NATIONAL UNIVERSITY OF ENGINEERING
SCIENCE FACULTY
PROFESSIONAL SCHOOL OF PHYSICS

Teaching Quantum Chemistry to a Deep Learning Model Transformers for the many body Schrodinger Equation

Jorge A. Muñoz Laredo*, Daniel A. Paredes Cabrel[1] [1]Docente, Escuela Profesional de Fisica, Facultad de Ciencias
[2]Curso: Algebra Lineal Para Fisicos II
*e-mail jorge.munoz.l@uni.pe

ABSTRACT

Accurate electronic-structure solutions are difficult due to the exponential complexity of the many-electron Schrödinger equation. Neural wavefunction methods offer a flexible alternative but can be sensitive to optimization and numerical stability. In this work we implement **PsiFormer**, a Transformer-based neural wavefunction that takes nuclear charges and positions as input and is trained via variational Monte Carlo. I evaluate the model on the first 8 atoms and report correlation error together with training stability indicators. While the current implementation does not match state-of-the-art accuracy, our experiments identify the dominant bottlenecks, determinant conditioning, and computational weight—and provide a reproducible baseline for future improvements..

INTRODUCTION

The many-electron Schrödinger equation provides an exact description of atomic and molecular systems, but solving it directly becomes computationally intractable due to the exponential growth of the Hilbert space.
Traditional electronic-structure methods such as Hartree–Fock and density functional theory trade accuracy for efficiency, and can struggle in regimes with strong electron correlation.
Recently, Neural Quantum States have emerged as a promising alternative by representing wavefunctions with deep learning models. In this work, I propose **PsiFormer**, a Transformer-based ansatz that combines physical structure with learning to approximate many-electron wavefunctions within a variational Monte Carlo framework.

THEORETICAL FRAMEWORK

I work under the Born–Oppenheimer aproximation:

Schrodinger equation: $\hat{H}\Psi(\mathbf{R})=E\,\Psi(\mathbf{R}),\ \hat{H}=\hat{T}+\hat{V}_{eN}+\hat{V}_{ee}$

Variational principle: $E(\theta)=\dfrac{\langle\Psi_\theta|\hat{H}|\Psi_\theta\rangle}{\langle\Psi_\theta|\Psi_\theta\rangle}\ \geq\ E_0$

Local energy: $E_L(\mathbf{R})=\dfrac{\hat{H}\Psi_\theta(\mathbf{R})}{\Psi_\theta(\mathbf{R})}$

Optimize the loss: $\mathcal{L}_\theta=\mathbb{E}_{\mathbf{R}\sim|\Psi_\theta|^2}[E_L(\mathbf{R})]\approx\dfrac{1}{n}\sum_k E_L(\mathbf{R}_k)$

- Sampling $\mathbf{R}\sim|\Psi_\theta|^2$ via Metropolis-Hasting algorithm.
- Electrons are fermions: swapping two electrons must flip the sign of the wavefunction (Pauli principle).
- Electrons avoid each other: Coulomb repulsion creates strong short-range correlation

METODOLOGY

Goal: Obtain accurate ground state-energies using an Ansatz created with **Torch** library

I parameterize the many-electron wavefunction Ψ_θ with an attention-based **Slater–Jastrow ansatz (PsiFormer)** and optimize it using variational Monte Carlo.

$$\exp\big(\mathcal{J}_\theta(\mathbf{x})\big)\sum\det\big[\mathbf{\Phi}^k_\theta(\mathbf{x})\big]$$

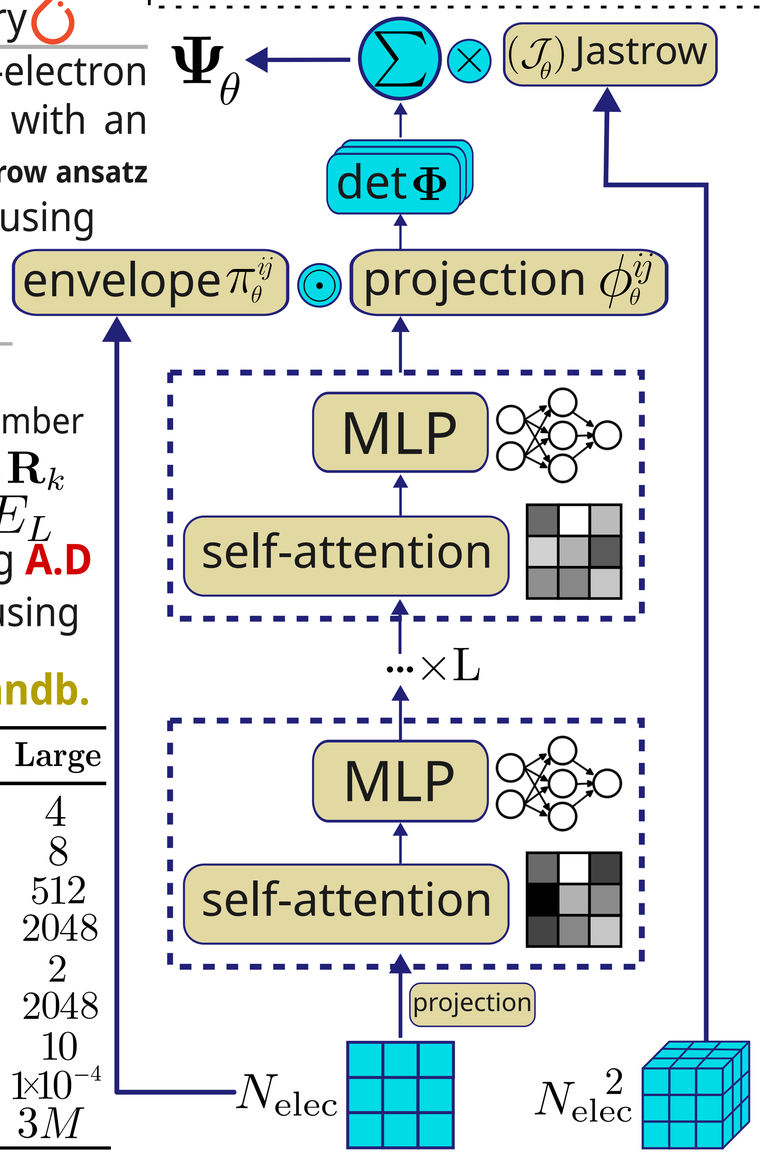

Figure 1. Psiformer VMC pipeline: attention-based ansatz

Training Loop:
Set the spins and electron number
1.Sample configurations $\mathbf{R}_k$
2.Estimate local energy E_L
3.Backpropagation using **A.D**
4.Update parameters θ using **AdamGrad**.
5. Track metrics with **Wandb.**

Hyperparameter	Small	Large
Layers L	2	4
Heads H	4	8
Model Dim d	256	512
MLP Dim d_{ff}	1024	2048
Determinants K	1	2
MCMC walkers N_w	1024	2048
MCMC steps / iter	10	10
Learning rate	2×10^{-4}	1×10^{-4}
Total parameters	50441	$3M$

Table 1. Psiformer hyperparameters (Small vs Large)

RESULTS

PsiFormer (Large) achieves lower VMC energies than PsiFormer (Small) and the baseline across He–O.

Atom	E_b	E_s	E_l	Δ_s	Δ_l	Δ_{l-s}
H	−0.500	−0.492	**-0.498**	0.008	0.002	**-0.006**
He	−2.903	−2.801	**-2.893**	0.102	0.010	**-0.092**
Li	−7.478	−7.097	**-7.243**	0.381	0.235	**-0.146**
Be	−14.667	−13.901	**-14.237**	0.766	0.430	**-0.336**
B	−24.653	−24.042	**-24.567**	0.611	0.086	**-0.525**
C	−37.845	−35.492	**-36.457**	2.353	1.388	**-0.965**
N	−54.589	−50.492	**-51.700**	4.097	2.889	**-1.208**
O	−75.067	−63.492	**-72.139**	11.575	2.928	**-8.647**

Table 2. VMC ground state energies (Ha) for H-O: baseline vs PsiFormer

- Largest gains appear for heavier atoms (C–O), consistent with stronger correlation/complexity.

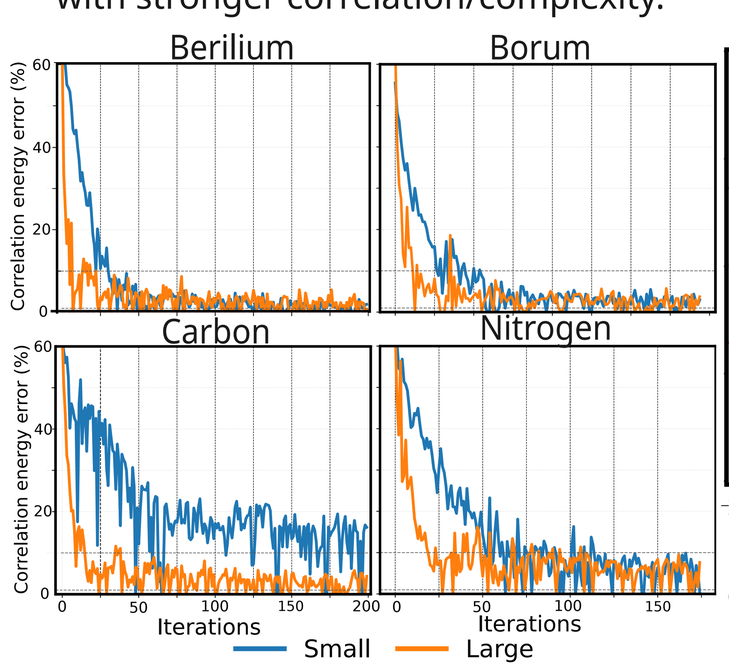

Figure 2. Large converges faster and reaches lower energies than Small

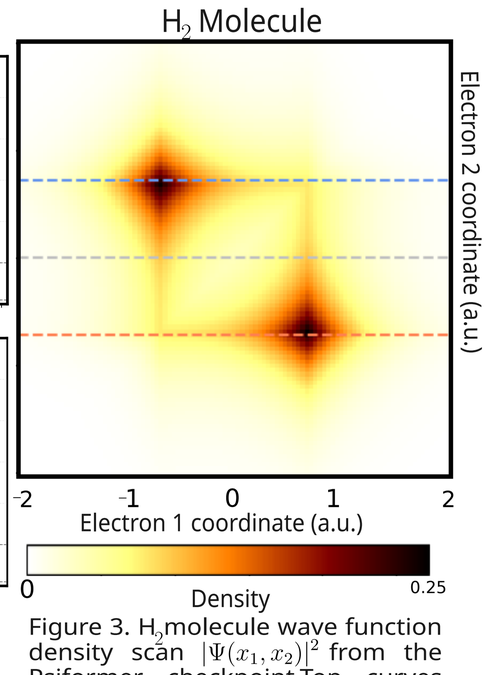

Figure 3. H_2molecule wave function density scan $|\Psi(x_1,x_2)|^2$ from the Psiformer checkpoint.Top curves are density slices at the dashed lines. Image source: **QMC Torch**

CONCLUSIONS

- PsiFormer combines Transformer attention with a Slater–Jastrow wavefunction, preserving key physical constraints (antisymmetry via determinants; correlation via Jastrow).
- Training with VMC + Metropolis sampling produces competitive ground-state energies for light atoms, with PsiFormer (Large) consistently improving over PsiFormer (Small).
- Current limitations include optimization stability and scaling to larger systems; future work will focus on improved sampling/optimization (e.g., better proposals, preconditioning/natural-gradient-style updates) and extension to molecules.

REFERENCES

- David Pfau, James S. Spencer, Alexand G.D. G. Ab initio solution of the many electron Schrodinger equation with deep neural networks, 2020.
- James S. Glehn, Ingrid von Spencer and David Pfau. A self-attention ansatz for ab-initioquantum chemistry, 2023.
- Ashish Vaswani, Noam Shazeer, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. 30, 2017.

ACKNOWLEDGEMENTS

I thank the Computer Science Faculty for providing access to GPU resources, in particular an NVIDIA GeForce RTX 4080 Super, ("Monstrito") which enabled the training and evaluation of PsiFormer models reported in this work.

¡CARRERA ACREDITADA INTERNACIONALMENTE POR ABET 2020-2026!