

<https://doi.org/10.1038/s42005-024-01613-w>

Neural-network quantum states for ultra-cold Fermi gases

Check for updates

Jane Kim^{1,9}, Gabriel Pescia^{2,3}, Bryce Fore⁴, Jannes Nys^{2,3}, Giuseppe Carleo^{2,3}, Stefano Gandolfi⁵, Morten Hjorth-Jensen^{1,6} & Alessandro Lovato^{4,7,8} ✉

Ultra-cold Fermi gases exhibit a rich array of quantum mechanical properties, including the transition from a fermionic superfluid Bardeen-Cooper-Schrieffer (BCS) state to a bosonic superfluid Bose-Einstein condensate (BEC). While these properties can be precisely probed experimentally, accurately describing them poses significant theoretical challenges due to strong pairing correlations and the non-perturbative nature of particle interactions. In this work, we introduce a Pfaffian-Jastrow neural-network quantum state featuring a message-passing architecture to efficiently capture pairing and backflow correlations. We benchmark our approach on existing Slater-Jastrow frameworks and state-of-the-art diffusion Monte Carlo methods, demonstrating a performance advantage and the scalability of our scheme. We show that transfer learning stabilizes the training process in the presence of strong, short-ranged interactions, and allows for an effective exploration of the BCS-BEC crossover region. Our findings highlight the potential of neural-network quantum states as a promising strategy for investigating ultra-cold Fermi gases.

The study of ultra-cold Fermi gases has received considerable experimental and theoretical attention in recent years due to their unique properties and potential applications in fields ranging from condensed matter physics to astrophysics. These systems can be created and manipulated in the laboratory with high precision, providing a versatile platform for investigating various phenomena. By tuning the s -wave scattering length a via external magnetic fields near a Feshbach resonance, one can smoothly transition from a fermionic superfluid Bardeen-Cooper-Schrieffer (BCS) state with $a < 0$, characterized by long-range Cooper pairs to a bosonic superfluid Bose-Einstein condensate (BEC) with $a > 0$, consisting of tightly-bound, repulsive dimers. Given their diluteness, the behavior of these systems is mainly governed by a and the s -wave effective range of the potential r_e , with natural units provided by the Fermi momentum k_F , the Fermi energy $\epsilon_F = \frac{\hbar^2}{2m} k_F^2$, and the Fermi gas energy per particle in the thermodynamic limit $E_{FG} = \frac{3}{5} \epsilon_F$ (see ref. 1 and references therein).

The region between the BCS and BEC states, known as the unitary limit, is particularly interesting as a diverges and r_e approaches zero. The unitary Fermi gas (UFG) is a strongly interacting system that exhibits stable

superfluid behavior. Studying the BCS-BEC crossover near the unitary limit can reveal critical aspects of the underlying mechanism behind superfluidity in fermionic matter. The UFG is also universal, meaning its properties are independent of the details of the two-body potential. This universality allows for robust comparisons and predictions between seemingly disparate quantum systems. For instance, the UFG is relevant for neutron stars, as they provide a means to study superfluid low-density neutron matter^{2,3}, whose properties are crucial for the phenomenology of glitches⁴ and the cooling of these stars via neutrino emission⁵⁻⁷.

Due to the onset of strong pairing correlations and the non-perturbative nature of the interaction, the theoretical study of ultra-cold Fermi gases is particularly challenging for quantum many-body methods. Despite these challenges, quantum Monte Carlo (QMC) has proven to be efficient in calculating various properties with high accuracy, including the energy⁸, pairing gap⁹, and other quantities related to the so-called contact parameter¹⁰. The simplest QMC method is variational Monte Carlo (VMC), often used as a preliminary step preceding more accurate methods, such as auxiliary-field quantum Monte Carlo (AFQMC) or diffusion Monte Carlo

¹Department of Physics and Astronomy and Facility for Rare Isotope Beams, Michigan State University, East Lansing, MI 48824, USA. ²Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. ³Center for Quantum Science and Engineering, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. ⁴Physics Division, Argonne National Laboratory, Argonne, IL 60439, USA. ⁵Theory Division T-2, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. ⁶Department of Physics and Center for Computing in Science Education, University of Oslo, N-0316 Oslo, Norway. ⁷Computational Science Division, Argonne National Laboratory, Argonne, IL 60439, USA. ⁸INFN-TIFPA Trento Institute of Fundamental Physics and Applications, 38123 Trento, Italy. ⁹Present address: Institute of Nuclear and Particle Physics and Department of Physics and Astronomy, Ohio University, Athens, OH 45701, USA. ✉e-mail: lovato@anl.gov

(DMC). In unpolarized unitary Fermi gases, AFQMC is renowned for providing the most accurate ground-state energies as it is sign-problem free⁸. However, its applicability to other problems is somewhat limited, as addressing odd- N systems, spin-polarized systems, or repulsive interactions necessitates approximations that risk violating the variational principle. Crucially, direct comparisons between continuum-based methods like VMC and DMC, and lattice-based methods like AFQMC are valid only in the limit of zero lattice spacing and zero effective range, due to both discretization and finite effective-range effects. Therefore, we opt to use state-of-the-art DMC calculations as the main benchmark in this initial investigation.

To control the fermion sign problem, DMC calculations typically rely on the fixed-node approximation, keeping the nodes of the wave function in the positions determined by the prior VMC calculation. The analytical form of the variational ansatz is usually tailored to specific problems of interest and biased by the physical intuition of the researchers. While the fixed-node approximation provides a rigorous upper bound to the ground-state energy that agrees well with other methods and experiments^{11,12}, the resulting wave function carries residual dependence on the starting variational wave function¹³. This dependence is particularly significant in DMC calculations of expectation values of operators that do not commute with the Hamiltonian, such as spatial and momentum distributions¹⁴.

In this work, we mitigate the biases introduced in traditional VMC (and, by extension, DMC) approaches by performing VMC calculations with highly flexible neural-network quantum states (NQS)¹⁵. After their initial application to quantum-chemistry problems^{16,17}, continuous-space NQS have been successfully employed to study quantum many-body systems in the presence of spatial periodicities, such as interacting quantum gases of bosons¹⁸, the homogeneous electron gas^{19,20}, and dilute neutron matter²¹. Recent works have also used NQS to solve the nuclear Schrödinger equation in both continuous space^{22–26} and the occupation number formalism²⁷, as well as spinless trapped fermion systems²⁸. When dealing with fermions, the antisymmetry is usually enforced by generalized Slater determinants, the expressivity of which can be augmented with either backflow transformations²⁹ or by adding hidden degrees of freedom³⁰.

Strong pairing correlations in fermionic systems motivate adopting an antisymmetric wave function constructed from pairing orbitals rather than single-particle orbitals. In the context of QMC studies of ultra-cold Fermi gases, this wave function is typically constructed as an antisymmetrized product of BCS spin-singlet pairs^{31–34}. It goes by a variety of names, such as the geminal wave function^{35,36}, the singlet pairing wave function³⁷, and the (number-projected) BCS wave function³⁴, just to name a few. Although geminal wave functions have demonstrated significant improvements over single-determinant wave functions of single-particle orbitals, the energy gains are typically smaller for partially spin-polarized systems³⁶, as contributions from the spin-triplet channel are missing. This naturally leads to the singlet-triplet-unpaired (STU) Pfaffian wave function^{37,38}, in which the pairing orbitals are explicitly decomposed into singlet and triplet channels. Then, the STU ansatz is expressed as the Pfaffian of a block matrix, with the singlet, triplet, and unpaired contributions partitioned into separate blocks. When the triplet blocks are zero, the STU wave function reduces to the geminal wave function.

Both the geminal and the STU wave functions rely on fixing the spin ordering of the interacting fermions. Consequently, they are not amenable to potentials that exchange spin, such as those used to model the interaction among nucleons³⁹. In neutron-matter calculations, for instance, the pairing orbital for the Pfaffian wave function can be taken as a product of a radial part and a spin-singlet part^{40,41}. The spin-triplet pairing has so far been neglected in neutron-matter calculations, but they can be treated similarly without requiring spin ordering. Pfaffian wave functions combined with neural-network Jastrow correlators⁴² have also successfully modeled lattice fermions, even revealing the existence of a quantum spin liquid phase in the J1–J2 models on two-dimensional lattices⁴³.

We propose a NQS that extends the conventional Pfaffian-Jastrow³⁷ ansatz by incorporating neural backflow transformations into a fully

trainable pairing orbital. Other than the assumption of strong pairing correlations, our ansatz incorporates only the most essential symmetries and boundary conditions. Permutation-equivariant backflow transformations are generated by a message-passing neural network (MPNN), recently introduced to model the homogeneous electron gas⁴⁴. In addition to being a significant departure from generalized Slater determinants, our Pfaffian-Jastrow NQS naturally encodes pairing in the singlet and triplet channels without stipulating a particular form for the pairing orbital. In view of this, it is broadly applicable to other strongly interacting systems with the same symmetries and boundary conditions.

We demonstrate the representative power of our NQS by computing ground-state properties of ultra-cold Fermi gases in the BCS-BEC crossover. Our Pfaffian-Jastrow NQS outperforms Slater-Jastrow NQS by a large margin, even when generalized backflow transformations are included in the latter. We find lower energies than those obtained with state-of-the-art DMC methods, which start from highly accurate BCS-like variational wave functions. Analysis of pair distribution functions and pairing gaps reveal the emergence of strong pairing correlations around unitarity. Transfer learning is utilized to investigate the BCS-BEC crossover region and approach the thermodynamic limit, all within a unified ansatz. Our results underscore the viability of utilizing NQS in the ab initio study of ultra-cold Fermi gases.

Results

Hamiltonian

We simulate the infinite system using a finite number of fermions N in a cubic simulation cell with side length L , equipped with periodic boundary conditions (PBCs) in all $d=3$ spatial dimensions. We use $\mathbf{r}_i \in \mathbb{R}^d$ and $s_i \in \{\uparrow, \downarrow\}$ to denote the positions and spin projections on the z -axis of the i -th particle, and the length L can be determined from the uniform density of the system $\rho = N/L^3 = k_F^3/(3\pi^2)$.

The dynamics of the gas is governed by the non-relativistic Hamiltonian

$$H = -\frac{\hbar^2}{2m} \sum_i \nabla_i^2 + \sum_{ij} v_{ij}, \quad (1)$$

where the attractive two-body interaction

$$v_{ij} = (\delta_{s_i, s_j} - 1) v_0 \frac{2\hbar^2}{m} \frac{\mu^2}{\cosh^2(\mu r_{ij})}, \quad (2)$$

acts only between opposite-spin pairs, making the interaction mainly in s -wave for small values of r_e . In the above equations, ∇_i^2 is the Laplacian with respect to \mathbf{r}_i and $r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|$ is the Euclidean distance between particles i and j . The Pöschl-Teller interaction potential of Eq. (2) provides an analytic solution of the two-body problem and has been employed in several previous QMC calculations^{10,32,33,45}. The parameters v_0 and μ tune the scattering length a and the effective range r_e in the s -wave channel. In the unitary limit $|a| \rightarrow \infty$, the zero-energy ground state between two particles corresponds to $v_0 = 1$ and $r_e = 2/\mu$. In order to analyze the crossover between the BCS and BEC phases, we will use different combinations of v_0 and μ that correspond to the same effective range. In addition, we will consider various values of μ with fixed $v_0 = 1$ to extrapolate the zero effective range behavior at unitarity.

Neural-network quantum states

We solve the Schrödinger equation associated with the Hamiltonian of Eq. (1) using two different families of NQS. All ansätze have the general form

$$\Psi(X) = e^{J(X)} \Phi(X), \quad (3)$$

where the Jastrow correlator $J(X)$ is symmetric under particle exchange and $\Phi(X)$ is antisymmetric. Here, we have introduced $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$, with

$\mathbf{x}_i = (\mathbf{r}_i, s_i)$, to compactly represent the set of all single-particle positions and spins.

To address the limitations of the works mentioned in the Introduction, we take the most general form of the Pfaffian wave function^{37,38} as the antisymmetric part of our ansatz,

$$\Phi_{PJ}(X) = \text{pf } P, \tag{4}$$

where P is an $N \times N$ matrix defined as

$$P \equiv \begin{bmatrix} 0 & \phi(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \phi(\mathbf{x}_1, \mathbf{x}_N) \\ \phi(\mathbf{x}_2, \mathbf{x}_1) & 0 & \cdots & \phi(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\mathbf{x}_N, \mathbf{x}_1) & \phi(\mathbf{x}_N, \mathbf{x}_2) & \cdots & 0 \end{bmatrix}. \tag{5}$$

Unlike the determinant, which is well-defined for all square matrices, the Pfaffian, denoted as pf , is only well-defined for even-dimensional, skew-symmetric matrices. Therefore, the above construction applies to even N only, but it can be easily extended to odd N , as we will discuss shortly. To ensure skew-symmetry, the pairing orbital $\phi(\mathbf{x}_i, \mathbf{x}_j)$ is defined as

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = \nu(\mathbf{x}_i, \mathbf{x}_j) - \nu(\mathbf{x}_j, \mathbf{x}_i), \tag{6}$$

where $\nu(\mathbf{x}_i, \mathbf{x}_j)$ is a dense feed-forward neural network (FNN)⁴⁶. This approach capitalizes on the universal approximation property of FNNs. We do not mandate a specific functional form for the pairing orbital, as in traditional VMC approaches, nor do we keep the spins fixed, as with the geminal or STU wave functions discussed in the Introduction. Instead, our pairing orbital discovers the spin-singlet and spin-triplet correlations on its own.

If N is odd, we augment the above construction by introducing one unpaired single-particle orbital $\psi(\mathbf{x}_i)$, represented by another dense feed-forward neural network. Instead of the $N \times N$ matrix in Eq. (5), we take the Pfaffian of an $(N + 1) \times (N + 1)$ matrix,

$$\Phi_{PJ}(X) = \text{pf} \begin{bmatrix} P & \mathbf{u} \\ -\mathbf{u}^T & 0 \end{bmatrix}, \tag{7}$$

where the vector $\mathbf{u} \in \mathbb{R}^N$ is defined by applying the unpaired single-particle orbital on each of the single-particle degrees of freedom

$$\mathbf{u} = \begin{bmatrix} \psi(\mathbf{x}_1) \\ \psi(\mathbf{x}_2) \\ \vdots \\ \psi(\mathbf{x}_N) \end{bmatrix}. \tag{8}$$

Now, our Pfaffian-Jastrow (PJ) ansatz is well-defined for both even and odd N , and the pairing orbital $\phi(\mathbf{x}_i, \mathbf{x}_j)$ used to construct the matrix P remains consistent in both scenarios.

Our PJ ansatz is agnostic to any particular form of the interaction and systematically improvable by simply increasing the size of the network representing $\nu(\mathbf{x}_i, \mathbf{x}_j)$ (and also $\psi(\mathbf{x}_i)$, if N is odd). The input dimensions of both $\nu(\mathbf{x}_i, \mathbf{x}_j)$ and $\psi(\mathbf{x}_i)$ only depend on the spatial dimension d and not the total number of particles N , leading to a scalable ansatz. Given the generality of our formulation, the Pfaffian ansatz calculation cannot be reduced to a determinant of singlet pairing orbitals, as in the geminal wave function. Thus, the efficient computation of the Pfaffian is crucial to the scalability of our approach. To this aim, we implement the Pfaffian computation according to ref. 47, an algorithm based on the well-known Gaussian elimination process.

In addition to the antisymmetry of the fermionic wave function, the periodic boundary conditions, and the translational symmetry (which will be discussed in Message-passing neural network subsection of the Methods section), we also enforce the discrete parity and time-reversal symmetries as

prescribed in ref. 25—see the Discrete symmetries subsection in the Methods for a detailed discussion. We further improve the nodal structure of our PJ ansatz through backflow (BF) transformations⁴⁸. To our knowledge, this is the first time neural BF transformations have been used in a Pfaffian wave function, although they have demonstrated their superiority over traditional BF transformations within the Slater-Jastrow formalism in numerous applications^{16,17,29}. We replace the original single-particle degrees of freedom \mathbf{x}_i by new ones $\tilde{\mathbf{x}}_i(X)$, such that correlations generated by the presence of all particles are incorporated into the pairing orbital. To ensure that the Pfaffian remains antisymmetric, the backflow transformation must be permutation equivariant with respect to the original \mathbf{x}_i , i.e. $\tilde{\mathbf{x}}_i$ depends on \mathbf{x}_i and is invariant with respect to the set $\{\mathbf{x}_j\}_{j \neq i}$. In the Message-passing neural network subsection of the Methods, we discuss in detail how the backflow correlations are encoded via a permutation-equivariant message-passing neural network. All calculations labeled as PJ-BF assume that we apply the transformations $\nu(\mathbf{x}_i, \mathbf{x}_j) \rightarrow \nu(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j)$ to the FNN in Eq. (6).

For comparison, we also report results obtained using a Slater-Jastrow (SJ) ansatz, where the antisymmetric part of the wave function is a Slater determinant of single-particle states,

$$\Phi_{SJ}(X) = \det \begin{bmatrix} \varphi_1(\mathbf{x}_1) & \varphi_1(\mathbf{x}_2) & \cdots & \varphi_1(\mathbf{x}_N) \\ \varphi_2(\mathbf{x}_1) & \varphi_2(\mathbf{x}_2) & \cdots & \varphi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_N(\mathbf{x}_1) & \varphi_N(\mathbf{x}_2) & \cdots & \varphi_N(\mathbf{x}_N) \end{bmatrix}. \tag{9}$$

In the fixed-node approximation, the single-particle states are the products of spin eigenstates with definite spin projection on the z -axis s_α and plane wave (PW) orbitals with discrete momenta $\mathbf{k}_\alpha = 2\pi\mathbf{m}_\alpha/L$, $\mathbf{n}_\alpha \in \mathbb{Z}^d$,

$$\varphi_\alpha(\mathbf{x}_i) = e^{i\mathbf{k}_\alpha \cdot \mathbf{r}_i} \chi_\alpha(s_i), \tag{10}$$

where $\chi_\alpha(s_i) = \delta_{s_i, s_\alpha}$. Here, $\alpha = (\mathbf{k}_\alpha, s_\alpha)$ denotes the quantum numbers characterizing the state. We will label Slater-Jastrow NQS calculations using the above plane wave orbitals as SJ-PW.

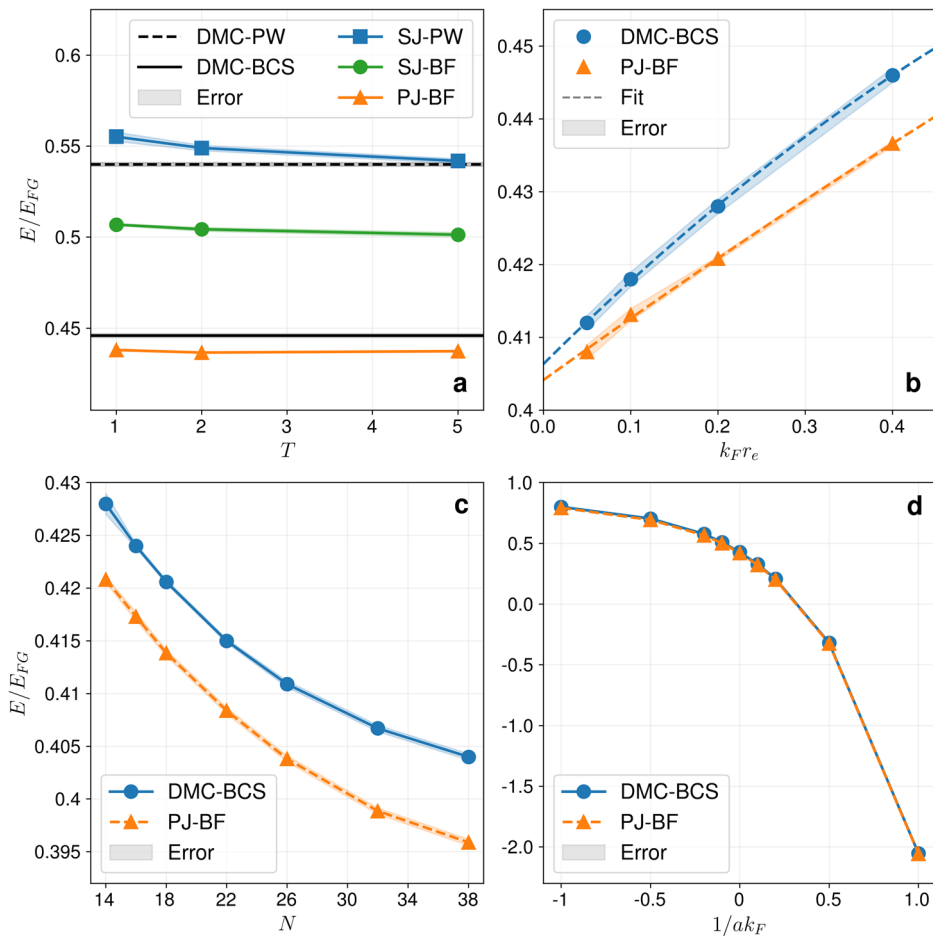
As in the Pfaffian case, we improve the nodal structure of the Slater determinant using backflow transformations generated by the MPNN, discussed in the Message-passing neural network subsection of the Methods. These transformations modify both the spatial and spin degrees of freedom, as detailed in the Backflow transformations for the SJ ansatz subsection of the Methods. Slater-Jastrow calculations using the backflow orbitals will be labeled as SJ-BF.

Energy benchmarks

We first benchmark the performance of the various NQS, outlined in the Neural-network quantum states subsection of the Results, against the state-of-the-art, fixed-node diffusion Monte Carlo (DMC) method, presented in the Diffusion Monte Carlo subsection of the Methods. The final converged energies for our NQS are calculated by averaging over the last 100 optimization steps and error bands are reported as their standard deviation. We opt for the standard deviation over the standard error to account for oscillations in the variational parameters around their converged values, treating the energy at each iteration as measurements from slightly different models. The training is considered converged when the changes in the energy are smaller than the error. For the DMC calculations, the error bands are the standard errors of the block-averaged energies^{49,50}. All assessments comparing our NQS with DMC are conducted under identical conditions, including the same number of particles, density, scattering length a , and effective range r_e . In addition, all plots include error bands as shaded regions, even if they are not easily visible. The overall convergence pattern is discussed in more detail in the Variational Monte Carlo subsection of the Methods.

In Fig. 1a, we aim to understand how the NQS respond as we vary the depth T of the MPNN, discussed in the Message-passing neural network subsection of the Methods. As shown, the final converged energies per particle for the Slater-Jastrow ansatz with plane wave orbitals (SJ-PW)

Fig. 1 | Ground-state energies per particle. The converged energies per particle for the neural-network quantum state (NQS) calculations are obtained by averaging over the last 100 optimization steps, and the corresponding error bands, represented by shaded regions, are the standard deviations. The error bands for the diffusion Monte Carlo (DMC) calculations are the standard errors of the block-averaged energies. The Pfaffian-Jastrow ansatz with backflow (PJ-BF) is represented by orange triangles in all panels. **a** Initial comparison among three different NQS and two DMC benchmarks as a function of the message-passing neural network depth T . The Slater-Jastrow ansatz with plane wave orbitals (SJ-PW) is represented by blue squares, and the Slater-Jastrow ansatz with backflow orbitals (SJ-BF) is represented by green circles. The interaction parameters are set to $v_0 = 1$ and $\mu = 5$, corresponding to an effective range of $k_F r_e = 0.4$. The DMC benchmark energies with (DMC-BCS) and without pairing (DMC-PW) are displayed as solid and dashed lines, respectively. **b** Energies at unitarity as a function of effective range $k_F r_e$. The DMC-BCS benchmark energies (blue circles) and the PJ-BF energies (orange triangles) are extrapolated to zero effective range using quadratic fits (dashed lines). See Supplementary Table 2 for the computed and extrapolated energies. **c** Energies at unitarity as a function of the number of particles N , for even N only. The effective range is fixed at $k_F r_e = 0.2$. See Supplementary Table 3 for the data depicted in the plot. **d** Energies in the BCS-BEC crossover region as a function of inverse scattering length $1/ak_F$ for a fixed effective range $k_F r_e = 0.2$. Refer to Supplementary Table 4 for the values of the interaction parameters v_0 and μ , along with the corresponding numerical values of the energies.



decreases monotonically towards the corresponding DMC-PW value, with agreement at $T = 5$. This behavior echoes the findings of ref. 51, and demonstrates the impact of the MPNN on the flexibility of our Jastrow.

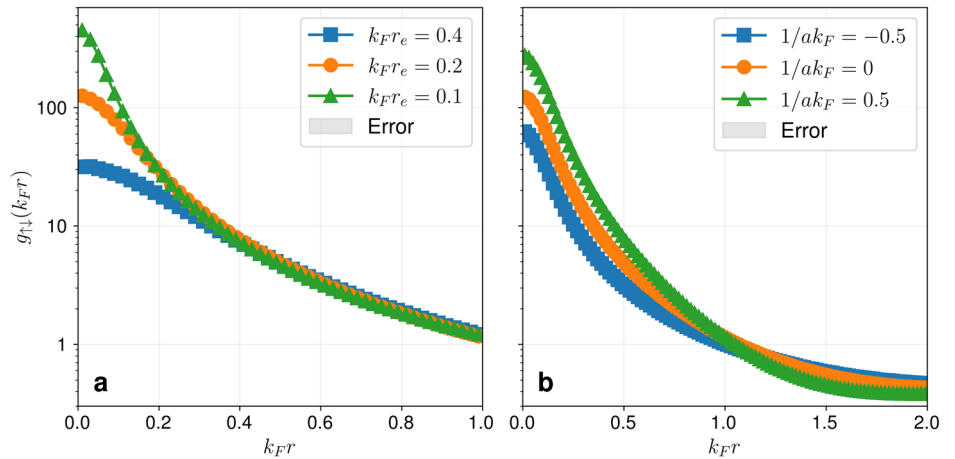
Incorporating backflow correlations into the Slater-Jastrow ansatz (SJ-BF) significantly improves results compared to the fixed-node approach with plane waves, but more than half of the discrepancy between the two DMC energies remains. Our SJ-BF ansatz presents a weak dependence on the MPNN depth T in Fig. 1a, suggesting it is unlikely that further increasing T would yield substantial improvements in energy. Other changes in the structure, such as increasing the number of nodes in a given hidden layer or increasing the depth of the individual FNNs comprising the MPNN, could theoretically provide more flexibility to the SJ-BF ansatz. However, it is commonly observed that achieving high accuracy using a generalized Slater determinant often requires the use of multiple Slater determinants¹⁷ or hidden degrees of freedom^{25,30}. In extreme cases, an entirely different ansatz may become necessary, like the generalized Pfaffian we study here.

Therefore, we turn our attention to our Pfaffian-Jastrow-Backflow (PJ-BF) ansatz. Even with a single MPNN layer, the PJ-BF ansatz easily outperforms DMC-BCS while also possessing fewer parameters than the single-layer SJ-BF ansatz (~ 5600 v.s. ~ 6200). For reference, the analytical expressions for the number of parameters in each NQS are listed in Supplementary Table 1, along with the specific numbers of parameters involved in this work. The overall dependence on the MPNN depth is weak, with $T = 2$ giving a slightly lower energy and variance than $T = 5$. For the remainder of our analysis, we will use the PJ-BF ansatz with $T = 2$, which contains about 8500 variational parameters. In this initial investigation, we have employed the same Jastrow correlator in all of our NQS to ensure a fair comparison between the different architectures. Our future research will explore whether the Jastrow component is essential or if we can exclusively rely on the generalized Pfaffian.

Given that the unitary limit is characterized by a vanishing effective range, we examine how the ground-state energy varies with r_e in Fig. 1b. To expedite and stabilize the training process for smaller values of r_e , we employ a technique called transfer learning. Initially, we train the PJ-BF ansatz with $T = 2$ using random initial parameters for $k_F r_e = 0.4$, which corresponds to approximately 20% of the average interparticle distance. We then fine-tune this model as we progressively reduce $k_F r_e$ to 0.2, 0.1, and finally, 0.05. With each decrease, the wave function becomes increasingly more challenging to learn.

The PJ-BF ansatz gives energies approximately $0.004E_{FG} - 0.009E_{FG}$ lower than DMC-BCS, with the largest differences occurring for the largest values of $k_F r_e$. This behavior is somewhat expected, as the DMC calculations rely on the nodes of the geminal wave function, which only considers contributions from the singlet channel. For small values of $k_F r_e$, this is a reasonable assumption to make, but the effects of the approximation are more apparent for larger values. Extrapolating to zero effective range using quadratic fits (see Supplementary Table 2) suggests that PJ-BF and DMC-BCS energies might eventually converge or at least approach each other, mainly due to the dominance of s -wave contributions. However, in our Pfaffian ansatz, we do not make any assumptions about the character of the pairs. The Pfaffian is a strict generalization of the geminal, and when it is coupled with a fully trainable pairing orbital, it is able to capture both singlet and triplet contributions without guidance. The true flexibility of our method originates from the ability to include completely generalized backflow correlations while remaining sign-problem free. Even if backflow correlations are included in the starting variational wave function for the DMC calculations, the fixed-node approximation and the underlying biases inherent to the variational ansatz would limit the final converged energy. As a result, we anticipate that our ansatz will outperform DMC at even smaller effective ranges than those explored in this work, although it will likely require fine-tuning of the network structure.

Fig. 2 | Opposite-spin pair distributions. Error bands are computed as the standard deviation over the last 100 iterations, but are not easily visible in the plot. **a** Distributions at unitarity for different effective ranges: $k_{F}r_e = 0.4$ (blue squares), $k_{F}r_e = 0.2$ (orange circles), and $k_{F}r_e = 0.1$ (green triangles). **b** Distributions in the crossover region for the BCS phase $1/ak_F = -0.5$ (blue squares), unitarity $1/ak_F = 0$ (orange circles), and the BEC phase $1/ak_F = 0.5$ (green triangles). The effective range is fixed at $k_{F}r_e = 0.2$. See Supplementary Table 4 for the corresponding values of the interaction parameters v_0 and μ .



One of the most appealing aspects of our PJ-BF ansatz is that it does not depend on the particle number N . In light of this, we can apply the transfer learning procedure again to accelerate the training of the larger N cases by starting with the parameters obtained from the smaller N cases. We use this process to investigate the ground-state energies at unitarity as a function of N , as shown in Fig. 1c. See Supplementary Table 3 for the numerical values of the energies and errors. Our PJ-BF ansatz gives energies about $0.007 - 0.008E_{FG}$ lower than DMC for all values of N tested, with the largest discrepancy at $N = 38$. Further investigations will be required to make conclusions about the thermodynamic limit.

As an initial comparison with existing AFQMC results, we consider the case of $N = 38$ and $k_{F}r_e = 0.2$. Our PJ-BF ansatz yields ground-state energies $0.0061(5)E_{FG}$ higher than the AFQMC value of $0.3897(4)E_{FG}$ that we extracted from Fig. 2 of ref. 8. A portion of this discrepancy is likely due to using different finite effective-range potentials⁵², discretization effects present in AFQMC also contribute. In addition, the refinement of our network through hyperparameter tuning could mitigate some of these discrepancies. In future work, we plan to carry out detailed benchmarks with the AFQMC. To this aim, we will carefully extrapolate to zero effective range using the transfer learning technique outlined here, so as to enable comparison with the AFQMC zero lattice spacing limit.

Finally, we explore the BCS-BEC crossover region for a fixed effective range $k_{F}r_e = 0.2$ in Fig. 1d. See Supplementary Table 4 for the values of the interaction parameters v_0 and μ , as well as the corresponding DMC-BCS benchmarks and the PJ-BF results. Once again, we employ transfer learning, wherein we use cases closer to unitarity to pretrain the cases that are further away. In the BCS regime, our PJ-BF ansatz consistently yields energies $\sim 0.01E_{FG}$ lower than those obtained from DMC-BCS, albeit with slightly inferior performance in the BEC regime. We attribute the minor difference in the performance of our PJ-BF ansatz between the BCS and BEC phases to the requirement for greater flexibility in capturing the short-range behavior of pairs in the BEC regime. We anticipate that enlarging the size of the FNN that defines the pairing orbital would help alleviate the small discrepancies in performance observed between the BCS and BEC phases.

Pair distribution functions

The spin-dependent two-body radial distribution functions capture the probability density of finding two particles with specific spin orientations at a given separation distance. Hence, they provide a quantitative description of the spatial correlations and pairing phenomena between fermionic particles. Note that NQS offer a significant advantage over DMC methods when computing expectation values of quantum mechanical operators that do not commute with the Hamiltonian. The accuracy of DMC estimates of two-body distribution functions heavily relies on the quality of the starting variational wave function, and extrapolations have to be performed to remove such dependencies that are particularly apparent at small interparticle distances¹⁰.

We compute the opposite-spin pair distributions similar to ref. 40, by first defining

$$g_c(r) = \frac{1}{2\pi r^2 \rho N} \sum_{i < j}^N \frac{\langle \Psi | \delta(r_{ij} - r) | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \quad (11)$$

$$g_z(r) = \frac{1}{2\pi r^2 \rho N} \sum_{i < j}^N \frac{\langle \Psi | \delta(r_{ij} - r) \sigma_i^z \sigma_j^z | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \quad (12)$$

where σ_i^z is the Pauli Z-operator acting on the i -th particle. Since the eigenvalues of $\sigma_i^z \sigma_j^z$ are $+1$ if particles i and j have the same spin and -1 if they have opposite spin, the opposite-spin pair distribution function can be computed as

$$g_{\uparrow\downarrow}(r) = \frac{1}{2} (g_c(r) - g_z(r)). \quad (13)$$

In Fig. 2a, we show the opposite-spin pair distribution functions at unitarity for $k_{F}r_e = 0.4, 0.2$, and 0.1 . Notice how the peaks of the distributions at $k_{FR} = 0$ grow roughly quadratically as $k_{F}r_e$ is reduced by half, demonstrating the presence of strong pairing correlations as we approach the unitary limit $r_e \rightarrow 0$. Clearly, the short-range character of the distributions are important to capture at unitarity, as they begin to overlap when $r > r_e$.

Figure 2 b presents a complementary set of opposite-spin pair distribution functions in the crossover region with fixed effective range of $k_{F}r_e = 0.2$. When leaning towards the BCS phase $1/ak_F = -0.5$, the long-range tail of the density is enhanced compared to the unitary case $1/ak_F = 0$. On the other hand, the tail is diminished in the BEC phase $1/ak_F = 0.5$, suggesting the initiation of dimer formation. The differences in the peaks of the distributions are not as dramatic as in Fig. 2a, but they are consistent with the expected behavior in the BCS and BEC regimes near unitarity.

Pairing gap

The pairing gap is computed using odd-even staggering⁵³

$$\Delta(N) = E(N) - \frac{1}{2}(E(N+1) + E(N-1)), \quad (14)$$

where N is taken to be odd and $E(N)$ denotes the total energy of N particles. Since the MPNN and the pairing orbital ϕ is independent of the system size N , the trained ansatz for $N = 14$ can serve as the initial state for the $N = 16$ calculation. For the $N = 15$ calculation, the pairing orbital from the $N = 16$ calculation was utilized as the starting point, while the unpaired single-particle orbital ψ was initialized with random parameters. The final energies per particle for $N = 14, 15, 16$ are listed in Table 1, along with the differences

Table 1 | Energies for even and odd particle numbers

$1/ak_F$	N	DMC-BCS	PJ-BF	Diff.
-0.5	14	0.7052 (3)	0.6937 (3)	-0.0115 (4)
	15	0.7546 (3)*	0.7433 (4)*	-0.0113 (5)
	16	0.7074 (4)	0.6980 (3)	-0.0094 (5)
0	14	0.428 (1)	0.4208 (3)	-0.007 (1)
	15	0.4900 (2)*	0.4766 (5)*	-0.0134 (5)
	15	0.5357 (2)	0.5209 (5)	-0.0148 (5)
	16	0.4240 (2)	0.4173 (4)	-0.0067 (4)
0.5	14	-0.3198 (3)	-0.3244 (4)	-0.0046 (5)
	15	-0.2020 (3)*	-0.2176 (5)*	-0.0156 (6)
	16	-0.3194 (3)	-0.3245 (3)	-0.0051 (4)

Ground-state energies per particle for even and odd particle numbers N , in units of the Fermi gas energy in the thermodynamic limit E_{FG} . The effective range is fixed at $k_{F\theta} = 0.2$. Values with asterisks (*) indicate calculations without translational symmetry. The last column lists the differences between the Pfaffian-Jastrow with backflow (PJ-BF) energies and the diffusion Monte Carlo with pairing (DMC-BCS) benchmarks.

between the PJ-BF and DMC-BCS methods. The pairing gaps in Table 2 are computed using the energies in Table 1.

In our investigations of even- N systems, we have enforced translational invariance in our ansatz to better approximate an infinite system of correlated pairs. This is achieved by excluding the absolute particle positions \mathbf{r}_i in the inputs to the MPNN — see the Message-passing neural network subsection of the Methods for the full description. For odd- N systems, however, breaking translational symmetry is necessary due to an unpaired fermion occupying the minimum of the quasiparticle dispersion at finite k^2 , yielding nonzero total momentum. To highlight the versatility of our approach, we present two different PJ-BF calculations for the unitary Fermi gas with $N = 15$ — one that enforces translational symmetry, and another that breaks it. In the former, we simply take the one-body output of the MPNN $\mathbf{h}_i^{(T)}$ as input to the single-particle pairing orbital ψ , since it is translationally invariant by construction. In the latter, we concatenate the positions $(\mathbf{r}_i, \mathbf{h}_i^{(T)})$ with the MPNN output, keeping in mind the spins have already been included $\mathbf{h}_i^{(T)}$.

In the DMC-BCS calculations, translational symmetry is enforced by placing the unpaired particle in the $k = 0$ momentum state. To break translational symmetry, we instead place the unpaired particle in different momentum states $k \neq 0$ and select the one giving the lowest energy. Notice that symmetry breaking and enforcement in the PJ-BF calculations is more general and efficient than in the DMC-BCS calculations, as we do not have to specify the momentum state of the unpaired particle in advance, nor do we have to run the calculations repeatedly for energy comparisons.

For $N = 15$ at unitarity, our PJ-BF energies are $\sim 0.015E_{FG}$ and $\sim 0.013E_{FG}$ lower than DMC-BCS in the cases with and without translational symmetry, respectively. Meanwhile, the PJ-BF energies are just $\sim 0.007E_{FG}$ lower for both the $N = 14$ and $N = 16$ cases, leading to a smaller pairing gap than predicted by DMC-BCS. Specifically, we predict a pairing gap that is $\sim 0.07\epsilon_F$ smaller with translational symmetry, and $\sim 0.06\epsilon_F$ smaller without translational symmetry.

At unitarity, our PJ-BF ansatz produces a pairing gap of $\Delta(15) = 0.519(8)\epsilon_F$, just $0.07(5)\epsilon_F$ larger than the reported experimental value $\Delta_{exp} = 0.45(5)\epsilon_F$ ⁹. Conversely, DMC-BCS calculations yield a larger pairing gap of $\Delta(15) = 0.577(8)\epsilon_F$, exceeding Δ_{exp} by $0.13(5)\epsilon_F$. Considering the influence of finite-size effects in both PJ-BF and DMC-BCS calculations, further investigation involving larger particle numbers is warranted. The pairing gap we computed for this small test system aligns, within errors, with previous DMC calculations of the pairing gap at $\Delta(66) = 0.84(4)E_{FG} = 0.50(2)\epsilon_F$ ⁵⁴.

The difference in performance between the odd- N cases and the even- N cases is even more noticeable on the BEC side, with the $N = 15$ calculation being $\sim 0.016E_{FG}$ lower, and the $N = 14$ and $N = 16$ calculations being

Table 2 | Pairing gaps

$1/ak_F$	DMC-BCS	PJ-BF	Diff.
-0.5	0.434 (6)*	0.426 (7)*	-0.008 (9)
0	0.577 (8)*	0.519 (8)*	-0.06 (1)
0	0.988 (8)	0.918 (8)	-0.07 (1)
0.5	1.058 (6)*	0.962 (8)*	-0.10 (1)

The pairing gap $\Delta(15)$, in units of the Fermi energy ϵ_F , in the BCS-BEC crossover with fixed $k_{F\theta} = 0.2$. Errors are propagated from Table 1. Values with asterisks (*) indicate using an $N = 15$ calculation without translational symmetry. The last column lists the differences between the Pfaffian-Jastrow with backflow (PJ-BF) pairing gaps and the corresponding diffusion Monte Carlo with pairing (DMC-BCS) benchmarks.

just $\sim 0.005E_{FG}$ lower than the DMC-BCS energy. This leads to a pairing gap that is about $0.10\epsilon_F$ smaller than DMC-BCS predicts. On the BCS side, the PJ-BF finds an improvement in energy of $0.009E_{FG}$ to $0.011E_{FG}$ compared to DMC-BCS, with no significant difference in performance between the odd and even N cases. Still, the gap predicted by PJ-BF is $\sim 0.008\epsilon_F$ smaller than DMC-BCS.

Discussion

In this study, we propose a neural-network quantum state based on the Pfaffian-Jastrow (PJ) framework that utilizes a message-passing neural network (MPNN) to encode pairing and backflow (BF) correlations. We evaluate its performance against comparable Slater-Jastrow (SJ) ansätze with identical MPNN architectures. Our results indicate that increasing the depth of the MPNN systematically improves the performance of the SJ ansätze, but backflow correlations within the single-particle picture are still insufficient in capturing all pairing correlations. However, a simple and compact PJ-BF ansatz outperforms state-of-the-art DMC-BCS benchmarks in both the BCS and BEC phases. This improvement remains significant in the limit of zero effective range, which is particularly challenging to compute with many-body methods.

Transfer learning has proven to be an essential tool in this work. It enables the realization of the unitary limit in a controlled manner, mitigating the risk of becoming trapped in local minima. It also allows for the efficient exploration of regions beyond unitarity, unlocking new avenues for studying the BCS-BEC crossover. Transfer learning also played a major role in moving towards the thermodynamic limit, with up to $N = 38$. All even- N systems can be treated with a single architecture, while the odd- N systems can be treated by introducing one additional FNN to represent the unpaired single-particle orbital. This small modification makes the calculation of the pairing gap more easily accessible and enables concrete assessments of superfluid behavior. Our PJ-BF ansatz performs particularly well for odd- N with $1/ak_F \geq 0$, but its prediction of a smaller pairing gap compared to DMC-BCS is consistent throughout the crossover region.

In future work, we intend to perform a more careful extrapolation to the $r_e \rightarrow 0$ limit with larger N since, with the exception of the unitary limit, we have used relatively large values of $k_{F\theta}$ in this initial investigation of the BCS-BEC crossover. More hyperparameter tuning will be needed, because the smaller effective ranges will demand greater flexibility in our ansatz to accurately capture the short-distance physics. In particular, we plan to explore changes in the width of the hidden layers, the depth of the individual FNNs, and the necessity of the Jastrow. With a more thoroughly tuned network structure and a system size closer to the thermodynamic limit, we plan to conduct a direct comparison with other state-of-the-art methods, particularly AFQMC.

Our Pfaffian-Jastrow-Backflow NQS displays immense potential in the study of ultra-cold Fermi gases. Unlike conventional methods, this ansatz is not subject to biases arising from physical intuition or a lack thereof, as it does not require specifying a particular form for the pairing orbitals. For this reason, it can be readily applied to other strongly correlated systems, including molecules and other systems relevant for quantum chemistry applications. In contrast to the commonly used geminal wave function, our ansatz does not rely on ordering the spin of the interacting fermions, and it is

therefore amenable to Hamiltonians that exchange spin, such as those modeling nuclear dynamics. In this regard, we anticipate calculations of atomic nuclei and low-density isospin-asymmetric nucleonic matter and carry out detailed investigations on the nature of nuclear pairing⁵⁵.

When the stochastic reconfiguration algorithm and transfer learning techniques are combined with the enforcement of translational, parity, and time-reversal symmetries, highly non-perturbative correlations can be encoded in a small number of parameters by modern standards. This approach will pave the way for future developments in the study of many-body systems, as it offers a powerful tool for encoding correlations in a compact and computationally feasible manner.

Methods

Message-passing neural network

Implementing the families of NQS, $\Psi(X)$, discussed in the Neural-network quantum states subsection of the Results is possible using X as direct inputs to the appropriate FNNs and Deep-Sets⁵⁶. Still, it is advantageous to devise new inputs that already capture a large portion of the correlations. As in ref. 44, we employ a permutation-equivariant message-passing neural network (MPNN) to iteratively build correlations into new one-body and two-body features from the original “visible” features. The visible features are chosen to be

$$\mathbf{v}_i = (s_i), \quad (15)$$

$$\mathbf{v}_{ij} = (\mathbf{r}_{ij}, \|\mathbf{r}_{ij}\|, s_{ij}), \quad (16)$$

with the separation vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and distances $\|\mathbf{r}_{ij}\| = r_{ij}$ replaced by their L -periodic surrogates

$$\mathbf{r}_{ij} \mapsto (\cos(2\pi r_{ij}/L), \sin(2\pi r_{ij}/L)), \quad (17)$$

$$\|\mathbf{r}_{ij}\| \mapsto \|\sin(\pi r_{ij}/L)\|, \quad (18)$$

and the quantity $s_{ij} \equiv 2\delta_{s_i, s_j} - 1$ assigned a value of $+1$ for aligned spins and -1 for anti-aligned spins. Note that we have excluded explicit dependence on the particle positions \mathbf{r}_i in the visible one-body features, thereby enforcing translational invariance in the new features. Linear transformations are applied to and concatenated with each feature to obtain the initial hidden features

$$\mathbf{h}_i^{(0)} = (\mathbf{v}_i, A\mathbf{v}_i), \quad (19)$$

$$\mathbf{h}_{ij}^{(0)} = (\mathbf{v}_{ij}, B\mathbf{v}_{ij}). \quad (20)$$

The main purpose of the linear transformations A and B is to preprocess the input data, but they also help simplify the implementation by keeping the dimension of the hidden features $\mathbf{h}_i^{(t)}, \mathbf{h}_{ij}^{(t)}$ constant for all t . In each iteration, $t = 1, \dots, T$ of the MPNN, information is exchanged between the one- and two-body streams through a so-called message

$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t(\mathbf{h}_i^{(t-1)}, \mathbf{h}_j^{(t-1)}, \mathbf{h}_{ij}^{(t-1)}). \quad (21)$$

For a given particle i , relevant messages are collected and pooled together to destroy the ordering with respect to all other particles $j \neq i$,

$$\mathbf{m}_i^{(t)} = \text{pool}(\{\mathbf{m}_{ij}^{(t)} \mid j \neq i\}). \quad (22)$$

The pooling operation, pool , collapses the order of the elements in the set it acts upon and produces a vector with the same dimension as an individual element. Throughout this work, we use logsumexp -pooling, the smooth

variation of max-pooling. More explicitly, the resulting vector of the logsumexp -pooling function acting on a set of vectors $\{z_i\}$ is given by

$$\text{logsumexp}(\{z_i\}) = \log\left(\sum_i \exp(z_i)\right), \quad (23)$$

where both the logarithm and exponential are applied element-wise.

The pairwise messages $\mathbf{m}_{ij}^{(t)}$ and the implied particle messages $\mathbf{m}_i^{(t)}$ are then used to update the hidden features

$$\mathbf{h}_i^{(t)} = (\mathbf{v}_i, \mathbf{F}_t(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)})), \quad (24)$$

$$\mathbf{h}_{ij}^{(t)} = (\mathbf{v}_{ij}, \mathbf{G}_t(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)})). \quad (25)$$

The functions $\mathbf{M}_t, \mathbf{F}_t$, and \mathbf{G}_t are all unique FNNs with the same output dimension as the linear preprocessors A and B . By incorporating concatenated skip connections to the visible features, we guarantee that the signal originating from the raw data remains discernible even as the MPNN depth T increases. Finally, we combine the resulting outputs $\mathbf{h}_i^{(T)}$ and $\mathbf{h}_{ij}^{(T)}$ into pairwise feature vectors

$$\mathbf{g}_{ij} = (\mathbf{h}_i^{(T)}, \mathbf{h}_j^{(T)}, \mathbf{h}_{ij}^{(T)}) \quad (26)$$

to feed into subsequent networks. The flow of information through the MPNN can be visualized in Fig. 3. Notice how the hidden features in a given layer depend on the hidden features of the previous layer and the original visible features.

For all our NQS, we use a Jastrow correlator based on a Deep-Set⁵⁶ to enforce permutation invariance over the set of all pairwise features

$$J(X) = \rho\left(\text{pool}\left(\{\zeta(\mathbf{g}_{ij}) \mid i \neq j\}\right)\right). \quad (27)$$

Here, ρ and ζ are FNNs, and the pooling operation is the same as in Eq. (22). While many Jastrow functions are typically designed to satisfy Kato’s cusp condition⁵⁷ for specific systems, we take a different approach and allow our neural networks to learn the cusp fully. The short-range behavior of the ground state is particularly important for the UFG, so leaving our NQS completely unbiased serves as an important test for evaluating the overall capabilities of NQS.

The Slater-Jastrow ansatz with plane wave orbitals (SJ-PW) does not require any additional neural networks beyond ρ and ζ , so it establishes a baseline for the number of trainable parameters in this work. On the other hand, the backflow variables \mathbf{q}_i and θ_i for the Slater-Jastrow ansatz with backflow orbitals (SJ-BF) are the outputs of another Deep-Set

$$(\text{Re}(\mathbf{q}_i), \text{Im}(\mathbf{q}_i), \theta_i) = \rho_{\text{bf}}\left(\text{pool}\left(\{\zeta_{\text{bf}}(\mathbf{g}_{ij}) \mid j \neq i\}\right)\right), \quad (28)$$

which is permutation invariant with respect to all $j \neq i$ by construction. The size of ρ_{bf} and ζ_{bf} determines the number of extra variational parameters present in the SJ-BF ansatz compared to the SJ-PW ansatz. For the PJ-BF ansatz, the pairing orbital ν in Eq. (6) simply takes \mathbf{g}_{ij} as input in place of $(\mathbf{x}_i, \mathbf{x}_j)$. Therefore, the number of additional variational parameters in the PJ-BF ansatz relative to the SJ-PW ansatz is determined by the size of ν .

The individual FNNs within our NQS are solely dependent on the spatial dimension d and not the system size N . The analytical expressions for the number of parameters for each of the NQS are shown in Supplementary Table 1, where the individual FNNs are assumed to have constant width. The hyperparameters are the MPNN depth T , the depth of the individual FNNs D , and the number of hidden nodes H in a single hidden layer. All of the FNNs mentioned throughout this work use $H = 16$ and $D = 1$. The

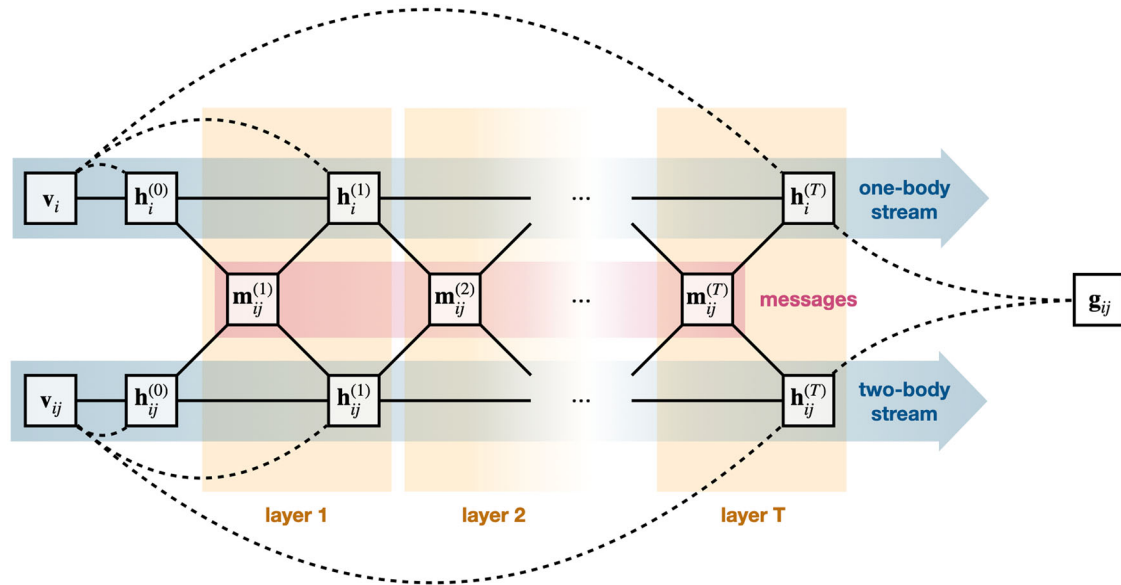


Fig. 3 | Schematic representation of the message-passing neural network. Each of the T total iterations of the network is represented by a yellow box. Dashed lines represent the concatenation operations, while solid lines represent the parameterized transformations (linear transformations and nonlinear feedforward neural networks). Messages, highlighted in pink, mediate the exchange of information between the one- and two-body streams, in blue.

activation function is GELU⁵⁸ and the weights/biases are initialized with glorot normal/zeros unless pretrained parameters are used.

Backflow transformations for the SJ ansatz

When employing the SJ ansatz in the fixed-node approximation (SJ-PW), we take the single-particle states to be products of momentum eigenstates with definite wave vector $\mathbf{k}_\alpha = 2\pi\mathbf{n}_\alpha/L$, $\mathbf{n}_\alpha \in \mathbb{Z}^d$, and eigenspinors with definite spin projection s_α ,

$$|\varphi_\alpha\rangle = |\mathbf{k}_\alpha\rangle|s_\alpha\rangle. \quad (29)$$

Omitting overall normalization constants, the probability amplitude of measuring particle i in state α is given by the plane wave orbital

$$\varphi_\alpha(\mathbf{x}_i) = \langle \mathbf{x}_i | \varphi_\alpha \rangle = \langle \mathbf{r}_i | \mathbf{k}_\alpha \rangle \langle s_i | s_\alpha \rangle = e^{i\mathbf{k}_\alpha \cdot \mathbf{r}_i} \delta_{s_\alpha, s_i}, \quad (30)$$

where $|\mathbf{x}_i\rangle = |\mathbf{r}_i\rangle|s_i\rangle$. Let us transform to a new basis with modified position eigenstates and a superposition of eigenspinors $|\tilde{\mathbf{x}}_i\rangle = |\tilde{\mathbf{r}}_i\rangle|\chi_i\rangle$. While permutation equivariance is the sole essential property required for the backflow transformation $|\mathbf{x}_i\rangle \mapsto |\tilde{\mathbf{x}}_i\rangle$, to preserve the antisymmetry of the fermionic wave function, an additional property is desirable for computational convenience. Specifically, when the transformation depends on certain parameters, we aim to have $|\tilde{\mathbf{x}}_i\rangle = |\mathbf{x}_i\rangle$ when the parameters are identically zero. Then, nonzero parameters signify deviations from the original plane wave orbitals, such that less training is required compared to completely trainable orbitals.

An appropriate spatial transformation is trivial. We simply define new parameters $\mathbf{q}_i \in \mathbb{C}^d$, called the backflow displacement, and shift the coordinates as

$$\mathbf{r}_i \rightarrow \mathbf{r}_i + \mathbf{q}_i(X). \quad (31)$$

The above backflow displacement is complex, allowing for changes in both the phases and amplitudes of the original plane wave orbitals.

For the spin part of the transformation, we look to spinors on the Bloch sphere for inspiration,

$$|\chi_i\rangle = \cos\left(\frac{\theta_i(X)}{2}\right)|s_i\rangle + \sin\left(\frac{\theta_i(X)}{2}\right)\sigma_i^x|s_i\rangle. \quad (32)$$

In the above, we have introduced another backflow variable $\theta_i \in \mathbb{R}$ akin to the polar angle of a Bloch spinor, and we have excluded the relative phase in favor of a completely real-valued wave function. We also write the superposition in terms of $|s_i\rangle$ and the Pauli X -operator σ_i^x , which flips the spin of the i -th particle, rather than $|\uparrow\rangle$ and $|\downarrow\rangle$. This way, it is obvious that $|\chi_i\rangle = |s_i\rangle$ when $\theta_i = 0$, as desired. The overlap of two spinors is given by

$$\begin{aligned} \langle \chi_i | \chi_j \rangle &= \cos\left(\frac{\theta_i}{2}\right) \cos\left(\frac{\theta_j}{2}\right) \langle s_i | s_j \rangle \\ &+ \sin\left(\frac{\theta_i}{2}\right) \cos\left(\frac{\theta_j}{2}\right) \langle s_i | \sigma_i^{x\dagger} | s_j \rangle \\ &+ \cos\left(\frac{\theta_i}{2}\right) \sin\left(\frac{\theta_j}{2}\right) \langle s_i | \sigma_j^x | s_j \rangle \\ &+ \sin\left(\frac{\theta_i}{2}\right) \sin\left(\frac{\theta_j}{2}\right) \langle s_i | \sigma_i^{x\dagger} \sigma_j^x | s_j \rangle \\ &= \cos\left(\frac{\theta_i - \theta_j}{2}\right) \delta_{s_i, s_j} + \sin\left(\frac{\theta_i - \theta_j}{2}\right) (1 - \delta_{s_i, s_j}). \end{aligned} \quad (33)$$

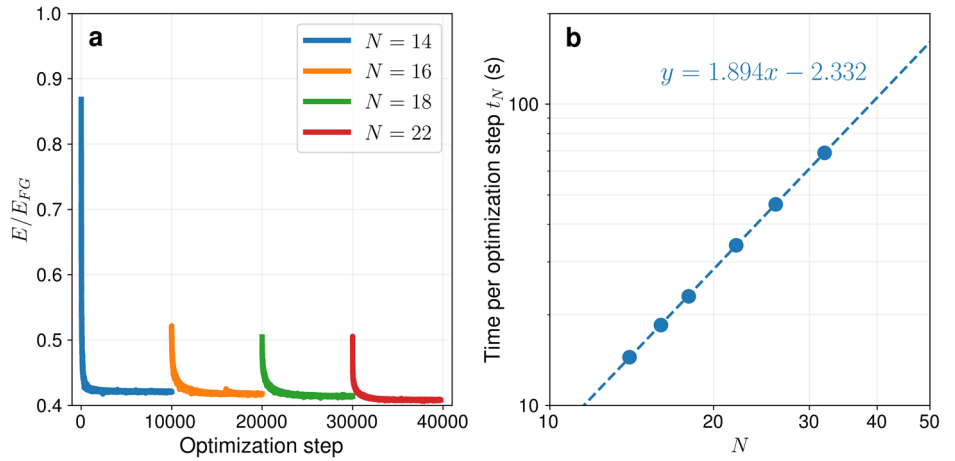
Therefore, if $s_i = s_j$, the overlap is 1 for $\theta_i = \theta_j$ and 0 for $\theta_i - \theta_j = \pm\pi$. On the other hand, if $s_i \neq s_j$, the overlap is 0 for $\theta_i = \theta_j$ and ± 1 for $\theta_i - \theta_j = \pm\pi$.

Finally, we can compute the backflow orbitals with the transformed degrees of freedom

$$\begin{aligned} \varphi_\alpha(\tilde{\mathbf{x}}_i) &= \langle \tilde{\mathbf{x}}_i | \varphi_\alpha \rangle \\ &= \langle \tilde{\mathbf{r}}_i | \mathbf{k}_\alpha \rangle \langle \chi_i | s_\alpha \rangle \\ &= e^{i\mathbf{k}_\alpha \cdot \tilde{\mathbf{r}}_i} \left(\cos\left(\frac{\theta_i}{2}\right) \langle s_i | s_\alpha \rangle + \sin\left(\frac{\theta_i}{2}\right) \langle s_i | \sigma_i^{x\dagger} | s_\alpha \rangle \right) \\ &= e^{i\mathbf{k}_\alpha \cdot (\mathbf{r}_i + \mathbf{q}_i)} \left(\cos\left(\frac{\theta_i}{2}\right) \delta_{s_\alpha, s_i} + \sin\left(\frac{\theta_i}{2}\right) (1 - \delta_{s_\alpha, s_i}) \right). \end{aligned} \quad (34)$$

In Eqs. (31) and (32), we use the notation $\mathbf{q}_i(X)$ and $\theta_i(X)$ to emphasize that the backflow parameters we define here are not variational parameters, but a function of all other particles. More specifically, they are permutation-equivariant functions of the original \mathbf{x}_i , whose functional forms depend on the outputs of the permutation-equivariant MPNN described in the Message-passing neural network subsection of the Methods.

Fig. 4 | Computational details. **a** To demonstrate the transfer learning strategy, we depict the energy per particle, in units of the Fermi gas energy E_{FG} , plotted against the optimization step, with the training curves for increasing N concatenated together. The final parameters from the previous training session were used as the initial parameters for the subsequent training session. **b** Log-log plot of wall time per optimization step t_N as a function of particle number N . The slope of the linear fit (dashed line) for $N \leq 32$ gives an approximate scaling of $O(N^{1.894})$ when using 4 NVIDIA-A100 GPUs.



Discrete symmetries

In addition to fermion antisymmetry, periodic boundary conditions, and translational invariance, we also enforce parity and time-reversal symmetries as prescribed in ref. 25. In the unpolarized case, we enforce the discrete parity and time-reversal symmetries using $\Psi^{PT}(R, S)$, which is given by

$$\Psi^P(R, S) = \Psi(R, S) + \Psi(-R, S), \quad (35)$$

$$\Psi^{PT}(R, S) = \Psi^P(R, S) + (-1)^n \Psi^P(R, -S), \quad (36)$$

where $n = N/2$ and we have used the notation $R = \{r_1, \dots, r_N\}$ and $S = \{s_1, \dots, s_N\}$ for the set of all positions and spins, respectively. Enforcing these symmetries has been shown to accelerate the convergence of ground-state energies for both atomic nuclei²⁵ and dilute neutron matter²¹.

Variational Monte Carlo and training

We train our NQS by minimizing the energy

$$E(\mathbf{p}) \equiv \frac{\langle \Psi(\mathbf{p}) | H | \Psi(\mathbf{p}) \rangle}{\langle \Psi(\mathbf{p}) | \Psi(\mathbf{p}) \rangle} \quad (37)$$

with respect to the variational parameters \mathbf{p} . To compute the energy and its gradient $\nabla_{\mathbf{p}} E$ using Monte Carlo integration, we sample positions R and spins S from $|\Psi(R, S)|^2$ in a way that preserves periodicity and total spin projection on the z -axis, as in refs. 18,25. Since the ordering of the spins is not fixed, our ansätze can be immediately applied to any continuous-space Hamiltonian that exchange spin, such as ref. 59.

A sophisticated optimization technique is critical for achieving an ansatz that is both compact and expressive. In this work, we employ the stochastic reconfiguration⁶⁰ (SR) algorithm with regularization based on the RMSprop method, introduced in ref. 25. The parameters are updated as

$$\mathbf{p} \leftarrow \mathbf{p} - \eta G^{-1} \nabla_{\mathbf{p}} E, \quad (38)$$

where η is a constant learning rate and G is the quantum geometric tensor⁶¹.

Due to the strong and short-range nature of the interaction in Eq. (2), it is likely for the optimization process to get trapped in a local minimum when initialized with random parameters, especially for small values of $k_{F}r_e$. For instance, the energy may initially decrease, but then abruptly begin to oscillate around a constant value, with a very large variance. To avoid this problem, we pretrain the NQS on a softer interaction ($k_{F}r_e = 0.4$) before proceeding to harder ones ($k_{F}r_e = 0.2, 0.1, 0.05$). Not only does this approach improve the final converged energy, but the efficiency of the optimization process overall, since the larger values of $k_{F}r_e$ can be trained more aggressively.

For the initial phase of training the softer interaction $k_{F}r_e = 0.4$ with random initial parameters, the SJ-PW and SJ-BF ansätze reached

convergence within 5000 iterations and the PJ-BF ansatz reached convergence within 10000 iterations. The differences in the required number of training iterations can be attributed to the pairing orbital being fully trainable in the PJ ansatz, while the SJ ansätze are designed to closely approximate the free Fermi gas wave function for small values of the variational parameters. During this phase, we used a learning rate of $\eta = 0.001$ and only 8000 Monte Carlo samples divided into 40 block averages.

After comparing the performance of the various NQS, in Fig. 1a, we then chose a specific model (PJ-BF with an MPNN depth of $T = 2$) to fine-tune at $k_{F}r_e = 0.2$ through the transfer learning process. In this stage, we reduced the learning rate to $\eta = 0.0001$ and increased the number of samples to 16000. Convergence was reached within 10000 iterations. Proceeding in this way, we continued to reduce $k_{F}r_e$ by half, reduce the learning rate, and increased the number of samples, until we reached $k_{F}r_e = 0.05$.

This transfer learning process is particularly useful as we decrease r_e , but we also use it to increase N . In Fig. 4a, we present the energy per particle throughout the training for different N . When we transition to a larger N , there is a slight initial increase in energy, but it is noticeably less pronounced compared to starting with random parameters. Transfer learning for $N > 14$ with a fixed effective range $k_{F}r_e = 0.2$ required approximately 10000 training iterations using learning rates in the range of $\eta \in [0.0001, 0.0004]$.

Similarly, when investigating the BCS-BEC crossover for $1/ak_F \in [-1, 1]$ under the same fixed effective range $k_{F}r_e = 0.2$, our approach demanded the same number of iterations and range of learning rates. However, the step size of the Metropolis sampler had to change drastically in the crossover region. We tuned this step size such that the acceptance rate was approximately 60%.

Supplementary Table 5 presents the time per optimization step for both the SJ-BF and PJ-BF ansätze as ratios compared to the time per optimization step for the SJ-PW ansatz with the same MPNN depth T . The computational costs of both the SJ and PJ ansätze are primarily determined by the evaluation of the shared MPNN, and of course, the details of the implementation, computing resources, and hyperparameters. For an MPNN depth of $T = 1$, our implementation of the SJ-BF ansatz required 22% more time per optimization step than the SJ-PW ansatz, while the PJ-BF ansatz required 52% more time compared to the SJ-PW ansatz. The differences in the time required decrease as T increases, with an 8% increase for SJ-BF and a 22% increase for PJ-BF at a depth of $T = 5$.

Supplementary Table 6 lists the time per optimization step t_N for the PJ-BF ansatz and different N as ratios compared to the $N = 14$ case, t_{14} . In addition, the times t_N are shown as a function of N in Fig. 4b. The algorithm we used to compute the Pfaffian, as referenced in refs. 37,47, is theoretically expected to scale as $O(N^3)$. The nearly perfect linear fit indicates that the computational time scales as $O(N^{1.894})$ for the PJ-BF ansatz with $T = 2$. Our more favorable scaling is most likely due to the parallelization we have implemented and the use of 4 NVIDIA-A100 GPUs during most of our simulations. The data point for $N = 38$ utilizes 8 NVIDIA-A100 GPUs, as

opposed to 4 GPUs used for other points, and was consequently excluded from the fit.

Diffusion Monte Carlo

The fixed-node DMC calculations are performed as described in ref. 62. All comparisons between our NQS and the DMC calculations involve identical conditions, including the same number of particles, density, scattering length a , and effective range r_e . The initial state is prepared using a variational wave function with the same general form as Eq. (3). Note that, within the fixed-node approximation, DMC provides a strict upper bound to the energy of the system. While DMC is a precise method, its accuracy relies on the choice of nodal surface and the quality of the preceding VMC calculation.

The symmetric Jastrow factor is given by

$$J(X) = \sum_{i\bar{i}}^n u(r_{i\bar{i}}), \quad (39)$$

$$u(r) = K \tanh(\mu_j r) \cosh(\gamma r)/r, \quad (40)$$

where $n = N/2$ and the unprimed and primed indices denote the spin-up and spin-down particles, respectively. The parameters K and γ are adjusted so that $u(d) = 0$ and $u'(d) = 0$, and μ_j and d are variational parameters. Considering that the s -wave channel dominates the interaction, the antisymmetric part is given by the number-projected BCS wave function

$$\Phi_{BCS}(X) = \det \begin{bmatrix} \phi(\mathbf{r}_{11'}) & \phi(\mathbf{r}_{12'}) & \cdots & \phi(\mathbf{r}_{1n'}) \\ \phi(\mathbf{r}_{21'}) & \phi(\mathbf{r}_{22'}) & \cdots & \phi(\mathbf{r}_{2n'}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(\mathbf{r}_{n1'}) & \phi(\mathbf{r}_{n2'}) & \cdots & \phi(\mathbf{r}_{nn'}) \end{bmatrix}, \quad (41)$$

with the pairing orbitals

$$\phi(\mathbf{r}) = \tilde{\beta}(\mathbf{r}) + \sum_i a(k_i^2) e^{i\mathbf{k}_i \cdot \mathbf{r}}, \quad (42)$$

$$\tilde{\beta}(\mathbf{r}) = \beta(\mathbf{r}) + \beta(L - r) - 2\beta(L/2), \quad (43)$$

$$\beta(\mathbf{r}) = (1 + cbr)(1 - e^{-dbr}) \frac{e^{-br}}{dbr}. \quad (44)$$

The parameters $a(k_i^2)$, b and d are obtained by minimizing the energy, and c is chosen so that the function β has zero slope at the origin. If we instead let $\beta = 0$ and restrict the sum in Eq. (42) to momentum states filled up to k_F , the antisymmetric part is equivalent to the Slater determinant with single-particle plane waves as in Eqs. (9) and (10). Since this approach does not involve pairing, we will refer to the related DMC results as DMC-PW. Conversely, the approach that accounts for pairing will be identified as DMC-BCS.

It should be emphasized that the BCS wave function of Eq. (41) is a special case of the generalized Pfaffian of Eq. (4). In fact, it can be easily shown³⁷ that by only retaining the spin-singlet blocks, the calculation of the Pfaffian reduces to the determinant of spin-singlet block.

Note Added: A work very recently appeared in pre-print⁶³ introduces neural backflow transformations in a geminal wave function and studies the unitary Fermi gas. We leave systematic comparisons between the two approaches to future works while already observing that the Pfaffian wave function is a strict generalization of the geminal wave function^{37,64}.

Data availability

The data supporting the findings of this study are available in the tables, figures, and Supplementary Information. Additional raw data used to

generate the plots can be made available from the corresponding author upon request.

Code availability

Simulation codes are available from the corresponding author upon request.

Received: 21 June 2023; Accepted: 1 April 2024;

Published online: 08 May 2024

References

- Gandolfi, S. Quantum Monte Carlo study of strongly interacting Fermi gases. *J. Phys. Confer. Series* **529**, 012011 (2014).
- Gezerlis, A. & Carlson, J. Strongly paired fermions: Cold atoms and neutron matter. *Phys. Rev. C* **77**, 032801 (2008).
- Gandolfi, S., Gezerlis, A. & Carlson, J. Neutron matter from low to high density. *Ann. Rev. Nuclear Particle Sci.* **65**, 303–328 (2015).
- Monrozeau, C., Margueron, J. & Sandulescu, N. Nuclear superfluidity and cooling time of neutron-star crust. *Phys. Rev. C* **75**, 065807 (2007).
- Yakovlev, D. G. & Pethick, C. J. Neutron star cooling. *Ann. Rev. Astron. Astrophys.* **42**, 169–210 (2004).
- Page, D., Prakash, M., Lattimer, J. M. & Steiner, A. W. Rapid cooling of the neutron star in cassiopeia a triggered by neutron superfluidity in dense matter. *Phys. Rev. Lett.* **106**, 081101 (2011).
- Ho, W. C. G., Elshamouty, K. G., Heinke, C. O. & Potekhin, A. Y. Tests of the nuclear equation of state and superfluid and superconducting gaps using the cassiopeia a neutron star. *Phys. Rev. C* **91**, 015806 (2015).
- Carlson, J., Gandolfi, S., Schmidt, K. E. & Zhang, S. Auxiliary-field quantum Monte Carlo method for strongly paired fermions. *Phys. Rev. A* **84**, 061602 (2011).
- Carlson, J. & Reddy, S. Superfluid pairing gap in strong coupling. *Phys. Rev. Lett.* **100**, 150403 (2008).
- Gandolfi, S., Schmidt, K. E. & Carlson, J. BEC-BCS crossover and universal relations in unitary Fermi gases. *Phys. Rev. A* **83**, 041601 (2011).
- Forbes, M. M., Gandolfi, S. & Gezerlis, A. Resonantly interacting fermions in a box. *Phys. Rev. Lett.* **106**, 235303 (2011).
- Ku, M. J. H., Sommer, A. T., Cheuk, L. W. & Zwierlein, M. W. Revealing the superfluid lambda transition in the universal thermodynamics of a unitary Fermi gas. *Science* **335**, 563 (2012).
- Foulkes, W. M. C., Mitas, L., Needs, R. J. & Rajagopal, G. Quantum Monte Carlo simulations of solids. *Rev. Mod. Phys.* **73**, 33–83 (2001).
- Carlson, J. et al. Quantum Monte Carlo methods for nuclear physics. *Rev. Mod. Phys.* **87**, 1067 (2015).
- Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
- Hermann, J., Schätzle, Z. & Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **12**, 891–897 (2020).
- Pfau, D., Spencer, J. S., Matthews, A. G. D. G. & Foulkes, W. M. C. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Phys. Rev. Res.* **2**, 033429 (2020).
- Pescia, G., Han, J., Lovato, A., Lu, J. & Carleo, G. Neural-network quantum states for periodic systems in continuous space. *Phys. Rev. Res.* **4**, 023138 (2022).
- Wilson, M. et al. Neural network ansatz for periodic wave functions and the homogeneous electron gas. *Phys. Rev. B* **107**, 235139 (2023).
- Cassella, G. et al. Discovering quantum phase transitions with fermionic neural networks. *Phys. Rev. Lett.* **130**, 036401 (2023).
- Fore, B. et al. Dilute neutron star matter from neural-network quantum states. *Phys. Rev. Res.* **5**, 033062 (2023).
- Keeble, J. W. T. & Rios, A. Machine learning the deuteron. *Phys. Lett. B* **809**, 135743 (2020). 1911.13092.
- Adams, C., Carleo, G., Lovato, A. & Rocco, N. Variational Monte Carlo calculations of $A \leq 4$ nuclei with an artificial neural-network correlator ansatz. *Phys. Rev. Lett.* **127**, 022502 (2021).

24. Gnech, A. et al. Nuclei with up to $A = 6$ nucleons with artificial neural network wave functions. *Few-Body Syst.* **63**, 7 (2021).
25. Lovato, A., Adams, C., Carleo, G. & Rocco, N. Hidden-nucleons neural-network quantum states for the nuclear many-body problem. *Phys. Rev. Res.* **4**, 043178 (2022).
26. Yang, Y. L. & Zhao, P. W. A consistent description of the relativistic effects and three-body interactions in atomic nuclei. *Phys. Lett. B* **835**, 137587 (2022).
27. Rigo, M., Hall, B., Hjorth-Jensen, M., Lovato, A. & Pederiva, F. Solving the nuclear pairing model with neural network quantum states. *Phys. Rev. E* **107**, 025310 (2023).
28. Keeble, J. W. T., Drissi, M., Rojo-Francàs, A., Juliá-Díaz, B. & Rios, A. Machine learning one-dimensional spinless trapped fermionic systems with neural-network quantum states. *Phys. Rev. A* **108**, 063320 (2023).
29. Luo, D. & Clark, B. K. Backflow transformations via neural networks for quantum many-body wave functions. *Phys. Rev. Lett.* **122**, 226401 (2019).
30. Moreno, J. R., Carleo, G., Georges, A. & Stokes, J. fermionic wave functions from neural-network constrained hidden states. *Proc. Natl. Acad. Sci. USA* **119**, e2122059119 (2022).
31. Carlson, J., Chang, S. Y., Pandharipande, V. R. & Schmidt, K. E. Superfluid Fermi gases with large scattering length. *Phys. Rev. Lett.* **91**, 050401 (2003).
32. Chang, S. Y., Pandharipande, V. R., Carlson, J. & Schmidt, K. E. Quantum Monte Carlo studies of superfluid Fermi gases. *Phys. Rev. A* **70**, 043602 (2004).
33. Gezerlis, A., Gandolfi, S., Schmidt, K. E. & Carlson, J. Heavy-light fermion mixtures at unitarity. *Phys. Rev. Lett.* **103**, 060403 (2009).
34. Galea, A., Dawkins, H., Gandolfi, S. & Gezerlis, A. Diffusion Monte Carlo study of strongly interacting two-dimensional Fermi gases. *Phys. Rev. A* **93**, 023602 (2016).
35. Casula, M. & Sorella, S. Geminal wave functions with Jastrow correlation: A first application to atoms. *J. Chem. Phys.* **119**, 6500–6511 (2003).
36. Casula, M., Attaccalite, C. & Sorella, S. Correlated geminal wave function for molecules: An efficient resonating valence bond approach. *J. Comput. Phys.* **121**, 7110–7126 (2004).
37. Bajdich, M., Mitas, L., Wagner, L. K. & Schmidt, K. E. Pfaffian pairing and backflow wavefunctions for electronic structure quantum Monte Carlo methods. *Phys. Rev. B* **77**, 115112 (2008).
38. Bajdich, M., Mitas, L., Drobný, G., Wagner, L. K. & Schmidt, K. E. Pfaffian pairing wave functions in electronic-structure quantum Monte Carlo simulations. *Phys. Rev. Lett.* **96**, 130201 (2006).
39. Piarulli, M. & Tews, I. Local nucleon-nucleon and three-nucleon interactions within chiral effective field theory. *Front. Phys.* **7**, 245 (2020).
40. Gandolfi, S., Illarionov, A. Y., Pederiva, F., Schmidt, K. E. & Fantoni, S. Equation of state of low-density neutron matter, and the 1S_0 pairing gap. *Phys. Rev. C* **80**, 045802 (2009).
41. Gandolfi, S., Lonardonì, D., Lovato, A. & Piarulli, M. Atomic nuclei from quantum Monte Carlo calculations with chiral EFT interactions. *Front. Phys.* **8**, 117 (2020). 2001.01374.
42. Nomura, Y., Darmawan, A. S., Yamaji, Y. & Imada, M. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. *Phys. Rev. B* **96**, 205152 (2017).
43. Nomura, Y. & Imada, M. 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).
44. Pescia, G., Nys, J., Kim, J., Lovato, A. & Carleo, G. Message-passing neural quantum states for the homogeneous electron gas. *arXiv preprint arXiv:2305.07240* (2023).
45. Morris, A. J., López Ríos, P. & Methods, R. J. Ultracold atoms at unitarity within quantum Monte Carlo methods. *Phys. Rev. A* **81**, 033619 (2010).
46. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **4**, 251–257 (1991).
47. Wimmer, M. Algorithm 923: Efficient numerical computation of the Pfaffian for dense and banded skew-symmetric matrices. *ACM Trans. Math. Softw.* **38**, 1–17 (2012).
48. Feynman, R. P. & Cohen, M. Energy spectrum of the excitations in liquid helium. *Phys. Rev.* **102**, 1189–1204 (1956).
49. Flyvbjerg, H. & Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **91**, 461–466 (1989).
50. Jonsson, M. Standard error estimation by an automated blocking method. *Phys. Rev. E* **98**, 043304 (2018).
51. Schätzle, Z., Hermann, J. & Noé, F. Convergence to the fixed-node limit in deep variational Monte Carlo. *J. Chem. Phys.* **154**, 124108 (2021).
52. Forbes, M. M., Gandolfi, S. & Gezerlis, A. Effective-range dependence of resonantly interacting fermions. *Phys. Rev. A* **86**, 053603 (2012).
53. Palkanoglou, G., Diakonou, F. K. & Gezerlis, A. From odd-even staggering to the pairing gap in neutron matter. *Phys. Rev. C* **102**, 064324 (2020).
54. Carlson, J. & Reddy, S. Asymmetric two-component fermion systems in strong coupling. *Phys. Rev. Lett.* **95**, 060401 (2005).
55. Dean, D. J. & Hjorth-Jensen, M. Pairing in nuclear systems: From neutron stars to finite nuclei. *Rev. Mod. Phys.* **75**, 607–656 (2003).
56. Zaheer, M. et al. Deep Sets. *arXiv e-prints arXiv:1703.06114* (2018).
57. Kato, T. On the eigenfunctions of many-particle systems in quantum mechanics. *Commun. Pure Appl. Math.* **10**, 151–177 (1957).
58. Hendrycks, D. & Gimpel, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415* (2016).
59. Schiavilla, R. et al. Two- and three-nucleon contact interactions and ground-state energies of light- and medium-mass nuclei. *Physical Review C* **103**, 054003 (2021).
60. Sorella, S. Wave function optimization in the variational Monte Carlo method. *Phys. Rev. B* **71**, 241103 (2005).
61. Stokes, J., Izaac, J., Killoran, N. & Carleo, G. Quantum natural gradient. *Quantum* **4**, 269 (2020).
62. Pessoa, R., Gandolfi, S., Vitiello, S. A. & Schmidt, K. E. Contact interaction in a unitary ultracold Fermi gas. *Phys. Rev. A* **92**, 063625 (2015).
63. Lou, W. T. et al. Neural wave functions for superfluids. *arXiv preprint arXiv:2305.06989* (2023).
64. Genovese, C., Shirakawa, T., Nakano, K. & Sorella, S. General correlated geminal ansatz for electronic structure calculations: Exploiting Pfaffians in place of determinants. *J. Chem. Theory Comput.* **16**, 6114–6131 (2020).

Acknowledgements

We would like to thank Shiwei Zhang for enlightening discussions. The work of JK and MHJ is supported by the U.S. National Science Foundation Grants No. PHY-1404159 and PHY-2013047. AL and BF are supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under contracts DE-AC02-06CH11357, by the 2020 DOE Early Career Award number ANL PRJ1008597, by the NUCLEI SciDAC program, and Argonne LDRD awards. The work of SG is supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under contract No. DE-AC52-06NA25396, by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Scientific Discovery through Advanced Computing (SciDAC) NUCLEI program, and by the Department of Energy Early Career Award Program. The work of GP, JN, GC is supported by the Swiss National Science Foundation under Grant No. 200021_200336, and by Microsoft Research. Computer time was provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy National Nuclear Security Administration under Contract No. 89233218CNA000001.

Author contributions

J. K. and B.F. developed the code and performed the numerical simulations. G.P., J.N., and G.C. developed the message-passing architecture. A.L. designed the Pfaffian ansatz. S.G. carried out the DMC calculations. M.H.J. provided overall supervision and guidance throughout the project. All authors contributed significantly to writing the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-024-01613-w>.

Correspondence and requests for materials should be addressed to Alessandro Lovato.

Peer review information *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© UChicago Argonne, LLC, Operator of Argonne National Laboratory 2024