6th IFERC Workshop on the usage of GPU based systems for fusion applications
24th June 2025

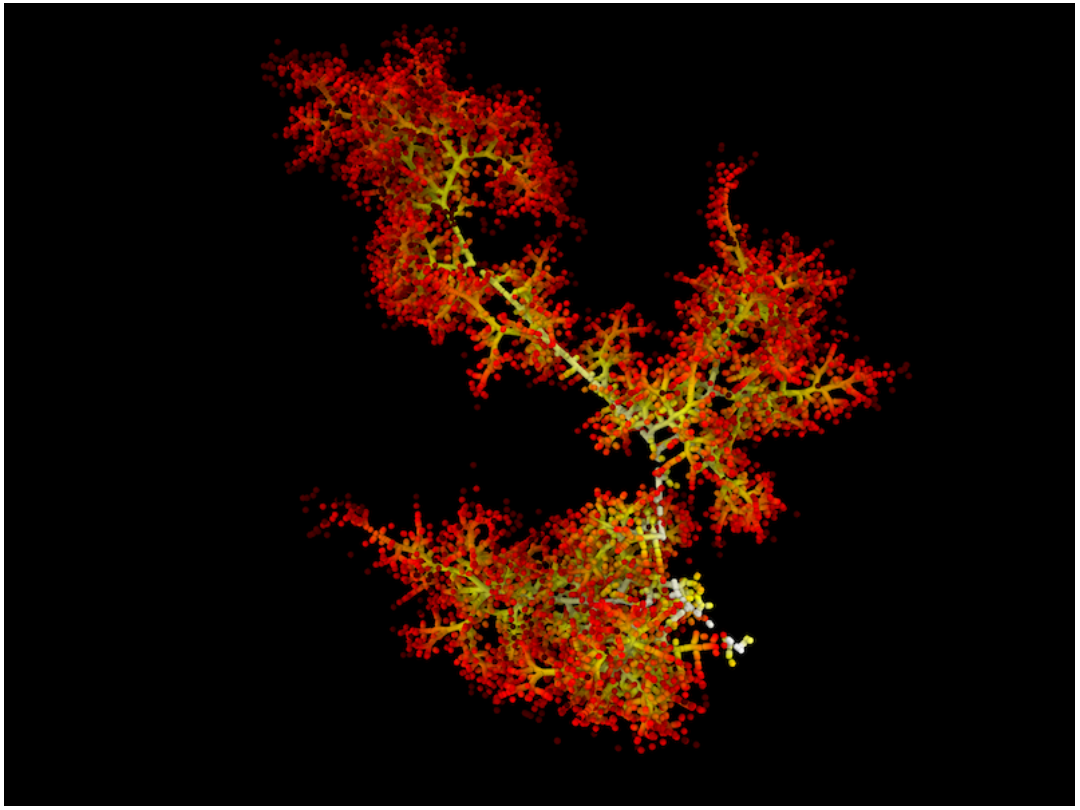
IMPERIAL

Approximating Many-Electron Wave Functions using Artificial Neural Networks

Gino Cassella Andres Perez Fadon Matthew Foulkes
Wan Tong Lou Halvard Sutterud David Pfau†
James Spencer† Sergei Dudarev*

†Google DeepMind Ltd. *UKAEA

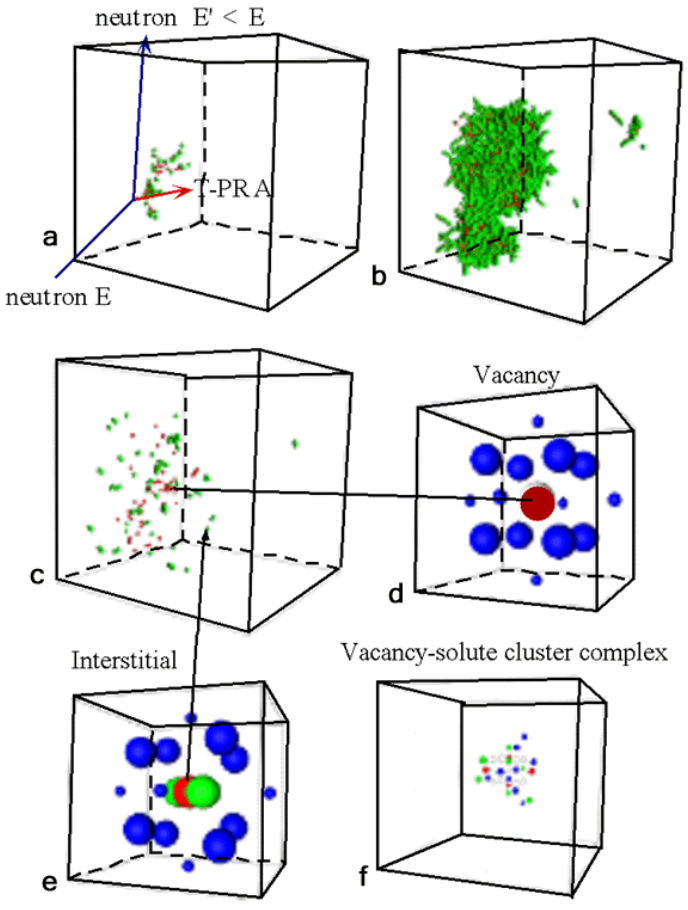
Radiation Damage



200 keV radiation damage cascade in tungsten

(cascadesdb.iaea.org)

Cascade Physics



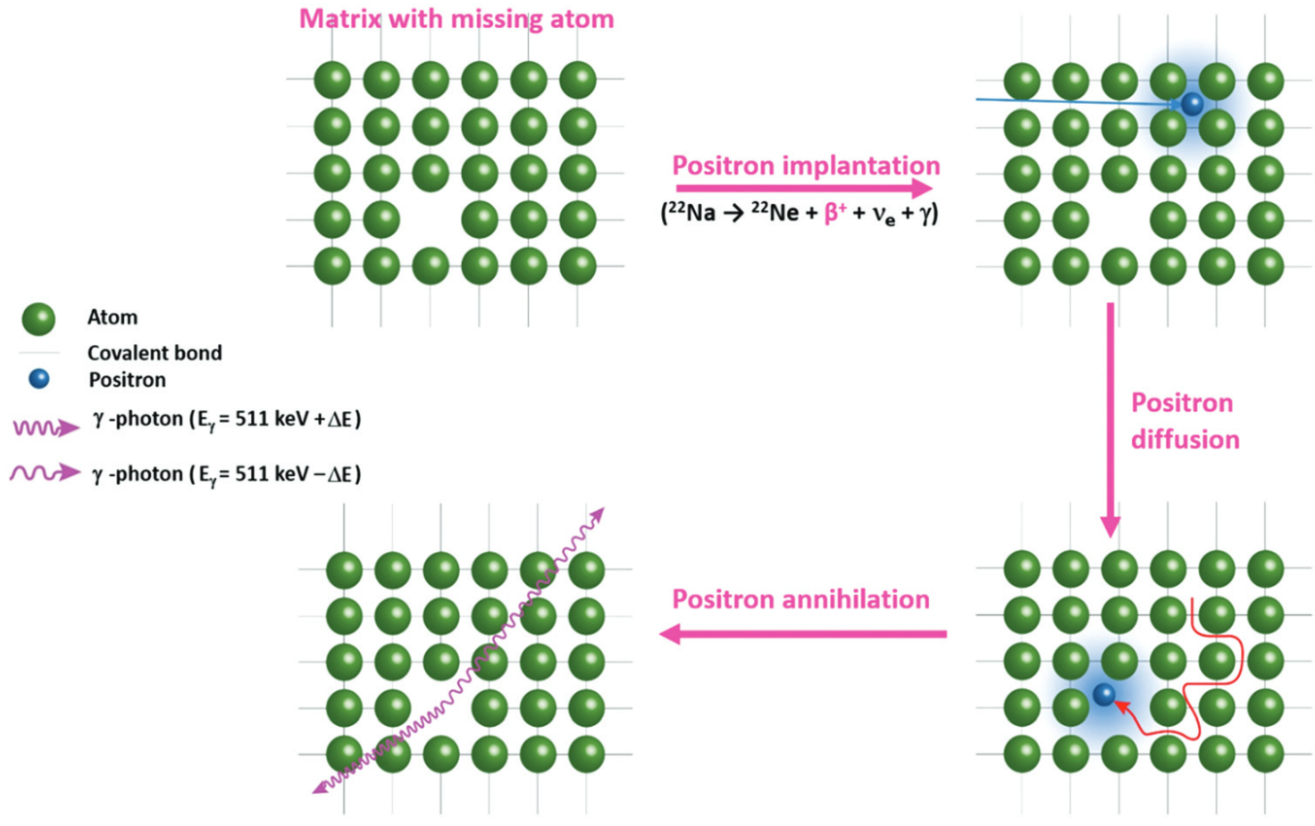
- ps time scales
- nm length scales
- electronic T $\sim 10^4$K

- ~100 ps

- hours or days

(www.tms.org/pubs/journals/JOM/0107/fig2.gif)

Positron Physics



(pubs.rsc.org/en/content/articlehtml/2019/tc/c8tc06330c)

PALS

- Voids and vacancies are hard to image.
- Positron annihilation lifetime spectroscopy is useful.
- Interpreting results requires quantum mechanical simulations.
- DFT may not be adequate.

Aim

Develop a beyond-DFT approach for computing annihilation rates and annihilating-pair momentum densities in fusion materials.

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.
- $\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \quad \longrightarrow \quad \mathcal{A}\phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\ldots\phi_N(\mathbf{r}_N)$
- There exists a mean field that yields exact ground-state energies and densites.

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.

$f($  $) \rightarrow dog$

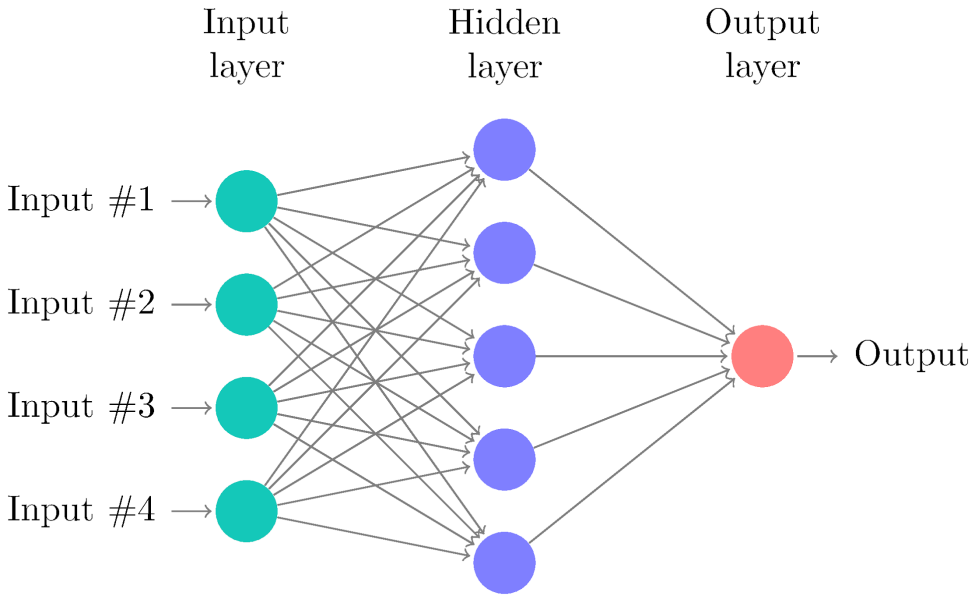
$f($  $) \rightarrow cat$

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

A Simple Neural Network

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

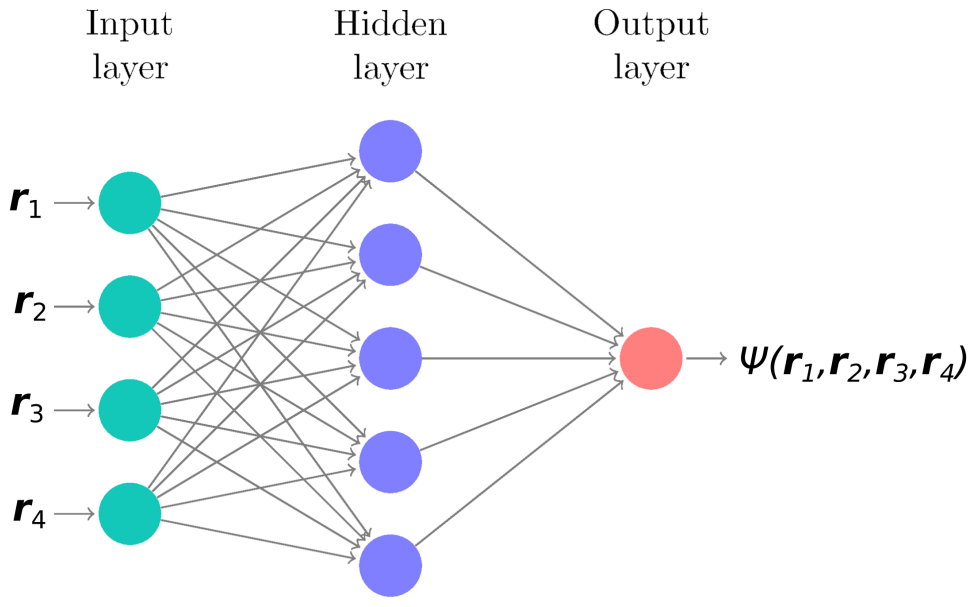
Learn network parameters using automatic differentiation and back propagation.



A Simple Neural Wave Function

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

Learn wavefunction parameters using automatic differentiation and back propagation.



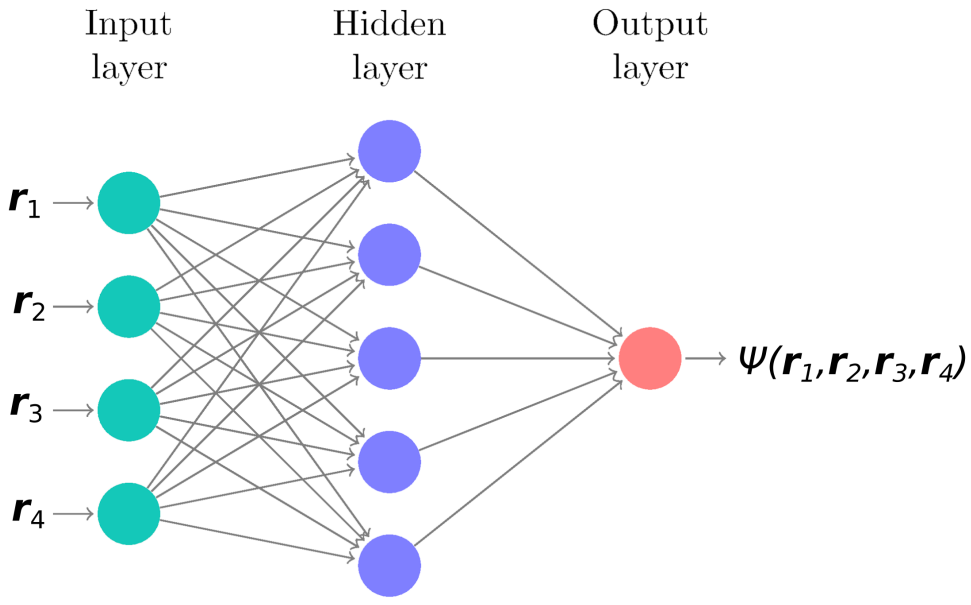
Variational Principle

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

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

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.
- 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:

$$\Psi_{\boldsymbol{\theta}}(\mathbf{r}_1, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_N) = -\Psi_{\boldsymbol{\theta}}(\mathbf{r}_1, \ldots, \mathbf{r}_j, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N)$$

for all electron labels i and j.

Hartree-Fock Theory

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

FermiNet

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

$$\Psi \approx \begin{vmatrix} \varphi_1(\mathbf{r}_1, \{\mathbf{r}_{/1}\}) & \varphi_1(\mathbf{r}_2, \{\mathbf{r}_{/2}\}) & \dots & \varphi_1(\mathbf{r}_N, \{\mathbf{r}_{/N}\}) \\ \varphi_2(\mathbf{r}_1, \{\mathbf{r}_{/1}\}) & \varphi_2(\mathbf{r}_2, \{\mathbf{r}_{/2}\}) & \dots & \varphi_2(\mathbf{r}_N, \{\mathbf{r}_{/N}\}) \\ . & . & \dots & . \\ . & . & \dots & . \\ \varphi_N(\mathbf{r}_1, \{\mathbf{r}_{/1}\}) & \varphi_N(\mathbf{r}_2, \{\mathbf{r}_{/2}\}) & \dots & \varphi_N(\mathbf{r}_N, \{\mathbf{r}_{/N}\}) \end{vmatrix}$$

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)

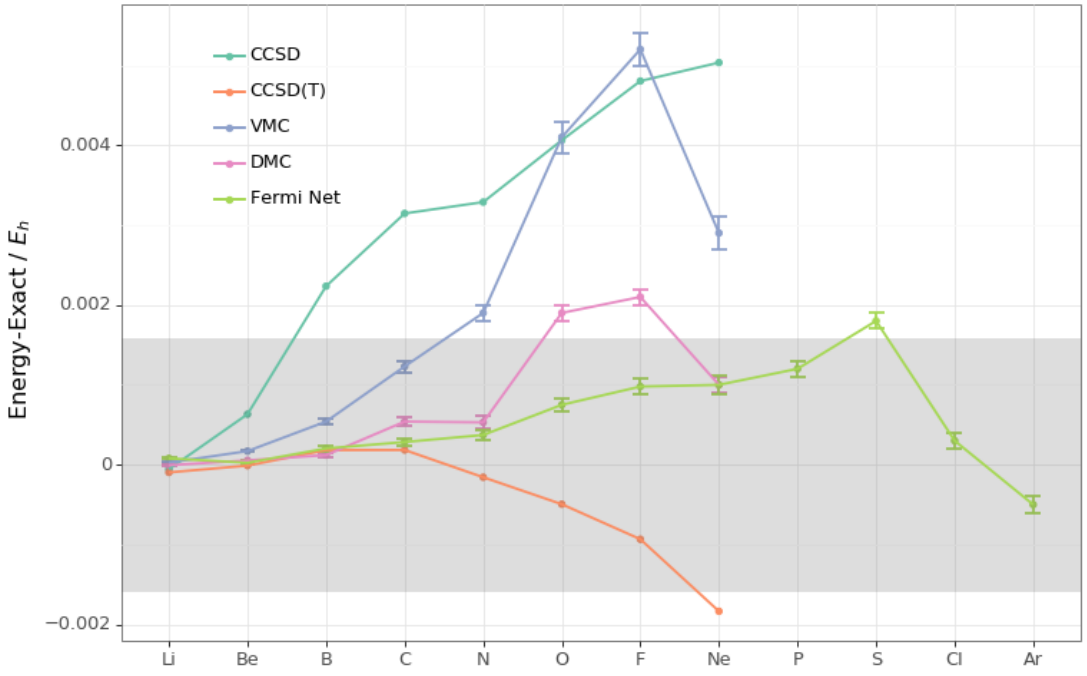
Psiformer

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

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)

Atoms

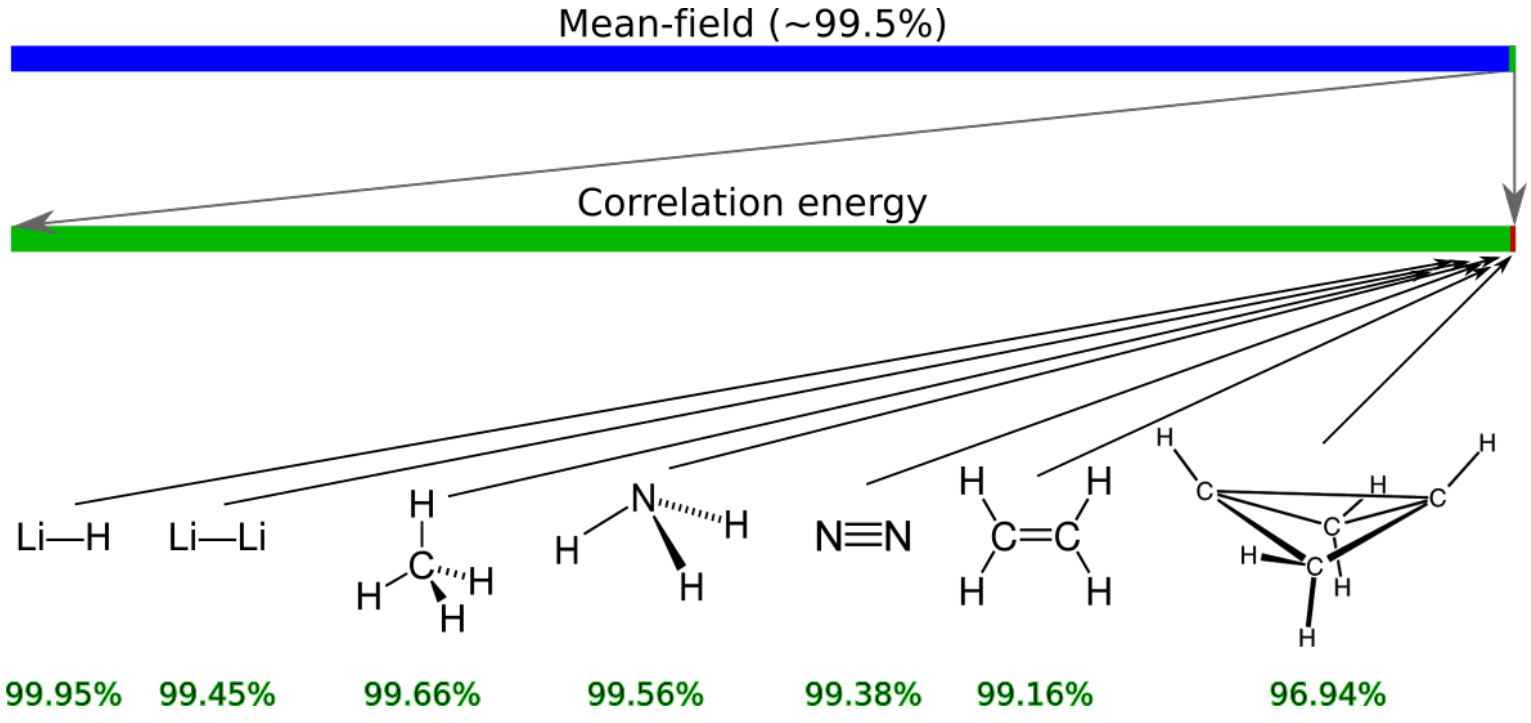


- 16 FermiNet determinants.
  (conventional VMC & DMC used 50–100 CSFs.)
- FermiNet consistently captures 99.7% of correlation energy.

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

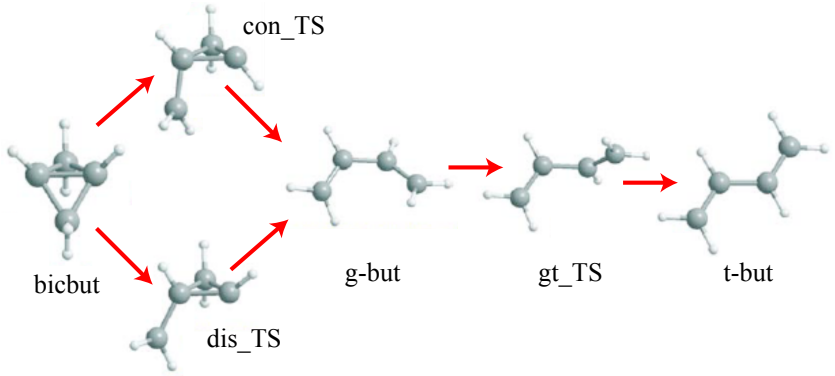
Exact:
S.J. Chakravorty et al.
Phys. Rev. B **47**, 3649 (1993)

Molecules


Mean-field (~99.5%)
Correlation energy
Li—H
Li—Li
99.95%
99.45%
99.66%
99.56%
99.38%
99.16%
96.94%

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 ± 0.5	58.6 ± 0.5	-25.2 ± 0.5	-22.2 ± 0.5	-27.9 ± 0.5
FermiNet	40.2 ± 0.1	57.7 ± 0.1	-25.3 ± 0.1	-22.5 ± 0.1	-28.4 ± 0.1
Experiment	40.6 ± 2.5	-	-	-	-25.9 ± 0.4

(Energies in kcal/mol relative to bicyclobutane)

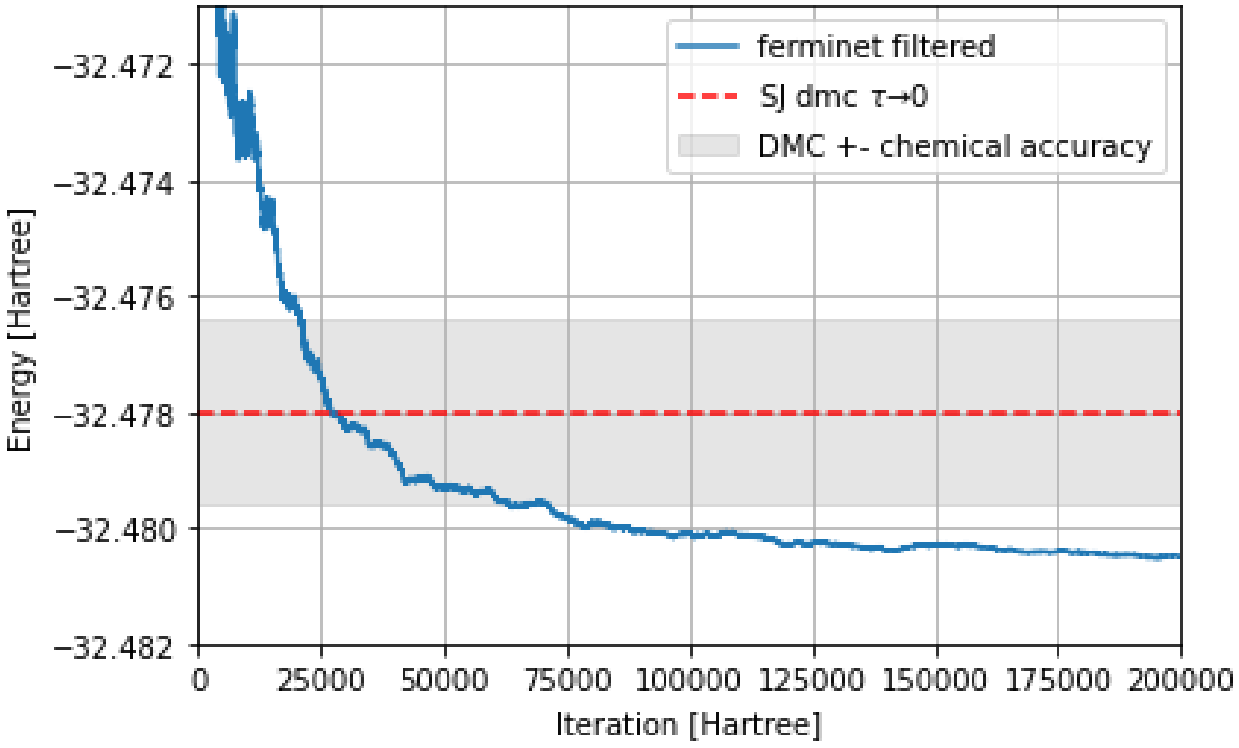
Solids

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

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

- 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)

The Electron Gas

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

Wigner crystal

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

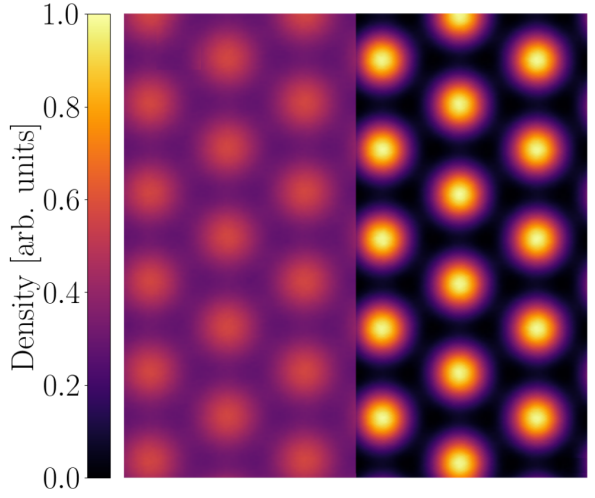
(ETH Zurich, July 2021)

Simulations

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

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.

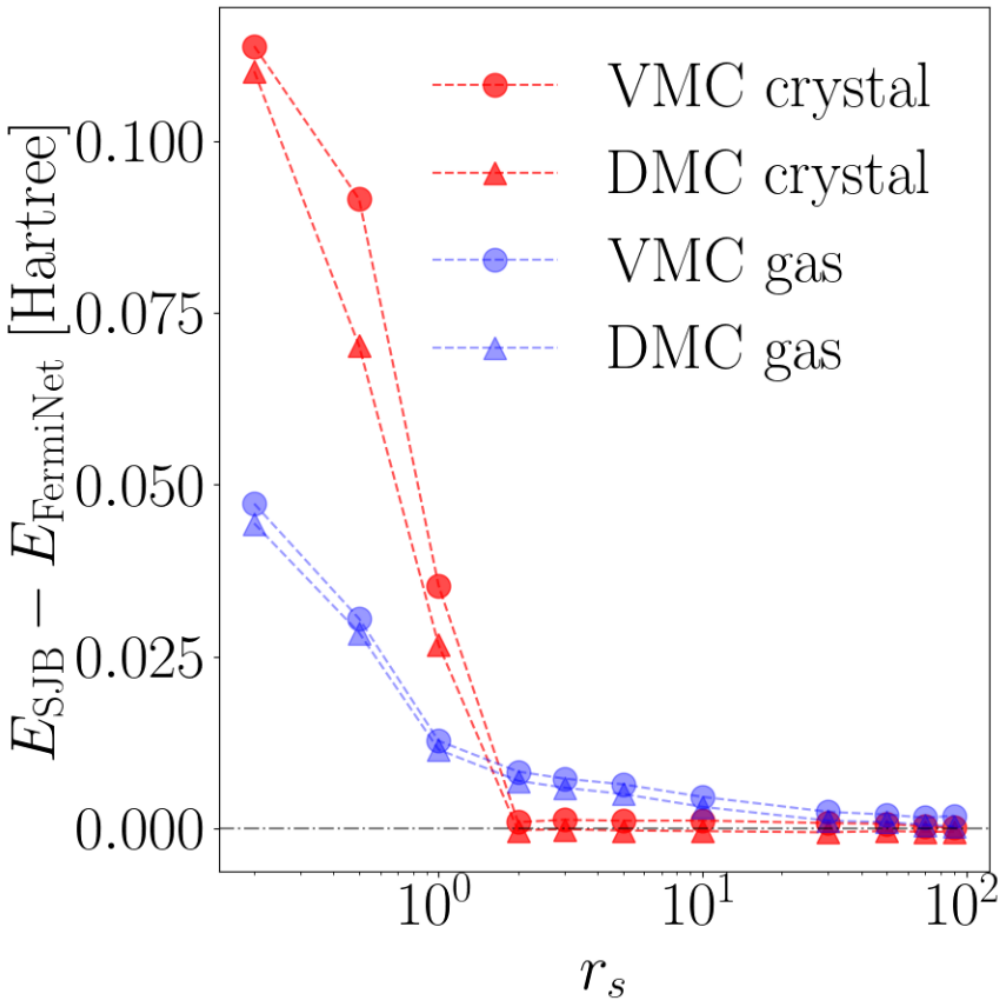


$r_s = 10$ $r_s = 70$

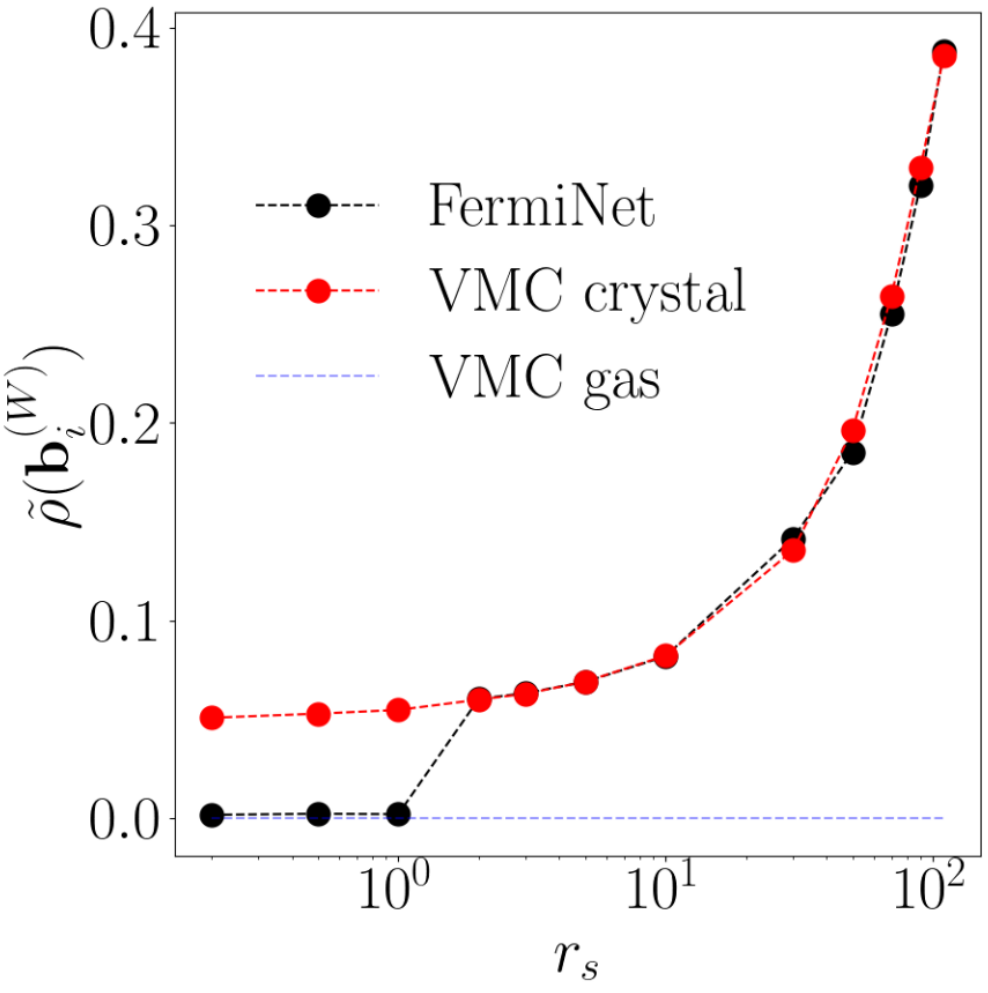
[Phys. Rev. Lett. **130**, 036401 (2023)]

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

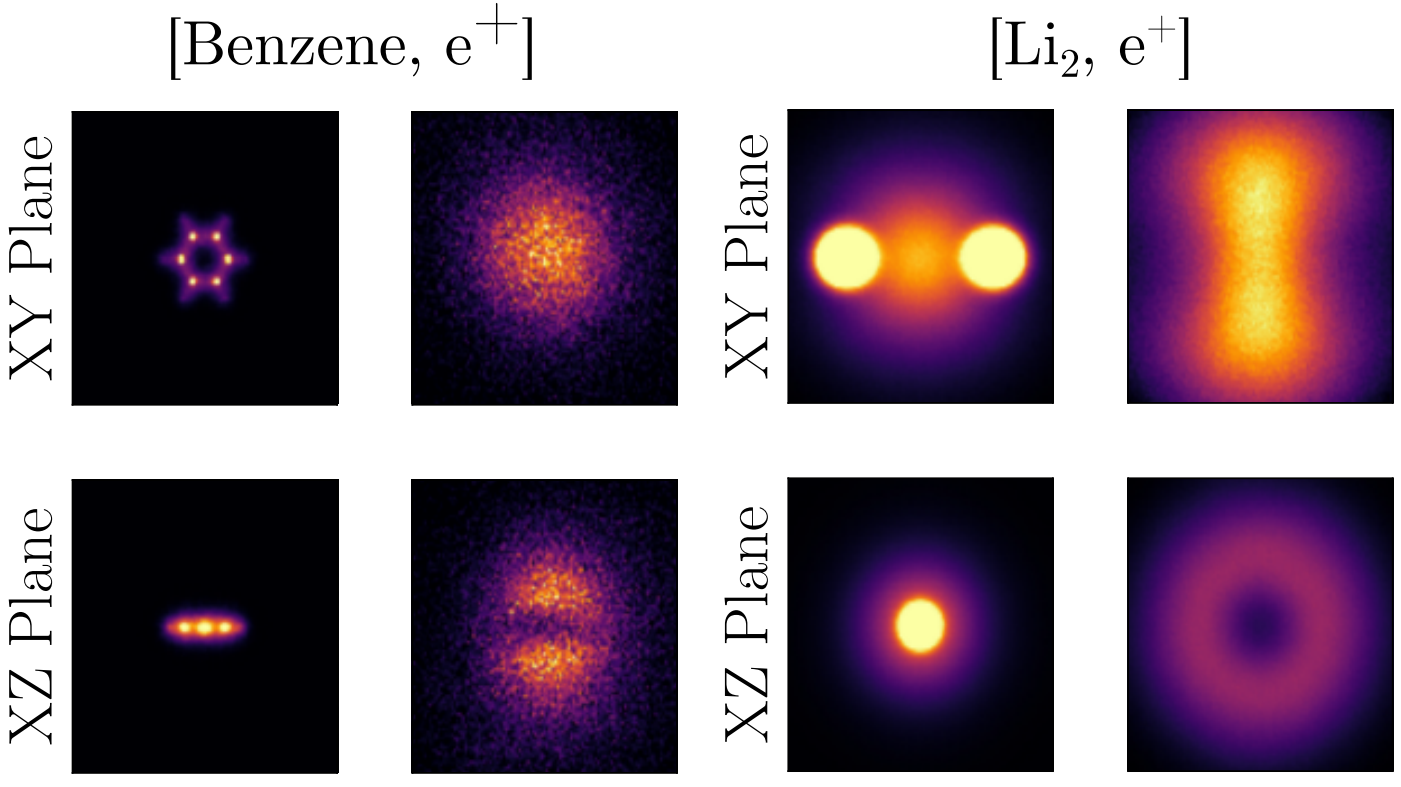
Relative Energies



Order Parameter



Positrons



Nat. Commun. **15**, 5214 (2024)

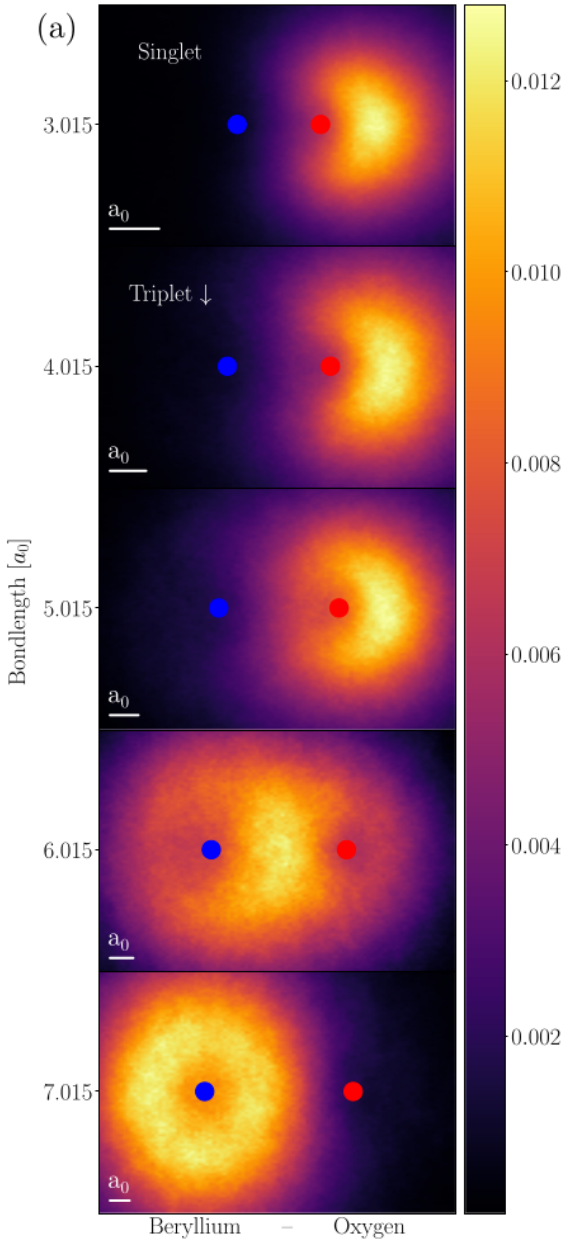
e^+ binding energy (mHa)

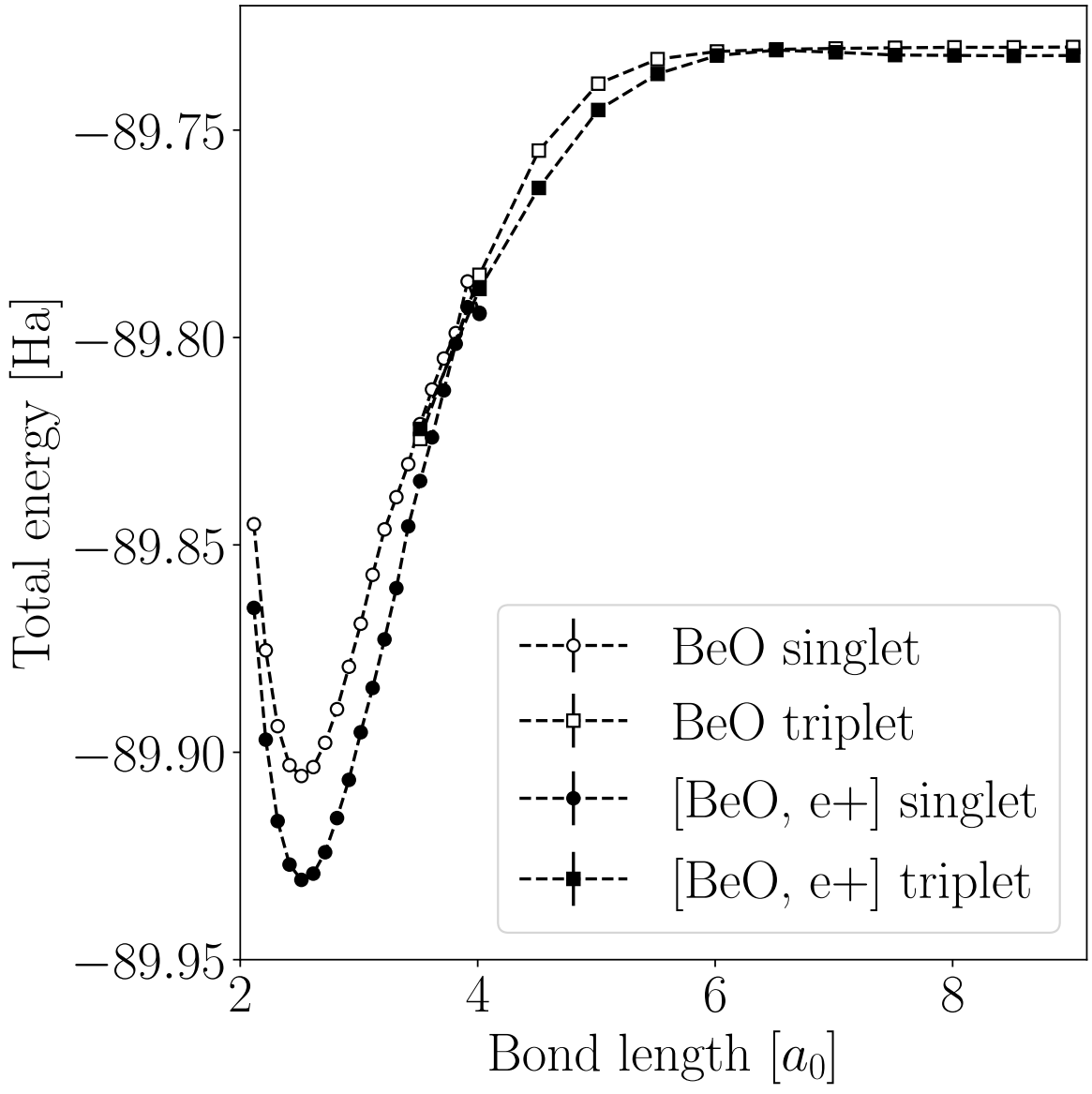
Method	LiH	BeO
FermiNet	37.23	25.10
SJ-VMC	17.27	19
FN-DMC	37.1	28.0
CISD	17	
MRD-CI	29.37	13.78
ECG-SVM	36.93	
GW	39	

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^+]


(a)
Singlet
Triplet ↓
3.015
4.015
5.015
6.015
7.015
a_0
Bondlength [a_0]
Beryllium – Oxygen
0.012
0.010
0.008
0.006
0.004
0.002


−89.75
−89.80
−89.85
−89.90
−89.95
Total energy [Ha]
2
4
6
8
Bond length [a_0]
BeO singlet
BeO triplet
[BeO, e+] singlet
[BeO, e+] triplet

Positives

- Neural wave functions are way better than other approximate wave functions for molecules with more than a few atoms.
- Solid-state simulation cells with $\sim$100 electrons have been simulated with equal success.
- Rival and sometimes outperform the best quantum chemical methods, many of which have been in development for 50+ years.
- Good at dealing with unusual systems (positronic molecules and solids; Wigner crystals) and strong correlations where conventional methods fail or need painful modification.
- Can sometimes discover new phases.

Questions and Negatives

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