation property of NNs, NN quantum states (NQS) can, in theory, accurately represent any quantum many-body state [13,14]. NQS have been extended to fermionic degrees of freedom in a discrete basis [15–17], by incorporating the indistinguishability of quantum particles. More recently, advancements to ground and excited state searches for fermionic and bosonic continuous degrees of freedom with open [18–20] and periodic boundary conditions (PBCs) [21–23] have been introduced.

The flexibility of NQS, compared with more traditional models, allows for the representation of multiple phases of matter, and even different physical systems, with a single *Ansatz*. To exemplify this point, we refer to NQS studies on the ground state of molecular systems [18,19], solutions to effective field theory Hamiltonians describing atomic nuclei [24–26], bulk studies of fermionic and bosonic extended systems [21,23,27,28] as well as NQS simulations of low-density neutron matter found in neutron stars [29]. The downside of this flexibility, especially in continuous space, is that NQS typically need a significant amount of variational parameters to reach a given accuracy. This makes optimization challenging and costly, preventing the usage of refined optimization schemes, e.g., second-order optimization procedures [30,31]. As a result, the accessible system sizes are limited to a few tens of particles. However, studying larger system sizes is of utmost importance to estimate physical properties in the thermodynamic limit [32–35]. To remedy the situation, NQS architectures must be developed that significantly reduce the parameter complexity while retaining high accuracy.

In this paper, we introduce a NN wave function suitable for simulating strongly interacting fermionic quantum systems in continuous space with 1–2 orders of magnitude fewer parameters than the current state-of-the-art NQS. The general form of the *Ansatz* is motivated by an analytical argument, relating the exact ground-state wave function to a many-body coordinate transformation of the electronic coordinates. It uses a permutation-equivariant message-passing architecture on a graph, inherently implementing the indistinguishability of quantum particles of the same species [36]. As an application, we study the homogeneous electron gas (HEG) in three spatial dimensions without explicitly breaking any of the fundamental symmetries of the system, such as translations and spin-inversion symmetry. We study the onset of Wigner crystallization at low densities.

*Contact author: gabriel.pescia@epfl.ch

## II. METHODS

### A. Exact backflow transformations

Throughout this paper, we consider a nonrelativistic Hamiltonian of identical particles with mass $m$ in $d$ spatial dimensions:

$$H = -\frac{\hbar^2}{2m} \sum_i^N \nabla_{\mathbf{r}_i}^2 + V(\mathbf{X}), \qquad (1)$$

where the potential and interaction energy $V$ is assumed to be diagonal in position representation, defined by the particle coordinates $\mathbf{X} = (\mathbf{r}_1, ..., \mathbf{r}_N)$, $\mathbf{r}_i \in \mathbb{R}^d$. In the following, we derive an analytic functional form of the ground-state wave function and relate it to our variational *Ansatz*.

Given a suitable reference state $|\Phi_0\rangle$, as the initial condition for the imaginary-time ($\tau$) evolution induced by the Hamiltonian $\Phi_\tau(\mathbf{X}) = \langle \mathbf{X}| \exp(-\tau H)|\Phi_0\rangle$, the exact ground state is obtained in the large imaginary-time limit: $\lim_{\tau \to \infty} \Phi_\tau(\mathbf{X}) \propto \Psi_0(\mathbf{X})$, provided $|\Phi_0\rangle$ is nonorthogonal to the exact ground state $|\langle\Psi_0|\Phi_0\rangle| > 0$. For fermions, nonorthogonality implies that the wave function must be at least antisymmetric with respect to the exchange of two particles, i.e., $\Phi_0[\mathcal{P}_{ij}(\mathbf{X})] = -\Phi_0(\mathbf{X})$ ($\mathcal{P}_{ij}$ permutes particles $i$ and $j$). Assuming a valid quantum reference state (twice differentiable and integrable) and finite matrix elements of the propagator [37], we can apply the mean value theorem to the imaginary-time evolved state, yielding the following:

$$\Phi_\tau(\mathbf{X}) = \int_\Omega d\mathbf{X}' G_\tau(\mathbf{X}, \mathbf{X}')\Phi_0(\mathbf{X}') \qquad (2)$$

$$= \text{Vol}(\Omega) \times G_\tau[\mathbf{X}, \mathbf{Y}(\mathbf{X})]\Phi_0[\mathbf{Y}(\mathbf{X})], \qquad (3)$$

where $\Omega$ is the (convex) integration domain for the positional degrees of freedom, and $G_\tau(\mathbf{X}, \mathbf{X}') = \langle\mathbf{X}|\exp(-\tau H)|\mathbf{X}'\rangle$ is the matrix element of the imaginary-time propagator, which is bounded for finite $\tau$. In Eq. (3), we introduced the mean value point $\mathbf{Y}(\mathbf{X}) = [\mathbf{y}_1(\mathbf{X}), \ldots, \mathbf{y}_N(\mathbf{X})] \in \Omega$, depending parametrically on the coordinates $\mathbf{X}$.

For general Hamiltonians in the form of Eq. (1), we have $G_\tau(\mathbf{X}, \mathbf{X}') \geqslant 0$, for all $\mathbf{X}, \mathbf{X}'$. Moreover, $G_\tau(\mathbf{X}, \mathbf{X}')$ is invariant under the exchange of particle coordinates: $G_\tau[\mathcal{P}_{ij}(\mathbf{X}), \mathcal{P}_{ij}(\mathbf{X}')] = G_\tau(\mathbf{X}, \mathbf{X}')$. In the fermionic case, the latter implies that $\mathbf{Y}(\mathbf{X})$ must be equivariant under particle exchange $\mathbf{Y}[\mathcal{P}_{ij}(\mathbf{X})] = \mathcal{P}_{ij}[\mathbf{Y}(\mathbf{X})]$ to ensure antisymmetry of the total wave function. Equation (3) therefore yields the product between a permutation symmetric, positive semidefinite function $J(\mathbf{X}) = G_\tau[\mathbf{X}, \mathbf{Y}(\mathbf{X})] \times \text{Vol}(\Omega)$ and a reference state computed at modified coordinates $\mathbf{Y}(\mathbf{X})$:

$$\Phi_\tau(\mathbf{X}) = J(\mathbf{X}) \times \Phi_0[\mathbf{Y}(\mathbf{X})]. \qquad (4)$$

Identification of the mean value point $Y(X)$ with a many-body coordinate transformation gives an alternative justification for backflow transformations [38] of single-particle coordinates. With a Slater determinant of given spin orbitals $\phi_\mu(\mathbf{r}_i)$ as the initial state $\Phi_0(\mathbf{X}) = \det \phi_\mu(\mathbf{r}_i)/\sqrt{N!}$, Eq. (4) is structurally related to the heuristic Jastrow-Backflow variational form [39,40]. We note that the symmetric contribution $J(X)$ can be incorporated into the determinant:

$$\Phi_\tau(\mathbf{X}) = \mathcal{K} \times \det \varphi_\mu[\mathbf{y}_i(\mathbf{X})], \qquad (5)$$
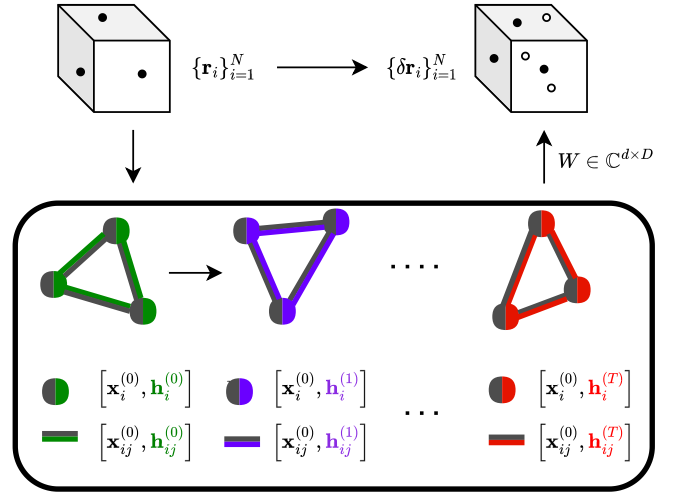


FIG. 1. Illustration of the backflow transformation using message-passing neural quantum state (MP-NQS). Single-particle coordinates $\mathbf{r}_i \in \mathbb{R}^d$ (black dots, top left) are mapped to quasiparticle coordinates $\delta\mathbf{r}_i \in \mathbb{C}^d$ (black/white dots for real/imaginary part, top right). From an initial graph consisting of an initial feature vector (dark gray) and a hidden state (green), a new graph, consisting of the initial features and an updated hidden state (indicated by different coloring), is constructed via messages. The final node states are linearly transformed back to $\delta\mathbf{r}_i = W \cdot \mathbf{g}_i^{(T)}$, containing information about all particles ($D$ is the dimension of the nodes of the last graph).

with $\varphi_\mu[\mathbf{y}_i(\mathbf{X})] = \phi_\mu[\mathbf{y}_i(\mathbf{X})] \times \sqrt[N]{J(\mathbf{X})}$, and $\mathcal{K}$ a normalization constant.

The functional form in Eqs. (4) and (5) is exact, provided that the symmetric factor $J(\mathbf{X})$ and the mean value coordinates $\mathbf{Y}(\mathbf{X})$ satisfy Eq. (3), and the reference state is not orthogonal to the exact ground state. An approximate but explicit form for the coordinate transformation $\mathbf{Y}(\mathbf{X})$ can be obtained by repeatedly applying the imaginary-time propagator to the reference state in the limit of small $\tau$. This process gives rise to the iterative backflow transformation, as introduced in Ref. [41,42].

### B. Message-passing neural quantum states

Motivated by Eq. (5), we use single-particle orbitals $\{\phi_\mu\}_{\mu=1}^N$ evaluated at many-body backflow coordinates $\mathbf{Y}(\mathbf{X})$ to construct the variational *Ansatz*. The backflow transformation $[\mathbf{Y}(\mathbf{X})]$ is parameterized with permutation-equivariant message-passing NNs (MPNNs) [36] (see Fig. 1); hence, we name it message-passing NQS (MP-NQS). In the MPNN, an all-to-all connected graph, encoding effective particle positions (nodes) and their interactions (edges):

$$\mathbf{g}_i^{(t)} = [\mathbf{x}_i^{(0)}, \mathbf{h}_i^{(t)}] \in \mathbb{R}^{D_1}, \qquad (6)$$

$$\mathbf{g}_{ij}^{(t)} = [\mathbf{x}_{ij}^{(0)}, \mathbf{h}_{ij}^{(t)}] \in \mathbb{R}^{D_2}, \qquad (7)$$

is updated iteratively. Here, we have introduced a discrete step index $t \geqslant 0$, initial feature vectors $\mathbf{x}_i^{(0)} \in \mathbb{R}^{D_1^{(0)}}$, $\mathbf{x}_{ij}^{(0)} \in \mathbb{R}^{D_2^{(0)}}$, and we denote concatenation with $[\cdot, \cdot]$. The iterative equations above also contain auxiliary variables known as hidden states $[\mathbf{h}_i^{(t)} \in \mathbb{R}^{D_1^h}, \mathbf{h}_{ij}^{(t)} \in \mathbb{R}^{D_2^h}]$ with suitably chosen *feature*

dimensions such that $D_1^h + D_1^{(0)} = D_1$, and $D_2^h + D_2^{(0)} = D_2$. The construction of the initial feature vectors $[\mathbf{x}_i^{(0)}, \mathbf{x}_{ij}^{(0)}]$ is system dependent and will be discussed after the description of the physical system.

The hidden states are initialized, independently of their index, with vectors of learnable parameters. They are updated using permutation-equivariant *messages*, obtained from a variant of the attention mechanism [43], we dub here *particle attention*. Specifically, the messages are given by weighted transformations of the edges of the graphs: $\mathbf{m}_{ij}^{(t+1)} = \boldsymbol{\omega}_{ij}^{(t)}[\mathbf{g}_{ij}^{(t)}] \odot \boldsymbol{\phi}[\mathbf{g}_{ij}^{(t)}]$, where $\odot$ represents element-wise multiplication along the feature dimension, and $\boldsymbol{\omega}_{ij}^{(t)}[\mathbf{g}_{ij}^{(t)}] \in \mathbb{R}^{D_2}$ are weight vectors, given by

$$\boldsymbol{\omega}_{ij}^{(t)} = \mathrm{GELU}\left[ \sum_l \mathbf{Q}_{il}^{(t)} \mathbf{K}_{lj}^{(t)} \right]. \tag{8}$$

where we used *query/key* matrices given by $\mathbf{Q}_{ij}^{(t)} = W_Q \cdot \mathbf{g}_{ij}^{(t)}$ and $\mathbf{K}_{ij}^{(t)} = W_K \cdot \mathbf{g}_{ij}^{(t)}$ (with $W_Q, W_K \in \mathbb{R}^{D_2 \times D_2}$) and applied an element-wise GELU nonlinearity [44] to the overlap between queries and keys along the *particle dimension* (as opposed to the feature dimension [43,45]), resulting in permutation-equivariant weights. Particle attention compares the environments of particles $i$ and $j$ and effectively increases the order of correlations that can be embedded in a single iteration of the network. This is crucial to reduce the total number of network iterations (parameters) and capture many-body effects. A comparison with the self-attention mechanism introduced in [43] can be found in Appendix C.

We update the hidden states using the nodes and edges of the current graphs together with the messages:

$$\mathbf{h}_i^{(t+1)} = \mathbf{f}\left[ \mathbf{g}_i^{(t)}, \sum_{j \neq i} \mathbf{m}_{ij}^{(t+1)} \right] \tag{9}$$

$$\mathbf{h}_{ij}^{(t+1)} = \tilde{\mathbf{f}}\left[ \mathbf{g}_{ij}^{(t)}, \mathbf{m}_{ij}^{(t+1)} \right] \tag{10}$$

The functions $\boldsymbol{\phi}$, $\mathbf{f}$, and $\tilde{\mathbf{f}}$ are parameterized by multilayer perceptrons (MLPs). The final updated graph then has the same structure as the former but with updated hidden states:

$$\mathbf{g}_i^{(t+1)} = \left[ \mathbf{x}_i^{(0)}, \mathbf{h}_i^{(t+1)} \right] \in \mathbb{R}^{D_1}, \tag{11}$$

$$\mathbf{g}_{ij}^{(t+1)} = \left[ \mathbf{x}_{ij}^{(0)}, \mathbf{h}_{ij}^{(t+1)} \right] \in \mathbb{R}^{D_2}. \tag{12}$$

Inclusion of the initial inputs, referred to as a *skip connection* in ML literature [46], mitigates the vanishing gradient problem and allows a more efficient capture of correlations. For a comparison of the MP-NQS and conventional MPNN, see Appendix B.

To obtain the backflow coordinates $\mathbf{Y}(\mathbf{X})$, we compute $\mathbf{y}_i(\mathbf{X}) = \mathbf{r}_i + \delta\mathbf{r}_i(\mathbf{X})$, where the displacements $\delta\mathbf{r}_i(\mathbf{X})$ are obtained by a linear transformation of the final node states to $d$ dimensions, i.e., $\delta\mathbf{r}_i(\mathbf{X}) = W \cdot \mathbf{g}_i^{(T)}$, with $W \in \mathbb{C}^{d \times D_1}$. The complex-valued backflow transformation allows changing the degree of localization, determined by the chosen single-particle orbitals and representing complex-valued wave functions in general.

Following Eq. (5), we further augment the orbitals with a permutation-invariant factor $J$ of the form:

$$J(\mathbf{Y}, \mu) = \sum_i j(\mathbf{y}_i, \mu), \tag{13}$$

where $j$ is parameterized with a MLP, and $\mu$ denotes the quantum numbers of the orbitals, yielding

$$\Psi(\mathbf{X}) = \det \varphi_\mu[\mathbf{y}_i(\mathbf{X})], \tag{14}$$

with $\varphi_\mu(\mathbf{y}_i) = \exp[J(\mathbf{Y}, \mu)] \times \phi_\mu(\mathbf{y}_i)$.

### C. Electron gas

We now study the case of the HEG in $d = 3$ spatial dimensions, a prototypical model for electronic structure in solids. It includes Coulomb interactions among the electrons of the solids while treating its positively charged ions as uniform, static, positive background [47]. Despite this simplification, the HEG exhibits different phases of matter and captures properties of real solids, particularly of alkali metals. The Hamiltonian (in units of Hartree), for a system of $N$ electrons with uniform density $n = \frac{N}{V}$, is given by

$$H = -\frac{1}{2r_s^2} \sum_i^N \nabla_{\mathbf{r}_i}^2 + \frac{1}{r_s} \sum_{i<j}^N \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|} + \mathrm{const.}, \tag{15}$$

where we introduced the Wigner-Seitz radius $r_s = \sqrt[3]{3/(4\pi n)}$ and a constant arising from the electron-background interaction [39]. The conditionally convergent series of pairwise Coulomb interactions is evaluated using the Ewald summation technique, as is standard for extended systems in quantum Monte Carlo (QMC) [48–50]. We will assume a fixed spin polarization with $N = N_\uparrow + N_\downarrow$, where $N_{\uparrow/\downarrow}$ denotes the number of up/down spins. Furthermore, $s_i \in \{\uparrow, \downarrow\}$ denotes the spin of the $i$th electron. We equip the cubic simulation cell of side length $L$ with PBCs in all spatial directions to access the bulk of the system.

As in Eq. (5), we use a single Slater determinant as a reference state $\Phi_0(\mathbf{X})$. We expect a liquid-crystal phase transition for the HEG, as a function of the density $n$. The dominating kinetic energy in Eq. (15) ($\sim 1/r_s^2$) for large $n$ leads to the well-known Fermi liquid behavior. For small $n$, the potential energy ($\sim 1/r_s$) dominates and enforces a crystalline body-centered cubic (BCC) structure among the electrons, known as a Wigner crystal [51]. To model the liquid state, we use plane-wave orbitals as a natural and physically motivated reference state: $\phi_{\mathbf{k}}(\mathbf{r}) = \exp[i\mathbf{k} \cdot \mathbf{r}]$, with $\mathbf{k} = \frac{2\pi}{L}\mathbf{n}$, where $\mathbf{n} \in \mathbb{Z}^d$. These orbitals allow modeling of the HEG at fixed total momentum $\mathbf{k}_{\mathrm{tot}} = \sum_{i=1}^N \mathbf{k}_i$. To account for the spin $s$ of a particle located at $\mathbf{r}$, we use spin orbitals $\phi_\mu(\mathbf{r}, s) = \phi_{\mathbf{k}_\mu}(\mathbf{r})\delta_{s_\mu, s}$, where each spin orbital is characterized by the quantum numbers $\mu = (\mathbf{k}_\mu, s_\mu)$. This choice of spin orbitals lets the determinant factorize into a product of determinants of up- and down-spin orbitals. To study the Wigner crystal phase, we employ localized Gaussian orbitals centered around the BCC lattice sites $\mathbf{R}_\mu$: $\phi_\mu(\mathbf{r}, s) = \sum_{\mathbf{R}_n} \exp[-\alpha(\mathbf{r} - \mathbf{R}_\mu + \mathbf{R}_n)^2]\delta_{s_\mu, s}$, where $\alpha$ is a variational parameter, $\mu = (\mathbf{R}_\mu, s_\mu)$, and we sum over simulation cell lattice vectors $\mathbf{R}_n$ to ensure periodicity. Note that it is also possible to use linear combinations of plane-waves to describe the crystalline phase (as guaranteed by Fourier

theory). We have experimented with these orbitals and saw agreement with predictions from Gaussian orbitals. However, we found that optimizing the plane-waves is hard, such that Gaussian orbitals are the more practical choice. For all densities, we use a simple cubic simulation cell.

We further specialize the MP-NQS to the HEG by defining initial feature vectors. To have the backflow transformation respect the spin inversion and translation symmetries of the HEG, we ignore single-particle positions and spins. Instead, we initialize the nodes to a learnable embedding vector $\mathbf{e} \in \mathbb{R}^{D_1 - D_1^h}$, which does not depend on the particle index $i$. For edge features, we use the translation-invariant particle distances $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and their norm. Same- and opposite-spin pairs are distinguished using products of the form $s_i \cdot s_j = \pm 1$ in the edge features. Overall, we obtain the following initial features:

$$\mathbf{x}_i^{(0)} = \mathbf{e}, \quad \mathbf{x}_{ij}^{(0)} = [\mathbf{r}_{ij}, \|\mathbf{r}_{ij}\|, s_i \cdot s_j]. \quad (16)$$

Notice that this choice preserves the spin-quantum number of each particle.

Following Ref. [21], PBCs of the simulation box are incorporated by mapping the components of the vectors $\mathbf{r} \in \mathbb{R}^d$ (where $\mathbf{r} = \mathbf{r}_i$ or $\mathbf{r} = \mathbf{r}_{ij}$) to a Fourier basis $\mathbf{r} \mapsto [\sin(\frac{2\pi}{L}\mathbf{r}), \cos(\frac{2\pi}{L}\mathbf{r})] \in \mathbb{R}^{2d}$, and the norm of the distance between two particles $\|\mathbf{r}_{ij}\|$ to a periodic surrogate $\|\mathbf{r}_{ij}\| \mapsto \| \sin(\frac{\pi}{L}\mathbf{r}_{ij})\|$. In summary, our *Ansatz* exhibits translation and spin-inversion invariance and maintains a fixed total momentum $\mathbf{k}_{\text{tot}}$. Its number of parameters is independent of system size (here, $\sim 19\,000$), and using stochastic reconfiguration (SR) [31], only $O(10^3)$ optimization steps are needed to reach convergence. More about optimizing the network and an additional comparison with other NQS approaches is given in Appendixes A and D.

## III. RESULTS

We study fully spin-polarized and unpolarized HEGs in different density regimes $r_s \in [1, 200]$ and up to system sizes of $N = 128$ electrons. We compare our ground-state energies against state-of-the-art NQS architectures—FermiNet [23] and WAPNet [22]—for small system sizes $N \in \{14, 19\}$ and against diffusion Monte Carlo (DMC) with backflow (BF-DMC) [52,53] for $N = 54$ electrons (see Appendix E). Additional benchmarks with respect to common quantum chemistry methods, including the transcorrelation augmented full configuration interaction method (FCI) and distinguishable clusters with doubles (DCD) method [34], are included in Appendix F. The effect of the backflow transformation on the nodal surface is studied by comparing with fixed-node DMC (FN-DMC) results. We use an energy of 1.5 mHa per particle (chemical accuracy) to assess the significance of energy differences between the different methods. An overview of all results and benchmarks for the various system sizes and densities is provided in Appendix F.

### A. Number of backflow transformations

We show in Fig. 2, for a system of $N = 14$ particles at the highest densities $r_s = 1, 2, 5$, that increasing the number of backflow iterations (and therefore the number of
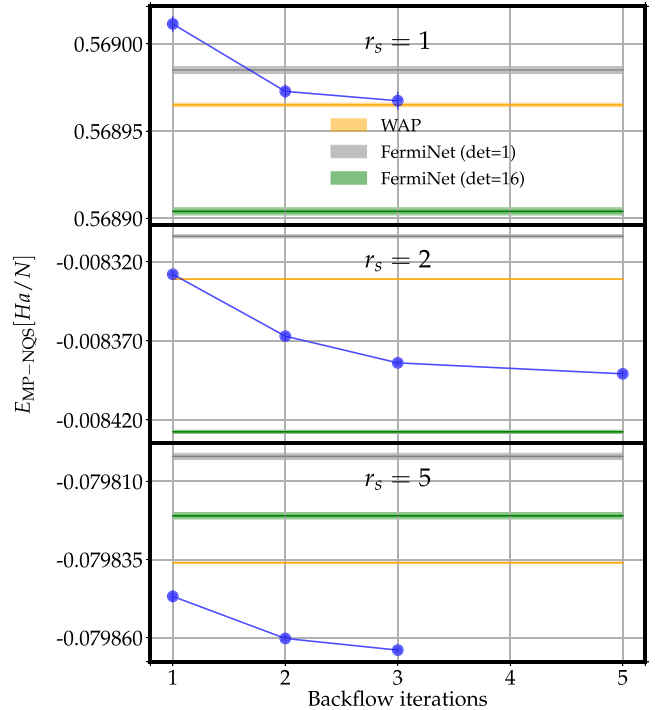


FIG. 2. Energy obtained with our variational *Ansatz* for $N = 14$ particles at densities $r_s = 1, 2, 5$ as a function of the number of message-passing backflow iterations. We compare to WAPNet and FermiNet with a single Slater determinant as well as FermiNet with 16 determinants.

variational parameters) systematically improves the accuracy of our *Ansatz*. At $r_s = 1$, we score slightly worse than WAPNet for a single backflow iteration while reaching the same accuracy for three backflow iterations. Similarly, for $r_s = 2$, we obtain a slightly higher energy than WAPNet for a single backflow iteration, but we surpass the performance of WAPNet after two backflow iterations. For $r_s = 5$, we obtain more precise results than WAPNet for a single backflow iteration. Upon comparison with FermiNet, our results demonstrate that we consistently outperform the unrestricted, single Slater-determinant version of FermiNet at these densities. Interestingly, for lower densities ($r_s = 5$), our *Ansatz* even outperforms the unrestricted FermiNet with 16 determinants with a single message-passing iteration.

### B. Small system size

The best available results for $N = 14$ are obtained with the FCI method, compared with which we find an energy difference of $<1.5$ mHa per particle. State-of-the-art NQS architectures perform comparably with the MP-NQS: The unrestricted FermiNet performs slightly better [$O(10^{-5})$ Ha/N] than both MP-NQS and WAPNet for $r_s \leqslant 2$, while MP-NQS and WAPNet improve over this version of FermiNet for $r_s = 5$. The restricted FermiNet yields ground-state energies worse than the MP-NQS over all probed densities [23] (see Fig. 3). Note that the energy differences for this system are of the order of machine precision. For $r_s \geqslant 5$, we find slightly better performance than WAPNet for all reported densities. All
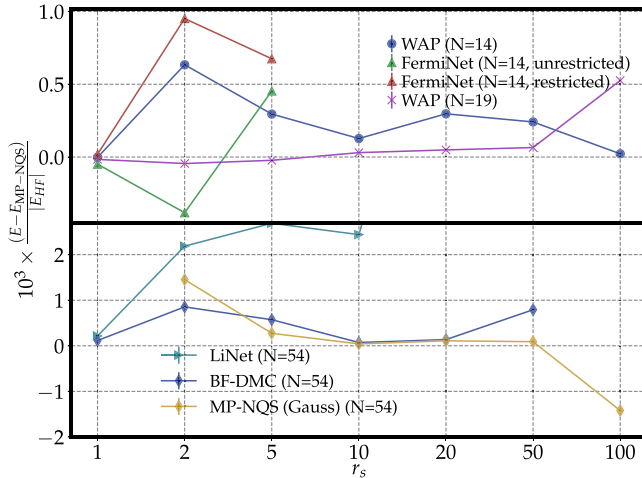
FIG. 3. Comparison of ground-state energies obtained with different methods for various densities, polarizations, and system sizes with respect to the Hartree-Fock energy $E_{HF}$ on the same system. Each line shows the difference in energy of the respective method to the message-passing neural quantum state (MP-NQS) energy with plane-wave orbitals $E_{MP\text{-}NQS}$ (values above zero are worse than the MP-NQS baseline). (top) $N = 14, 19$; (bottom) $N = 54$ particles. Error bars are too small to be visible for most densities. The corresponding numerical data can be found in Appendix F.

differences lie within a range of 1.5 mHa per particle. Similarly, for $N = 19$, the MP-NQS obtains slightly higher energies than WAPNet for large densities ($r_s \leqslant 5$) and marginally lower ones at smaller densities ($r_s \geqslant 5$) with differences $<1.5$ mHa per particle.

### C. Large system size

For $N = 54$ particles, most accurate results are obtained with the FCI method [34]. At this system size, the energy differences per particle are smaller than for $N = 14$, suggesting size consistency of our *Ansatz*, provided that the FCI method is also size consistent. Compared with purely variational methods, we obtain ground-state energies significantly better than those of FN-DMC and BF-DMC, especially at high densities. This is in stark contrast with the (FermiNet-based) architecture of Ref. [27] (dubbed LiNet in the following), which does not improve on BF-DMC energies over the whole density regime (see Fig. 3, bottom). At $r_s > 50$, we observe improved results using the MP-NQS with Gaussian orbitals compared with plane waves (see Fig. 3, bottom panel, yellow line). This strongly suggests a transition from a delocalized Fermi liquid to a localized crystalline phase, as expected from previous studies [35,54]. To analyze finite size effects, we also simulate a larger system of $N = 128$ electrons at $r_s = 50, 110, 200$ and confirm that Gaussian orbitals lead to lower ground-state energies compared with the plane waves for $r_s > 50$ (see Appendix F). Furthermore, for $r_s = 110, 200$, the crystalline character of the variational state can be clearly seen in the radial distribution functions and the corresponding structure factors shown in Fig. 4. The prominent peak in the structure factor and the pronounced density fluctuations in the radial distribution function up to the maximum distance of
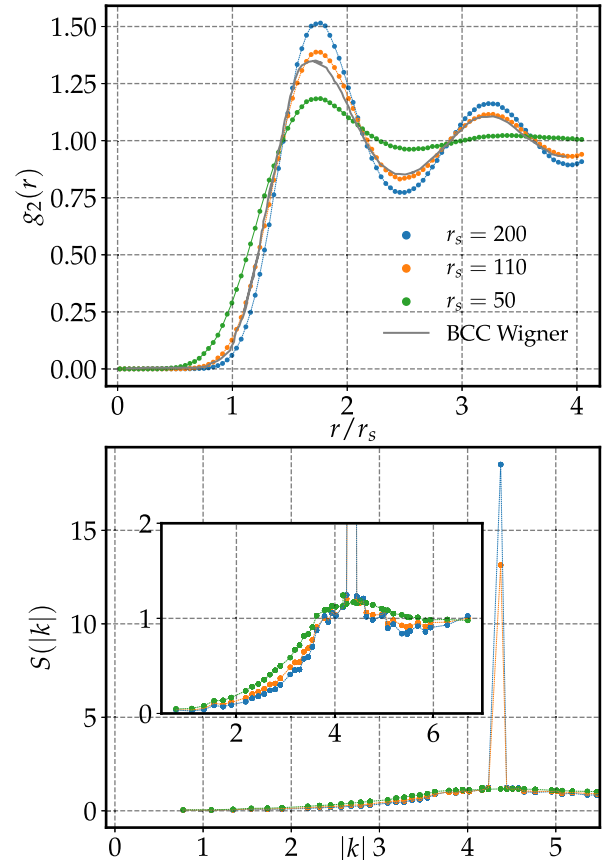


FIG. 4. Spin-averaged radial distribution function (top) and corresponding structure factor (bottom) for the homogeneous electron gas (HEG) system with $N = 128$ electrons at $r_s = 50, 110, 200$. For $r_s = 50$, we used plane-waves as the reference state, while for $r_s = 110, 200$, Gaussian orbitals centered at body-centered cubic (BCC) lattice sites were used. Error bars are smaller than the symbols. The crystal and liquid benchmarks are obtained from Ref. [54] for $r_s = 110$.

$L/2$ indicate the crystalline nature of the represented state. Note that these are absent for $r_s = 50$, suggesting a fluid state.

### IV. CONCLUSIONS

We have introduced MP-NQS, a NQS architecture that leverages MPNNs to build highly expressive backflow coordinates. We demonstrate its power on the HEG system, reducing the number of parameters by orders of magnitude compared with state-of-the-art NQS in continuous space while reaching at par or better accuracy. We also show improvement over state-of-the-art BF-DMC results on large systems. The favorable scaling allows us to accurately simulate large periodic electronic systems, previously inaccessible to state-of-the-art NQS models. In this paper, we increase the size of the reachable system from $N = 27$ and 54 electrons in periodic systems [22,23,27] to $N = 128$ electrons. Hence, we open the door to extrapolation methods to the thermodynamic limit for extended systems. Furthermore, we reproduce the liquid-crystal phase transition of the HEG around $r_s = 100$, matching previous predictions on the

transition density [35,54,55], showing the ability of MP-NQS to quantitatively describe different phases of matter better than previous studies of the HEG based on neural wave functions [23]. In addition to the numerical results, we also introduced an analytical argument, justifying commonly adopted backflow transformations. Our argument shows that a backflow transformation over a reference state is sufficient to obtain the exact ground-state wave function. It will be of particular interest to characterize the geometrical properties of these transformations and understand in what cases NN parametrizations can efficiently describe them.

## APPENDIX A: OPTIMIZATION

From the Rayleigh-Ritz principle, we obtain a lower bound on the expectation value of the Hamiltonian $\langle \Psi | H | \Psi \rangle / \langle \Psi | \Psi \rangle \equiv E[\Psi] \geqslant E_0$. We use the energy expectation value with respect to our variational *Ansatz* as a cost function to gauge the proximity of the variational state to the ground state of the Hamiltonian. Formally, the ground state is given by

$$\Psi_0 = \underset{\Psi}{\operatorname{argmin}} \, E[\Psi]. \tag{A1}$$

where $H | \Psi_0 \rangle = E_0 | \Psi_0 \rangle$.

Since the exact expected value of the Hamiltonian requires us to analytically solve a high-dimensional integral, which in general is not feasible, we resort to Monte Carlo sampling and integration. The energy expectation is evaluated as the average of a set of local energies $E_{\mathrm{loc}}(\mathbf{X}) = \langle \mathbf{X} | H | \Psi \rangle / \langle \mathbf{X} | \Psi \rangle$, where the samples $\mathbf{X}$ are obtained from the probability distribution $|\Psi(\mathbf{X})|^2 / \langle \Psi | \Psi \rangle$ using the Metropolis-Hastings algorithm.

To update the variational parameters in our *Ansatz* so that it exhibits progressively lower energy expectation values, we apply the SR algorithm [31], which can be shown to be equivalent to imaginary-time evolution on the variational manifold and is related to the natural gradient descent method. The update rule for the variational parameters is given by $\delta\theta = -\eta S^{-1} \mathbf{F}$, where $\mathbf{F}$ is the force vector consisting of the log-derivatives of the energy with respect to the variational parameters $\theta$ given by

$$F_i = 2 \left( \frac{\langle \partial_i \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} - E[\Psi] \frac{\langle \partial_i \Psi | \Psi \rangle}{\langle \Psi | \Psi \rangle} \right), \tag{A2}$$

and the so-called quantum geometric tensor $S$ is

$$S_{ij} = \frac{\langle \partial_i \Psi | \partial_j \Psi \rangle}{\langle \Psi | \Psi \rangle} - \frac{\langle \partial_i \Psi | \Psi \rangle \langle \Psi | \partial_j \Psi \rangle}{\langle \Psi | \Psi \rangle \langle \Psi | \Psi \rangle}, \tag{A3}$$

and $\eta$ is the learning rate [59]. To regularize the computation of $S^{-1}$, we apply a small diagonal shift such that we compute $(S + \epsilon \mathbb{I})^{-1}$. For all experiments, we used $\epsilon = 10^{-4}$. The learning rate was chosen according to the density as $\eta = \{0.05, 0.05, 0.05, 0.1, 0.1, 0.5, 1, 2.5\}$ for $r_s = \{1, 2, 5, 10, 20, 50, 100, 110\}$.

## APPENDIX B: COMPARISON OF MPNN AND MP-NQS

In the following, we provide an overview of the general structure of MPNNs [36,60] and relate it to the MP-NQS. First, the relevant data must be encoded in a graph structure, where the nodes of the graph $\mathbf{g}_i^{(t)} = [\mathbf{x}_i^{(0)}, \mathbf{h}_i^{(t)}]$ describe single-body information and the edges between nodes $\mathbf{g}_{ij}^{(t)} = [\mathbf{x}_{ij}^{(0)}, \mathbf{h}_{ij}^{(t)}]$ contain relational information about the connected nodes. The graph $[\mathbf{g}_i^{(t)}, \mathbf{g}_{ij}^{(t)}]$ then gets successively updated over $T$ iterations, using a two-step message-passing procedure. In the first step, the messages are typically constructed for all the nodes of the graphs, but not the edges:

$$\mathbf{m}_i^{(t+1)} = \sum_{j \in \mathcal{N}(i)} \mathbf{M}^{(t)} \big[ \mathbf{g}_i^{(t)}, \mathbf{g}_j^{(t)}, \mathbf{g}_{ij}^{(t)} \big], \tag{B1}$$

where $\mathbf{M}^{(t)}$ is a learnable function, and the sum runs over all nodes in the neighborhood of node $i$, denoted by $\mathcal{N}(i)$. The second step then consists of updating the graph nodes, using the constructed messages:

$$\mathbf{g}_i^{(t+1)} = \mathbf{U}^{(t)} \big[ \mathbf{g}_i^{(t)}, \mathbf{m}_i^{(t+1)} \big], \tag{B2}$$

where $\mathbf{U}^{(t)}$ is again a learnable function. In the MP-NQS, the message construction is done via the particle attention mechanism. Additionally, the explicit dependence on the nodes of the graph is removed, and an update of the graph edges is added:

$$\mathbf{m}_{ij}^{(t+1)} = \mathbf{M}^{(t)} \big[ \mathbf{g}_{ij}^{(t)} \big], \tag{B3}$$

$$\mathbf{m}_i^{(t+1)} = \sum_{j \neq i} \mathbf{m}_{ij}^{(t+1)}, \tag{B4}$$

where $\mathcal{N}(i)$ is chosen to be all other particles to retain permutation equivariance. The nodes and edges of the graph are then updated as follows:

$$\mathbf{g}_i^{(t+1)} = \mathbf{U}_1 \big[ \mathbf{g}_i^{(t)}, \mathbf{m}_i^{(t+1)} \big], \tag{B5}$$

$$\mathbf{g}_{ij}^{(t+1)} = \mathbf{U}_2 \big[ \mathbf{g}_{ij}^{(t)}, \mathbf{m}_{ij}^{(t+1)} \big], \tag{B6}$$

where we introduced another two learnable functions $\mathbf{U}_1, \mathbf{U}_2$.

In traditional MPNN applications, the nodes of the intermediate and final graphs are processed, in a so-called *readout phase*, to predict a scalar quantity, e.g., the energy of an atom in a specific molecular configuration:

$$E_i = \sum_t R^{(t)} \big[ \mathbf{g}_i^{(t)} \big], \tag{B7}$$

where $R^{(t)}$ is again a learnable function. In the MP-NQS, we replace this step with a transformation of the nodes of the final

graph to the physical configuration space of the particles:

$$\delta\mathbf{r}_i = \mathbf{W} \cdot \mathbf{g}_i^{(T)} \in \mathbb{R}^d. \tag{B8}$$

## APPENDIX C: COMPARISON BETWEEN DIFFERENT ATTENTION MECHANISMS

In this section, we summarize the original self-attention mechanism introduced in Ref. [43] and provide a comparison with the particle attention mechanism introduced in this paper. The attention mechanism builds on the query/key/value concept, which is notably recognized in search algorithms. Given a query vector (e.g., search phrase), it is compared with a set of key vectors associated with candidates of a database (e.g., words on Websites). The comparison of the query and key provides a weight determining the relevance of a corresponding value (e.g., Website) given the query. Given (single-body) input data $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, $\mathbf{x}_i \in \mathbb{R}^D$, keys/queries/values are formally computed using learnable weight matrices $W_i^{(Q)} \in \mathbb{R}^{D \times D_k}, W_i^{(K)} \in \mathbb{R}^{D \times D_k}, W_i^{(V)} \in \mathbb{R}^{D \times D_v}$:

$$\mathbf{Q}_i = W_i^{(Q)}\mathbf{x}_i, \quad \mathbf{K}_i = W_i^{(K)}\mathbf{x}_i, \quad \mathbf{V}_i = W_i^{(V)}\mathbf{x}_i. \tag{C1}$$

To obtain permutation-invariant self-attention from the above, the weight matrices cannot depend on the index $i$ but are the same, independent of the input data:

$$\mathbf{Q}_i = W^{(Q)}\mathbf{x}_i, \quad \mathbf{K}_i = W^{(K)}\mathbf{x}_i, \quad \mathbf{V}_i = W^{(V)}\mathbf{x}_i. \tag{C2}$$

Comparing query and key vectors is commonly performed using the (scaled) dot-product along their feature dimension (with dimension $D_k$), such that the self-attention map can be written as

$$\mathbf{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}^T\mathbf{K}}{\sqrt{d}}\right)\mathbf{V} \in \mathbb{R}^{N \times D_v}, \tag{C3}$$

where $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ are matrices of stacked queries, keys, and values, respectively $[\mathbf{Q} = (\mathbf{Q}_1, ..., \mathbf{Q}_N) \in \mathbb{R}^{N \times D_k}]$. The above can also be written as a weighted sum of value vectors, where the weights are given by the inner product between queries and keys.

In the MP-NQS, we develop and extend the idea of self-attention to two-body input data $\{\mathbf{x}_{11}, \ldots, \mathbf{x}_{1N}, \ldots \mathbf{x}_{NN}\}$, $\mathbf{x}_{ij} \in \mathbb{R}^D$. As in self-attention, we utilize query/key weight matrices to compute query/key pairs:

$$\mathbf{Q}_{ij} = W^{(Q)}\mathbf{x}_{ij} \in \mathbb{R}^{D_k}, \tag{C4}$$

$$\mathbf{K}_{ij} = W^{(K)}\mathbf{x}_{ij} \in \mathbb{R}^{D_k}. \tag{C5}$$

The comparison between queries and keys is again performed taking their inner product but, importantly, along the *particle dimension* rather than the feature dimension. Additionally, we replace the softmax activation with a GELU activation:

$$\boldsymbol{\omega}_{ij} = \text{GELU}\left(\sum_l \mathbf{Q}_{il}\mathbf{K}_{lj}\right), \tag{C6}$$

where we introduced the weights $\boldsymbol{\omega}_{ij}$. Instead of using a linear map for the values (as is done for queries and keys), we use a full MLP to compute the values $\mathbf{V}_{ij} = \boldsymbol{\phi}(\mathbf{x}_{ij})$ from the input data. We obtain the following attention mechanism, called particle attention:

$$\mathbf{PAtt}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \boldsymbol{\omega}_{ij} \odot \mathbf{V}_{ij} \in \mathbb{R}^{N \times N \times D_v}, \tag{C7}$$

where $\odot$ denotes element-wise multiplication along the feature dimension. Using the above as messages for the MPNN, Eq. (B4), results in weighted permutation-equivariant sums of edge features:

$$\mathbf{m}_{ij}^{(t)} = \mathbf{PAtt}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \boldsymbol{\omega}_{ij} \odot \mathbf{V}_{ij}, \tag{C8}$$

$$\mathbf{m}_i^{(t)} = \sum_{j \neq i} \boldsymbol{\omega}_{ij} \odot \mathbf{V}_{ij}. \tag{C9}$$

## APPENDIX D: COMPARISON WITH OTHER NQS ARCHITECTURES

The most thoroughly investigated NQS architectures for continuous space are FermiNet and PauliNet [18,19]. Both have been shown to provide good approximations to the ground-state energies of molecular systems. FermiNet has also been applied to the HEG system with very accurate results [23]. Although FermiNet shows slightly better results on molecular systems than PauliNet, differences in the architectures allow PauliNet to reach comparable energies with considerably fewer parameters than FermiNet. Here, we compare our MP-NQS with these two *Ansätze* and highlight similarities and differences.

Both FermiNet and PauliNet construct the variational wave function *Ansatz* by using the antisymmetry of the Slater determinant. PauliNet uses physically motivated orbital functions as input to the Slater determinant and alters these by an orbital-aware multiplicative symmetric factor. Additionally, an overall Jastrow factor is used to increase the expressiveness of the *Ansatz*. FermiNet imposes no physical knowledge on the orbitals (up to an overall envelope function) and does not need an overall Jastrow factor. To construct the multiplicative factor in PauliNet and the orbitals in FermiNet, the number of parameters scales linearly with the size of the system. In addition, both *Ansätze* construct multiple Slater determinants to increase their accuracy, thereby multiplying the number of parameters by a constant factor. Both *Ansätze* treat the different spin species by utilizing distinct parameter sets for the different possible spin projections.

Like PauliNet, we use physically motivated orbitals, which are evaluated at backflowed coordinates. These coordinates are constructed using a powerful MPNN that leverages the expressiveness of the attention mechanism. There exists yet another version of FermiNet, called PsiFormer, which uses the original self-attention mechanism introduced in Ref. [43] but has not been applied to extended systems. While PsiFormer applies the attention mechanism along the *feature dimension* of the single-particle stream, we find substantially improved results by applying attention along the *particle dimension* instead. This approach introduces additional correlations between the particles rather than between the features of the edge-states of our graphs. As PauliNet, we use a multiplicative factor to alter the orbitals even further, which allows us to omit the overall Jastrow factor that is present in PauliNet. By inputting orbital information to the multiplicative factor directly, we can keep the number of parameters independent of
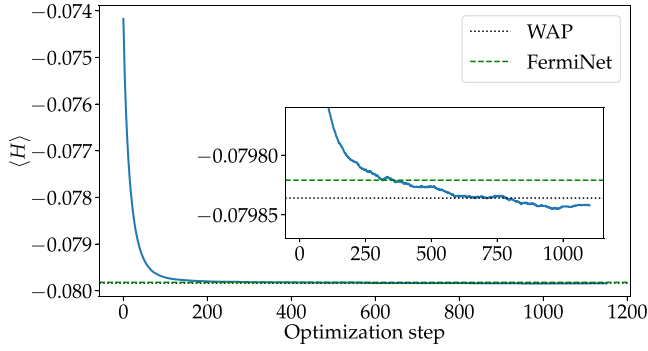
FIG. 5. Moving average of the optimization curve (energy per particle in Hartree averaged over 100 optimization iterations) for $N = 14$ particles at $r_s = 5$ obtained with a learning rate $\eta = 0.05$ and a diagonal shift $\epsilon = 10^{-4}$ for stochastic reconfiguration (SR). We compare to results from WAP-net [22] and FermiNet [23] to our message-passing neural quantum state (MP-NQS). Note that the physically motivation orbitals used in our *Ansatz* makes us start close to the ground-state energy with our optimization.

the number of particles. Notably, our *Ansatz* does not require multiple Slater determinants.

As mentioned above, FermiNet has been applied to study the HEG in continuous space. The orbitals are constructed from scratch with plane-wave envelope functions. In contrast with our approach, FermiNet uses large networks (512 width) to process the positions of the simulated particles and considerably smaller networks (32 width) to act on the distance vectors between the particles. The usage of single-particle coordinates breaks the translation invariance of the system. In our MP-NQS, we only use the distance information between the particles and ignore the single-particle information to stay translation invariant. We only use small networks (32 width) to construct the messages $\mathbf{m}_{ij}^{(t+1)}$ [Eq. (B4)] that are used to update the nodes and edges of the particle graph. Overall, we use ∼19 000 parameters in our *Ansatz*, which consists of a single Slater determinant. PauliNet usually uses 70 000– 100 000 parameters and up to 36 Slater determinants, while FermiNet uses on the order of a million parameters with up to 32 Slater determinants. The fact that we have considerably fewer parameters allows us to use SR to optimize our variational *Ansatz* (see Appendix A). Both FermiNet and PauliNet use $O(10^5)$ optimization steps, but we can reduce this number to less than $O(10^3)$ iterations. To illustrate this point, we show the optimization curve for $N = 14$ particles at $r_s = 5$ in Fig. 5.

## APPENDIX E: DMC

The DMC method projects the variational states $|\Psi_V\rangle$ in imaginary time to filter out its ground-state component:

$$|\Psi_0\rangle \propto \lim_{\tau \to \infty} \exp(-H\tau)|\Psi_V\rangle. \quad (E1)$$

The variational state is taken to be of the Slater-Jastrow form $\Psi_V = \exp[J(\mathbf{X})]S_\uparrow S_\downarrow$, where $S_\uparrow$ and $S_\downarrow$ are the Slater determinants of single-electron states for the spin-up and spin-down species, respectively. The Jastrow factor captures correlation between electrons and hence is a function of electron-electron

separation coordinates $\mathbf{r}_{ij} = \{x_{ij}, y_{ij}, z_{ij}\}$:

$$J(\mathbf{X}) = \sum_{\substack{i<j \\ \sigma,\tau\in\{\uparrow,\downarrow\}}} J_{\sigma\tau}(\mathbf{r}_{ij}). \quad (E2)$$

The Pauli principle requires the functional form of $J_{\sigma\tau}$ must be even under exchange of particles $i$ and $j$. In addition, $J_{\sigma\tau}$ must satisfy the PBCs at the edge of the simulation box. To fulfill the above requirements, we utilize the parameterization introduced in Ref. [61]:

$$J_{\sigma,\tau}(\mathbf{r}_{ij}) = \sum_{n=1}^{N_j} c_{n,\sigma\tau}[j(x_{ij})^2 + j(y_{ij})^2 + j(z_{ij})^2]^{n/2}, \quad (E3)$$

where $j(x) = |x|[1 - 2(|x|/L)^3]$. By taking $c_{1,\uparrow\uparrow} = c_{1,\downarrow\downarrow} = \frac{1}{4}$ and $c_{1,\uparrow\downarrow} = c_{1,\downarrow\uparrow} = \frac{1}{2}$, the Kato cusp conditions at particle coalescence are also automatically fulfilled. The parameters $c_{n>1,\uparrow\uparrow}$ are determined by minimizing the energy of the systems by using the linear optimization scheme [30]. We find that using more than $N_J = 6$ variational parameters does not improve the energy noticeably.

The imaginary-time diffusion is broken into many small steps $\delta\tau$. The Trotter-Suzuki decomposition can be applied to the short-time propagator as $\exp(-H\delta\tau) \simeq \exp(-V\delta\tau/2)\exp(-T\delta\tau)\exp(-V\delta\tau/2)$, where $T$ and $V$ are the kinetic and potential energy operators, respectively. At each imaginary-time step, we use the free propagator:

$$G_0(\mathbf{X}', \mathbf{X}) \equiv \langle \mathbf{X}' | \exp(-T\delta\tau) | \mathbf{X} \rangle$$
$$= \left[ \sqrt{\frac{m}{2\pi\hbar^2\delta\tau}} \right]^{3N} \exp\left[ -\frac{(\mathbf{R}' - \mathbf{R})^2}{2\hbar^2\delta\tau/m} \right], \quad (E4)$$

to sample the new coordinates $\mathbf{X}'$ of all particles. As routinely done in nuclear physics applications [62,63], to remove the linear terms coming from the exponential of Eq. (E4), we use two mirror samples $\mathbf{X}' = \mathbf{X} \pm \delta\mathbf{X}$, and we consider the corresponding importance-sampled weights:

$$w_\pm = \frac{\Psi_V(\mathbf{X} \pm \delta\mathbf{X})}{\Psi_V(\mathbf{X})} \exp\left\{ -\frac{[V(\mathbf{X} \pm \delta\mathbf{X}) + V(\mathbf{X})]\delta\tau}{2} \right\}. \quad (E5)$$

Only one of the two samples is kept in the propagation according to a heat-bath sampling among the two normalized weights $w_\pm/(\Sigma_\pm w_\pm)$, and the average weight $\Sigma_\pm w_\pm/2$ is assigned to the propagated configuration.

The fermion-sign problem is controlled by employing the fixed-node approximation, which amounts to evaluating the weights of Eq. (E5) with the replacement:

$$\frac{\Psi_V(\mathbf{X} \pm \delta\mathbf{X})}{\Psi_V(\mathbf{X})} \to \text{Re}\frac{\Psi_V(\mathbf{X} \pm \delta\mathbf{X})}{\Psi_V(\mathbf{X})}. \quad (E6)$$

Note that, if the real part of the above ratio is negative, the weight of the configuration is set to zero only after computing the average weight.

## APPENDIX F: NUMERICAL RESULTS

In Tables I–V we show the ground-state energies of the unpolarized and fully polarized HEG for different system sizes $N = \{14, 19, 27, 54, 128\}$ and densities $r_s = \{1, 2, 5, 10, 20, 30, 50, 70, 90, 100, 110, 200\}$.

TABLE I. Total energy per particle in Hartree for unpolarized system of $N = 14$ particles. WAPNet and FermiNet are alternative NQS architectures optimized via VMC. We include FCI and DCD results as benchmarks from quantum chemistry.

| N | $r_s$ | MP-NQS | WAP [22] | FermiNet [23] | FCI*/DCD** [34] |
|---|---|---|---|---|---|
| 14 | 1 | 0.568 967(6) | 0.568 965(1) | 0.568 904(1) | 0.568 61(1)* |
| | 2 | −0.008 391(1) | −0.008 331 0(3) | −0.008 427(1) | −0.008 68(2)* |
| | 5 | −0.079 854 4(4) | −0.079 836 0(1) | −0.079 821(1) | −0.080 02(2)* |
| | 10 | −0.055 212 6(6) | −0.055 203 80(3) | N/A | −0.055 09** |
| | 20 | −0.032 455 3(2) | −0.032 443 4(1) | N/A | −0.032 01** |
| | 50 | −0.014 626 31(6) | −0.014 622 11(4) | N/A | −0.013 84** |
| | 100 | −0.007 730 18(3) | −0.007 729 980(2) | N/A | N/A |

TABLE II. Total energy per particle in Hartree for the unpolarized system of $N = 54$ particles. FN-DMC results are obtained using the method in Appendix E. We include FCI and DCD results as benchmarks from quantum chemistry.

| N | $r_s$ | MP-NQS | MP-NQS (Gauss) | LiNet [27] | FN-DMC | BF-DMC [52,53] | FCI*/DCD** [34] |
|---|---|---|---|---|---|---|---|
| 54 | 1 | 0.529 73(1) | N/A | 0.530 019(1) | 0.530 94(2) | 0.529 89(4) | 0.529 73(3)* |
| | 2 | −0.014 046(8) | −0.013 90(1) | −0.013 840(1) | −0.013 26(2) | −0.013 966(2) | −0.013 79** |
| | 5 | −0.079 090(2) | −0.079 064(4) | −0.078 835 4(2) | −0.078 67(1) | −0.079 036(3) | −0.078 37** |
| | 10 | −0.054 448(1) | −0.054 445(1) | −0.054 278 5(1) | −0.054 269(8) | −0.054 443(2) | −0.053 22** |
| | 20 | −0.032 052 4(5) | −0.032 048 0(6) | −0.031 688 6(1) | −0.031 976(8) | −0.032 047(2) | −0.031 13** |
| | 50 | −0.014 501 5(1) | −0.014 499 9(1) | N/A | −0.013 87(2) | −0.014 487 7(1) | −0.012 81** |
| | 100 | −0.007 679 3(1) | −0.007 692 03(5) | N/A | −0.007 674(3) | N/A | N/A |

TABLE III. Total energy per particle in Hartree for the unpolarized system of $N = 128$ particles. FN-DMC results are obtained using the method in Appendix E.

| N | $r_s$ | MP-NQS | MP-NQS (Gauss) | FN-DMC |
|---|---|---|---|---|
| 128 | 110 | −0.007 020 3(1) | −0.007 112 1(2) | −0.007 009 518(1) |
| | 200 | −0.004 002 3(2) | −0.004 061 5(1) | −0.003 992 3(7) |

TABLE IV. Total energy per particle in Hartree for the polarized system of $N = 19$ particles. We compare to results obtained using WAPNet.

| N | $r_s$ | MP-NQS | WAP [22] |
|---|---|---|---|
| 19 | 1 | 1.046 262(2) | 1.046 241(3) |
| | 2 | 0.096 307(1) | 0.096 303(1) |
| | 5 | −0.067 248 9(3) | −0.067 251 05(3) |
| | 10 | −0.052 862 4(2) | −0.052 860 35(1) |
| | 20 | −0.032 010 1(1) | −0.032 008 2(1) |
| | 50 | −0.014 566 85(3) | −0.014 565 71(2) |
| | 100 | −0.007 718 32(2) | −0.007 713 62(2) |

TABLE V. Total energy per particle in Hartree for the polarized system of $N = 27$ particles. FN-DMC results are obtained using the method in Appendix E.

| N | $r_s$ | MP-NQS | FN-DMC |
|---|---|---|---|
| 27 | 1 | 1.051 771(4) | 1.052 00(3) |
| | 2 | 0.098 460(2) | 0.098 66(2) |
| | 5 | −0.066 552 3(6) | −0.066 429(4) |
| | 10 | −0.052 601 5(2) | −0.052 540(6) |
| | 30 | −0.022 818 4(1) | −0.022 808 0(8) |
| | 50 | −0.014 579 76(5) | −0.014 575 2(6) |
| | 70 | −0.010 749 86(3) | −0.010 749(2) |
| | 90 | −0.008 531 56(2) | −0.008 531 4(5) |

[1] B. M. Austin, D. Y. Zubarev, and W. A. Lester, Jr., Quantum Monte Carlo and related approaches, Chem. Rev. **112**, 263 (2012).

[2] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Quantum Monte Carlo methods for nuclear physics, Rev. Mod. Phys. **87**, 1067 (2015).

[3] A. J. Leggett, A theoretical description of the new phases of liquid $^3$He, Rev. Mod. Phys. **47**, 331 (1975).

[4] M. Tinkham, *Introduction to Superconductivity* (Dover Publications, Inc., Mineola, 2004).

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

[6] J. Carrasquilla, Machine learning for quantum matter, Adv. Phys.: X **5**, 1797528 (2020).

[7] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and many-body excitations with neural-network quantum states, Phys. Rev. Lett. **121**, 167204 (2018).

[8] F. Ferrari, F. Becca, and J. Carrasquilla, Neural Gutzwiller-projected variational wave functions, Phys. Rev. B **100**, 125131 (2019).

[9] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, Phys. Rev. Res. **2**, 023358 (2020).

[10] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, SciPost Phys. **10**, 147 (2021).

[11] Y. Nomura and M. Imada, Dirac-type nodal spin liquid revealed by refined quantum many-body solver using neural-network wave function, correlation ratio, and level spectroscopy, Phys. Rev. X **11**, 031034 (2021).

[12] N. Astrakhantsev, T. Westerhout, A. Tiwari, K. Choo, A. Chen, M. H. Fischer, G. Carleo, and T. Neupert, Broken-symmetry ground states of the Heisenberg model on the pyrochlore lattice, Phys. Rev. X **11**, 041021 (2021).

[13] D.-L. Deng, X. Li, and S. Das Sarma, Quantum entanglement in neural network states, Phys. Rev. X **7**, 021021 (2017).

[14] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum entanglement in deep learning architectures, Phys. Rev. Lett. **122**, 065301 (2019).

[15] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, Phys. Rev. Lett. **122**, 226401 (2019).

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

[17] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, Proc. Natl. Acad. Sci. USA **119**, e2122059119 (2022).

[18] 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).

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

[20] M. Entwistle, Z. Schätzle, P. A. Erdman, J. Hermann, and F. Noé, Electronic excited states in deep variational Monte Carlo, Nat. Commun. **14**, 274 (2023).

[21] 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).

[22] M. Wilson, S. Moroni, M. Holzmann, N. Gao, F. Wudarski, T. Vegge, and A. Bhowmik, Neural network *Ansatz* for periodic wave functions and the homogeneous electron gas, Phys. Rev. B **107**, 235139 (2023).

[23] 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).

[24] C. Adams, G. Carleo, A. Lovato, and N. Rocco, Variational Monte Carlo calculations of $a \leqslant 4$ nuclei with an artificial neural-network correlator *Ansatz*, Phys. Rev. Lett. **127**, 022502 (2021).

[25] A. Gnech, C. Adams, N. Brawand, G. Carleo, A. Lovato, and N. Rocco, Nuclei with up to $A = 6$ nucleons with artificial neural network wave functions, Few-Body Syst. **63**, 7 (2022).

[26] A. Lovato, C. Adams, G. Carleo, and N. Rocco, Hidden-nucleons neural-network quantum states for the nuclear many-body problem, Phys. Rev. Res. **4**, 043178 (2022).

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

[28] H. Xie, L. Zhang, and L. Wang, $m^*$ of two-dimensional electron gas: A neural canonical transformation study, SciPost Phys. **14**, 154 (2023).

[29] B. Fore, J. M. Kim, G. Carleo, M. Hjorth-Jensen, A. Lovato, and M. Piarulli, Dilute neutron star matter from neural-network quantum states, Phys. Rev. Res. **5**, 033062 (2023).

[30] J. Toulouse and C. J. Umrigar, Optimization of quantum Monte Carlo wave functions by energy minimization, J. Chem. Phys. **126**, 084102 (2007).

[31] S. Sorella, Green function Monte Carlo with stochastic reconfiguration, Phys. Rev. Lett. **80**, 4558 (1998).

[32] 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).

[33] N. D. Drummond and R. J. Needs, Diffusion quantum Monte Carlo calculation of the quasiparticle effective mass of the two-dimensional homogeneous electron gas, Phys. Rev. B **87**, 045131 (2013).

[34] 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).

[35] S. Azadi, N. D. Drummond, and S. M. Vinko, Correlation energy of the paramagnetic electron gas at the thermodynamic limit, Phys. Rev. B **107**, L121105 (2023).

[36] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, Neural message passing for quantum chemistry, in *Proceedings of the 34th International Conference on Machine Learning*, edited by D. Precup and Y. W. Teh, Proceedings of Machine Learning Research (PMLR, 2017), Vol. 70, pp. 1263–1272.

[37] T. M. Apostol, *Mathematical Analysis: A Modern Approach to Advanced Calculus* (Addison-Wesley Publishing Co., Inc., Reading, 1957).

[38] R. P. Feynman and M. Cohen, Energy spectrum of the excitations in liquid helium, Phys. Rev. **102**, 1189 (1956).

[39] Y. Kwon, D. M. Ceperley, and R. M. Martin, Effects of backflow correlation in the three-dimensional electron gas: Quantum Monte Carlo study, Phys. Rev. B **58**, 6800 (1998).

[40] Y. Kwon, D. M. Ceperley, and R. M. Martin, Effects of three-body and backflow correlations in the two-dimensional electron gas, Phys. Rev. B **48**, 12037 (1993).

[41] 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).

[42] M. Ruggeri, S. Moroni, and M. Holzmann, Nonlinear network description for many-body quantum systems in continuous space, Phys. Rev. Lett. **120**, 205302 (2018).

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, Attention is all you need, in *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*, edited by I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), Vol. 30.

[44] D. Hendrycks and K. Gimpel, Gaussian error linear units (gelus), arXiv:1606.08415.

[45] I. von Glehn, J. S. Spencer, and D. Pfau, A self-attention *Ansatz* for *ab-initio* quantum chemistry, arXiv:2211.13672.

[46] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2016), pp. 770–778.

[47] D. Pines, *Elementary Excitations in Solids* (CRC Press, Boca Raton, 1999).

[48] P. P. Ewald, Die Berechnung optischer und elektrostatischer Gitterpotentiale, Ann. Phys. **369**, 253 (1921).

[49] L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, and A. J. Williamson, Finite-size effects and Coulomb interactions in quantum Monte Carlo calculations for homogeneous systems with periodic boundary conditions, Phys. Rev. B **53**, 1814 (1996).

[50] A. Y. Toukmaji and J. A. Board Jr, Ewald summation techniques in perspective: A survey, Comput. Phys. Commun. **95**, 73 (1996).

[51] E. Wigner, On the interaction of electrons in metals, Phys. Rev. **46**, 1002 (1934).

[52] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs, Inhomogeneous backflow transformations in quantum Monte Carlo calculations, Phys. Rev. E **74**, 066701 (2006).

[53] S. Azadi and N. D. Drummond, Low-density phase diagram of the three-dimensional electron gas, Phys. Rev. B **105**, 245135 (2022).

[54] N. D. Drummond, Z. Radnai, J. R. Trail, M. D. Towler, and R. J. Needs, Diffusion quantum Monte Carlo study of three-dimensional Wigner crystals, Phys. Rev. B **69**, 085116 (2004).

[55] D. M. Ceperley and B. J. Alder, Ground state of the electron gas by a stochastic method, Phys. Rev. Lett. **45**, 566 (1980).

[56] G. Carleo, K. Choo, D. Hofmann, J. E. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin *et al.*, NetKet: A machine learning toolkit for many-body quantum systems, SoftwareX **10**, 100311 (2019).

[57] D. Häfner and F. Vicentini, mpi4jax: Zero-copy MPI communication of jax arrays, J. Open Source Softw. **6**, 3419 (2021).

[58] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev *et al.*, NetKet 3: Machine learning toolbox for many-body quantum systems, SciPost Phys. Codebases 7 (2022).

[59] 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).

[60] I. Batatia, D. P. Kovacs, G. Simm, C. Ortner, and G. Csányi, MACE: Higher order equivariant message passing neural networks for fast and accurate force fields, *Advances in Neural Information Processing Systems*, edited by S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., 2022), Vol. 35, 11423–11436.

[61] T. M. Whitehead, M. H. Michael, and G. J. Conduit, Jastrow correlation factor for periodic systems, Phys. Rev. B **94**, 035157 (2016).

[62] B. S. Pudliner, V. R. Pandharipande, J. Carlson, S. C. Pieper, and R. B. Wiringa, Quantum Monte Carlo calculations of nuclei with $A \lesssim 7$, Phys. Rev. C **56**, 1720 (1997).

[63] S. Gandolfi, D. Lonardoni, A. Lovato, and M. Piarulli, Atomic nuclei from quantum Monte Carlo calculations with chiral EFT interactions, Front. Phys. **8**, 117 (2020).