DeepMind

A Self-Attention Ansatz for *Ab-initio* Quantum Chemistry

Ingrid von Glehn, James Spencer, David Pfau
{ingridvg, jamessspencer, pfau}@google.com

Quantum Chemistry

The behaviour of molecules and materials at the quantum scale is governed by how electrons and nuclei interact.

The (time-independent) **Schrödinger equation** describes this interaction:

$$\hat{H}\Psi = E\Psi \qquad \hat{H} = -\tfrac{1}{2}\nabla^2 + \sum_{i>j}\frac{1}{r_{ij}} - \sum_{iI}\frac{Z_I}{r_{iI}} + \sum_{I>J}\frac{Z_I Z_J}{r_{IJ}}$$

$\hat{H}$ **Hamiltonian** - kinetic energy of electrons, potential due to electrons and nuclei

Ψ **wavefunction** - function all of electron positions, Ψ^2 is joint probability

E **energy** - energy of the molecule

Problem Many-electron Schrödinger equation is NP hard [1]
Goal find accurate approximate solutions

Idea Use a neural network as a function approximator for the wavefunction [2]

This work explores attention-based network architecture for quantum wavefunctions we call the **Psiformer**: the **most accurate** ansatz for large molecular systems.

Optimization

Variational Monte Carlo (VMC) training loop [3]:

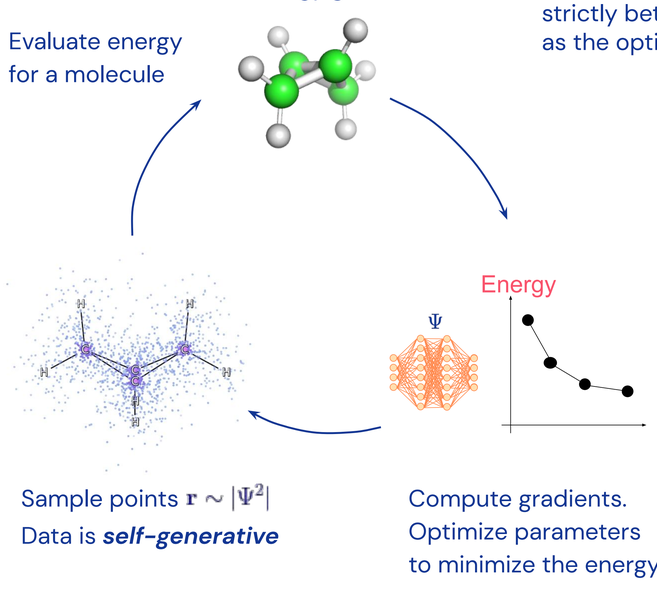
$$E = \frac{\langle\Psi\hat{H}\Psi\rangle}{\langle\Psi^2\rangle} = \mathbb{E}_{\mathbf{r}\sim\Psi^2}\left[\Psi^{-1}(\mathbf{r})\hat{H}\Psi(\mathbf{r})\right] = \mathbb{E}_{\mathbf{r}\sim\Psi^2}\left[E_L(\mathbf{r})\right]$$

Monte Carlo estimate of energy

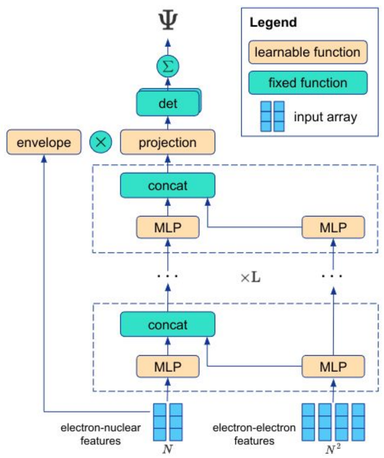
$$\nabla E = \mathbb{E}_{\mathbf{r}\sim\Psi^2}\left[\left(E_L(\mathbf{r}) - \langle E_L\rangle\right)\nabla\Psi(\mathbf{r})\right]$$

Unbiased estimate of energy **gradients**

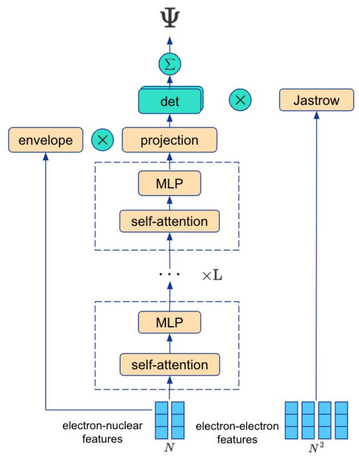
Variational: Minimize upper bound on the energy. Lower is strictly better. We use K-FAC as the optimizer [4].



Architectures



FermiNet [5]



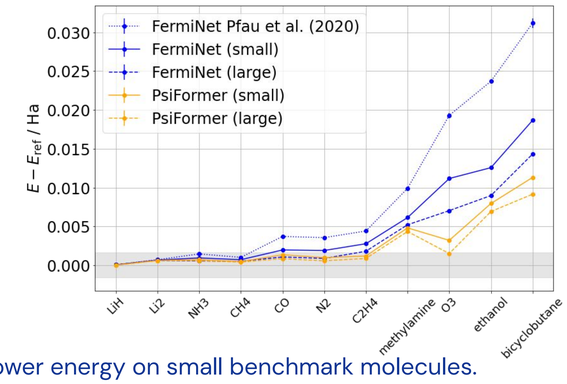
Psiformer

The Pauli exclusion principle requires that the network is antisymmetric with respect ot the electrons

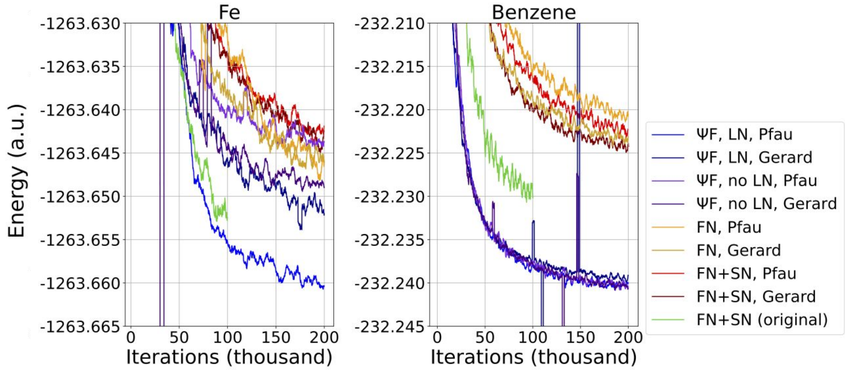
This is achieved using a permutation-equivariant network, with a final determinant.

Code available at https://github.com/deepmind/ferminet

Psiformer vs FermiNet

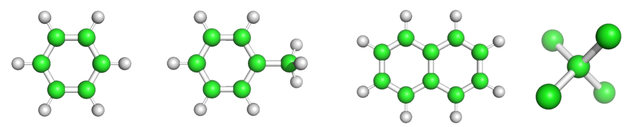


Lower energy on small benchmark molecules.



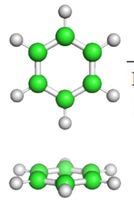
Psiformer performance gap grows for larger systems, even compared to newer FermiNet extensions [6]

Large molecules



System	FermiNet	Psiformer	
		No LayerNorm	LayerNorm
Benzene	-232.2205(2)	**-232.2400(1)**	-232.2393(1)
Toluene	-271.5274(2)	-271.5494(1)	**-271.5538(1)**
Naphthalene	-385.8147(4)	-385.8679(2)	**-385.8685(2)**
CCl_4	-1878.684(1)	-1878.734(1)	**-1878.804(1)**

Ground-state energy of large molecules (lower is better). The Psiformer also outperforms external SotA results.



	FermiNet	Psiformer	
		No LayerNorm	LayerNorm
Equilibrium	-464.3770(5)	-464.4624(2)	**-464.4667(2)**
Dissociated	-464.3724(6)	**-464.4674(2)**	-464.4660(2)
ΔE_{mono}	-0.0640(6)	-0.0176(2)	-0.0119(2)
$\Delta E_{10\text{Å}}$	**0.0046(8)**	-0.0050(3)	0.0007(3)

DMC-level accuracy on the benzene dimer.

Future Work and References

Architecture exploration: how to beat the $O(n^2)$ scaling of self-attention and $O(n^3)$ scaling of determinants without sacrificing accuracy?

Materials: how to handle finite size effects?

[1] M. Troyer and U.-J. Wiese, *Phys. Rev. Lett.* (2005)
[2] J. Hermann *et al.*, arXiv:2208.12590 (2022)
[3] W. M. C. Foulkes *et al.*, *Rev. Mod. Phys.* (2001)
[4] J. Martens, R. Grosse, *ICML* (2015)
[5] D. Pfau, J. S. Spencer *et al.*, Phys. Rev. Research (2020)
[6] L. Gerard, M. Scherbela *et al.*, NeurIPS (2022)