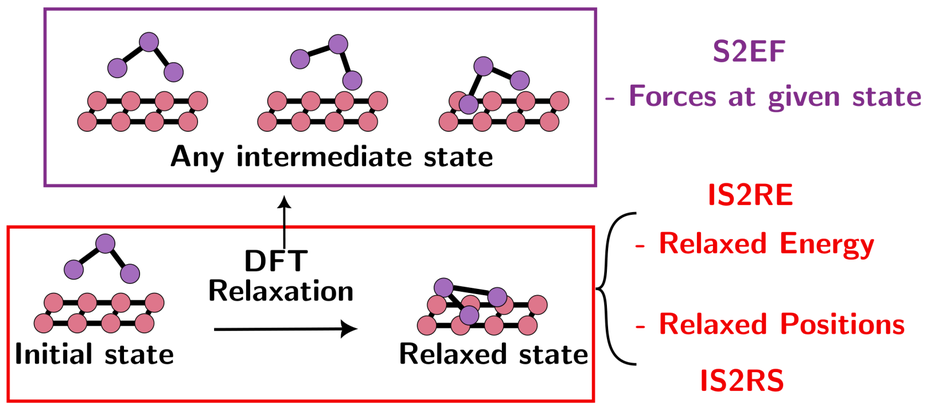
Improving Molecular Modeling with Geometric GNNs: an Empirical Study

Ali Ramlaoui* Théo Saulus* Basile Terver*
Victor Schmidt David Rolnick Fragkiskos D. Malliaros Alexandre Duval

Introduction

- Recent advancements in geometric Graph Neural Networks (GNNs) have shown promising results in molecular modeling.
- Large-scale datasets like OC20 [1] and QM7-X [2] facilitate fast and accurate molecular property predictions.
- This study investigates the impact of:
  - Canonicalization methods
  - Graph creation strategies
  - Auxiliary tasks
- Aim: Improve performance, scalability, and symmetry enforcement in molecular modeling using geometric GNNs.



Canonicalization methods

- $E(3)$-equivariance is desirable to learn representations suited for tasks such as force predictions on atoms. It can be enforced on unconstrained GNNs with a coordinate-preprocessing step referred to as canonicalization.
- We evaluate several canonicalization procedures with FAENet backbone architecture [3] on OC20 and QM9:
  - Vector Neurons Network (VNN) [4] using the VN re-implementation of PointNet [5] and DGCNN [6]
  - Stochastic Frame Averaging (SFA) [3], which approximates Frame Averaging [7] by sampling one canonical orientation per epoch.
  - SFA+SignNet, which handles the sign ambiguity problem in SFA with a sign-invariant network [8]. We use two versions of SignNet, either using VNNs or MLPs (resp. exactly and approx. equivariant).

Results

- Heuristic approximations of equivariance can perform as well as exact equivariance in some practical applications.
- In terms of symmetry enforcement, non-exact methods are nearly as effective as fully invariant methods, suggesting that the FAENet backbone implicitly learns to handle symmetries.
- For exact canonicalization methods, training or not the network, and swapping methods has little impact.

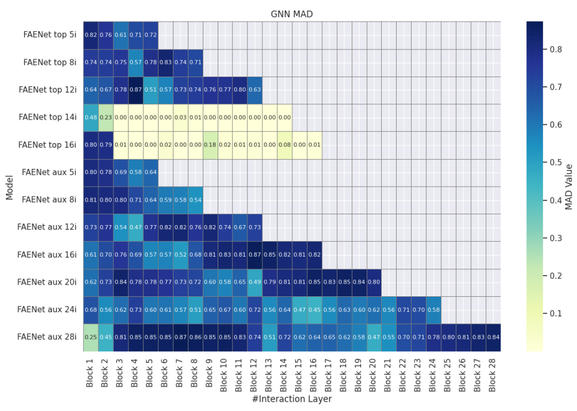
Canonicalization	Cano. trained parameters	avg. MAE (meV) ↓	EwT (ID) (%) ↑	3D Rotation Invariance ↓
SFA	0	594	4.40	$1.30 \cdot 10^{-2}$
(U) SFA+MLP-SignNet	0	580	4.48	$9.71 \cdot 10^{-2}$
(T) SFA+MLP-SignNet	454	583	4.46	$4.00 \cdot 10^{-2}$
(U) SFA+VN-SignNet	0	592	4.69	$7.58 \cdot 10^{-3}$
(T) SFA+VN-SignNet	2,620	599	4.25	$2.57 \cdot 10^{-2}$
(U) VN-Pointnet	0	605	4.09	$4.62 \cdot 10^{-3}$
(T) VN-Pointnet	1,310	598	4.12	$\mathbf{3.80 \cdot 10^{-3}}$
(U) VN-DGCNN	0	600	4.31	$3.11 \cdot 10^{-2}$
(T) VN-DGCNN	663,804	593	4.42	$9.10 \cdot 10^{-3}$

Invariance comparison of canonicalization methods on OC20 IS2RE dataset. (U) (resp. (T)) indicates an untrained (resp. trained) canonicalization network.

Auxiliary Tasks

Noisy Nodes

- To address oversmoothing, [9] propose to add an auxiliary node-level denoising task encouraging diversity in latent representations of nodes.
- Implementation on IS2RE: adding position decoding head in addition to the original energy prediction head + adding Gaussian noise to input positions of atoms.

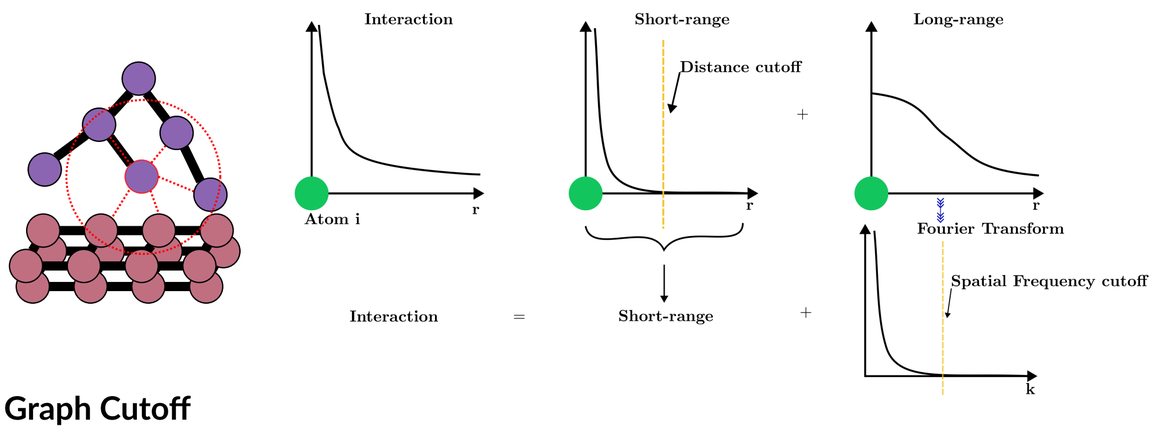


Results

- Models trained with Noisy Nodes IS2RS auxiliary task do not suffer from oversmoothing (i.e. MAD going to zero) even with 28 interaction layers.
- The improvements are only observed when using canonicalization methods, showing that equivariance is a beneficial inductive bias and allows for robustness to noising.

Graph creation strategies

- Long-range interactions between atoms are essential for property predictions [10].
- Need for architectures and graph models that allow to correctly model these interactions.
- Traditional graph creation strategies used on SOTA models consist in defining a distance cutoff between atoms to decide whether to create a link.



Graph Cutoff

Model	ID EwT (%) ↑	ID MAE (eV) ↓
Cutoff 30 - Max neighbours 40	2.65	0.697
Cutoff 20 - Max. neighbours 40	3.08	0.673
Cutoff 20 - Max. neighbours 10	2.25	0.768
Cutoff 10 - Max. neighbours 50	4.17	0.553
Cutoff 10 - Max. neighbours 10	4.49	0.553
Cutoff 6 - Max. neighbours 40	4.31	0.553
Cutoff 1 - Max. neighbours 40	1.35	1.069

- Need to be careful about not connecting extremity or isolated atoms because it would create unwanted interactions.
- Complete graph strategies are way too expressive leading to poor performance.

Ewald Message Passing (EMP) [11]

- Physics-Inspired message passing seem to inform expressivity-limited models such as SchNet but not already expressive models.

Model	ID EwT (%) ↑	ID MAE (eV) ↓
FAENet	4.05	0.551
FAENet + Ewald	4.12	0.562
SchNet	2.93	0.654
SchNet + Ewald	3.48	0.597

Iterative Relaxation

- Iterative relaxation is competitive with direct IS2RE for non-symmetry constraining models!
- The subsurface atoms of the catalyst crystals (tag 0 atoms, although periodic and repetitive) are crucial to correctly compute the forces but they can be ignored for direct IS2RE [12].

Model	IS2RE EwT (%) ↑	IS2RE MAE (eV) ↓	IS2RS DwT (%) ↑	IS2RS Pos. MAE ↓
FAENet (Direct)	4.05	0.551	-	-
FAENet (SFA)	4.92	0.587	31.1	0.390
FAENet (UTPN)	5.64	0.560	33.7	0.381

Discussion and Conclusion

- Approximative heuristics for symmetry enforcements seem to yield similar performance as exact methods. Thus, how can we design canonicalization methods for practical settings beyond theoretical guarantees?
- Need for new graph creation strategies that are not necessarily Physics-Inspired but architecture oriented for expressive models.
- Future research should explore pre-training strategies inspired by techniques like Noisy Nodes [13] or design helpful auxiliary tasks.

References

[1] L. Chanussot et al. “The Open Catalyst 2020 (OC20) Dataset and Community Challenges”. In: *ACS Catalysis* (2020).
[2] Johannes Hoja et al. “QM7-X: A comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules”. In: *Scientific Data* (2021).
[3] Alexandre Duval et al. “FAENet: Frame Averaging Equivariant GNN for Materials Modeling”. In: *International Conference on Machine Learning* (2023).
[4] Congyue Deng et al. “Vector neurons: A general framework for so (3)-equivariant networks”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, pp. 12200–12209.
[5] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. *PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation*. 2017.
[6] Yue Wang et al. *Dynamic Graph CNN for Learning on Point Clouds*. 2019.
[7] Omri Puny et al. “Frame Averaging for Invariant and Equivariant Network Design”. In: *International Conference on Learning Representations*. 2021.
[8] Derek Lim et al. “Sign and Basis Invariant Networks for Spectral Graph Representation Learning”. In: *The Eleventh International Conference on Learning Representations*. 2023.
[9] Jonathan Godwin et al. *Simple GNN Regularisation for 3D Molecular Property Prediction & Beyond*. 2022.
[10] Johannes Gasteiger et al. “Fast and Uncertainty-Aware Directional Message Passing for Non-Equilibrium Molecules”. In: *arXiv preprint arXiv: 2011.14115* (2020).
[11] Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. “Ewald-based Long-Range Message Passing for Molecular Graphs”. In: *arXiv preprint arXiv:2303.04791* (2023).
[12] Alexandre Duval et al. “PhAST: Physics-Aware, Scalable, and Task-specific GNNs for Accelerated Catalyst Design”. In: *arXiv preprint arXiv: 2211.12020* (2022).
[13] Nima Shoghi et al. *From Molecules to Materials: Pre-training Large Generalizable Models for Atomic Property Prediction*. 2024.