OAK RIDGE National Laboratory

Predicting Quantum Monte Carlo Charge Densities using Graph Neural Networks

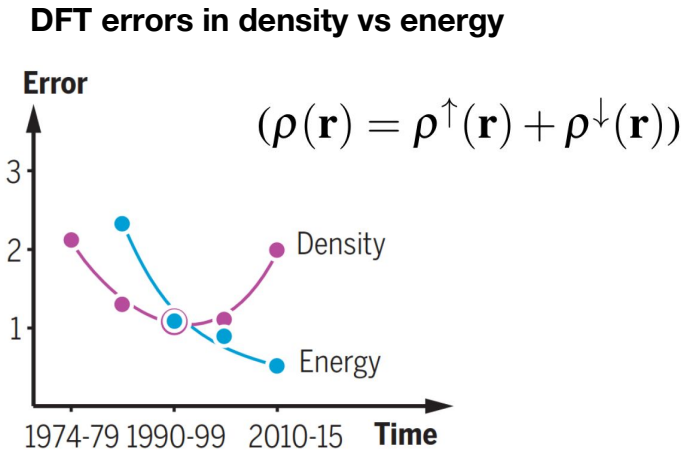
Gani Annaberdiyev[1], Fan Shu[2], Victor Fung[2], P. Ganesh[1]

1) Oak Ridge National Laboratory 2) Georgia Institute of Technology

2024 Artificial Intelligence for Materials Science (AIMS) Workshop | July 17 - 18, 2024
National Cybersecurity Center of Excellence (NCCoE) | Rockville, MD

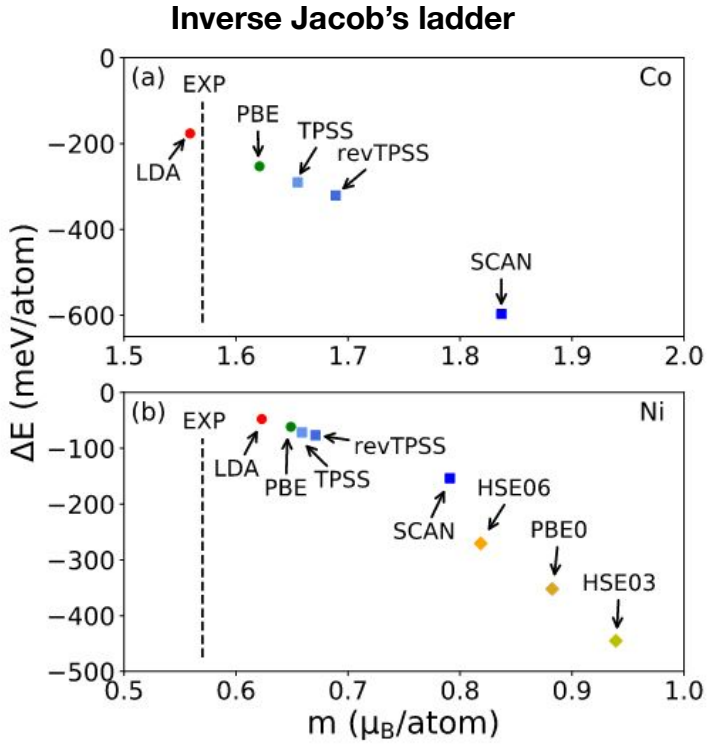
Why Charge Densities?

- Electronic charge density is a fundamental quantity from which all ground state properties can be obtained. In other words, the true electronic density uniquely delivers the ground state energy (Hohenberg-Kohn, 1964 [1]).



$$(\rho(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) + \rho^{\downarrow}(\mathbf{r}))$$

Figure: DFT errors of energy have improved over time yet yield lower accuracy for corresponding electron densities [average median normalized error from literature]. Taken from Ref. [2].



$$(m(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) - \rho^{\downarrow}(\mathbf{r}))$$

Figure: Magnetic energy and spin magnetization per atom of hcp Co (top) and fcc Ni (bottom) as obtained with different functionals. The LDA, PBE GGA, and meta-GGA calculations were performed with VASP , and the hybrid functional calculations were performed with WIEN2K. Taken from Ref. [3].

Advanced density functionals do not seem to always produce improved charge- and spin-densities -> need for more accurate methods at comparable computational cost -> ML surrogates for densities with high accuracy.

1. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. Phys. Rev. 136, B864–B871 (1964).
2. Huang, B., von Rudorff, G. F. & von Lilienfeld, O. A. The central role of density functional theory in the AI age. Science 381, 170–175 (2023).
3. Fu, Y. & Singh, D. J. Density functional methods for the magnetism of transition metals: SCAN in relation to other functionals. Phys. Rev. B 100, 045126 (2019).

Applications of accurate charge density

Spin gaps of $Fe[(NCH)_6]^{2+}$ complex

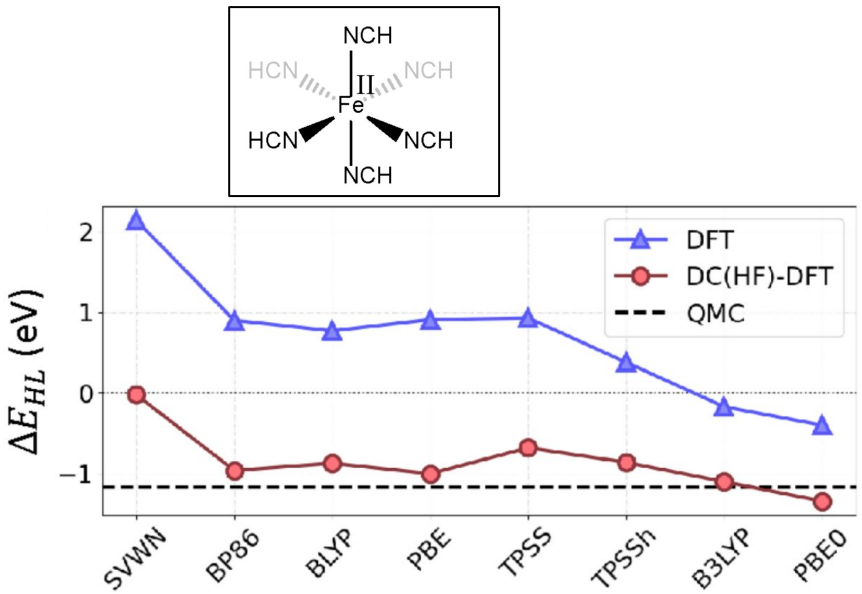


Figure: Self-consistent DFT vs DC(HF)-DFT results for the spin gap of $Fe[(NCH)_6]^{2+}$ complex. Energy difference between high- and low-spin state ($\Delta E_{HL} = E_{HS} - E_{LS}$) for various functionals. Basis set cc-pVQZ. Taken from Ref. [1]

Speed of converging SCF using various initializations

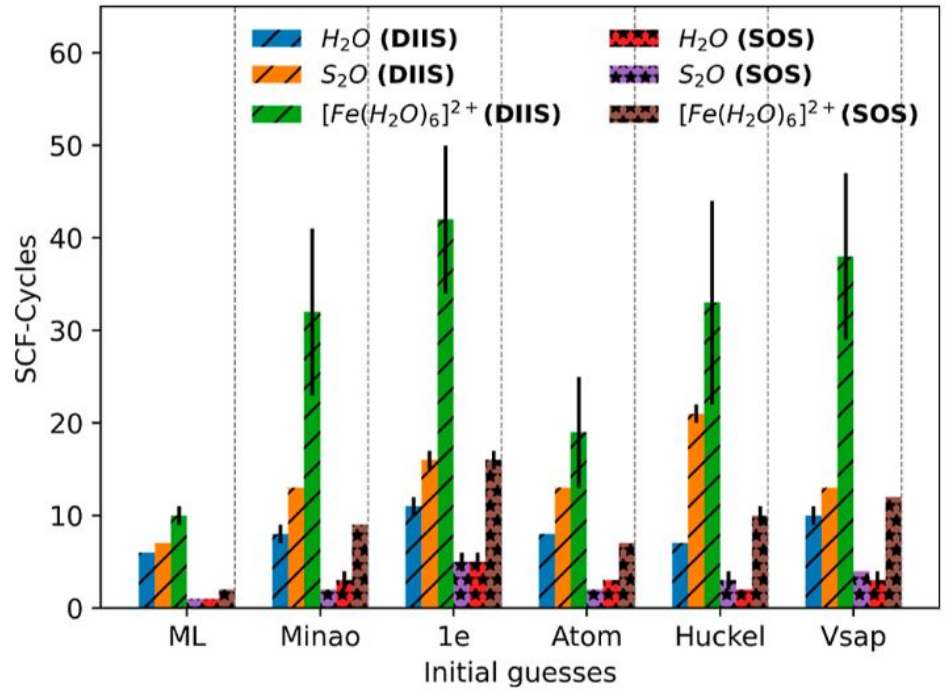


Figure: Total average number of SCF steps taken to achieve convergence for different starting DMs and mixing schemes. Here, the convergence criterion is on the total energy between two consecutive SCF steps that should be lower than 10^{-9} Ha. The black bars around the mean indicate the variance. Variances lower than one iteration are not displayed.Taken from Ref. [2]

Accurate ML charge-densities can facilitate a wide range of properties, such as 1RDM, orbital occupancies, magnetism, spin-gaps, etc.

1. Sim, E., Song, S., Vuckovic, S. & Burke, K. Improving Results by Improving Densities: Density-Corrected Density Functional Theory. J. Am. Chem. Soc. 144, 6625–6639 (2022).
2. Hazra, S., Patil, U. & Sanvito, S. Predicting the One-Particle Density Matrix with Machine Learning. J. Chem. Theory Comput. 20, 4569–4578 (2024).

Reaching accurate charge densities using Quantum Monte Carlo (QMC)

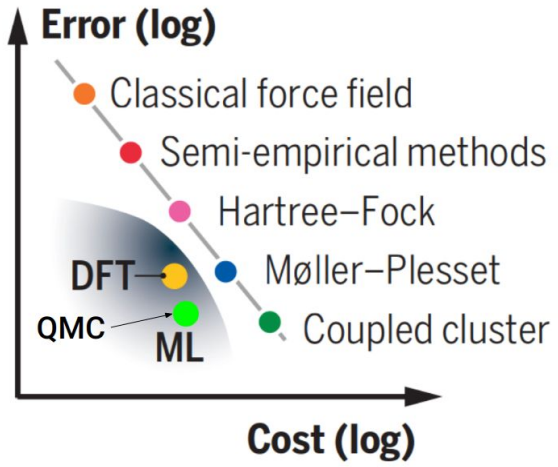


Figure: Budget-aware computing strategies for sampling chemical compound space. QMC has the same scaling as DFT ($O(n^3)$). Taken and modified from Ref. [1]

Table: Cohesive energy [eV] of **Si** obtained from DMC/PBE0 data compared with independent calculations and with the experimental value. All values were corrected by zero-point energy (0.06 eV) to correspond to the bottom of the interaction well (D_e). Taken from Ref. [2]

Method	Cohesion (eV)
DMC/PBE0	4.629(2)
DMC/PBE0	4.683(3)
DFT/LDA	5.1
MP2	5.05
MP2	4.96
CCSD	4.15
DMC/LDA	4.57(3)
VMC/LDA	4.54(1)
DMC/LDA	4.69(1)
DMC	4.68(2)
Experiment	**4.68(8)**

Table: Cohesive Energies [eV] of **Al** Predicted by Various Methods. The theoretical values do not include the zero-point-energy (ZPE) contributions. The experimentally reported number (327.320 ± 4.2 kJ/mol) was corrected for ZPE (0.04 eV) for consistency with the theory. Taken from Ref. [3]

method	E_{coh} [eV]	refs
LDA	3.884	this work
HF	1.388	ref 52
CCSD	2.966	ref 52
$CCSD(T)_{SR}$	3.102	ref 52
CCSD-SVC	3.347	ref 52
DMC/SJ	3.341(1)	ref 8
DMC/BF	3.403(1)	ref 8
DMC/SJ	3.438(2)	this work
experiment	3.432(44)	ref 59

In addition to isolated systems, QMC is routinely applied to solids, where other high-accuracy methods such as CCSD are still challenging (*k*-point convergence).

1. Huang, B., von Rudorff, G. F. & von Lilienfeld, O. A. The central role of density functional theory in the AI age. Science 381, 170–175 (2023).
2. Annaberdiyev, A., Wang, G., Melton, C. A., Bennett, M. C. & Mitas, L. Cohesion and excitations of diamond-structure silicon by quantum Monte Carlo: Benchmarks and control of systematic biases. Phys. Rev. B 103, 205206 (2021).
3. Annaberdiyev, A., Ganesh, P. & Krogel, J. T. Enhanced Twist-Averaging Technique for Magnetic Metals: Applications Using Quantum Monte Carlo. J. Chem. Theory Comput. 20, 2786–2797 (2024).

Efficient evaluation and prediction of charge densities using ML

Machine learning approaches to learn electron densities:

- Expand the density in terms of a basis, and predict coefficients:
  - Kernel ridge regression
  - Multi-centered atomic basis

- Directly predicting the density on a grid:
  - Invariant graph neural networks (inGNN)
  - Equivariant graph neural networks (**eqGNN**)
  - eqGNN predicts charge densities with **high accuracy** and **linear scaling, O(n)**.

- We take inspiration from Jorgenson's [1, 2] message-passing networks, and implement a new ML approach in pytorch to learn densities (Ref [3]).

1. Jørgensen, P. B. & Bhowmik, A. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. npj Comput Mater 8, 1–10 (2022).
2. Jørgensen, P. B. & Bhowmik, A. DeepDFT: Neural Message Passing Network for Accurate Charge Density Prediction. Preprint at https://doi.org/10.48550/arXiv.2011.03346 (2020).
3. https://github.com/Fung-Lab/MatDeepLearn_dev/tree/feature/chg_predict_fan

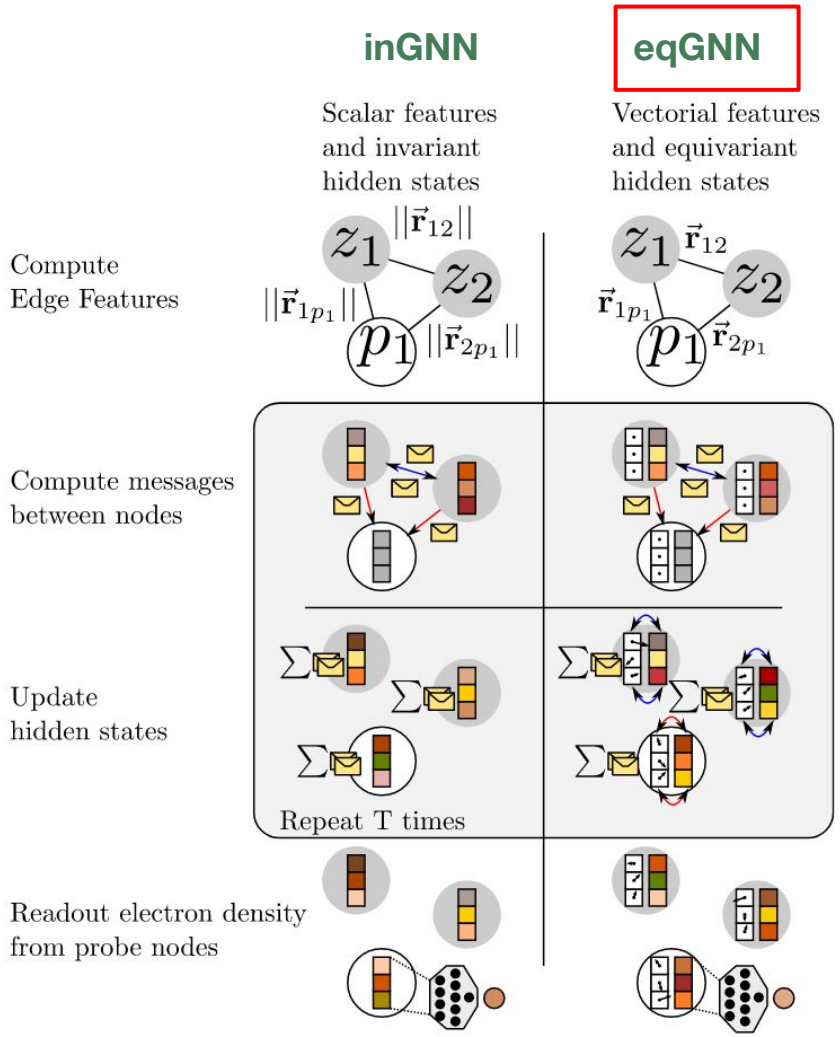


Figure: Conceptual overview of the two message passing architectures used in DeepDFT. Taken from Ref. [1]

Architecture of the Implemented ML Model

- The GNN is purely data-driven: inputs are geometry and atomic numbers only.
- The message passing model architecture used is an equivariant transformer based on the TorchMD-NET model. Ref. [2].
- The model uses 5 message passing layers with a hidden dimension of 64 and SiLU activation functions.
- The readout function is a fully connected neural network with 1 layer and a hidden dimension of 64.
- The model was trained with a learning rate of 0.001 with the Adam optimizer and trained for 500 epochs.
- A cutoff radius of 12 Ang was used to construct the graphs.

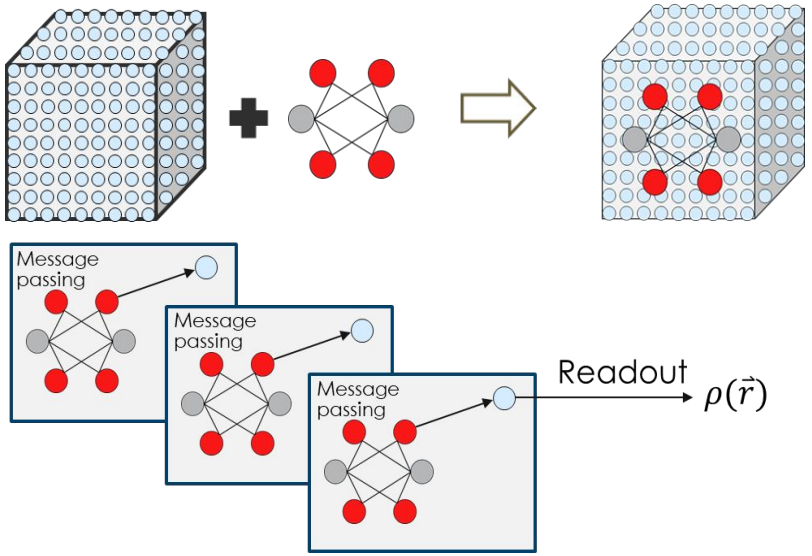


Figure: Conceptual overview of the GNN architecture used in this work.

How well can we learn the different densities (charge/spin/energy)? Can a single architecture be used to learn all quantities? Can we find insights from these ML-models into how density improvements might also lead to energy improvements in constructing functionals?

1. Thölke, P. & De Fabritiis, G. TorchMD-NET: Equivariant Transformers for Neural Network based Molecular Potentials. Preprint at https://doi.org/10.48550/arXiv.2202.02541 (2022).

Training data for learning charge densities

- QM9 dataset. H, C, N, O elements. 2305 molecules in total. Dense 0.04 Ang grid length.
- We targeted closed-shell singlet (1316) and open-shell triplet (989) states to obtain spin-densities in addition to charge densities.
- Available in our dataset: charge and spin densities, DFT/QMC total and kinetic energies, QMC variances and errors.
- Although DFT is commonly used for training dataset, Jørgensen et al. mention [1, 2] that the **eqGNN errors in QM9 are lower than the DFT density variations due to various XC functionals!** This calls for generating datasets with more accurate methods.
- Use PBE to generate the reference orbitals.
- We use real-space variational Monte Carlo (**VMC**) with single-reference Slater-Jastrow wave functions, which recovers ~80% of the correlation energy.

Variational Monte Carlo

$$\Psi(\mathbf{X}) = e^{J(\mathbf{X})} D(\mathbf{X})$$

$$E_T = \frac{\langle \Psi_T | \hat{H} | \Psi_T \rangle}{\langle \Psi_T \mid \Psi_T \rangle} = \frac{\int d\mathbf{R} \Psi_T^*(\mathbf{R}) \hat{H} \Psi_T(\mathbf{R})}{\int d\mathbf{R} \, |\Psi_T(\mathbf{R})|^2} = \int \left(\frac{|\Psi_T|^2}{\int |\Psi_T|^2 \, d\mathbf{R}} \right) \frac{\hat{H}\Psi_T}{\Psi_T} d\mathbf{R}$$

$$\Psi_T(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \in \mathbb{R}^{3N}$$

$$E_T = \frac{1}{M} \sum_{m=1}^{M} \frac{\hat{H}\Psi_T(\mathbf{R}_m)}{\Psi_T(\mathbf{R}_m)} + \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \geq E_{exact}$$

Density Sampling

$$\rho_{\mathrm{DFT}}^{\uparrow/\downarrow}(\mathbf{r}) = \sum_{i}^{N_{\uparrow/\downarrow}} |\phi_i(\mathbf{r})|^2 \qquad \tilde{\rho}_{\mathrm{QMC}}^{\uparrow/\downarrow}(\mathbf{r}) = \sum_{i}^{M_{\uparrow/\downarrow}} \frac{\delta(\mathbf{r} - \mathbf{r}_i)}{L^3}$$

1. Jørgensen, P. B. & Bhowmik, A. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. npj Comput Mater 8, 1–10 (2022).
2. Jørgensen, P. B. & Bhowmik, A. DeepDFT: Neural Message Passing Network for Accurate Charge Density Prediction. Preprint at https://doi.org/10.48550/arXiv.2011.03346 (2020).

Comparison of DFT vs. QMC energetics for QM9 data

Kinetic energy/electron

$$K_{\mathrm{DFT}} = -\frac{1}{2}\sum_{i=1}^{N_{\mathrm{occ}}} \langle \psi_i | \nabla^2 | \psi_i \rangle \qquad K_{\mathrm{QMC}} = -\frac{1}{2}\frac{\langle \Psi_T | \nabla^2 | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle}$$

Singlet-triplet gaps

$$G = E_{\mathrm{triplet}} - E_{\mathrm{singlet}}$$

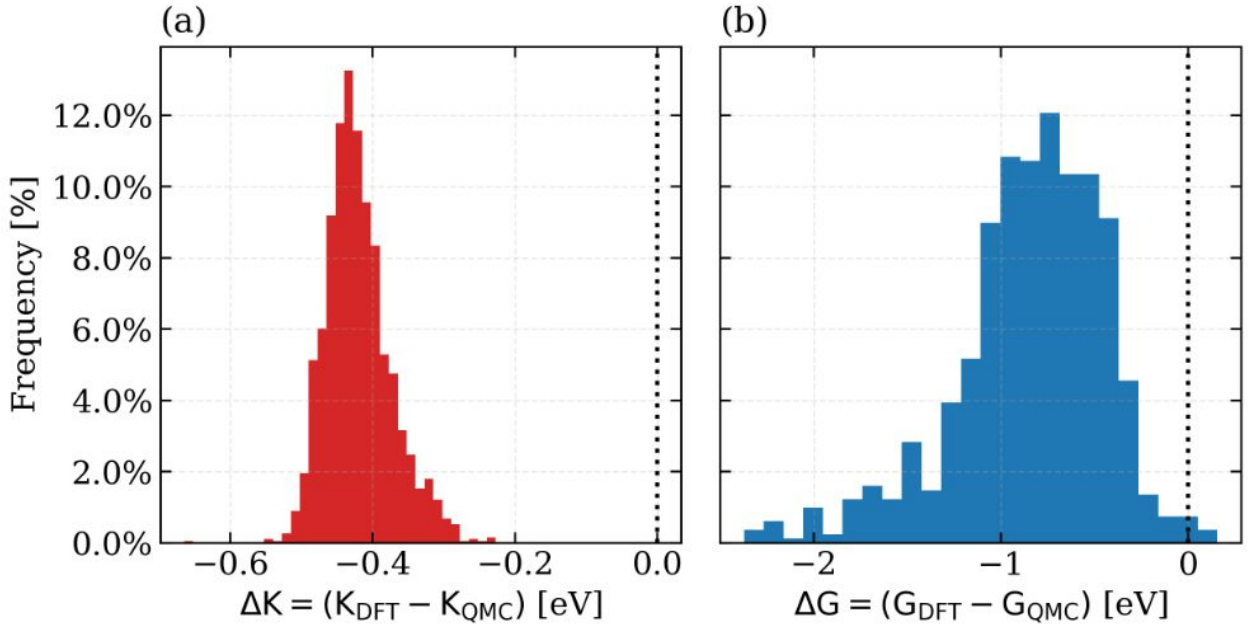
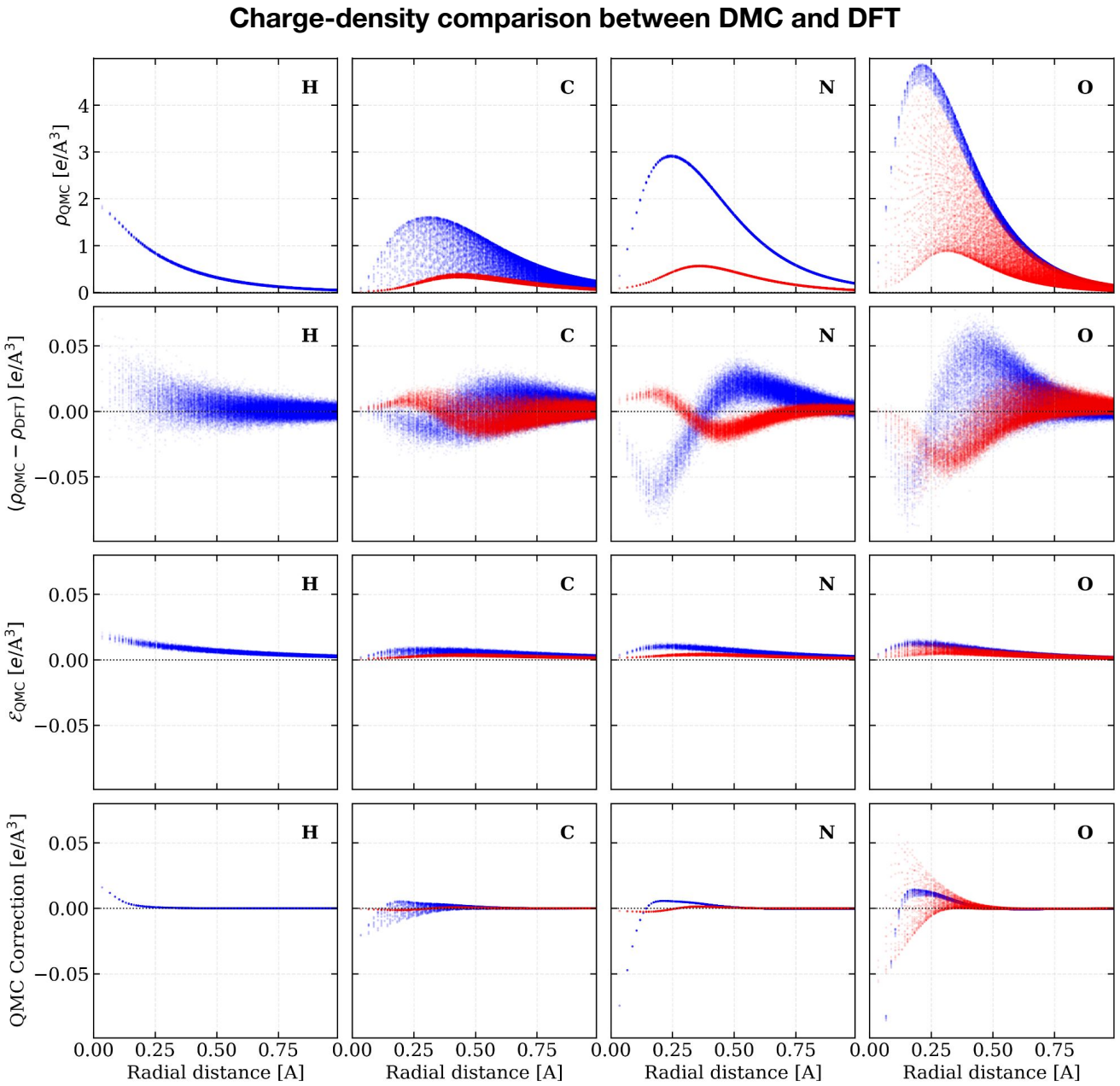


Figure: Comparison of QMC and DFT energetics over the whole dataset. (a) Difference of kinetic energies per electron, (b) difference of singlet-triplet gaps.

Kinetic energies and singlet-triplet gaps are underestimated in PBE.

Accurate charge densities from QMC



Charge-density comparison between DMC and DFT

Figure: A study of H, C, N, and O isolated atomic QMC charge densities with respect to radial distance from the atom. Each dot represents a grid point. Blue/red colors correspond to up/down spin channels, respectively. Each row represents a different quantity, and each column is an element.

The QMC densities were corrected as follows:

$$\rho_{\mathrm{QMC}}^{\uparrow/\downarrow}(\mathbf{r}) = \tilde{\rho}_{\mathrm{QMC}}^{\uparrow/\downarrow}(\mathbf{r}) + \alpha[\rho_{\mathrm{DFT}}^{\uparrow/\downarrow}(\mathbf{r}) - \tilde{\rho}_{\mathrm{DFT}}^{\uparrow/\downarrow}(\mathbf{r})]$$

$$\tilde{\rho}_{\mathrm{DFT}}^{\uparrow/\downarrow}(\mathbf{r}) = \sum_{i,j,k} \frac{\rho_{\mathrm{DFT}}^{\uparrow/\downarrow}(\mathbf{r} + \Delta x_i + \Delta y_j + \Delta z_k)}{27}$$

Most DFT vs. QMC discrepancies in the density occur closer to the core regions (r < 0.5 Ang). This points to relatively larger error in the kinetic energy density.

Learning QMC densities using our eqGNN model

We observed that the eqGNNs prediction errors decrease rapidly with the training dataset size, with apparent plateau at ~1000 molecules.

Learning rate of our eqGNN model for charge-density dataset

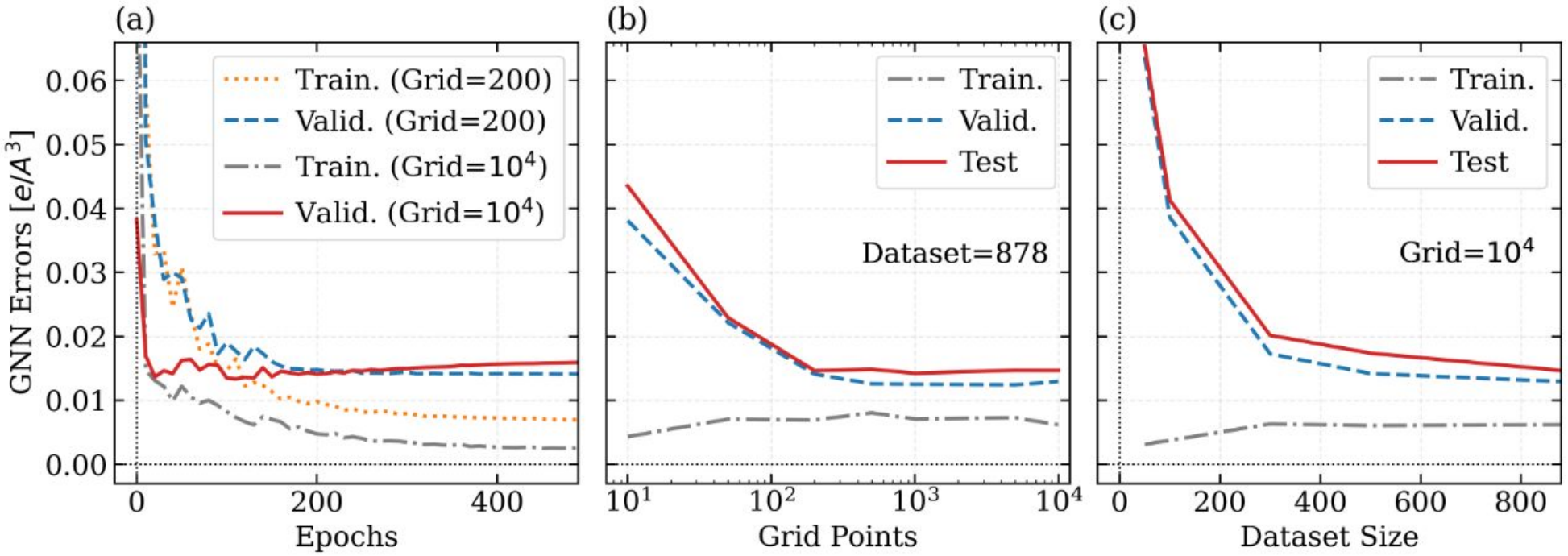


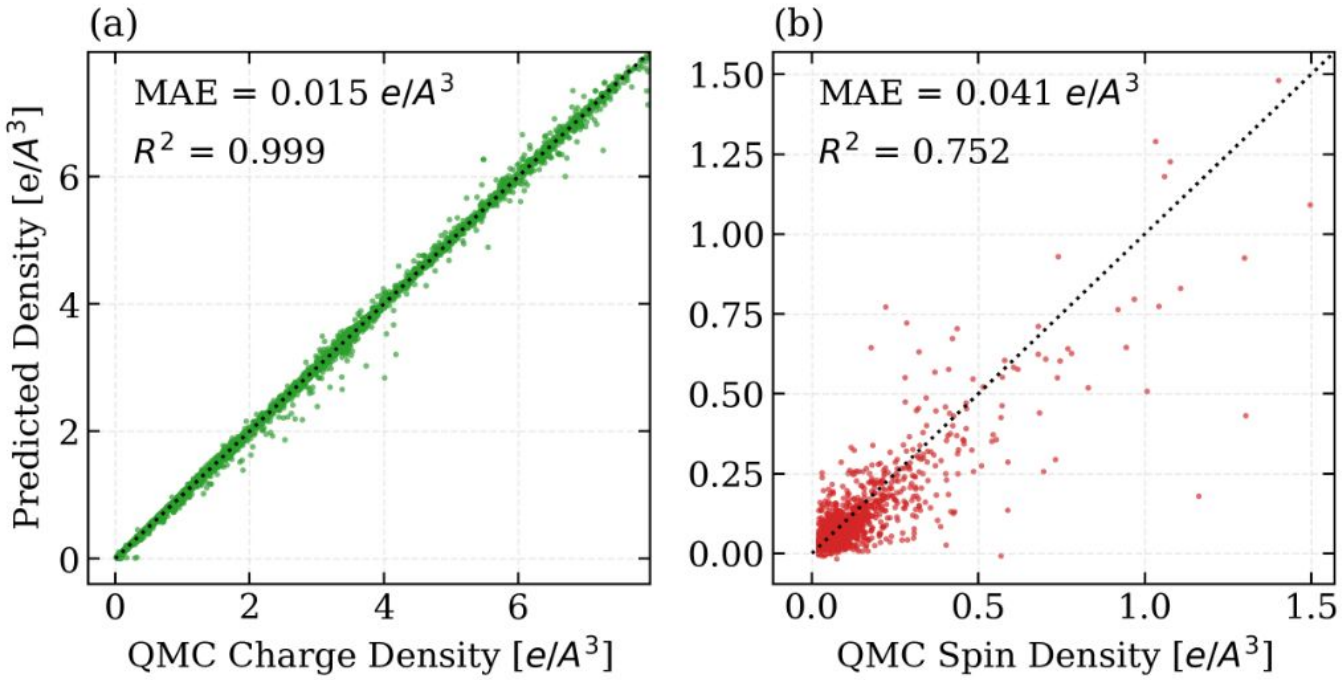
FIG. 4: GNN charge-density errors for training, validation, and test datasets for various optimization parameters. (a) Training and validation dataset errors during the optimization as the epoch number increases. Curves for sampling with 200 and 10^4 grid points for each molecule are shown. (b) The dependence of GNN errors on the size of the grid sampling for each molecule. 878 molecules were used in this training. (c) The dependence of GNN errors on the size of the training dataset. 10^4 grid points were sampled from each molecule.

Performance of the QMC-trained eqGNN in QM9

Charge Density

$$(\rho(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) + \rho^{\downarrow}(\mathbf{r}))$$

A low mean absolute error (MAE) of ~0.015 e/A^3 was achieved despite the small dataset size (~1000 singlet molecules). Here, MAE is evaluated over the test dataset (90/5/5)%.



Spin Density

$$(m(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) - \rho^{\downarrow}(\mathbf{r}))$$

The errors in the spin density are much larger when compared to charge density.

FIG. 5: Correlation plots for GNN density prediction errors versus the QMC densities using the test datasets. (a) Accuracy of the charge density predictions for the singlet molecules, (b) accuracy of the spin density predictions for the triplet molecules. The dashed diagonal line represents the ideal case. Only densities larger than 0.02 e/A^3 were considered for these plots.

Origin of spin density learning challenge

Consider spherically-averaged densities around the atoms from randomly chosen molecules:

Charge Density $(\rho(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) + \rho^{\downarrow}(\mathbf{r}))$

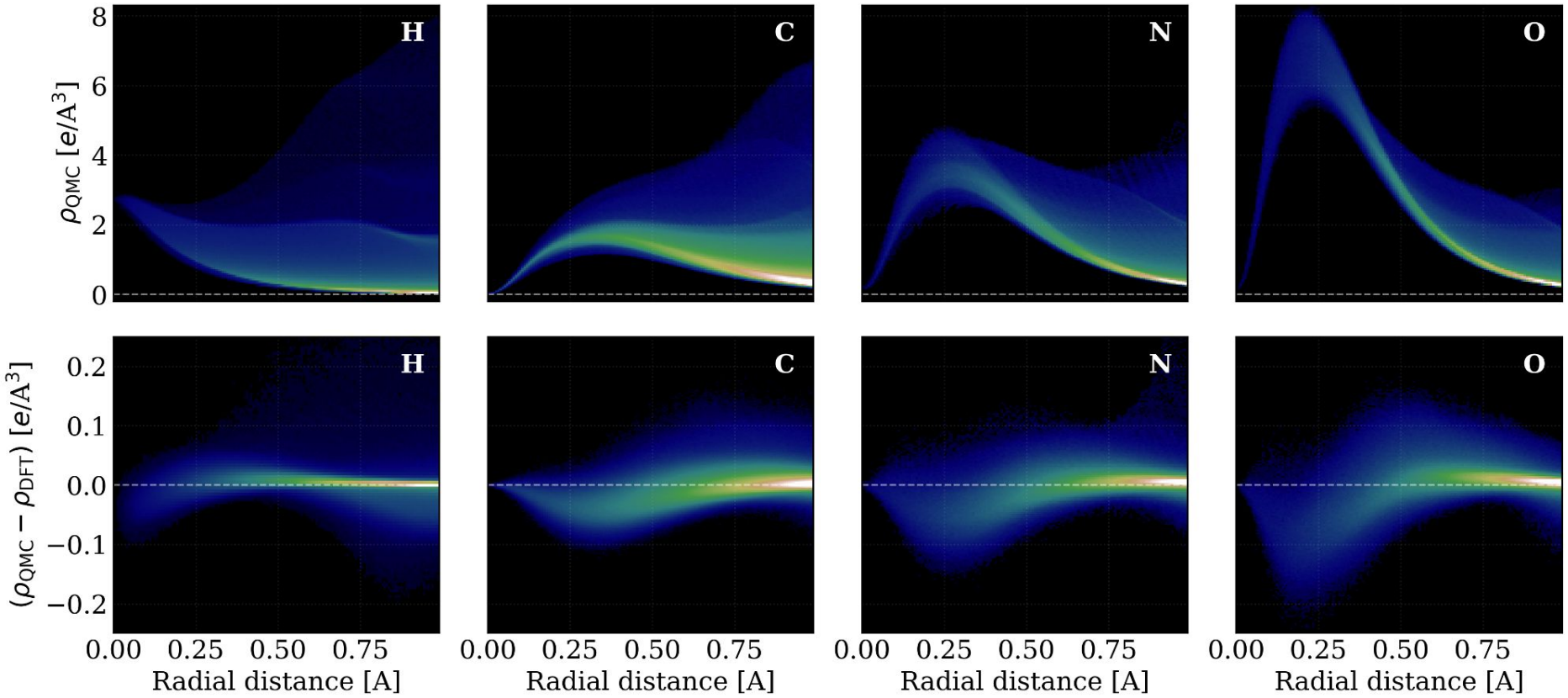


Figure: A study of H, C, N, and O charge densities with respect to radial distance from the atom in a molecular environment. The heatmap plot shows the frequency of errors obtained by accumulating densities over the test dataset (64 random molecules). Each row represents a different quantity, and each column is an element.

Origin of spin density learning challenge

Spin Density $(m(\mathbf{r}) = \rho^{\uparrow}(\mathbf{r}) - \rho^{\downarrow}(\mathbf{r}))$

Charge within R_c: ~constant (Charge conservation)

Moment within R_c: ~$2/N_{atoms}$ (2S=2 for triplets)

High variability in spin-density within the QM9 dataset renders it problematic to learn

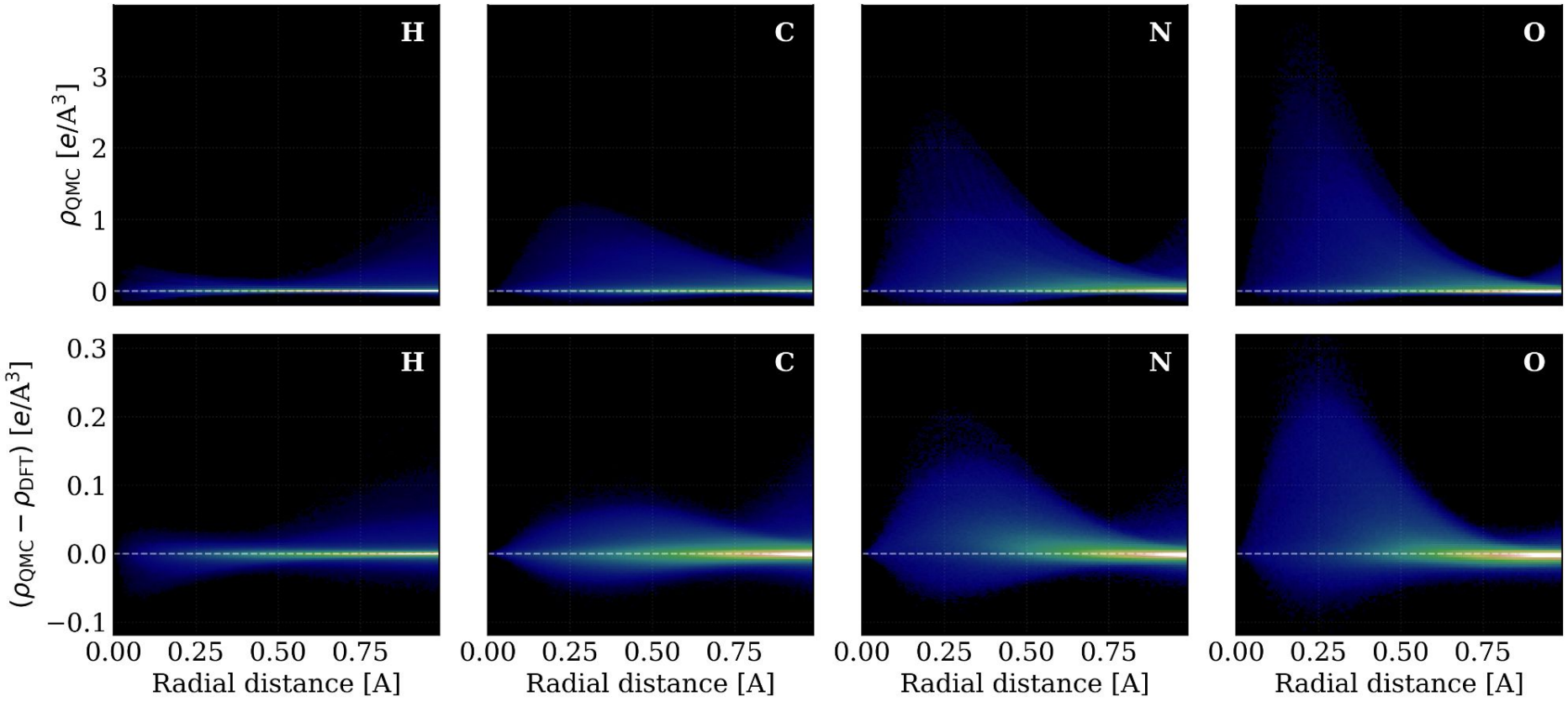


Figure: A study of H, C, N, and O spin densities with respect to radial distance from the atom in a molecular environment. The heatmap plot shows the frequency of errors obtained by accumulating densities over the test dataset (64 random molecules). Each row represents a different quantity, and each column is an element.

Challenges in QMC and ML methods

- Sampling the electronic density using delta functions $\delta(\boldsymbol{r}-\boldsymbol{r}_i)$ results in averaged density in the histogram grid, which makes it challenging to compare against DFT or determine the density value at a specific point.
- A related challenge is the density regions far away from the molecule, where the QMC walkers will seldomly visit. If all-electron is used, density in the core will have cusps, again making histogram accumulation challenging.
- Regularization of the delta function as proposed by Assaraf et al [1]. will help to address the above issues.
- The accuracy of Slater-Jastrow QMC in bulk, magnetic metallic systems is still not well known.
- Although QMC scales as $O(n^3)$, same as DFT, the prefactor is much larger. Thus the cost of QMC calculations should still be considered/estimated before large scale ML training.

$$\delta(\mathbf{r}_i - \mathbf{r}) = \frac{-f(\mathbf{r}_i;\mathbf{r})}{4\pi}\nabla_i^2 \frac{1}{|\mathbf{r}_i - \mathbf{r}|},$$

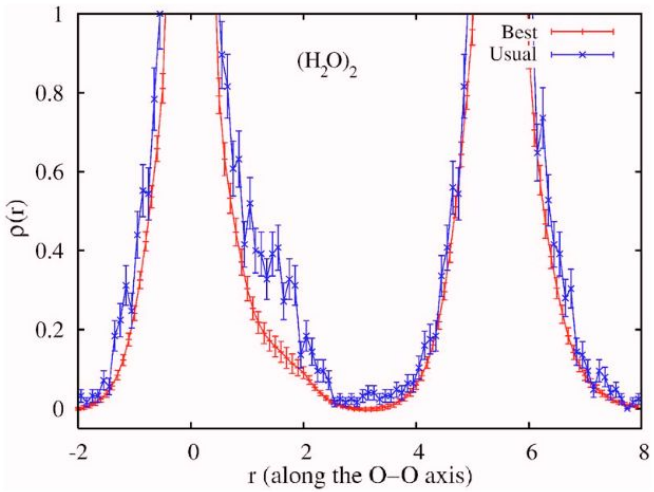


FIG. 3. (Color online) Cut of the one-electron density along the O-O axis of the H_2O dimer. Data for the best and usual estimators. Solid lines are simple linear extrapolations of the data.

1. Assaraf, R., Caffarel, M. & Scemama, A. Improved Monte Carlo estimators for the one-body density. Phys. Rev. E 75, 035701 (2007).

Conclusions

- We emphasized the importance of studying the electronic density in addition to the total energy.
- Real-space QMC is well-positioned as a highly-accurate method to generate accurate charge densities (assuming delta functions are regularized).
- Equivariant graph neural networks were used as state-of-the-art methods that is sufficiently flexible to accurately and efficiently predict electron densities.
- Our modified eqGNN model can learn from a small dataset (~1000 molecules) and produce accurate charge densities.
- We have shown that the main errors (DFT vs. QMC) arise from density regions that are close to atoms (QM9 and PBE).
- The errors in the spin-density are larger when trained on similar dataset size due to larger variability in spin density when compared to charge density.

Acknowledgements

This work was supported by the **Center for Predictive Simulation of Functional Materials**, a DOE funded Computational Materials Sciences Center, where in addition to this research we develop and make available the open source **QMCPACK** code.

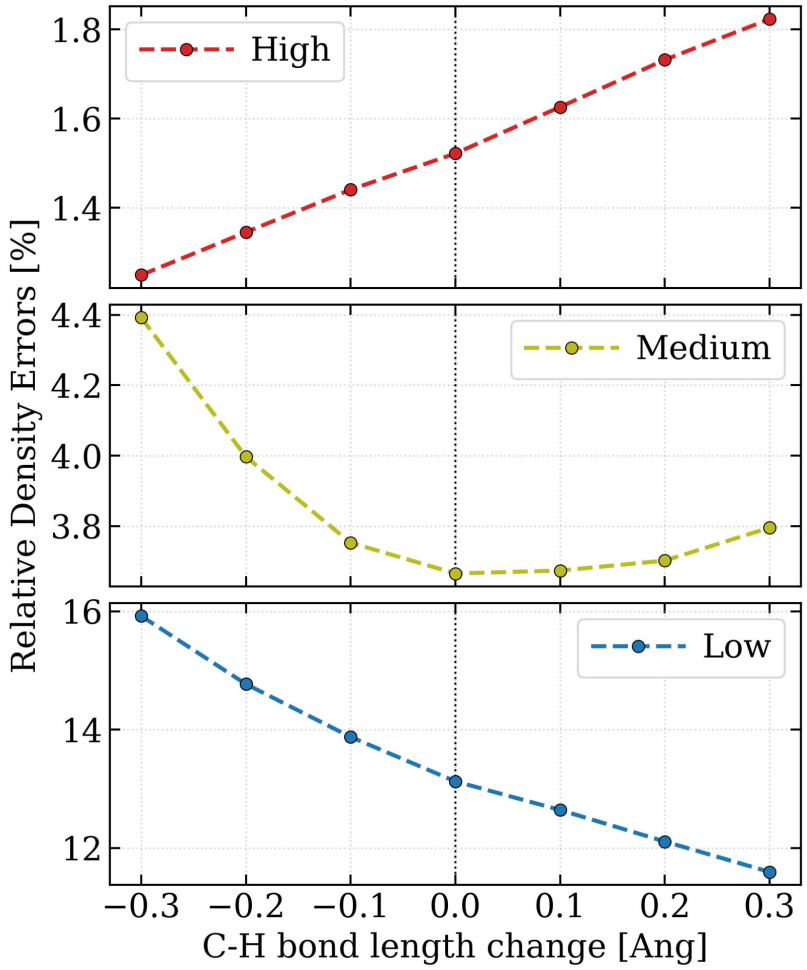
Fan Shu

Victor Fung

P. Ganesh

Appendix

To Do: Transferability Tests



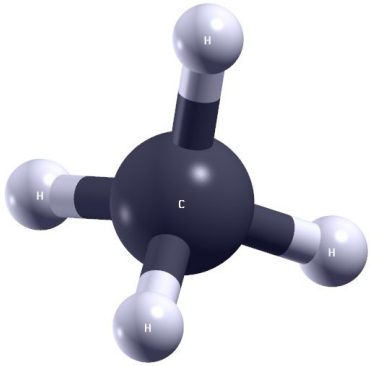


Figure: A study of (DFT-QMC) density deviations ($\Delta\rho/\rho_{QMC}$) of CH_4 as the C-H bond length is varied. Low density ($10^{-3} < \rho < 10^{-2}$ e/A^3), medium density ($10^{-2} < \rho < 10^{-1}$ e/A^3), and high density ($10^{-1} < \rho$ e/A^3) regions are shown.

Details of the employed ML architecture

- An augmented graph with (virtual) nodes for each charge density grid point is created on top of an existing graph of real nodes for each atom in the molecule
- Message passing is performed between real nodes to real nodes, and real nodes to virtual nodes
- A readout function is then used to obtain the density from the virtual node embedding at the last layer
- The message passing model architecture used is an equivariant transformer based on the TorchMD-NET model
- The model uses 5 message passing layers with a hidden dimension of 64 and SiLU activation functions.
- The readout function is a fully connected neural network with 1 layer and a hidden dimension of 64.
- The model was trained with a learning rate of 0.001 with the Adam optimizer and trained for 500 epochs.
- A cutoff radius of 12A was used to construct the graphs.
- 10k points learned from each molecule.

Methods

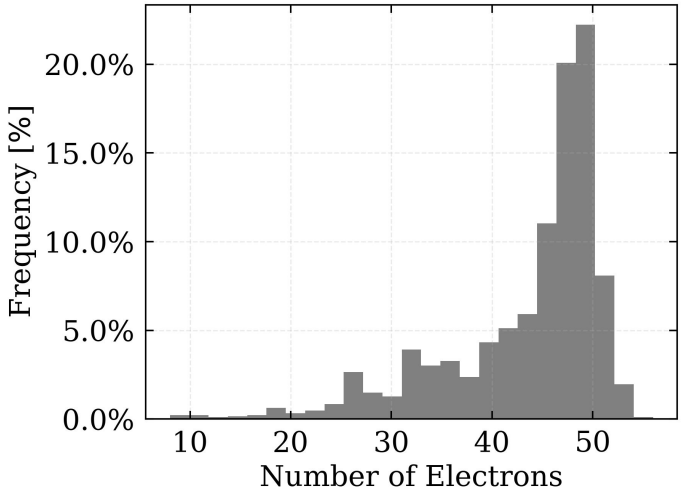


Figure: Distribution of the randomly selected molecules from QM9 with respect to the number of electrons.

Method	Scaling
non-hybrid DFT	N^3
Hartree-Fock	N^4
MP2	N^5
CCSD	N^6
MP4,CCSD(T)	N^7
...	
FCI	N!

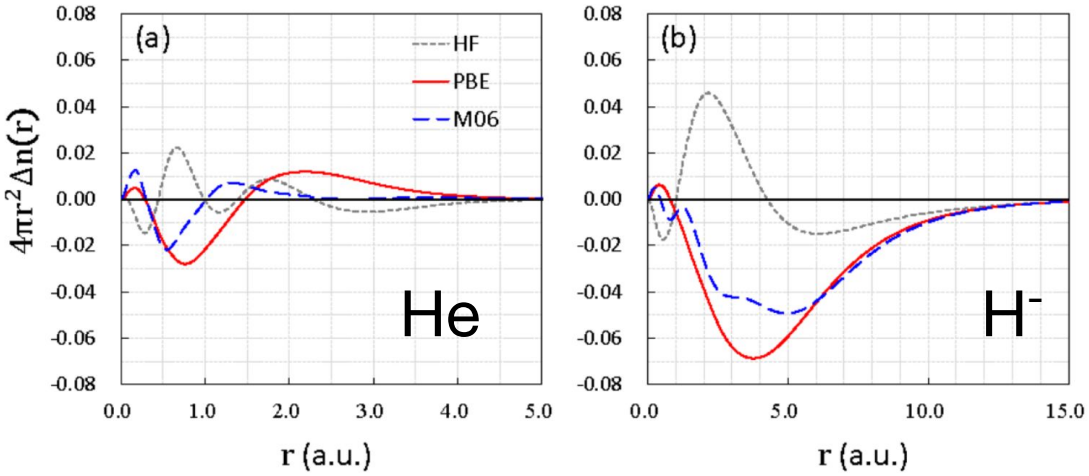


Figure: Density errors for a density-insensitive case (a, He atom) and an extremely density-sensitive case (b, H^-), for HF, PBE, and M06. An aug-cc-pV5Z basis is used. Reference value from QMC density. Taken from Ref. [1].

Notes

- Graph: A set of points (vertices) connected by lines (edges). “Vertex” and “edge” terms are used since a graph can represent a 3D object. A planar graph has non-intersecting edges.
- Bhowmik et al: Atoms are vertices in a graph. Query points are special vertices that only receive messages.
- DeepDFT scales linearly since r_cut is fixed, and message propagation layer number is linear.
- DeepDFT is purely data-driven: inputs are geometry and atomic numbers only.
- How to propagate the QMC errors? One way is to re-sample the density using the QMC error bars and form an ensemble of eqGNN models to estimate the error bars.