

Discovering Quantum Phase Transitions with Fermionic Neural Networks

Gino Cassella^{1,*}, Halvard Sutterud¹, Sam Azadi⁴, N.D. Drummond³,

David Pfau^{2,1}, James S. Spencer², and W.M.C. Foulkes¹

¹*Dept. of Physics, Imperial College London, London SW7 2AZ, United Kingdom*

²*DeepMind, London N1C 4DJ, United Kingdom*

³*Dept. of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom and*

⁴*Dept. of Physics, University of Oxford, Oxford OX1 3PU, United Kingdom*

(Dated: February 3, 2023)

Deep neural networks have been very successful as highly accurate wave function ansätze for variational Monte Carlo calculations of molecular ground states. We present an extension of one such ansatz, FermiNet, to calculations of the ground states of periodic Hamiltonians, and study the homogeneous electron gas. FermiNet calculations of the ground-state energies of small electron gas systems are in excellent agreement with previous initiator full configuration interaction quantum Monte Carlo and diffusion Monte Carlo calculations. We investigate the spin-polarized homogeneous electron gas and demonstrate that the same neural network architecture is capable of accurately representing both the delocalized Fermi liquid state and the localized Wigner crystal state. The network converges on the translationally invariant ground state at high density and spontaneously breaks the symmetry to produce the crystalline ground state at low density, despite being given no *a priori* knowledge that a phase transition exists.

The correlated motion of electrons in condensed matter gives rise to rich emergent phenomena. Although these are governed by fundamental quantum mechanical principles known for almost a century, they remain difficult to understand and even harder to predict theoretically or computationally. One of the major themes of modern condensed matter physics is the study of phase transitions caused by electron correlation.

The difficulty of solving the Schrödinger equation scales exponentially with particle number in general, so exact solutions for interacting many-electron systems are rarely accessible. This explains why approximate numerical techniques have become such vital tools in the search for exotic zero-temperature phases, providing accurate predictions of experimentally observable quantities in phases already understood qualitatively. Most computational approaches, however, encode prior assumptions about the appropriate phase, which poses a substantial difficulty in predicting previously unknown electronic states. Changes in symmetry or topology are rarely discovered computationally before they have been seen experimentally or proposed on theoretical grounds.

In this Letter, we introduce a neural-network-based approach to predicting the qualitative nature of electronic ground states in condensed matter. We utilize a representation of the wave function, the fermionic neural network (FermiNet) [1], which is capable of representing *any* antisymmetric state [2], and requires no *a priori* knowledge of the system being studied. Guided by the quantum mechanical variational principle alone, without reference to experimental data, FermiNet can learn the ground state of a many-body interacting Hamiltonian.

We extend FermiNet, which has previously only been applied to atoms and molecules [1, 3–5], to systems subject to periodic boundary conditions. Phase transitions are seen by studying changes in the ground state as the parameters of the system are varied.

A significant body of recent work has used machine learning to detect phase transitions in simulated classical [6–8] and quantum [9–11] systems, but these studies required a source of external data, looking for patterns characteristic of different phases. Our approach requires only the Hamiltonian. There has also been work using neural network ansätze to study lattice models and spin systems, including their phase transitions [12–16], but for applications to many real systems, the wave function must be treated, as in the present work, in continuous space.

The flexibility of FermiNet hinges on the universal approximation property of neural networks [17, 18], which makes them a versatile tool for approximating high-dimensional functions and has led to radical advances in many computational fields [19–22]. This success has motivated the application of neural networks to solving problems across the physical sciences, including quantum mechanics [12, 23–25]. Several neural-network-based wave functions in both first-quantized [1, 3, 26–28] and second-quantized [29] representations have recently been used to compute the ground-state energies of molecules to a level of accuracy rivaling, or in some cases exceeding, sophisticated quantum chemistry methods such as coupled cluster [30]. FermiNet and ansätze derived from it are the most accurate of these so far, gaining an advantage over second-quantized neural and most quantum chemical approaches because they are basis-set free. This, coupled with the flexibility of the neural representation, enables the application of FermiNet to generic

* g.cassella20@imperial.ac.uk

phases of matter.

We demonstrate the flexibility of the periodic FermiNet by studying the quantum phase transition between the Fermi liquid and Wigner crystal [31] in the three-dimensional interacting homogeneous electron gas (HEG) [32]. Two-dimensional Wigner crystals were very recently imaged for the first time [33–35], but three-dimensional Wigner crystals have not yet been observed in electronic systems and are thus less well understood. The zero-temperature properties of the three-dimensional HEG depend on a single dimensionless parameter, r_s , defined as the ratio of the radius of a sphere that contains one electron on average to the Bohr radius. At high density (small r_s), the ground state is a weakly interacting Fermi liquid. At low density (large r_s), the correlations are stronger and the translational symmetry breaks spontaneously, giving rise to a spatially ordered Wigner crystal [31].

We find that the same neural network architecture learns the appropriate ground-state wave function either side of the Wigner phase transition, spontaneously breaking continuous translational symmetry when the crystal phase is stable. As we give the network no information about the nature of the ground state, the degree of inductive bias in the determination is very low.

The Hamiltonian for a finite HEG of N electrons subject to periodic boundary conditions is

$$\mathcal{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + U_{\text{Coulomb}}, \quad (1)$$

where the indices i label the N electrons in the simulation cell and U_{Coulomb} is the Coulomb energy per simulation cell of an infinite periodic lattice of identical copies of that cell. In practice, the Coulomb energy is evaluated using the Ewald method [36, 37]. We work in Hartree atomic units, where energies are measured in Hartrees (1 Ha \approx 27.211 eV) and distances in Bohr radii.

The wave function represented by a FermiNet is a sum of determinants of many-electron functions [1, 3]:

$$\Psi(\{\mathbf{x}_j\}) = \sum_k^{n_{\text{det}}} \det [\psi_i^k(\mathbf{x}_j; \{\mathbf{x}_{/j}\})], \quad (2)$$

where $\mathbf{x} = (\mathbf{r}, \alpha)$ labels the spatial and spin coordinates of an electron, and the set $\{\mathbf{x}_{/j}\}$ includes all electron coordinates except \mathbf{x}_j . The orbital $\psi_i^k(\mathbf{x}_j; \{\mathbf{x}_{/j}\})$ depends on the coordinates \mathbf{x}_j of the j -th electron, and, in a permutation-invariant fashion, on the set of all other electron coordinates. The use of many-electron orbitals makes a FermiNet determinant much more flexible than a Slater determinant of one-electron orbitals and a linear combination of a small number of FermiNet determinants has a much greater representational capacity than a linear combination of a similar number of Slater determinants [1]. The original FermiNet architecture assumed,

as is conventional in variational Monte Carlo (VMC) [40], that the determinants in Eq. (2) can be factorized into spin-up and spin-down determinants. Removing this constraint has proven to be more accurate (see Supplementary Material) and is used here.

FermiNet uses a neural network to approximate the many-electron orbitals appearing in the determinants [1]. The network consists of two parallel streams, for processing one-electron and two-electron information. The one-electron stream is constructed of repeating blocks, where each block contains a nonlinear layer and a permutation-equivariant function. The two-electron stream is a comparatively small fully-connected feed-forward network. The outputs of the one- and two-electron streams at each layer are fed into the permutation-equivariant function. The multiple outputs of the one-electron stream are fed through a final linear layer to produce the required number of many-electron functions, $\{\phi_i^{k\alpha}\}$. Finally, the network outputs are multiplied by a parameterized envelope, \mathbf{f} , to produce the many-electron orbitals $\psi_i^{k\alpha}(\mathbf{r}) = f_i^{k\alpha}(\mathbf{r})\phi_i^{k\alpha}(\mathbf{r})$. The electron position vectors \mathbf{r}_i and norms $\|\mathbf{r}_i\|$, and the electron-electron separation vectors $(\mathbf{r}_i - \mathbf{r}_j)$ and norms $\|\mathbf{r}_i - \mathbf{r}_j\|$, are supplied as inputs to the network. Full details of the network architecture are given in Ref. [1] and the Supplementary Material.

To adapt the FermiNet architecture to periodic systems, it is sufficient to modify the input features to ensure that periodic boundary conditions are satisfied. Periodic input features are most easily expressed in the basis $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ of primitive Bravais lattice vectors of the simulation cell. For an arbitrary vector $\mathbf{r} \triangleq s_1\mathbf{a}_1 + s_2\mathbf{a}_2 + s_3\mathbf{a}_3$, the periodic input features are obtained from the fractional coordinates s_i via the component-wise transformation $s_i \rightarrow (\sin(2\pi s_i), \cos(2\pi s_i))$. A periodic analogue of the Euclidean norm may be defined as

$$\|s\|_p^2 = \sum_{ij} [1 - \cos(2\pi s_i)] S_{ij} [1 - \cos(2\pi s_j)] + \sin(2\pi s_i) S_{ij} \sin(2\pi s_j), \quad (3)$$

where $S_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j$ acts as a metric tensor in the fractional coordinate system. This definition of the norm is smooth, periodic with respect to the simulation cell, and proportional to the Euclidean norm as $\mathbf{s} \rightarrow 0$. Unlike the simpler norm introduced in [41], it retains these properties for non-cubic simulation cells. Convergence speed and asymptotic convergence are improved by including an envelope of the form

$$f_i^{k\alpha}(\mathbf{r}) = \sum_m [\nu_{im}^{k\alpha} \cos(\mathbf{k}_m \cdot \mathbf{r}) + \mu_{im}^{k\alpha} \sin(\mathbf{k}_m \cdot \mathbf{r})], \quad (4)$$

for real wave functions, or

$$f_i^{k\alpha}(\mathbf{r}) = \sum_m \nu_{im}^{k\alpha} \exp(i\mathbf{k}_m \cdot \mathbf{r}), \quad (5)$$

for complex wave functions. The \mathbf{k}_m are simulation-cell reciprocal lattice vectors up to the Fermi wavevector of

TABLE I: Correlation energy of the spin unpolarized $N = 14$ HEG with simple cubic boundary conditions. The i-FCIQMC energies [38] were calculated using a basis of 778 plane-wave orbitals for $r_s = 5.0$ or 2378 plane waves otherwise, corresponding to Hilbert spaces of 10^{24} and 10^{31} Slater determinants, respectively. The extrapolation of i-FCIQMC results to the complete basis set limit may yield correlation energies that are 1–2 mHa too negative [39].

Method	Correlation energy [Hartree]			
	$r_s = 0.5$	$r_s = 1.0$	$r_s = 2.0$	$r_s = 5.0$
SJB ($n_{\text{det}} = 1$)				
VMC	−0.58624(1)	−0.5254(1)	−0.437(3)	−0.30339(2)
DMC	−0.58778(1)	−0.5254(1)	−0.4385(3)	−0.30474(8)
FermiNet				
$n_{\text{det}} = 1$	−0.58895(6)	−0.52568(3)	−0.43881(1)	−0.30468(1)
$n_{\text{det}} = 16$	−0.59094(6)	−0.52682(3)	−0.44053(1)	−0.30495(1)
i-FCIQMC[38]				
finite basis	−0.5939(4)	−0.5305(5)	−0.4430(7)	−0.304(1)
basis set limit	−0.5969(3)	−0.5325(4)	−0.4447(4)	−0.306(1)

the noninteracting electron gas, and $\nu_{im}^{k\alpha}, \mu_{im}^{k\alpha}$ are learnable parameters. Finally, when simulating the electron gas, the absence of nuclei (and hence electron-nuclear cusps) removes the need to include the norms of the electron positions as inputs. The FermiNet wave function is optimized using VMC [40]. Unless specified, all calculations used the same training procedure and hyperparameters as in Ref. [1]. Further details are given in the Supplementary Material.

Table I shows the results of FermiNet calculations of the total energy of a 14-electron simple cubic simulation cell of unpolarized HEG at four different densities. This system is sufficiently small that near-exact initiator full configuration interaction quantum Monte Carlo (i-FCIQMC) benchmarks are available [38, 42].

However, the fermion sign problem in FCIQMC increases rapidly with r_s , rendering i-FCIQMC calculations at low densities with large basis sets impractical; the calculations at $r_s = 5$ were $\sim 10^4$ times more expensive than those at $r_s = 1$ [38]. There is no clear systematic trend with the disagreement between FermiNet and i-FCIQMC; FermiNet is 4–6mH above i-FCIQMC at $r_s < 5$ and 1mH above at $r_s = 5$. The variance in the FermiNet energies decreases with r_s , however the magnitude of the total energy of the $N=14$ HEG (see Supplementary Material) also decreases sharply with r_s . As a result, it is ambiguous if the relative accuracy of FermiNet is greater at lower density. Table I also includes VMC and fixed-node diffusion Monte Carlo (DMC) results calculated using a conventional Slater-Jastrow-backflow (SJB) wave function.

Although FermiNet is a VMC method, it achieves an accuracy similar to that of SJB-DMC, with both approaches obtaining 99% of the i-FCIQMC correlation energy extrapolated to the complete basis set limit (which may be 1–2 mHa too large [39]). FermiNet obtains a similar fraction of the correlation energy for molecular systems with a comparable number of electrons [1]. Again as in molecular systems, calculations using sixteen Fer-

miNet determinants are noticeably better than calculations using one FermiNet determinant.

To assess the performance of FermiNet as the strength of the correlation increases, we study the $N=27$ electron spin-polarized HEG in the density range from $r_s = 1$ to 90 in a body-centred cubic (bcc) cell, as this minimizes the packing density and has the lowest Madelung energy [31] of the Wigner crystal in the low density limit. Prior work [43, 44] had found Wigner crystallization to occur in the interval $r_s = [100, 110]$, although a recent study [45] lowers this estimate substantially. The 27-electron system studied here is very small and there are substantial finite-system-size effects that broaden the phase transition and move it to a much higher density.

Ground-state energies obtained using VMC with a FermiNet wave function and using VMC and DMC with SJB wave functions targeted at gas and crystal states are compared in Fig. 1(a) for the 27-electron system. For $r_s \leq 1$, real-valued FermiNets often become trapped in local minima during optimization, with energies typically $\sim 0.1\%$ higher than the SJB-DMC benchmarks. Complex-valued FermiNets do not become trapped in local minima at these densities, and converge to wave functions with a uniform-in-space complex phase, indicating that a complex wave function improves optimization, rather than simply increasing representational capacity. All other results were obtained with real wave functions. The SJB ansätze used to describe gases and crystals are detailed in the Supplementary Material and Ref. [44]. FermiNet VMC calculations produce a tighter variational lower bound than both the SJB gas and crystal wave functions at all densities. Furthermore, FermiNet outperforms fixed-node DMC calculations based on a SJB gas wave function across the entire density range, even at $r_s \leq 1$. In the low-density regime, fixed-node DMC calculations using the SJB crystal wave function give slightly better results than our FermiNet VMC calculations. These results suggest that the nodal surface of the SJB crystal wave function is highly accurate but that the

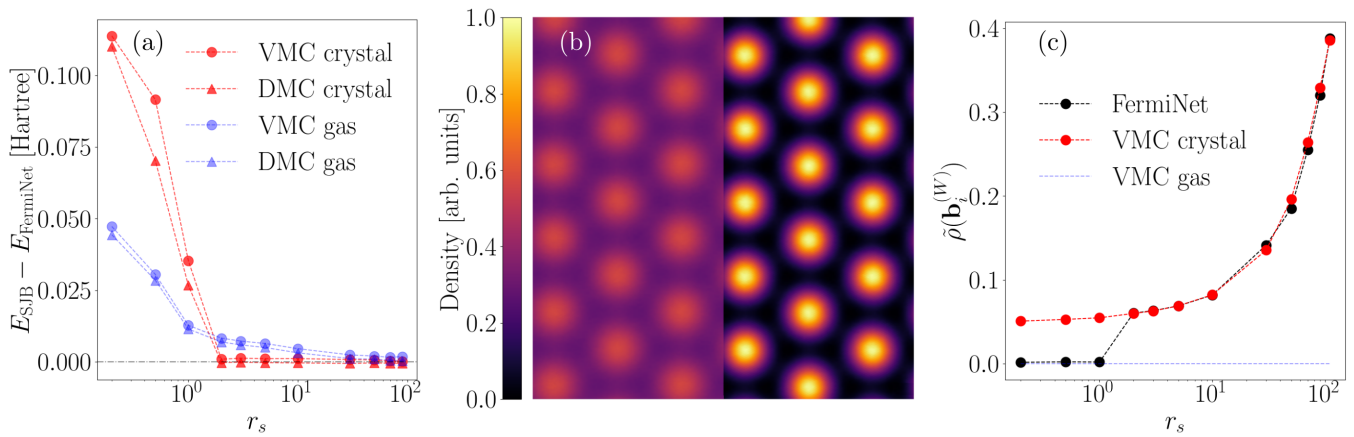


FIG. 1. The $N = 27$ spin-polarized homogeneous electron gas in a body-centered cubic (bcc) simulation cell. **(a)** Single-determinant Slater-Jastrow-backflow (SJB) ground-state total energies per electron relative to FermiNet. The “gas” and “crystal” results were obtained using SJB wave functions built using determinants of plane waves and Gaussian orbitals, respectively. Error bars are smaller than the markers. FermiNet results for $r_s \leq 1$ used complex wave functions. FermiNet-VMC yields a variational improvement over SJB-VMC and SJB-DMC in the gas phase and over SJB-VMC in the crystal phase. **(b)** One-electron density from FermiNet-VMC at $r_s = 10$ (left) and 70 (right), projected into the (011) plane of the conventional bcc structure. Four simulation cells are shown. Length scales are normalized by r_s , such that the apparent length scales are equivalent and crystal sites are superimposable. **(c)** Order parameter averaged over crystal axes for the bcc Wigner crystal state. Error bars are smaller than the markers. At small values of r_s , the order parameter is ~ 0 , corresponding to a uniform one-electron density (gas-like); the order parameter rises sharply to a finite value at $r_s = 2$, corresponding to the emergence of a crystalline state.

shape of the wave function away from the nodal surface is captured better by FermiNet. Figure 1(b) shows scans of the one-electron density of FermiNet wave functions at $r_s = 10$ and 70 in the $(\mathbf{a}_2, \mathbf{a}_3)$ plane, which is normal to the (011) direction of the conventional bcc cell. Figure 1(c) shows the Fourier component of the one-electron density, $\tilde{\rho}$ (see Supplementary Material), an order parameter appropriate for the Wigner crystal, as calculated from VMC simulations using FermiNet and SJB gas and crystal wave functions. These results show that FermiNet is capable of learning wave functions in both the gas and Wigner crystal states to very high accuracy without any hand-crafted features indicating whether the wave function should be localized or diffuse, any specific designation of crystal sites, or any other information that a transition should occur. Unlike the gas and crystal SJB trial wave functions required to describe the gaseous and crystalline states accurately, the FermiNet ansatz is identical across the entire density range.

The HEG Hamiltonian is symmetric under the simultaneous translation of all electron co-ordinates. Thus, the true ground state of the Wigner crystal is uniform in the one-electron density, with the crystal appearing only in the pair-correlation function. This is known as a “floating crystal” state [46, 47]. In Refs. [44, 45], it is shown

that the energy difference between the fixed and floating crystal is approximately $\Delta E = 0.055r_s^{-3/2}$. While FermiNet differs from Slater-type wave functions used to

derive ΔE , we expect a similar reduction in kinetic energy. At low r_s , ΔE is large (20mHa at $r_s = 2$), so we would expect FermiNet to learn the floating crystal state. Fig. 1 (b, c) show that FermiNet instead learns the fixed crystal. The notion of a fixed origin can be removed by removing the one-electron features, however we find this increases the energy obtained. This suggests that the two-electron stream is insufficiently flexible to fully describe the two-electron correlations in the Wigner crystal without help from the one-electron stream. Improving the flexibility of the two-electron stream will be the focus of future work. We do not believe that these issues impact the central conclusion of the present work, and stress that in real condensed matter systems the Hamiltonian does not possess continuous translational symmetry.

Other recent works have explored using neural networks to represent wavefunctions in real-space periodic systems. Pescia *et al.* use a DeepSets architecture to study bosons in 1D and 2D, obtaining results competitive with conventional DMC [41]. The iterative backflow network approach of Holzmann and co-workers [48–50] represents Jastrow and backflow functions in a manner equivalent to fully-connected neural networks. They observe liquid and solid phases of 2D (bosonic) ^4He [49], but the application to the HEG [50] gives a higher energy at low density than the crystal state observed by Drummond *et al.* [44]. Whilst this manuscript was in preparation, two related preprints were reported. Wilson *et al.* presented FermiNet-based and iterative back-

flow results for the 14-electron system studied here, plus 7- and 19-electron HEGs [51]. Although they used a more heavily modified version of the FermiNet architecture and include a backflow-based term in the orbitals, their results are similar to ours. Li *et al.* used a similar but smaller FermiNet architecture to ours, with simpler periodic inputs and no envelope function, to study the $N=54$ HEG along with other atomic solids [52], but do not outperform their SJB VMC or DMC baseline energies. Neither of these two preprints address Wigner crystallization. We expect that other FermiNet-derived models are capable of discovering phase transitions, as in the present work, however we are unsure if this applies to the heavily modified ansatz of Wilson *et al.* which introduces a higher degree of inductive bias via the inclusion of Hartree-Fock orbitals in the determinant. Applying any of the aforementioned approaches to fermionic systems, including FermiNet, incurs the same $\mathcal{O}(N^3)$ scaling of determinant evaluation. An additional factor of N is introduced if analytic gradients are not available and the Laplacian must be evaluated via automatic differentiation.

To summarize, we have extended the FermiNet neural wave function to calculations with periodic boundary conditions. This we accomplished by making minimal, physically-motivated, modifications to render the input features periodic, and by adding a periodic envelope function. As proof of concept, we have demonstrated the accuracy of the modified architecture on the $N=14$ HEG, where we obtained $\sim 99\%$ of the correlation energy and slightly outperformed VMC and DMC calculations using conventional one-determinant SJB trial wave functions. For the $N=27$ HEG, we see that FermiNet is capable of learning the localized Wigner crystal phase *a priori*, producing energies in excellent agreement with SJB trial wave functions which encode the qualitative nature of the ground state in their construction. This suggests that FermiNet may be capable of determining novel quantum phases in condensed matter given only the Hamiltonian.

To study quantum phase transitions in realistic strongly-correlated electronic systems, it will be necessary to scale to larger numbers of electrons to overcome finite-size effects. However, the scaling required to achieve a fixed accuracy per electron is not yet known. Achieving an acceptable accuracy at an affordable scaling may require additional innovations in the neural network architecture employed. FermiNet could also be used as a trial wave function for DMC calculations in periodic boundary conditions, an approach that yields small improvements in molecular systems [4]. More generally, we believe that the flexibility and accuracy offered by neural networks make them promising tools for studying complex correlation effects and other emergent phenomena. The advantages of neural-network-based methods are most compelling when the phenomena in question are unexpected or not yet understood.

This work was undertaken with funding from the UK Engineering and Physical Sciences Research Council (EP/T51780X/1) (GC) and the Aker scholarship (HS). Calculations were carried out with resources provided by the Baskerville Accelerated Compute Facility through a UK Research and Innovation Access to HPC grant. Via his membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/R029431), Foulkes used the UK Materials and Molecular Modelling Hub for computational resources, MMM Hub, which is partially funded by EPSRC (EP/T022213). Our SJB-DMC Wigner crystal calculations were performed using Lancaster University's High-End Computing cluster, and the ARCHER2 UK National Supercomputing Service (<http://www.archer2.ac.uk>) via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/R029431).

-
- [1] D. Pfau, J. S. Spencer, A. G. D. G. Matthews, and W. M. C. Foulkes, *Phys. Rev. Research* **2**, 033429 (2020).
 - [2] M. Hutter, arXiv:2007.15298 (2020).
 - [3] J. S. Spencer, D. Pfau, A. Botev, and W. M. C. Foulkes, arXiv:2011.07125 (2020).
 - [4] M. Wilson, N. Gao, F. Wudarski, E. Rieffel, and N. M. Tubman, arXiv:2103.12570 (2021).
 - [5] X. Li, C. Fan, W. Ren, and J. Chen, *Phys. Rev. Research* **4**, 013021 (2022).
 - [6] L. Wang, *Phys. Rev. B* **94**, 195105 (2016).
 - [7] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
 - [8] E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Nat. Phys.* **13**, 435 (2017).
 - [9] S. Arai, M. Ohzeki, and K. Tanaka, *J. Phys. Soc. Jpn.* **87**, 033001 (2018).
 - [10] J. Venderley, V. Khemani, and E.-A. Kim, *Phys. Rev. Lett.* **120**, 257204 (2018).
 - [11] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. S. Davis, E. Khatami, and E.-A. Kim, *Nature* **570**, 484 (2019).
 - [12] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
 - [13] H. Saito, *J. Phys. Soc. Jpn.* **86**, 093001 (2017).
 - [14] D. Luo and B. K. Clark, *Phys. Rev. Lett.* **122**, 226401 (2019).
 - [15] J. Stokes, J. R. Moreno, E. A. Pnevmatikakis, and G. Carleo, *Phys. Rev. B* **102**, 205122 (2020).
 - [16] N. Astrakhantsev, T. Westerhout, A. Tiwari, K. Choo, A. Chen, M. H. Fischer, G. Carleo, and T. Neupert, *Phys. Rev. X* **11**, 041021 (2021).
 - [17] G. Cybenko, *Math. Control Signal Systems* **2**, 303 (1989).
 - [18] K. Hornik, *Neural Networks* **4**, 251 (1991).
 - [19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, *NeurIPS* **25** (2012).
 - [20] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017).

- [21] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, *Nature* **529**, 484 (2016).
- [22] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, *Nature* **596**, 583 (2021).
- [23] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, *Nat. Phys.* **14**, 447 (2018).
- [24] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, *Nat. Phys.* **15**, 887 (2019).
- [25] K. T. Schütt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda, and K.-R. Müller, *Lecture Notes in Phys.* (2020).
- [26] J. Hermann, Z. Schätzle, and F. Noé, *Nat. Chem.* **12**, 891 (2020).
- [27] M. Scherbela, R. Reisenhofer, L. Gerard, P. Marquetand, and P. Grohs, arXiv:2105.08351 (2021).
- [28] N. Gao and S. Günnemann, arXiv:2110.05064 (2021).
- [29] K. Choo, A. Mezzacapo, and G. Carleo, *Nat. Comm.* **11**, 1 (2020).
- [30] I. Shavitt and R. J. Bartlett, *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*, Cambridge Molecular Science (Cambridge University Press, 2009).
- [31] E. Wigner, *Phys. Rev.* **46**, 1002 (1934).
- [32] G. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, 2005).
- [33] Y. Zhou, J. Sung, E. Brutschea, I. Esterlis, Y. Wang, G. Scuri, R. J. Gelly, H. Heo, T. Taniguchi, K. Watanabe, *et al.*, *Nature* **595**, 48 (2021).
- [34] T. Smoleński, P. E. Dolgirev, C. Kuhlenkamp, A. Popert, Y. Shimazaki, P. Back, X. Lu, M. Kroner, K. Watanabe, T. Taniguchi, *et al.*, *Nature* **595**, 53 (2021).
- [35] H. Li, S. Li, E. C. Regan, D. Wang, W. Zhao, S. Kahn, K. Yumigeta, M. Blei, T. Taniguchi, K. Watanabe, *et al.*, *Nature* **597**, 650 (2021).
- [36] P. Ewald, *Ann. Phys.* **369**, 253 (1921).
- [37] L. M. Fraser, W. M. C. Foulkes, G. Rajagopal, R. J. Needs, S. D. Kenny, and A. J. Williamson, *Phys. Rev. B* **53**, 1814 (1996).
- [38] J. J. Shepherd, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **136**, 244101 (2012).
- [39] V. A. Neufeld and A. J. W. Thom, *J. Chem. Phys.* **147**, 194105 (2017).
- [40] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, *Rev. Modern Phys.* **73**, 33 (2001).
- [41] G. Pescia, J. Han, A. Lovato, J. Lu, and G. Carleo, arXiv:2112.11957 (2021).
- [42] D. Cleland, G. H. Booth, and A. Alavi, *J. Chem. Phys.* **134**, 024112 (2011).
- [43] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- [44] N. D. Drummond, Z. Radnai, J. R. Trail, M. D. Towler, and R. J. Needs, *Phys. Rev. B* **69**, 085116 (2004).
- [45] S. Azadi and N. D. Drummond, *Phys. Rev. B* **105**, 245135 (2022).
- [46] R. F. Bishop and K. H. Lührmann, *Phys. Rev. B* **26**, 5523 (1982).
- [47] M. Lewin, E. H. Lieb, and R. Seiringer, *Phys. Rev. B* **100**, 035127 (2019).
- [48] M. Taddei, M. Ruggeri, S. Moroni, and M. Holzmann, *Phys. Rev. B* **91**, 115106 (2015).
- [49] M. Ruggeri, S. Moroni, and M. Holzmann, *Phys. Rev. Lett.* **120**, 205302 (2018).
- [50] M. Holzmann and S. Moroni, *Phys. Rev. Lett.* **124**, 206404 (2020).
- [51] M. Wilson, S. Moroni, M. Holzmann, N. Gao, F. Wudarski, T. Vegge, and A. Bhowmik, arXiv:2202.04622 (2022).
- [52] X. Li, Z. Li, and J. Chen, arXiv:2203.15472 (2022).
- [53] J. Martens and R. Grosse, *Proceedings of the 32nd International Conference on Machine Learning*, PMLR **37**, 2408 (2015).
- [54] S.-I. Amari, *Neural Comp.* **10**, 251 (1998).
- [55] S. Sorella, *Phys. Rev. B* **71**, 241103(R) (2005).
- [56] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, *Phys. Rev. B* **96**, 205152 (2017).
- [57] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, *Quantum* **4**, 269 (2020).
- [58] G. Rajagopal, R. J. Needs, A. James, S. D. Kenny, and W. M. C. Foulkes, *Phys. Rev. B* **51**, 10591 (1995).
- [59] T. Kato, *Communications on Pure and Applied Mathematics* **10**, 151 (1957).
- [60] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: composable transformations of Python+NumPy programs, <http://github.com/google/jax> (2018).
- [61] D. Pfau and J. Spencer, Ferminet JAX implementation, <http://github.com/deepmind/ferminet> (2020).
- [62] A. Botev, KFAC JAX implementation, https://github.com/deepmind/deepmind-research/tree/master/kfac_ferminet_alpha (2020).
- [63] H. Flyvbjerg and H. G. Petersen, *J. Chem. Phys.* **91**, 461 (1989).
- [64] R. J. Needs, M. D. Towler, N. D. Drummond, and P. L. Ríos, *Journal of Physics: Condensed Matter* **22**, 023201 (2010).
- [65] R. J. Needs, M. D. Towler, N. D. Drummond, P. López Ríos, and J. R. Trail, *J. Chem. Phys.* **152**, 154106 (2020).

SUPPLEMENTARY

FERMIONIC NEURAL NETWORKS

The FermiNet architecture maps a set of input features derived from the electron coordinates, $\{\mathbf{r}_j^\alpha\}$, where α labels the spin of the electron, to the set of functions $\psi_i^{k\alpha}(\mathbf{r}_j; \{\mathbf{r}/_j\})$. In the original FermiNet, this set of inputs to each layer of the network is

$$\mathbf{j}_i^{l\alpha} = \left(\mathbf{h}_i^\alpha, \frac{1}{n^\uparrow} \sum_{j=1}^{n^\uparrow} \mathbf{h}_j^{l\uparrow}, \frac{1}{n^\downarrow} \sum_{j=1}^{n^\downarrow} \mathbf{h}_j^{l\downarrow}, \frac{1}{n^\uparrow} \sum_{j=1}^{n^\uparrow} \mathbf{h}_{ij}^{l\alpha\uparrow}, \frac{1}{n^\downarrow} \sum_{j=1}^{n^\downarrow} \mathbf{h}_{ij}^{l\alpha\downarrow} \right) \quad (6)$$

where

$$\mathbf{h}_i^{1\alpha} = (\mathbf{r}_i^\alpha - \mathbf{r}_I, \|\mathbf{r}_i^\alpha - \mathbf{r}_I\| \quad \forall \quad I) \quad (7)$$

$$\mathbf{h}_{ij}^{1\alpha\beta} = \left(\mathbf{r}_i^\alpha - \mathbf{r}_j^\beta, \left\| \mathbf{r}_i^\alpha - \mathbf{r}_j^\beta \right\| \right), \quad (8)$$

with capitalized subscripts referring to atomic coordinates, and $\|\cdot\|$ the Euclidean norm. These input features are updated by consecutive transformations,

$$\mathbf{h}_i^{l+1\alpha} = \tanh \left(\underline{\mathbf{V}}^l \mathbf{j}_i^{l\alpha} + \mathbf{b}^l \right) + \mathbf{h}_i^{l\alpha} \quad (9)$$

$$\mathbf{h}_{ij}^{l+1\alpha\beta} = \tanh \left(\underline{\mathbf{W}}^l \mathbf{h}_{ij}^{l\alpha\beta} + \mathbf{c}^l \right) + \mathbf{h}_{ij}^{l\alpha\beta}. \quad (10)$$

The first transformation (one subscripted index) is referred to as the one-electron stream, and the second (two subscripted indices) the two-electron stream. The outputs from the L^{th} transformation are subject to a final, spin-dependent, linear transformation and multiplied by (in open boundary conditions) an exponentially decaying envelope [3] which enforces the decay of the wave function as $\mathbf{r}_i \rightarrow \infty$,

$$\psi_i^{k\alpha}(\mathbf{r}_j, \{\mathbf{r}/_j\}) = \phi_i^{k\alpha}(\mathbf{r}_j, \{\mathbf{r}/_j\}) f_i^{k\alpha}(\mathbf{r}_j) \quad (11)$$

where

$$\phi_i^{k\alpha}(\mathbf{r}_j, \{\mathbf{r}/_j\}) = (\mathbf{w}_i^{k\alpha} \cdot \mathbf{h}_j^{L\alpha} + g_i^{k\alpha}) \quad (12)$$

and

$$f_i^{k\alpha}(\mathbf{r}_j) = \left[\sum_m \pi_{im}^{k\alpha} \exp(-\sigma_{im}^{k\alpha} |(\mathbf{r}_j^\alpha - \mathbf{R}_m)|) \right]. \quad (13)$$

The functions $\psi_i^{k\alpha}$ are used as the inputs to the determinants, Eq. (2), in the main text. The determinant expansion in Eq. (2) does not require multiplicative coefficients as they can be trivially absorbed into the determinants[3]. Hutter showed that a single determinant of such many-body functions can represent any antisymmetric function[2], though the proof depends upon the construction of discontinuous functions that cannot be represented in practice by a finite network of

a reasonable size. Note that the pooling operations in Eq. (6) are chosen such that the feature vector $\mathbf{j}_i^{l\sigma}$ is only permutation-invariant with respect to the exchange of electrons of the same spin. Thus, the desired fermionic exchange statistics are enforced even with the dense determinant form, discussed below.

To construct complex wave functions, the number of functions $\phi_i^{k\alpha}$ output from the network is doubled, and the inputs to the determinant become

$$\psi_i^{k\alpha} = (\phi_{2i}^{k\alpha} + i\phi_{2i+1}^{k\alpha}) f_i^{k\alpha} \quad (14)$$

The set of parameters,

$$\theta = \{ \underline{\mathbf{V}}^l, \underline{\mathbf{W}}^l, \mathbf{w}_i^{k\alpha}, \mathbf{b}^l, \mathbf{c}^l, g_i^{k\alpha}, \pi_{im}^{k\alpha}, \sigma_{im}^{k\alpha} \}, \quad (15)$$

are all learnable. Pretraining these parameters to minimize the deviation between FermiNet orbitals and Hartree-Fock orbitals is possible, but we find that it is often unnecessary to achieve a well converged result and it is not performed in any calculations in the current study. The linear transformations specified by $\underline{\mathbf{V}}^l$ and $\underline{\mathbf{W}}^l$ are known as hidden layers. For a more extensive description of the FermiNet architecture, see Pfau *et al.* [1].

FermiNets are trained via the variational Monte Carlo (VMC) method: the parameters of the network are adjusted to lower the expectation value of the energy, which is calculated using Metropolis-Hastings Monte Carlo integration over the $3N$ -dimensional space of electron positions. A detailed description of the VMC approach is provided by Foulkes *et al.* [40]. The parameters θ are optimized via gradient descent to minimize $\langle \mathcal{H} \rangle$. This guides the wave function Ψ_θ toward the ground state as a result of the variational principle. Working in log-space, the gradient of $\langle \mathcal{H} \rangle$ with respect to the parameters θ is

$$\nabla_\theta \langle \mathcal{H} \rangle = \langle E_L \nabla_\theta \log \Psi^* + E_L^* \nabla_\theta \log \Psi - 2 \langle E_L \rangle \nabla_\theta \log |\Psi| \rangle, \quad (16)$$

where $E_L(\mathbf{r}) = \Psi^{-1}(\mathbf{r}) \mathcal{H} \Psi(\mathbf{r})$ and the expectation value is evaluated for samples of \mathbf{r} taken from the probability amplitude $|\Psi(\mathbf{r})|^2$. The kinetic component of the local energy is calculated in log-space via,

$$T_L(\mathbf{r}) = -\frac{1}{2} \sum_i \left[\left. \frac{\partial^2 \log \Psi}{\partial r_i^2} \right|_{\mathbf{r}} + \left(\left. \frac{\partial \log \Psi}{\partial r_i} \right|_{\mathbf{r}} \right)^2 \right]. \quad (17)$$

Gradients of the wavefunction are obtained using standard back-propagation techniques. We employ the Kronecker-factored approximate curvature algorithm (KFAC) [53] which approximates the Fisher information matrix for large neural networks, and enables scalable natural gradient descent [54]. Natural gradient descent is equivalent, up to a normalization constant, to the stochastic reconfiguration method [55] frequently used in VMC [1, 56]. For complex wave functions, the inverse Fisher matrix is not strictly the appropriate choice of

metric for natural gradient descent, and one should instead use the Fubini-Study metric [57]. In the present work, we found that natural gradient using the usual inverse Fisher was capable of finding the ground state, possibly because the Hamiltonian is real-valued. A full treatment of the optimization of complex wave functions is beyond the scope of the present work. Unlike in conventional VMC optimization, correlated sampling is not used due to the small batch sizes, frequent parameter updates, and implementation details of K-FAC.

PERIODIC BOUNDARY CONDITIONS

The ground state wave function of an interacting system possesses a macroscopically large number of degrees of freedom n , due to the many-body interactions between all of the charges in the system. Solving for the many-body wave function $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_n)$ in \mathbb{R}^{3n} is intractable for n approaching the—effectively infinite on a computational scale—number of electrons found in a real solid.

To approximate real solids by simulating a small number of electrons we employ periodic boundary conditions: a finite-sized simulation cell is embedded in a periodic array of images of all charges in the simulation cell. The resulting Hamiltonian possesses discrete translational symmetry: displacing any charge by a simulation cell lattice vector leaves the system invariant. The many-body eigenfunctions then have the property [58]

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_n) = \Psi(\mathbf{r}_1, \dots, \mathbf{r}_i + \mathbf{R}_S, \dots, \mathbf{r}_n), \quad (18)$$

where \mathbf{R}_S is a simulation cell lattice vector. As a result, the problem of finding eigenfunctions on \mathbb{R}^{3n} for extremely large n has been reduced to a problem of finding eigenfunctions on the torus \mathbb{T}^{3n} where n is a small number. The errors arising due to this approximation are known as finite-size effects. A full treatment of finite-size effects are beyond the scope of the present work, and do not alter the conclusions of the comparisons presented as all systems being compared are utilizing the same finite-size Hamiltonian.

FERMINET WITH PERIODIC BOUNDARY CONDITIONS

To impose the constraint (Eq. (18)) on the FermiNet with it is sufficient to choose an alternative set of input features to the first layer of the FermiNet which are invariant under the translation of any one electron coordinate by a simulation cell lattice vector. In the following sections we describe modifications to the coordinate and distance features which fulfill this requirement.

Fractional coordinates

Any vector in real space can be expressed as a linear combination of primitive simulation cell lattice vectors $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$,

$$\mathbf{v} = s_1 \mathbf{a}_1 + s_2 \mathbf{a}_2 + s_3 \mathbf{a}_3, \quad s_1, s_2, s_3 \in \mathbb{R}, \quad (19)$$

defining $\mathbf{s} = (s_1, s_2, s_3) \in \mathbb{R}^3$, which is an equally valid representation of a position on the lattice that we will refer to as fractional coordinates. There is a one-to-one mapping between positions in real space and positions in fractional coordinates,

$$\mathbf{s} = \underline{\underline{\mathbf{A}}}^{-1} \mathbf{v} \quad (20)$$

where,

$$\underline{\underline{\mathbf{A}}} = \begin{pmatrix} | & | & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \\ | & | & | \end{pmatrix}, \quad (21)$$

is a matrix whose columns consist of the primitive simulation cell lattice vectors. The simulation cell is a parallelepiped in real space but a unit cube in fractional coordinates. As a result, it is simple to construct maps that are periodic under translations by simulation cell lattice vectors using trigonometric functions.

All vectors \mathbf{v} in the original set of input features [Eqs. (7) and (8)] are replaced via the component-wise mapping

$$v_n \rightarrow (\sin(2\pi s_n), \cos(2\pi s_n)). \quad (22)$$

An additional subtlety is introduced by the fact that any *continuous*, unique labeling of points on a unit circle requires two numbers. This necessitates the use of both sine and cosine input features for each spatial dimension by recognizing that the torus \mathbb{T}^{3n} decomposes into a product of unit circles, $\mathbb{S}^1 \times \mathbb{S}^1 \times \mathbb{S}^1$. Displacing \mathbf{v} by a simulation cell lattice vector leaves the value of the right-hand side of (Eq. (22)) invariant as desired.

Periodic norm

The periodic analogue of $\|\mathbf{v}\|$ must retain a cusp as $\mathbf{v} \rightarrow \mathbf{0}$, resembling the Euclidean norm. The inclusion of the Euclidean norm features is known to have a substantial impact on the accuracy of the FermiNet [1]: the network is incapable of introducing discontinuities into the wave function, and thus cannot satisfy the Kato cusp conditions [59] on the derivatives of the wave function as electrons approach nuclei and each other without cusps being included explicitly in the input features. Similarly, the periodic norm must be continuous everywhere except at the cusps, as the network will be unable to remove these discontinuities from the wave function, resulting in an unphysical contribution to the kinetic energy. In

summary, we require a function of \mathbf{s} which behaves like $|\mathbf{v}| = \sqrt{v_x^2 + v_y^2 + v_z^2}$ as $\mathbf{v} \rightarrow \mathbf{R}$, is periodic on the domain $[0, 1]^3$, and whose derivative vanishes at the simulation cell boundaries to ensure continuity.

We proceed by considering the definition of the Euclidean norm as the Euclidean inner product of a vector with itself,

$$\|\mathbf{v}\|^2 = \mathbf{v} \cdot \mathbf{v} = (\underline{\mathbf{A}}\mathbf{s})^T (\underline{\mathbf{A}}\mathbf{s}) = \sum_{ij} s_i S_{ij} s_j, \quad (23)$$

where

$$S_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j. \quad (24)$$

We conjecture by analogy that the norm in terms of the periodic co-ordinates (Eq. (22)) should be

$$\begin{aligned} \|s\|_p^2 = \sum_{ij} [1 - \cos(2\pi s_i)] S_{ij} [1 - \cos(2\pi s_j)] \\ + \sin(2\pi s_i) S_{ij} \sin(2\pi s_j), \end{aligned} \quad (25)$$

This definition of the norm possesses all of the properties that we desired: as $s_i \rightarrow 0$ this expression reduces to the Euclidean norm [Eq. (23)] by considering the first-order Taylor expansions of sine and cosine; the periodicity is obvious as the expression is invariant to translations $s_i \rightarrow s_i \pm 1$; and, as a result of reducing to the Euclidean norm at the origin, this function retains the desired cusps, while also being differentiable in the rest of the unit cell. All distances, $\|\cdot\|$, in the original set of input features [Eqs. (7) and (8)] are replaced by the periodic norm.

Periodic multiplicative envelope

For the periodic envelope function introduced in Eqs. (4) and (5), ν_{im}^k and μ_{im}^k are strictly positive learnable parameters, optimized during training. The envelope eases the representation of highly oscillatory functions, while still being trivially capable of representing any function representable by the network with no envelope because $\mathbf{k}_0 = \mathbf{0}$. All ν_{im}^k and μ_{im}^k are initialized to small random values except $\nu_{i0}^k = 1$. Other initialization schemes may be more appropriate, but have not been studied here. Fig. 2 demonstrates the improved training performance due to the envelope.

ONE-ELECTRON DENSITY OPERATOR

The emergent localization of the wave function due to Wigner crystallization can be seen by accumulating the expectation value of the one-electron density operator,

$$\rho(\mathbf{r}) = \left\langle \frac{1}{N} \sum_i \delta(\mathbf{r}_i - \mathbf{r}) \right\rangle, \quad (26)$$

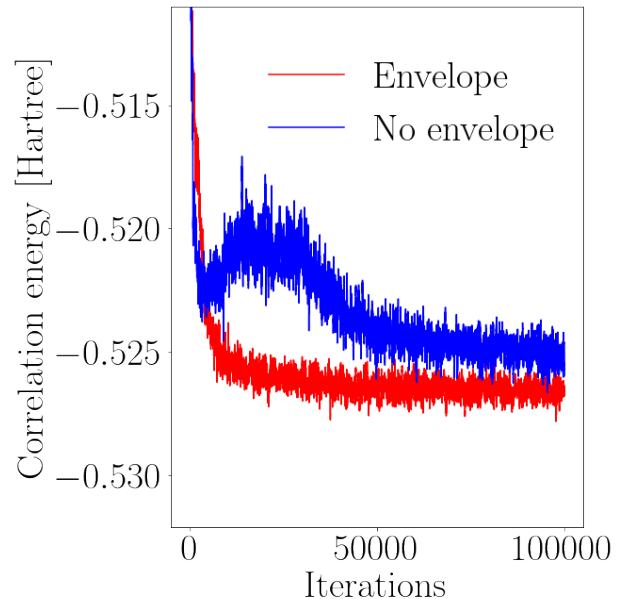


FIG. 2. Learning curves, given in terms of the correlation energy, for the $N = 14$ spin unpolarized HEG at $r_s = 1.0$ utilizing 16 dense determinants, with and without a sinusoidal envelope, using identical training parameters. An $N = 500$ moving average filter has been applied to both curves to improve visual clarity.

where the expectation value is taken over samples of one-electron coordinates. An order parameter for the broken-symmetry state is the Fourier component of $\rho(\mathbf{r})$ corresponding to any primitive reciprocal lattice vector, \mathbf{b}_i^W , of the emergent crystal:

$$\tilde{\rho}(\mathbf{b}_i^W) = \left\langle \frac{1}{N} \sum_j \exp(i\mathbf{b}_i^W \cdot \mathbf{r}_j) \right\rangle. \quad (27)$$

A state with $\tilde{\rho} = 0$ is gas-like, and a state with $\tilde{\rho} \neq 0$ is crystalline. If the simulation cell is bcc in shape and contains $N = M^3$ (spin-polarized) electrons at low enough density, it contains an $M \times M \times M$ Wigner lattice and $\mathbf{b}_i^W = M\mathbf{b}_i$, where the \mathbf{b}_i are the primitive reciprocal lattice vectors corresponding to the simulation cell.

EXPERIMENTAL SETUP

FermiNet calculations

Four A100 GPUs were used for all calculations presented. Calculations were carried out using single precision floating point numbers, as we found statistically identical results were achieved using double precision at the cost of approximately doubling runtime. The modifications to FermiNet were implemented using the JAX

TABLE II: FermiNet hyperparameters for all experiments in the paper. These are mostly the same parameters as used in the original FermiNet paper, except we omit pretraining and slightly increase the number of training iterations.

Kind	Parameter	Value
Optim	Batch size	4096
Optim	Training iterations	3e5
Optim	Pretraining iterations	0
Optim	Learning rate	$(1e4 + t)^{-1}$
Optim	Local energy clipping	5.0
KFAC	Momentum	0
KFAC	Covariance moving average decay	0.95
KFAC	Norm constraint	1e-3
KFAC	Damping	1e-3
MCMC	Proposal std. dev. (per dimension)	0.02
MCMC	Steps between parameter updates	10

Python library [60], extending a development version of the FermiNet [61]. Optimization used a JAX implementation of the Kronecker-factored approximate curvature (KFAC) gradient descent algorithm [3, 53, 62]. In the $N=14$ electron gas, a FermiNet with 4 layers of 256 units in the one-electron stream and 32 units in the two-electron stream was used for the 1 and 16 determinant calculations. In all calculations for the $N=27$ HEG, we used a FermiNet of four layers with 512/64 units in the one/two-electron streams respectively with 16 determinants. For both systems, the wave function was optimized over 3e5 training iterations and 5e4 additional samples of $\langle \mathcal{H} \rangle$ with the wave function parameters frozen were taken to obtain the final energies. The standard error associated with these energies was evaluated using a reblocking method [63] to account for sequential correlations introduced by the Monte Carlo sampling strategy.

For the majority of calculations, we employ a set of network and training hyperparameters identical to those previously used to obtain results on molecular systems [1]. For the $N=27$ HEG at $r_s = 2, 3$, and 5, we find that a more aggressive initial learning rate (base value 1e-2 versus the default 1e-4) is required to reliably avoid local minima. This is accompanied by an increased KFAC norm constraint and damping of 1.0 and 1e-1, respectively.

Slater-Jastrow-backflow calculations

The Slater-Jastrow-backflow (SJB) wave function ansatz used to produce the benchmark VMC and DMC calculations in the main text takes the form of a Slater determinant multiplied by a Jastrow factor

$$\Psi(\mathbf{r}) = \exp(J(\mathbf{r})) \det [\phi_i(\mathbf{x}(\mathbf{r}_j))]. \quad (28)$$

The Slater determinant is composed of one-electron functions and enforces the fermionic antisymmetry of the

wave function, just as in the FermiNet. This determinant is evaluated at coordinates which are modified by a backflow transformation,

$$\mathbf{x}(\mathbf{r}_j) = \mathbf{r}_j + \boldsymbol{\xi}(\mathbf{r}_j), \quad (29)$$

and multiplied by a Jastrow factor $\exp(J)$ which is a permutation-invariant function of the electronic coordinates. Here we will only provide a brief overview of the terms incorporated into these factors. A much more detailed account is provided in Ref. [45]. All SJB VMC and DMC calculations were performed using the CASINO program [64, 65]. The parameters of the SJB VMC wave functions were optimized using variance minimization and then energy minimization. The DMC results were extrapolated to zero time step.

For the Fermi fluid, the one-electron orbitals in the Slater determinant are the Hartree-Fock orbitals for the homogeneous electron gas,

$$\phi_{\mathbf{k}}(\mathbf{r}_j) = \exp(i\mathbf{k} \cdot \mathbf{r}_j), \quad (30)$$

where the $\{\mathbf{k}\}$ are the $N/2$ (spin unpolarized) or N (spin polarized) smallest simulation cell reciprocal lattice vectors. In the crystal, periodic one-electron orbitals are evaluated as sums over periodic images of site-centered Gaussian functions,

$$\phi_{\mathbf{R}_P}(\mathbf{r}_j) = \sum_{\mathbf{R}_S} \exp\left(-C|\mathbf{r}_j - \mathbf{R}_P - \mathbf{R}_S|^2\right), \quad (31)$$

where \mathbf{R}_P is a primitive-cell lattice point within the simulation cell, \mathbf{R}_S is a simulation cell lattice point, and C is an optimizable parameter controlling the width of the Gaussian. This sum is truncated when the contributions of the images of the Gaussian basis functions become smaller than 10^{-7} at the edge of the simulation cell in which the orbital is being evaluated. There are N primitive-cell lattice points within the simulation cell.

The Jastrow exponent consists of a sum of three terms,

$$J(\mathbf{r}) = \sum_{i < j}^N [u(r_{ij}) + p(\mathbf{r}_{ij})] + \sum_i^N q(\mathbf{r}_i), \quad (32)$$

where u is a power series in the electronic separations which includes fixed terms to impose the Kato cusp conditions [59]. This term is smoothly cut off at a radius less than or equal to the radius of the largest sphere that can be inscribed in the Wigner-Seitz cell of the simulation cell. The p term is

$$p(\mathbf{r}_{ij}) = \sum_A a_A \sum_{\mathbf{G} \in A^+} \cos(\mathbf{G} \cdot \mathbf{r}_{ij}), \quad (33)$$

where A consists of shells of simulation cell reciprocal lattice vectors, and A^+ excludes one from each pair of

vectors which are related by inversion symmetry. Similarly,

$$q(\mathbf{r}_i) = \sum_B b_B \sum_{\mathbf{G} \in B^+} \cos(\mathbf{G} \cdot \mathbf{r}_i), \quad (34)$$

where B consists of shells of Wigner crystal primitive cell reciprocal lattice vectors. The q term is omitted from the Fermi fluid wave function, as it does not retain continuous translational invariance with respect to the electronic center of mass.

The backflow transformation consists of two terms,

$$\xi(\mathbf{r}_i) = \sum_{j \neq i}^N \eta(r_{ij}) \mathbf{r}_{ij} + \sum_{j \neq i}^N \boldsymbol{\pi}(\mathbf{r}_{ij}), \quad (35)$$

where η is mathematically identical to the Jastrow u term, and $\boldsymbol{\pi}$ has the form of the gradient of the Jastrow p term:

$$\boldsymbol{\pi}(\mathbf{r}_{ij}) = - \sum_A c_A \sum_{\mathbf{G} \in A^+} \sin(\mathbf{G} \cdot \mathbf{r}_{ij}) \mathbf{G}. \quad (36)$$

The coefficients a_A , b_B , and c_A are all optimizable parameters.

All of these terms are evaluated using a minimum image convention.

BLOCK DETERMINANTS

We use, as is standard in VMC, spin-assigned wavefunctions[40], where we replace $\Psi(\mathbf{r}_1, \alpha_1; \dots, \mathbf{r}_N, \alpha_N)$ with a function of position alone: $\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \triangleq \Psi(\mathbf{r}_1, \uparrow; \dots, \mathbf{r}_{N_\uparrow}, \uparrow; \mathbf{r}_{N_\uparrow+1}, \downarrow; \dots, \mathbf{r}_{N_\downarrow}, \downarrow)$, where $N = N_\uparrow + N_\downarrow$ is the number of electrons and $N_\uparrow - N_\downarrow = 2S_z$ is the spin polarization. The spin-assigned wave function is only antisymmetric on interchange of the position coordinates of electrons of the same spin, but assigning the spins has no effect on expectation values of spin-independent operators. This enables a Slater determinant to be factorized into a product of spin-up and a spin-down determinants:

$$\det[\boldsymbol{\psi}] = \det[\boldsymbol{\psi}^\uparrow] \det[\boldsymbol{\psi}^\downarrow] \quad (37)$$

The original FermiNet architecture similarly assumed that the determinants were factorised into spin components. Relabeling the electron positions according to the assigned spins, this amounts to considering determinants to be in a block-diagonal form:

$$\det[\boldsymbol{\psi}] = \begin{vmatrix} \psi_i^\uparrow(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}^\uparrow\}, \{\mathbf{r}^\downarrow\}) & \mathbf{0} \\ \mathbf{0} & \psi_i^\downarrow(\mathbf{r}_j^\downarrow; \{\mathbf{r}^\uparrow\}, \{\mathbf{r}_{/j}^\downarrow\}) \end{vmatrix}, \quad (38)$$

where the diagonal blocks are of dimension $N_\uparrow \times N_\uparrow$ and $N_\downarrow \times N_\downarrow$.

TABLE III: Correlation energy of the spin unpolarized $N = 14$ HEG with simple cubic boundary conditions, compared between a FermiNet wave function constructed from block diagonal and dense determinants.

n_{det}	Correlation energy [Ha]			
	$r_s = 0.5$	$r_s = 1.0$	$r_s = 2.0$	$r_s = 5.0$
Block determinants				
1	-0.58831(7)	-0.52510(3)	-0.43842(1)	-0.30433(1)
16	-0.58962(6)	-0.52558(3)	-0.43876(1)	-0.30466(1)
Dense determinants				
1	-0.58895(6)	-0.52568(3)	-0.43881(1)	-0.30468(1)
16	-0.59094(6)	-0.52682(3)	-0.44053(1)	-0.30495(1)

TABLE IV: Hartree-Fock energy of the $N=14$ HEG with simple cubic boundary conditions.

r_s	E_{HF}
0.5	48.368516
1.0	8.491476
2.0	0.322545
5.0	-0.812549

If we do not assume the block-diagonal form, we obtain (dense) determinants

$$\det[\boldsymbol{\psi}] = \left| \psi_i^\uparrow(\mathbf{r}_j^\uparrow; \{\mathbf{r}_{/j}^\uparrow\}, \{\mathbf{r}^\downarrow\}) \quad \psi_i^\downarrow(\mathbf{r}_j^\downarrow; \{\mathbf{r}^\uparrow\}, \{\mathbf{r}_{/j}^\downarrow\}) \right|, \quad (39)$$

where the blocks are now of dimension $N \times N_\uparrow$ and $N \times N_\downarrow$. We have found that lifting the block-diagonal constraint provides a small but noticeable variational improvement. Dense determinants are hence used unless otherwise stated.

We provide a comparison with energies for the unpolarized $N=14$ HEG, obtained using block diagonal determinants. These values are presented in Table III, alongside the results obtained using dense determinants, copied from the main text for comparison. In all cases dense determinants offer a small variational improvement in the correlation energy obtained.

HARTREE-FOCK ENERGIES OF $N=14$ HEG

In table Table I, we have given our results in terms of the correlation energy to enable direct comparison with published i-FCIQMC benchmarks. The results of our calculations, however, are total energies. To obtain the correlation energies we have subtracted the Hartree-Fock energies given in Table IV, obtained using the CASINO program [64, 65].

TABLE V: Total energy per electron of the $N=27$ HEG in a body-centered cubic simulation cell at a range of densities, obtained via variational Monte Carlo. SJB values were obtained from our calculations using the CASINO package. Here, “gas” refers to a Slater determinant of plane-wave orbitals, and “crystal” refers to a Slater determinant of Gaussian orbitals. FermiNet results for $r_s \leq 1$ used complex wave functions.

r_s	Total energy [Hartree per electron]				
	FermiNet	SJB-VMC Gas	SJB-DMC Gas	SJB-VMC Crystal	SJB-DMC Crystal
0.2	44.18067(2)	44.18235(4)	44.18224(1)	44.184815(4)	44.18468(3)
0.5	6.318704(2)	6.31983(2)	6.319754(7)	6.322090(8)	6.32130(6)
1.0	1.261547(1)	1.262022(8)	1.261975(2)	1.26286(1)	1.26254(2)
2.0	0.152086(1)	0.152393(3)	0.152344(3)	0.1521205(4)	0.152077(4)
3.0	-0.006961(2)	-0.006693(2)	-0.0067419(7)	-0.006915(2)	-0.006966(2)
5.0	-0.057691(4)	-0.057453(1)	-0.0575023(4)	-0.0576496(9)	-0.057702(1)
10.0	-0.050320(5)	-0.0501501(6)	-0.0502038(1)	-0.0502782(7)	-0.0503349(8)
30.0	-0.022541(5)	-0.0224510(2)	-0.02249870(5)	-0.0225119(3)	-0.0225620(2)
50.0	-0.014478(4)	-0.0144033(1)	-0.01444409(5)	-0.0144528(1)	-0.0144918(1)
70.0	-0.010693(3)	-0.0106355(1)	-0.0106735(1)	-0.0106824(1)	-0.0107118(1)
90.0	-0.008496(1)	-0.0084277(1)	-0.0084862(1)	-0.00849274(7)	-0.0085137(2)