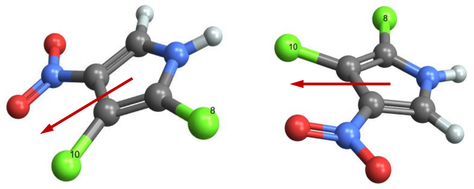
An Equivariant Geometric Machine Learning Model

for mean–field and correlated density matrices prediction

Kecai Xuan[1], Hong–Zhou Ye[1,2]

[1] Department of Chemistry and Biochemistry, University of Maryland, College Park

[2] Institute for Physical Science and Technology, University of Maryland, College Park



Introduction

Density matrices are fundamental to theoretical calculations, yet traditional methods are computationally expensive for large molecular systems. This work presents an equivariant machine learning framework inspired by Graph2Mat[1] that combines MACE[2] and E3NN[3] architectures to predict both mean–field and correlated density matrices directly from molecular geometries. We also introduce several key methodological advances:

- ☑ use of **Gaussian basis sets**;
- ☑ use of superposition of atomic densities (SAD) as a **prior**;
- ☑ extension to **correlated** density matrix learning.

We benchmark our approach through evaluation on Aspirin MD 17 and QM9 with one SCF cycle for mean–field DM and frozen natural orbital (FNO) method for correlated DM .

Theory

1–Particle Reduced Density Matrix (1RDM)

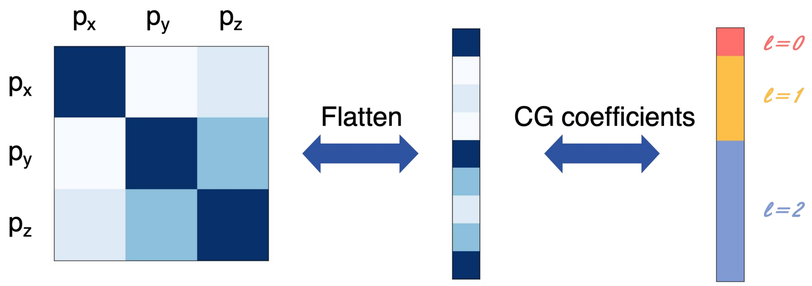
Definition: $P_{\mu\nu} = \langle a_\nu^\dagger a_\mu \rangle \in \mathbb{R}^{\text{basis}\times\text{basis}}$

Calculation: HF, MP2, CCSD, DFT...

Usage: Energy, Dipole, Band Gap, LNO, FNO, ...

Equivariance of the 1RDM

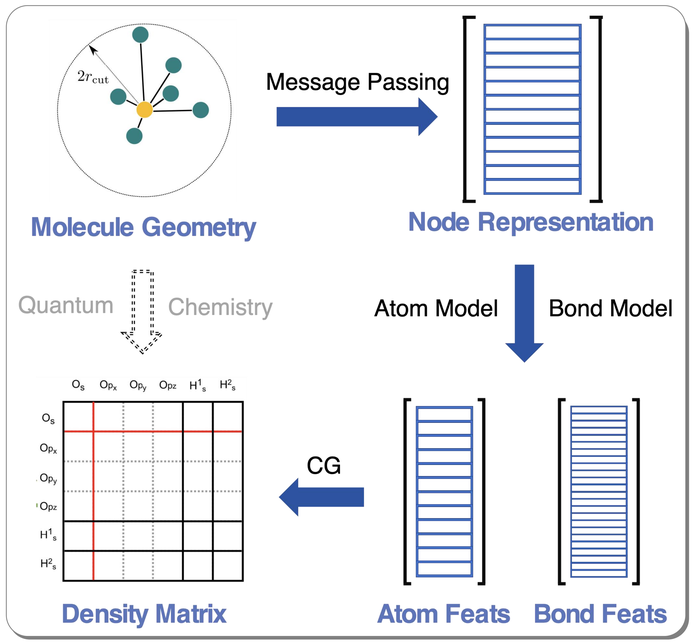
$$F(Q\cdot(\mathrm{r}_1,\mathrm{r}_2,\ldots,\mathrm{r}_n)) = D(Q)F(\mathrm{r}_1,\mathrm{r}_2,\ldots,\mathrm{r}_n)$$



Data Preparation: forward, $P \Longrightarrow \vec{q}$

Model Prediction: backward, $\hat{\vec{q}} \Longrightarrow \hat{P}$

Model — MACE + E3NN



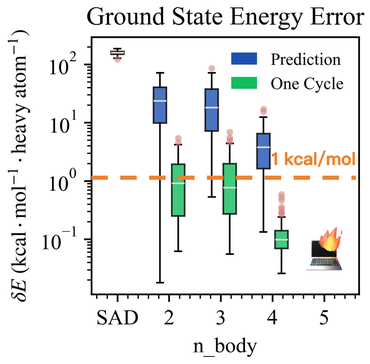
Results

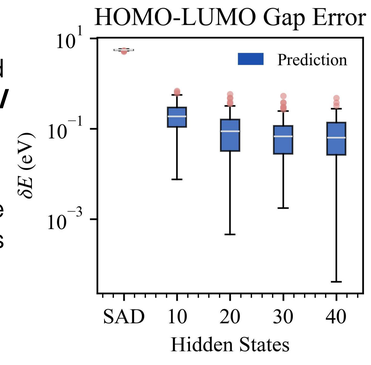
Mean–field 1RDM

Dataset: Aspirin MD 17

Label: HF DM (w/o SAD)

- **1 kcal/mol** (chemical accuracy) accuracy for total energy with up to **4–body** MACE features.
- **0.1 eV** accuracy in band gap using up to ***g–orbital*** (*l*=4) MACE features.
- Similar performance compared to previous work (Graph2Mat).



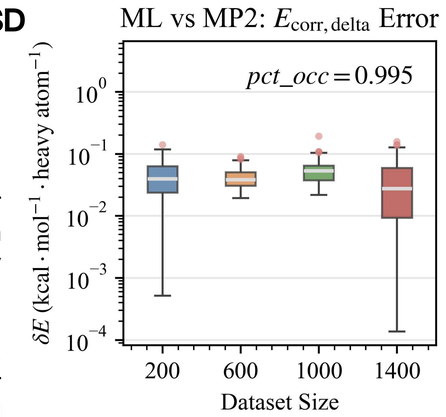


MP2 1RDM & FNO–CCSD

Dataset: Aspirin MD 17

Label: MP2 DM (vv block)

- **0.1 kcal/mol** accuracy for FNO–CCSD E_{corr} on **unseen** configurations using only **200** training configurations.
- Up to **d–orbital** (*l*=2) MACE features are sufficient, lower than HF DM learning (need *l*=4).

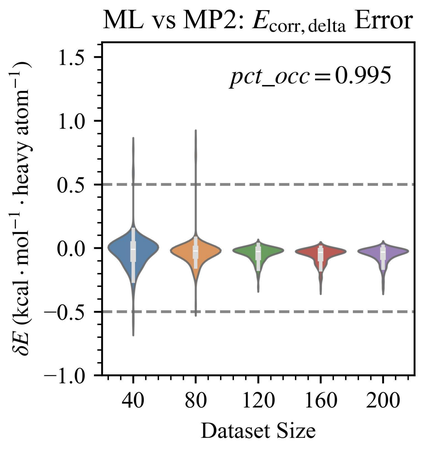


Transferability

Dataset: QM9

Label: MP2 DM (vv block)

- **1 kcal/mol** accuracy for FNO–CCSD E_{corr} on **unseen** molecules using only ~**100** training points.



Conclusion

- **MACE + E3NN** offers a general framework for 1RDM learning.
- **Validated** our framework for both mean–field (HF) and correlated (MP2) 1RDM learning.
- Model predicted HF 1RDM leads to accurate total *and* orbital energy after taking a single SCF step.
- Model predicted MP2 1RDM leads to accurate reduced–scaling CCSD calculations as demonstrated for FNO–CCSD.

References

[1] Pol Febrer *et al 2025 Mach. Learn.: Sci. Technol.* 6 025013

[2] Ilyes Batatia *et al*, *NeurIPS 2022. arXiv:2206.07697.*

[3] Mario Geiger, Tess Smidt, *arXiv:2207.09453*