

IMPERIAL

RPMBT22
Tsukuba, Japan
23rd September 2024

Approximating Many-Electron Wave Functions using Neural Networks

Gino Cassella Andres Perez Fadon Matthew Foulkes
Wan Tong Lou Halvard Sutterud David Pfau[†]
James Spencer[†]

[†]Google DeepMind Ltd.

Outline

- 1 Introduction
 - Simulating Many-Electron Systems
 - Neural Wave Functions
 - 2 Benchmarks
 - 3 Applications
 - Positrons
 - Wigner Crystals
 - Superfluidity
 - 4 Outlook
-

Outline

- 1 Introduction
 - Simulating Many-Electron Systems
 - Neural Wave Functions
 - 2 Benchmarks
 - 3 Applications
 - Positrons
 - Wigner Crystals
 - Superfluidity
 - 4 Outlook
-

Simulating Many-Electron Systems

The Many-Electron Schrödinger Equation

$$\left(-\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i,j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_a \sum_{i=1}^N \frac{Z_a e^2}{|\mathbf{d}_a - \mathbf{r}_i|} + \frac{1}{2} \sum_{a,b} \frac{Z_a Z_b e^2}{|\mathbf{d}_a - \mathbf{d}_b|} \right) \Psi = E\Psi$$

Simulating Many-Electron Systems

The Many-Electron Schrödinger Equation

$$\left(-\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i,j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_a \sum_{i=1}^N \frac{Z_a e^2}{|\mathbf{d}_a - \mathbf{r}_i|} + \frac{1}{2} \sum_{a,b} \frac{Z_a Z_b e^2}{|\mathbf{d}_a - \mathbf{d}_b|} \right) \Psi = E\Psi$$

The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.

P.A.M. Dirac, Proc. Roy. Soc. A, **123**, 714 (1929)

Density Functional Theory

- Deal with the electrons one by one.
- Effects of other electrons approximated by a mean field.
- $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \longrightarrow \mathcal{A}\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \dots \phi_N(\mathbf{r}_N)$

Density Functional Theory

- Deal with the electrons one by one.
 - Effects of other electrons approximated by a mean field.
 - $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) \longrightarrow \mathcal{A}\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2) \dots \phi_N(\mathbf{r}_N)$
 - There exists a mean field that yields exact ground-state energies and densities.
-

DFT works remarkably well, but:

- is not accurate enough for room-temperature chemistry;
 - cannot tell you about the correlations between electrons.
-

DFT works remarkably well, but:

- is not accurate enough for room-temperature chemistry;
- cannot tell you about the correlations between electrons.

We are going to tackle the full many-electron problem

DFT works remarkably well, but:

- is not accurate enough for room-temperature chemistry;
- cannot tell you about the correlations between electrons.

We are going to tackle the full many-electron problem

(. . . and prove Dirac wrong?)

Neural Wave Functions

Artificial neural networks

are flexible and efficient function approximators in high-dimensional spaces.

$f(\text{img of dog}) \rightarrow \text{dog}$

$f(\text{img of cat}) \rightarrow \text{cat}$

Neural Wave Functions

Artificial neural networks

are flexible and efficient function approximators in high-dimensional spaces.

$$f(\text{img}_{\text{dog}}) \rightarrow \text{dog}$$

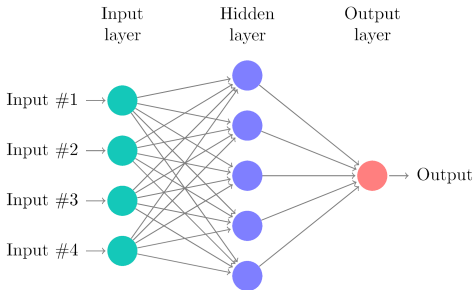
$$f(\text{img}_{\text{cat}}) \rightarrow \text{cat}$$

$$\mathbf{f}(\mathbf{r}_1, \dots, \mathbf{r}_N) \rightarrow \Psi$$

A Simple Neural Network

$$\mathbf{x}^{\ell+1} = \mathbf{A}(\mathbf{W}^{\ell}\mathbf{x}^{\ell} + \mathbf{b}^{\ell})$$

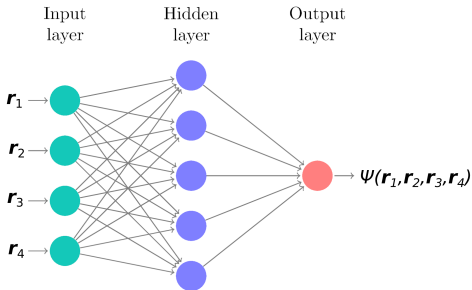
Learn network parameters using automatic differentiation and back propagation.



A Simple Neural Wave Function

$$\mathbf{x}^{\ell+1} = \mathbf{A}(\mathbf{W}^{\ell} \mathbf{x}^{\ell} + \mathbf{b}^{\ell})$$

Learn wavefunction parameters using automatic differentiation and back propagation.



Variational Principle

Given an approximate ground-state wavefunction $\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)$, we can improve it by minimizing

$$E(\theta) = \int \dots \int \Psi_{\theta}^* \hat{H} \Psi_{\theta} \, d\mathbf{r}_1 \dots d\mathbf{r}_N$$

Variational Principle

Given an approximate ground-state wavefunction $\Psi_\theta(\mathbf{r}_1, \dots, \mathbf{r}_N)$, we can improve it by minimizing

$$E(\theta) = \int \dots \int \Psi_\theta^* \Psi_\theta \frac{\hat{H} \Psi_\theta}{\Psi_\theta} d\mathbf{r}_1 \dots d\mathbf{r}_N$$

Variational Principle

Given an approximate ground-state wavefunction $\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)$, we can improve it by minimizing

$$E(\theta) = \int \dots \int \left(\frac{\hat{H}\Psi_{\theta}}{\Psi_{\theta}} \right) |\Psi_{\theta}|^2 d\mathbf{r}_1 \dots d\mathbf{r}_N$$

Variational Principle

Given an approximate ground-state wavefunction $\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)$, we can improve it by minimizing

$$E(\theta) = \mathbb{E}_{\Psi_{\theta}^2} \left[\frac{\hat{H}\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \right]$$

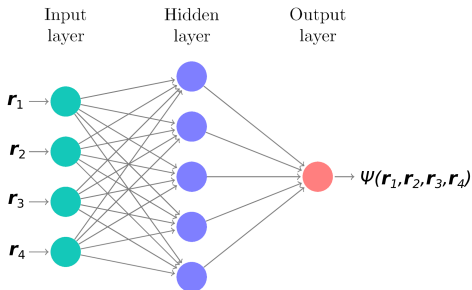
Variational Principle

Given an approximate ground-state wavefunction $\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)$, we can improve it by minimizing

$$E(\theta) = \mathbb{E}_{\Psi_{\theta}^2} \left[\frac{\hat{H}\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_N)} \right]$$

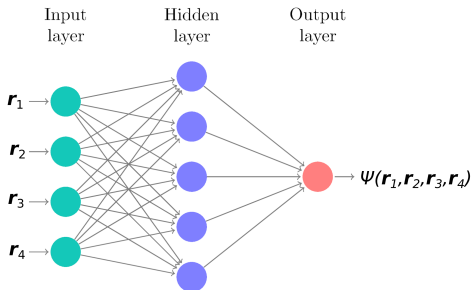
QM comes with a built-in loss function. We can use it to learn wavefunctions without recourse to external data.

Neural Variational Monte Carlo



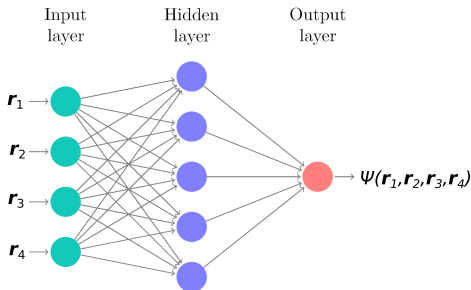
- Estimate the energy expectation value and its gradients with respect to the network parameters using Monte Carlo sampling.

Neural Variational Monte Carlo



- Estimate the energy expectation value and its gradients with respect to the network parameters using Monte Carlo sampling.
- Adapt the weights and biases of the neural network to lower the energy.

Neural Variational Monte Carlo



- Estimate the energy expectation value and its gradients with respect to the network parameters using Monte Carlo sampling.
- Adapt the weights and biases of the neural network to lower the energy.
- No training dataset is required. We are “learning from equations.”

The Pauli Principle

Many-electron wavefunctions must be totally antisymmetric:

$$\begin{aligned}\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_i, \dots, \mathbf{r}_j, \dots, \mathbf{r}_N) \\ = -\Psi_{\theta}(\mathbf{r}_1, \dots, \mathbf{r}_j, \dots, \mathbf{r}_i, \dots, \mathbf{r}_N)\end{aligned}$$

for all electron labels i and j .

Hartree-Fock Theory

$$\Psi \approx \begin{vmatrix} \varphi_1(\mathbf{r}_1) & \varphi_1(\mathbf{r}_2) & \dots & \varphi_1(\mathbf{r}_N) \\ \varphi_2(\mathbf{r}_1) & \varphi_2(\mathbf{r}_2) & \dots & \varphi_2(\mathbf{r}_N) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \varphi_N(\mathbf{r}_1) & \varphi_N(\mathbf{r}_2) & \dots & \varphi_N(\mathbf{r}_N) \end{vmatrix}$$

FermiNet

Nothing requires the orbitals to be functions of the coordinates of a single electron:

$$\Psi \approx \begin{vmatrix} \varphi_1(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_1(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_1(\mathbf{r}_N, \{\mathbf{r}/_N\}) \\ \varphi_2(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_2(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_2(\mathbf{r}_N, \{\mathbf{r}/_N\}) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \varphi_N(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_N(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_N(\mathbf{r}_N, \{\mathbf{r}/_N\}) \end{vmatrix}$$

FermiNet

Nothing requires the orbitals to be functions of the coordinates of a single electron:

$$\Psi \approx \begin{vmatrix} \varphi_1(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_1(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_1(\mathbf{r}_N, \{\mathbf{r}/_N\}) \\ \varphi_2(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_2(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_2(\mathbf{r}_N, \{\mathbf{r}/_N\}) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \varphi_N(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_N(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_N(\mathbf{r}_N, \{\mathbf{r}/_N\}) \end{vmatrix}$$

Ψ remains antisymmetric as long as $\varphi_i(\mathbf{r}_j; \{\mathbf{r}/_j\})$ is invariant under any change in the order of the arguments after \mathbf{r}_j .

FermiNet

Nothing requires the orbitals to be functions of the coordinates of a single electron:

$$\Psi \approx \begin{vmatrix} \varphi_1(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_1(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_1(\mathbf{r}_N, \{\mathbf{r}/_N\}) \\ \varphi_2(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_2(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_2(\mathbf{r}_N, \{\mathbf{r}/_N\}) \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \varphi_N(\mathbf{r}_1, \{\mathbf{r}/_1\}) & \varphi_N(\mathbf{r}_2, \{\mathbf{r}/_2\}) & \dots & \varphi_N(\mathbf{r}_N, \{\mathbf{r}/_N\}) \end{vmatrix}$$

Ψ remains antisymmetric as long as $\varphi_i(\mathbf{r}_j; \{\mathbf{r}/_j\})$ is invariant under any change in the order of the arguments after \mathbf{r}_j .

(A drastic generalisation of a backflow wavefunction)

FermiNet and PsiFormer

FermiNet

D. Pfau, J.S. Spencer, A.G.D.G. Matthews, and W.M.C. Foulkes
Phys. Rev. Res. **2**, 033429 (2020)

PsiFormer

I. von Glehn, J.S. Spencer and D. Pfau
arXiv:2211.13672 (2022)

FermiNet and PsiFormer

FermiNet

D. Pfau, J.S. Spencer, A.G.D.G. Matthews, and W.M.C. Foulkes
Phys. Rev. Res. **2**, 033429 (2020)

PsiFormer

I. von Glehn, J.S. Spencer and D. Pfau
arXiv:2211.13672 (2022)

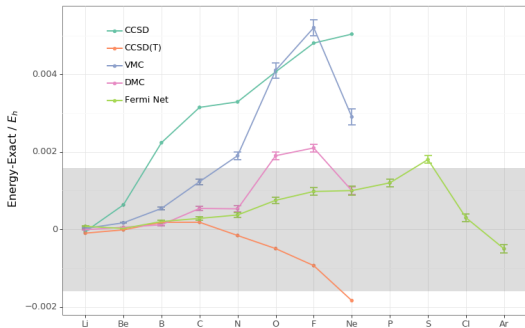
Recent Review

J. Hermann, J. Spencer, K. Choo, A. Mezzacapo, W.M.C. Foulkes,
D. Pfau, G. Carleo, and F. Noé
Nat. Rev. Chem. **7**, 692 (2023)

Outline

- 1 Introduction
 - Simulating Many-Electron Systems
 - Neural Wave Functions
 - 2 **Benchmarks**
 - 3 Applications
 - Positrons
 - Wigner Crystals
 - Superfluidity
 - 4 Outlook
-

Atoms

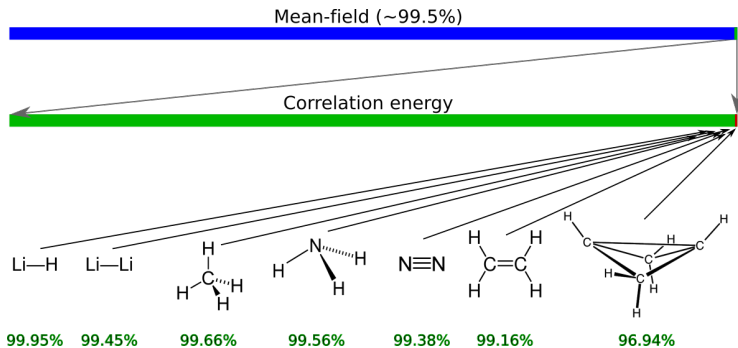


- 16 FermiNet determinants.
(conventional VMC & DMC used 50–100 CSFs.)
- FermiNet consistently captures 99.7% of correlation energy.

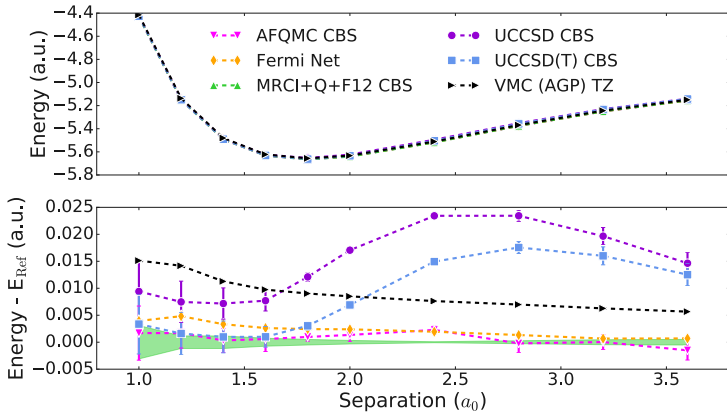
VMC, DMC:
P. Seth, P. López Ríos and R.J. Needs
J. Chem. Phys. **134**, 084105 (2011)

Exact:
S.J. Chakravorty et al.
Phys. Rev. B **47**, 3649 (1993)

Molecules



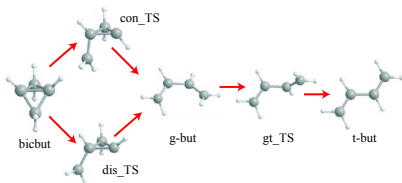
H₁₀



Motta et al., PRX 031059 (2017)

Chemical Reactions

Bicyclobutane \longrightarrow Butadiene



Method	con_TS	dis_TS	g-but	gt_TS	t-but
CCSD(T)	40.4	21.8	-25.1	-22.3	-28.0
CR-CC(2,3)	41.1	66.1	-24.9	-22.1	-27.9
CCSDt	40.1	59.0	-27.2	-25.3	-31.1
CC(t;3)	40.2	60.1	-25.3	-22.6	-28.3
DMC	40.4 \pm 0.5	58.6 \pm 0.5	-25.2 \pm 0.5	-22.2 \pm 0.5	-27.9 \pm 0.5
FermiNet	40.2 \pm 0.1	57.7 \pm 0.1	-25.3 \pm 0.1	-22.5 \pm 0.1	-28.4 \pm 0.1
Experiment	40.6 \pm 2.5	-	-	-	-25.9 \pm 0.4

(Energies in kcal/mol relative to bicyclobutane)

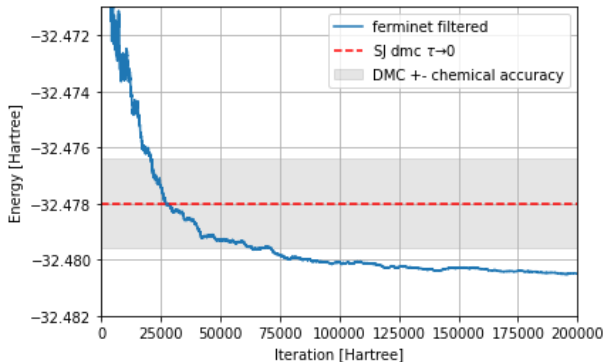
Solids

- Simulation cell with sides $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$.
- Any point in cell can be written

$$\mathbf{r} = r_1 \mathbf{a}_1 + r_2 \mathbf{a}_2 + r_3 \mathbf{a}_3$$

- Replace position inputs r_i by explicitly periodic functions $\sin(2\pi r_i)$ and $\cos(2\pi r_i)$.
-

LiH crystal

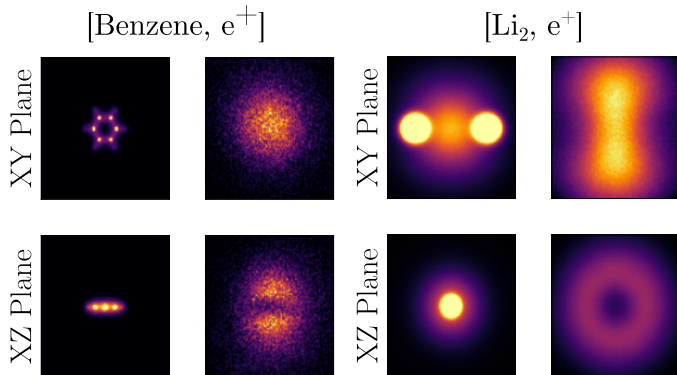


See also Li, Li and Chen, Nat. Commun. **13**, 7895 (2022)

Outline

- 1 Introduction
 - Simulating Many-Electron Systems
 - Neural Wave Functions
 - 2 Benchmarks
 - 3 Applications
 - Positrons
 - Wigner Crystals
 - Superfluidity
 - 4 Outlook
-

Positrons



e^+ binding energy (mHa)

Method	LiH	BeO
FermiNet	37.23	25.10
SJ-VMC	17.27	19
FN-DMC	37.1	28.0
CISD	17	
MRD-CI	29.37	13.78
ECG-SVM	36.93	
GW	39	

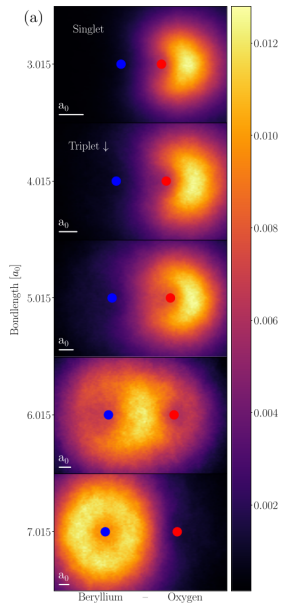
e^+ binding energy (mHa)

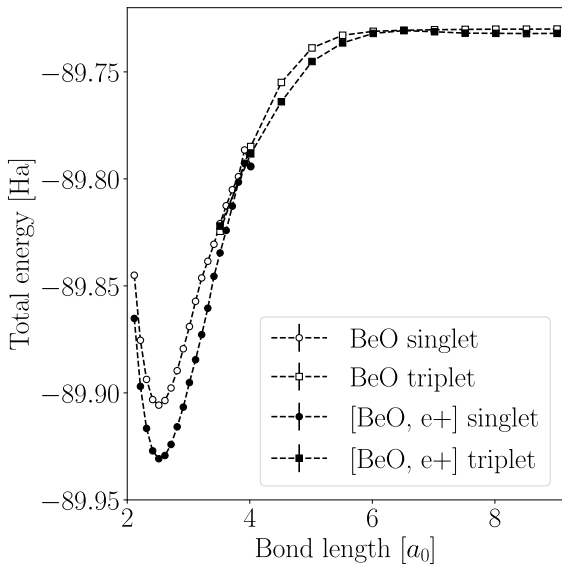
Method	LiH	BeO
FermiNet	37.23	25.10
SJ-VMC	17.27	19
FN-DMC	37.1	28.0
CISD	17	
MRD-CI	29.37	13.78
ECG-SVM	36.93	
GW	39	

e^+ annihilation rate (ns^{-1})

Method	PsH	Mg	LiH	BeO	Li ₂	Benzene
CI	2.0183	1.001	0.8947			
FN-DMC	2.32		1.3602			
ECG-SVG	2.4361	0.955	1.375			
	2.4722	1.0249				
	2.4685					
GW			2.083			0.666
FermiNet	2.440	1.076	1.3391	0.9533	1.962	0.5247

[BeO, e⁺]



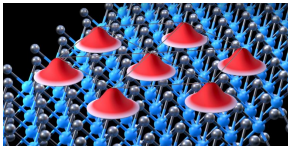


Wigner Crystals

- At very low densities, a uniform electron liquid freezes into a

Wigner crystal

- Wigner crystallisation has not yet been observed in 3D, but was recently imaged in 2D.



(ETH Zurich, July 2021)

Simulations

- Continuum QMC methods cannot easily find quantum phase transitions; you usually have to guess them first.
-

Simulations

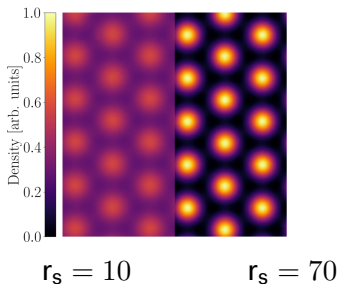
- Continuum QMC methods cannot easily find quantum phase transitions; you usually have to guess them first.
 - FermiNet found the 3D Wigner crystal spontaneously.
-

Results

- 27 electrons in a body-centered cubic simulation cell.
-

Results

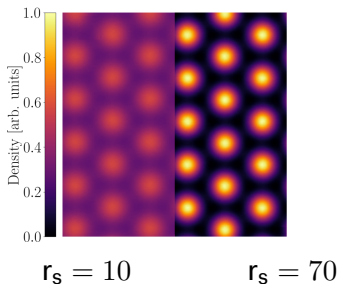
- 27 electrons in a body-centered cubic simulation cell.
- Simulation cell is far too small to get the right transition density. Formation of a $3 \times 3 \times 3$ Wigner crystal is strongly favoured.



Phys. Rev. Lett. **130**, 036401 (2023)

Results

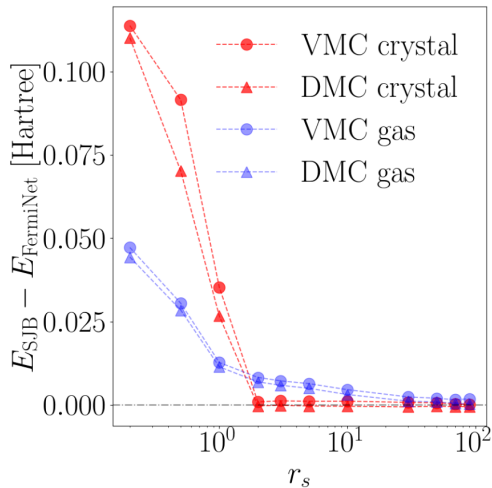
- 27 electrons in a body-centered cubic simulation cell.
- Simulation cell is far too small to get the right transition density. Formation of a $3 \times 3 \times 3$ Wigner crystal is strongly favoured.



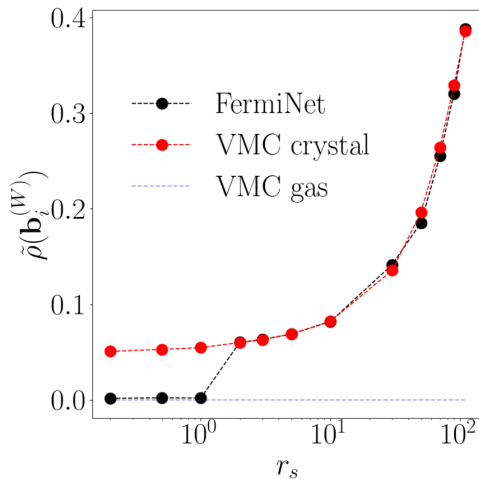
Phys. Rev. Lett. **130**, 036401 (2023)

- Comparing with conventional VMC and DMC simulations for the same simulation cell is nevertheless valid.

Relative Energies



Order Parameter



The Floating Crystal

The Wigner crystal was seen in the one-electron density. If we omit the one-electron input features we see a true floating crystal, but the energies are significantly higher.

Superfluidity

The Unitary Fermi Gas

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^{N^\uparrow} \sum_{j=1}^{N^\downarrow} U(\mathbf{r}_i^\uparrow - \mathbf{r}_j^\downarrow)$$

Superfluidity

The Unitary Fermi Gas

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^{N^\uparrow} \sum_{j=1}^{N^\downarrow} U(\mathbf{r}_i^\uparrow - \mathbf{r}_j^\downarrow)$$

- Attractive interaction U chosen such that s-wave scattering length a_s diverges.

Superfluidity

The Unitary Fermi Gas

$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^{N^\uparrow} \sum_{j=1}^{N^\downarrow} U(\mathbf{r}_i^\uparrow - \mathbf{r}_j^\downarrow)$$

- Attractive interaction U chosen such that s-wave scattering length a_s diverges.
- $(\text{Range of } U)/a_s \rightarrow 0$.

Superfluidity

The Unitary Fermi Gas

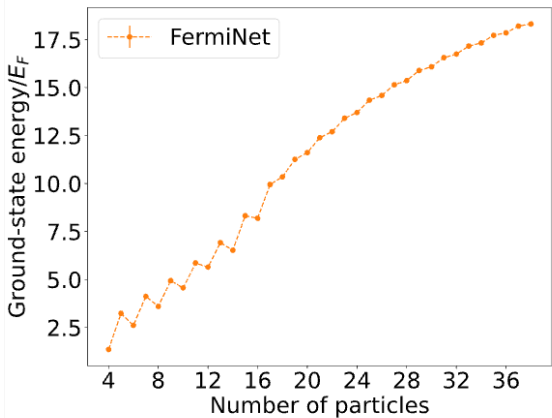
$$\hat{H} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 + \sum_{i=1}^{N^\uparrow} \sum_{j=1}^{N^\downarrow} U(\mathbf{r}_i^\uparrow - \mathbf{r}_j^\downarrow)$$

- Attractive interaction U chosen such that s-wave scattering length a_s diverges.
- $(\text{Range of } U)/a_s \rightarrow 0$.
- Only remaining length scale is $1/k_F$.

- All thermodynamic quantities depend on k_F only. For example

$$E = \xi E_{FG} = \xi \frac{3\hbar^2 k_F^2}{2m}$$

where ξ is a dimensionless constant known as the Bertsch parameter.



Energy as a function of number of particles

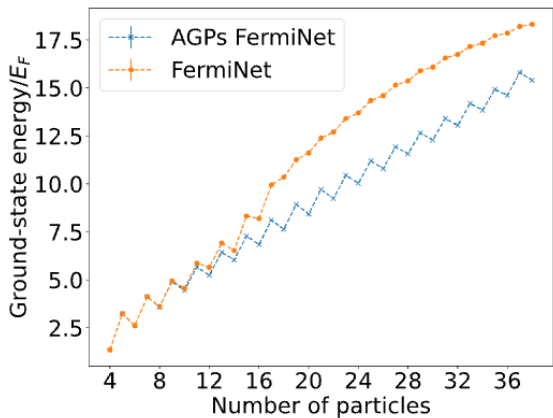
Ideas for Improvement?

- Use a bigger network or a linear combination of more terms?
 - Slightly increases maximum system size that works.
 - Costly.

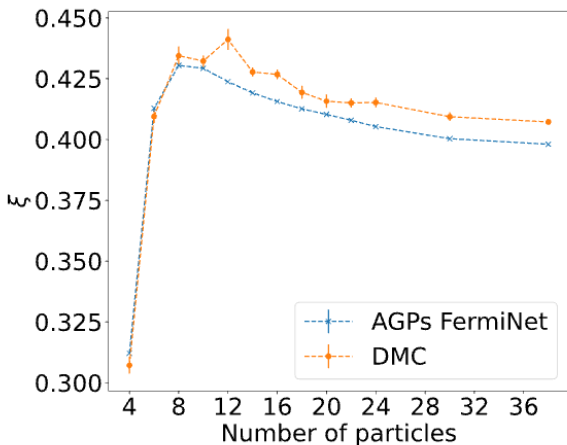
Ideas for Improvement?

- Use a bigger network or a linear combination of more terms?
 - Slightly increases maximum system size that works.
 - Costly.
- Build a neural generalisation of the BCS (AGPs) wave function:

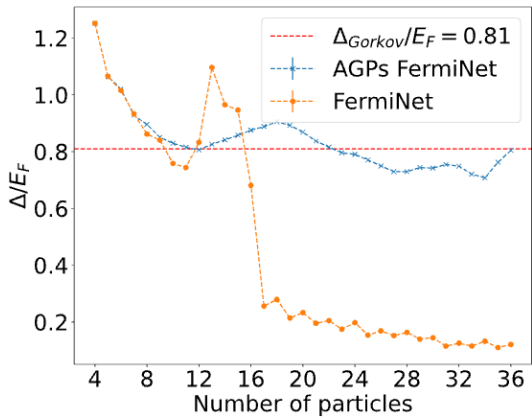
$$\Psi^{2N} = \mathcal{A} \left[\Phi(\mathbf{r}_1^\uparrow, \mathbf{r}_2^\downarrow) \Phi(\mathbf{r}_3^\uparrow, \mathbf{r}_4^\downarrow) \dots \Phi(\mathbf{r}_{2N-1}^\uparrow, \mathbf{r}_{2N}^\downarrow) \right]$$



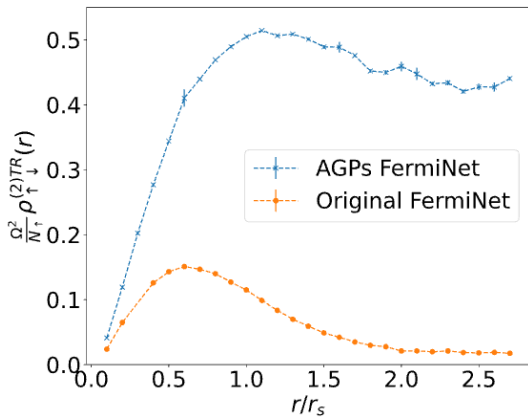
Energies calculated using pairing wave function



Bertsch parameter
(comparison with fixed-node DMC)



Superconducting gap



Superfluid fraction

Outline

- 1 Introduction
 - Simulating Many-Electron Systems
 - Neural Wave Functions
 - 2 Benchmarks
 - 3 Applications
 - Positrons
 - Wigner Crystals
 - Superfluidity
 - 4 Outlook
-

Positives

- Neural wave functions are way better than other approximate wave functions for molecules with more than a few atoms.
 - Rival and sometimes outperform the best quantum chemical methods, many of which have been in development for 50+ years.
 - Seem to be good at dealing with unusual systems (positronic molecules; Wigner crystals) where conventional methods need painful modification.
 - Might be able to deal with strong correlations.
 - Might be able to discover new phases.
-

Questions and Negatives

- How general is the FermiNet wavefunction?
 - Does it work for non-Fermi liquids?
 - Size consistency and extensivity?
 - Optimization is slow and sometimes problematic.
 - Scaling.
-

Range of Ideas and Approaches

	Systems	Parameterisation	Optimisation	Accuracy
NNB: Luo, Clark, PRL 122 , 226401 (2019)	Hubbard	backflow	first order	beats conventional backflow
PauliNet: Hermann et al., Nature Chemistry 12 , 891 (2020)	molecules continuum	Jastrow + backflow	ADAM	Boron: 97.3% c.e. H ₁₀ : 90-97% c.e.
NQS: Choo et al., Nature Commun. 11 , 2360 (2020)	molecules basis	map to spin system; approximate as RBM	stochastic reconfig.	< 1mH relative to FCI in STO-3G
FermiNet: Pfau et al., Phys. Rev. Res. 2 , 033429 (2020)	molecules continuum	everything	KFAC	Boron: 99.8% c.e. H ₁₀ : 98.5-99.3% c.e.
FermiNet+DMC: Wilson et al., arXiv:2103.12570 (2021)	molecules continuum	everything	KFAC	consistent with our FermiNet+DMC
Weight Sharing: Scherbela et al., arXiv:2105.08351 (2021)	multiple geometries	Jastrow + backflow	ADAM	similar to PauliNet

Typical Network Configuration

- 16 FermiNet determinants
 - 256 hidden units for the one-electron stream
 - 32 hidden units for the two-electron stream
 - Approximately 700,000 parameters
 - Batch size 4096
 - Training iterations 2×10^5
 - Steps between parameter updates 10
-

Scaling

