Rapidash: Scalable Molecular Modeling Through Controlled Equivariance Breaking

Sharvaree Vadgama [1] **Erik Bekkers** [1]

Abstract

Molecular generation and property prediction have traditionally relied on models with strict equivariance, particularly preserving E(3) symmetries through equivariant models. However, recent advances in large-scale models suggest that unconstrained architectures, trained on extensive datasets, can implicitly learn these symmetries. In this work, we explore whether strict equivariance is required and present Rapidash, an architecture that allows breaking exact equivariance constraints. Using this architecture, we achieve state-of-the-art performance in molecular generation and property prediction, surpassing traditional equivariant models.

Code is available on GitHub.

1. Introduction

There is an ongoing debate about the necessity of explicit structural priors, particularly group-equivariance. A growing line of work argues that strict equivariance may overconstrain a model and limit its scalability, and that increasing model capacity and training data (Qu & Krishnapriyan, 2024) can compensate for the lack of built-in symmetry. This perspective is supported by models like AlphaFold (Abramson et al., 2024), which train equivariance by data augmentation. In contrast, several recent works highlight the limitations of learning equivariance from data alone. Moskalev et al. (2023) show that learned symmetries can be unreliable and degrade quickly out-of-distribution. Petrache & Trivedi (2023) demonstrate that even partial symmetry alignment can improve generalization. Theoretical results by Perin & Deny (2024) further highlight that learning equivariance is fundamentally limited when class orbits are sparse or poorly separated. Given these competing perspectives, it remains unclear under what conditions equivariance leads to tangible benefits, and when unconstrained models may be preferable, highlighting the need for a deeper understanding of this trade-off.

To address this debate, we systematically investigate the impact of equivariant and non-equivariant models for molecular modeling. We begin by formalizing three core research questions: *(i)* how kernel constraints affect expressivity and generalization, *(ii)* whether increased representation capacity can close the performance gap between more and less constrained models, *(iii)* how different forms of equivariance breaking affect performance.

We approach these questions through both theoretical and empirical analysis. Theoretically, we show that equivariant models decompose into symmetry-preserving and symmetry-breaking components, and provide formal results linking generalization to pose entropy and inductive bias alignment. To conduct systematic empirical analysis, we introduce Rapidash, a unified group convolutional architecture that enables fine-grained control over equivariance constraints, input/output representations, and equivariance-breaking mechanisms. This design supports systematic comparisons across a wide range of model variants with varying degrees of equivariance. Experiments on molecular property prediction and generation show that more constrained equivariant models outperform less constrained alternatives when aligned with task geometry. Increasing representation capacity helps both strongly and weakly constrained models, but does not fully eliminate performance gaps (Fig. 2).

The main contributions of the paper can be summarized as:

- **Unified architecture**: We introduce Rapidash, a scalable and modular group convolutional architecture. This design supports multiple symmetry groups, input/output variants, allowing for equivariance-breaking and symmetry-breaking options. We achieve **state-of-the-art** performance on QM9 *molecule generation*.

- **Theoretical analysis**: We provide formal results showing when equivariance and symmetry breaking offer provable advantages in expressivity and generalization.

[1]AMLab, University of Amsterdam. Correspondence to: Sharvaree Vadgama <sharvaree.vadgama@gmail.com>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

2. Related Work

Wang et al. (2022); van der Ouderaa et al. (2022); Kim et al. (2023); Pertigkiozoglou et al. (2024) demonstrate that controlled equivariance breaking can significantly improve performance, also highlighting the benefits of relaxed group equivariance constraints. Similarly, work by Petrache & Trivedi (2023) suggests that models with partial or approximate equivariance can still exhibit improved generalization compared to fully non-equivariant ones. Equivariant models with higher-order tensor representations (e.g., E3GNN, NequIP (Batzner et al., 2022), SEGNN (Brandstetter et al., 2022)) have shown improved data efficiency and performance for tasks like molecular interatomic potentials, findings in other domains differ. Thais & Murnane (2023) reported that Lorentz equivariant models (Brehmer et al., 2023) offered no clear advantage over non-equivariant counterparts in particle physics. In materials science, though $\mathrm{SO}(3)$-equivariant Graph Transformers (Liao & Smidt, 2023; Liao et al., 2024) achieved strong results, Qu & Krishnapriyan (2024) (EScAIP) demonstrated that extensively scaled non-equivariant models can be competitive or even superior. Conversely, scaling experiments by Brehmer et al. (2024) using transformers for rigid-body simulations argued in favor of equivariant networks.

3. Method

Problem Statement We pose the problem statement as research questions listed as follows: How do different equivariance strategies—ordered from most to least constrained as previously outlined—impact task performance? If more constrained $\mathrm{SE}(3)$-equivariant models outperform $T(3)$-models (despite $T(3)$'s larger theoretical hypothesis space), can scaling $T(3)$-models (e.g., via channel capacity C or training duration) close this apparent generalization gap? Does incorporating explicit geometric information, such as pose R, to facilitate symmetry breaking lead to improved empirical performance compared to models that do not utilize such information?

3.1. Breaking exact equivariance

Our analysis of expressivity and equivariance constraints centers on our model, `Rapidash` (detailed in App. B). `Rapidash` employs regular group convolutions by processing feature fields $f : \mathbb{R}^3 \times S^2 \to \mathbb{R}^C$ over the position-orientation space. This involves convolution kernels on $\mathbb{R}^3 \times S^2$ that respect the quotient space symmetries ($\mathbb{R}^3 \times S^2 \equiv \mathrm{SE}(3)/\mathrm{SO}(2)$), a condition guaranteed to be met by conditioning message passing layers on the geometric invariants derived in (Bekkers et al., 2024).

A key theoretical aspect of `Rapidash`'s architecture is its fundamental connection to steerable tensor field networks.

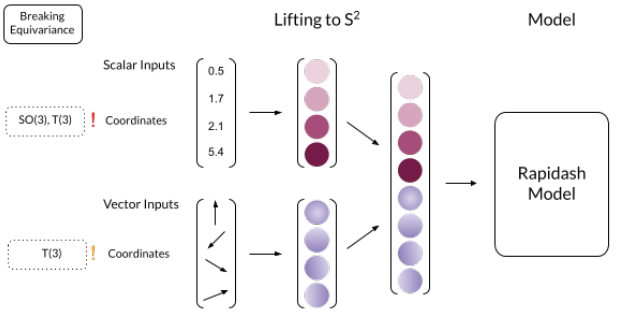

Figure 1. Input variations to Rapidash with base space $\mathbb{R}^3 \times S^2$ and $\mathrm{SE}(3)$-equivariant convolutions lifted to S^2. It shows equivariance breaking through different inputs and symmetry breaking by using a global frame input.

This relationship is established through Fourier analysis on the sphere S^2, where scalar fields can be decomposed into spherical harmonic coefficients that correspond to features transforming under irreducible representations of $\mathrm{SO}(3)$—the building blocks of steerable networks. This equivalence is formally stated as:

Proposition 3.1 (Equivalence via Fourier Transform)**.** *A regular group convolution operating on scalar fields $f : \mathbb{R}^n \times Y \to \mathbb{R}^C$ (where Y is S^{n-1} or $SO(n)$) can be equivalently implemented as a steerable group convolution operating on fields of Fourier coefficients $\hat{f} : \mathbb{R}^n \to V_\rho$ (where V_ρ is the space of combined irreducible representations) by performing point-wise Fourier transforms $\mathcal{F}_Y$ before the steerable convolution and inverse Fourier transforms $\mathcal{F}_Y^{-1}$ after.*

Remark 3.2. This equivalence (discussed further by Bekkers et al. (2024, Appx. A.1), Brandstetter et al. (2022), and Cesa et al. (2021)) implies that `Rapidash` could, in principle, be reformulated as a tensor field network. Thus, our architecture inherently represents the capabilities of this widely used class of steerable networks.

This connection to steerable networks also underpins its crucial universal approximation property. Building on established results for steerable GNNs (Dym & Maron, 2020) and related equivariant message passing schemes (Villar et al., 2021; Gasteiger et al., 2021), the $\mathrm{SE}(3)$-equivariant universal approximation for `Rapidash` is formally stated as:

Proposition 3.3 (Universal Approximation for Rapidash)**.** *`Rapidash`, as an instance of message passing networks over $\mathbb{R}^3 \times S^2$ with message functions conditioned on the bijective invariant attributes (derived in (Bekkers et al., 2024, Thm. 1)), is an $\mathrm{SE}(3)$-equivariant universal approximator. This specific universality follows from (Bekkers et al., 2024, Cor. 1.1), leveraging the sufficient expressivity of feature*

maps over $\mathbb{R}^3 \times S^2$.

Regardless of Rapidash' strong theoretical expressivity, practical performance must weigh computational cost against actual expressivity. G-convs on extended domains like $\mathbb{R}^3 \times S^2$ involve feature fields of size, e.g., $P \times O \times C$ per point cloud (Points $\times$ Orientations $\times$ Channels), compared to $P \times C$ for $\mathbb{R}^3$-based models. While matching total features (e.g., $O \times C$ in an $\mathbb{R}^3 \times S^2$ model to an enhanced C' in an $\mathbb{R}^3$ model) might suggest a nominally similar capacity, the O-axis in the former signifies a structured domain where features are correlated, not merely independent channels. The true expressive power, related to a model's hypothesis space size (Elesedy & Zaidi, 2021), varies with the imposed equivariance constraints: $T(3)$-equivariant $\mathbb{R}^3$ convolutions are least constrained, followed by $\mathrm{SE}(3)$-equivariant $\mathbb{R}^3 \times S^2$ convolutions, and isotropic $\mathrm{SE}(3)$-equivariant $\mathbb{R}^3$ convolutions being most constrained. For fair architectural comparison, we thus must consider scenarios where channel capacity (C) is maximized within practical limits. This leads to key research questions:

3.2. Symmetry Breaking and Generalization

This section analyzes generalization performance through the lens of probabilistic symmetry breaking, as introduced by Lawrence et al. (2025b) in their SymPE framework. Their work shows that incorporating informative auxiliary random variables Z (such as pose information $Z \in SO(3)$) into a stochastic model $f(X, Z)$—that e.g. takes a featurized point cloud X as input—can positively impact predictive inference, particularly when using jointly equivariant architectures. The ability of a model to leverage such an auxiliary variable Z is reflected in the effective conditional distribution $\mathbb{P}(Z|X)$ that the model implicitly defines. We adopt this perspective to interpret how a model's architectural constraints, like those in our Rapidash architecture, affect its ability to use global orientation information, satisfying joint invariance $f(gX, gZ) = f(X, Z)$ for $g \in SO(3)$.

We analyze model expressivity by considering two scenarios for the conditional distribution $\mathbb{P}(Z|X)$ of the auxiliary pose $Z \in SO(3)$. First, if the pose R_{known} is explicitly available , $\mathbb{P}(Z|X) = \delta(Z - R_{\text{known}})$. This provides zero-entropy pose information, enabling a model $f(X, R_{\text{known}})$ to directly leverage this specific orientation. Second, standard $SO(3)$-*invariant* models $f_{\text{inv}}(X)$, which do not receive explicit pose input, operate as if Z is drawn from a maximum entropy (uniform) distribution. This implies no specific orientation information is utilized, as derived in Proposition D.1. The disparity in available pose information—and thus in the effective entropy of Z—between these scenarios directly dictates model expressivity:

Corollary 3.4 (Expressivity Gain from Low-Entropy Pose Information)**.** *Let the optimal invariant mapping for a task, $f^*(X, R)$, depend non-trivially on the canonical orientation R. Based on the distinct informational content of Z outlined above: (a) Standard invariant models $f_{inv}(X)$ (maximum entropy pose) lack the expressivity to represent f^*; and (b) Pose-conditioned models $f_{cond}(X, R)$ (zero-entropy pose) can represent f^*.*

The enhanced expressivity afforded by conditioning on pose R, as established in Corollary 3.4, is fundamental for enabling symmetry breaking. By providing a determinate canonical reference R, models can disambiguate features arising from exact or approximate input symmetries, thereby producing more specific and potentially less symmetric outputs than standard equivariant models. Crucially, the well-known generalization advantages of equivariant architectures (Elesedy & Zaidi, 2021) also extend to models $f(\mathcal{X}, Z)$ that incorporate an auxiliary variable Z for symmetry breaking, as formalized by Lawrence et al. (2025b, Thm. 6.1). This general theorem confirms that structuring models $f(\mathcal{X}, Z)$ to be jointly equivariant is critical for realizing provable generalization gains when an auxiliary variable Z is introduced (e.g., for symmetry breaking). Our Proposition D.2 details the application of this principle to our setting where Z is a known, determinate pose R.

3.3. Rapidash and different variants

Our architecture, Rapidash, is designed with the flexibility to process both scalar and vector-valued features at its input and output stages. For instance, it can map input vectors v_i^{in} (e.g., initial velocities or normals) to spherical signals and predict output vectors v_i^{out} (e.g., displacements) by appropriately projecting from spherical representations (see App. B for architectural specifics). This capability allows Rapidash to naturally incorporate informative geometric inputs such as global pose, normal vectors, or velocities. If these inputs are treated as proper geometric objects that transform consistently under $\mathrm{SE}(3)$ actions, their inclusion can enhance the model's contextual understanding while preserving its overall $\mathrm{SE}(3)$-equivariance.

This same flexibility in handling inputs also allows for controlled deviations from strict $\mathrm{SE}(3)$-equivariance within Rapidash. For example, global coordinates or other geometric data can be supplied as fixed scalar features, which do not transform canonically under $\mathrm{SE}(3)$ actions as true geometric vectors would (Fig.1). This presents a critical trade-off: while more input data can be powerful, what is the impact if it compromises the $\mathrm{SE}(3)$-equivariance prior? The approximate and relaxed group equivariance approach suggests that such controlled deviations from strict symmetry can be beneficial, potentially improving training dynamics and performance. We explore this empirically with Rapidash, with the mechanisms for deviating from strict equivariance detailed in Appendix C.

Table 1. Ablation results on QM9 for property prediction with metric mean absolute error (MAE) ($\times 10^{-3}$).

Models	μ (D)	$\alpha(a_o^3)$	ϵ_{HOMO} (eV)
EGNN	29.0	71.0	29.0
DimeNet++	29.7	43.5	24.6
SE(3)-T	53.0	51.0	53.0
Rapidash (effective equivariance) in $\mathbb{R}^3$			
Rapidash (SE(3))	$17.41_{\pm 0.37}$	$53.03_{\pm 0.92}$	$22.29_{\pm 0.27}$
Rapidash (T_3)	$22.11_{\pm 1.03}$	$61.97_{\pm 1.05}$	$25.83_{\pm 0.25}$
Rapidash (none)	$17.77_{\pm 0.21}$	$52.22_{\pm 0.93}$	$22.44_{\pm 0.15}$
Rapidash (effective equivariance) in $\mathbb{R}^3 \times S^2$			
Rapidash (SE(3))	$10.39_{\pm 0.33}$	$42.20_{\pm 1.17}$	$19.30_{\pm 1.02}$
Rapidash (SO(3))	$10.44_{\pm 0.22}$	$42.68_{\pm 1.04}$	$19.64_{\pm 0.75}$
rapidash (none)	$10.53_{\pm 0.17}$	$40.21_{\pm 0.56}$	$18.69_{\pm 0.20}$

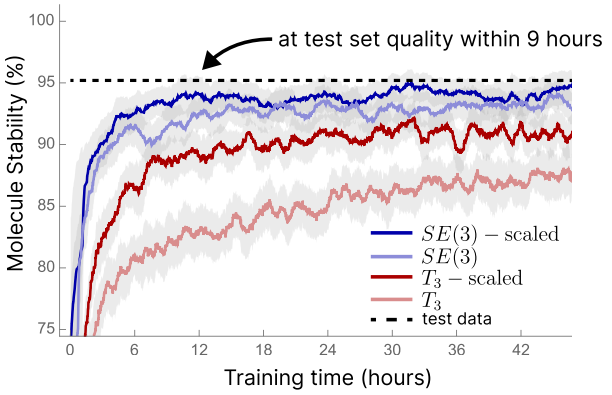

Figure 2. We plot molecular stability and training time for SE(3) and T_3 equivariant models with a hidden dimension of 256/512(inflated) for the molecular generation task on QM9. Inflated models have representations with higher hidden dimensions, hence increased capacity.

4. Experiments

For predicting molecular properties and generating molecules, we use QM9 (Ramakrishnan R., 2014). We evaluate the prediction of molecular properties using the MAE in Tab. 1 and compare these with EGNN (Satorras et al., 2021), Dimenet++ (Gasteiger et al., 2022), and SE(3)- Transformer (Fuchs et al., 2020). For the molecule generation, we train a generative model that uses equivariant denoising layers E. For molecular property prediction, SE(3)-equivariant Rapidash variants consistently outperform their T_3-equivariant counterparts, and the performance gap does not close with extended training. The models with no equivariance are obtained through passing co-ordinate information as scalar input, while models with effective equivariance SO(3) uses co-ordinate information as vector input. Among these model variants, SE(3) model, those utilizing the $\mathbb{R}^3 \times \mathbb{S}^2$ base type achieve superior accuracy over $\mathbb{R}^3$-only versions—highlighting the benefits of directional input and internal directional representations—and notably reach performance levels on par with established state-of-the-art methods on several regression tasks.

To address the effect of scaling the representation capacity of the models, we perform experiments on two models, one with $SE(3)$ equivariance and the other with T_3 and their respective inflated models (higher hidden dimensions), see Fig. 2.

We evaluate molecular generation, see Tab. 2, on metrics from Hoogeboom et al. (2022b) that include atomic stability, molecule stability, as well as a new aggregate metric *discovery*, which is a fraction of generated samples that are jointly valid, unique, and new. We compare our models with equivariant diffusion model like EDM (Hoogeboom et al., 2022a), PΘNITA (Bekkers et al., 2024), MuDiff (Hua et al., 2024), Geometric latent diffusion model (Xu et al., 2023a), END (Cornet et al., 2024), EquiFM (Song et al., 2024) and Clifford Diffusion Models (Liu et al., 2025). For

Table 2. Results for unconditional generation task on QM9 dataset.

Models	Atom Stability (%)	Mol Stability (%)	Discovery
GDM-AUG	97.6	71.6	–
EDM ((Hoogeboom et al., 2022a))	$98.7_{\pm 0.1}$	$82.0_{\pm 0.4}$	
GeoLDM ((Xu et al., 2023b))	$98.9_{\pm 0.1}$	$89.4_{\pm 0.5}$	
PΘNITA ((Bekkers et al., 2024))	98.9	87.8	-
MUDiff ((Hua et al., 2024))	$98.8_{\pm 0.2}$	$89.9_{\pm 1.1}$	
END ((Cornet et al., 2024))	$98.9_{\pm 0.2}$	$89.1_{\pm 0.1}$	
EquiFM((Song et al., 2024))	$98.9_{\pm 0.1}$	$88.3_{\pm 0.3}$	
CDM (all-grade) (Liu et al., 2025)	$99.0_{\pm 0.2}$	$89.7_{\pm 1.4}$	
Rapidash (effective equivariance) in $\mathbb{R}^3$			
Rapidash (T_3)	$98.57_{\pm 0.01}$	$81.62_{\pm 0.15}$	$91.83_{\pm 0.45}$
Rapidash (none)	$96.74_{\pm 0.07}$	$72.39_{\pm 0.50}$	$89.06_{\pm 0.39}$
Rapidash (effective equivariance) in $\mathbb{R}^3 \times S^2$			
Rapidash (SO(3))	$99.38_{\pm 0.02}$	$92.91_{\pm 0.41}$	$90.26_{\pm 0.20}$
Rapidash (SE(3))	$99.38_{\pm 0.01}$	$93.12_{\pm 0.28}$	$90.78_{\pm 0.11}$
Rapidash (none)	$99.33_{\pm 0.06}$	$92.71_{\pm 0.51}$	$91.54_{\pm 0.34}$

this task, the best results are decisively achieved by our SE(3)-equivariant Rapidash models, again favoring the $\mathbb{R}^3 \times \mathbb{S}^2$ base type, which demonstrate strong molecular stability in contrast to T_3 variants. Remarkably, on this generative task, these models significantly surpass existing state-of-the-art performance.

5. Conclusion

Through the introduction of Rapidash, a general regular group convolutional architecture enabling controlled comparisons—coupled with targeted theoretical analysis of equivariance, symmetry breaking, and generalization, we have sought to provide empirical insights into these critical trade-offs. Our findings offer a structured understanding of when and how different equivariance strategies and symmetry considerations affect model performance for molecular modeling tasks, thereby offering guidance for future archi-

tectural design and model selection in this domain.

Acknowledgements

Sharvaree Vadgama is supported by the Hybrid Intelligence Center, a 10-year program funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research (NWO). Authors thank Mohammad Islam, Domas Buracas, Artem Moskalev, and Christian Shewmake for valuable discussions.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., et al. Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature*, pp. 1–3, 2024.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1), May 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-29939-5. URL http://dx.doi.org/10.1038/s41467-022-29939-5.

Bekkers, E. J. B-spline cnns on lie groups. In *International Conference on Learning Representations*, 2019.

Bekkers, E. J., Vadgama, S., Hesselink, R., der Linden, P. A. V., and Romero, D. W. Fast, expressive se(n) equivariant networks through weight-sharing in position-orientation space. In *The Twelfth International Conference on Learning Representations*, 2024.

Brandstetter, J., Hesselink, R., van der Pol, E., Bekkers, E. J., and Welling, M. Geometric and Physical Quantities Improve E(3) Equivariant Message Passing. In *International Conference on Learning Representations*, 2022.

Brehmer, J., de Haan, P., Behrends, S., and Cohen, T. Geometric algebra transformer, 2023. URL https://arxiv.org/abs/2305.18415.

Brehmer, J., Behrends, S., de Haan, P., and Cohen, T. Does equivariance matter at scale?, 2024. URL https://arxiv.org/abs/2410.23179.

Cesa, G., Lang, L., and Weiler, M. A program to build e (n)-equivariant steerable cnns. In *International Conference on Learning Representations*, 2021.

Chollet, F. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

Cohen, T. and Welling, M. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Cornet, F., Bartosh, G., Schmidt, M. N., and Naesseth, C. A. Equivariant neural diffusion for molecule generation. In *38th Conference on Neural Information Processing Systems*, 2024.

Dym, N. and Maron, H. On the universality of rotation equivariant point cloud networks. In *International Conference on Learning Representations*, 2020.

Elesedy, B. and Zaidi, S. Provably strict generalisation benefit for equivariant models. In *International conference on machine learning*, pp. 2959–2969. PMLR, 2021.

Fey, M. and Lenssen, J. E. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

Fuchs, F. B., Worrall, D. E., Fischer, V., and Welling, M. Se(3)-transformers: 3d roto-translation equivariant attention networks, 2020. URL https://arxiv.org/abs/2006.10503.

Gasteiger, J., Becker, F., and Günnemann, S. Gemnet: Universal directional graph neural networks for molecules. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=HS_sOaxS9K-.

Gasteiger, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules, 2022. URL https://arxiv.org/abs/2011.14115.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. In Precup, D. and Teh, Y. W. (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1263–1272. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/gilmer17a.html.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d. In *International conference on machine learning*, 2022a.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3d, 2022b. URL https://arxiv.org/abs/2203.17003.

Hua, C., Luan, S., Xu, M., Ying, Z., Fu, J., Ermon, S., and Precup, D. Mudiff: Unified diffusion for complete molecule generation. In *Learning on Graphs Conference*, pp. 33–1. PMLR, 2024.

Karras, T., Aittala, M., Aila, T., and Laine, S. Elucidating the design space of diffusion-based generative models, 2022. URL https://arxiv.org/abs/2206.00364.

Kim, H., Lee, H., Yang, H., and Lee, J. Regularizing towards soft equivariance under mixed symmetries. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 16712–16727. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/kim23p.html.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Knigge, D. M., Romero, D. W., and Bekkers, E. J. Exploiting redundancy: Separable group convolutional networks on lie groups. In *International Conference on Machine Learning*, pp. 11359–11386. PMLR, 2022.

Kondor, R., Lin, Z., and Trivedi, S. Clebsch–gordan nets: a fully fourier space spherical convolutional neural network. *Advances in Neural Information Processing Systems*, 31, 2018.

Kuipers, T. P. and Bekkers, E. J. Regular se (3) group convolutions for volumetric medical image analysis. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2023*. arXiv preprint arXiv:2306.13960, 2023.

Lawrence, H., Portilheiro, V., Zhang, Y., and Kaba, S.-O. Improving equivariant networks with probabilistic symmetry breaking. In *The Thirteenth International Conference on Learning Representations*, 2025a. URL https://openreview.net/forum?id=ZE6lrLvATd.

Lawrence, H., Portilheiro, V., Zhang, Y., and Kaba, S.-O. Improving equivariant networks with probabilistic symmetry breaking, 2025b. URL https://arxiv.org/abs/2503.21985.

Liao, Y.-L. and Smidt, T. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs, 2023. URL https://arxiv.org/abs/2206.11990.

Liao, Y.-L., Wood, B., Das, A., and Smidt, T. Equiformerv2: Improved equivariant transformer for scaling to higher-degree representations, 2024. URL https://arxiv.org/abs/2306.12059.

Liu, C., Vadgama, S., Ruhe, D., Bekkers, E., and Forré, P. Clifford group equivariant diffusion models for 3d molecular generation, 2025. URL https://arxiv.org/abs/2504.15773.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.

Moskalev, A., Sepliarskaia, A., Bekkers, E. J., and Smeulders, A. On genuine invariance learning without weight-tying, 2023. URL https://arxiv.org/abs/2308.03904.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Perin, A. and Deny, S. On the ability of deep networks to learn symmetries from data: A neural kernel theory, 2024. URL https://arxiv.org/abs/2412.11521.

Pertigkiozoglou, S., Chatzipantazis, E., Trivedi, S., and Daniilidis, K. Improving equivariant model training via constraint relaxation, 2024. URL https://arxiv.org/abs/2408.13242.

Petrache, M. and Trivedi, S. Approximation-generalization trade-offs under (approximate) group equivariance. In Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., and Levine, S. (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 61936–61959. Curran Associates, Inc., 2023.

Qu, E. and Krishnapriyan, A. S. The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains, 2024. URL https://arxiv.org/abs/2410.24169.

Ramakrishnan R., Dral P.O., R. M. V. L. O. Quantum chemistry structures and properties of 134 kilo molecules, 2014.

Satorras, V. G., Hoogeboom, E., and Welling, M. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Song, Y., Gong, J., Xu, M., Cao, Z., Lan, Y., Ermon, S., Zhou, H., and Ma, W.-Y. Equivariant flow matching with hybrid probability transport for 3d molecule generation. *Advances in Neural Information Processing Systems*, 36, 2024.

Thais, S. and Murnane, D. Equivariance is not all you need: Characterizing the utility of equivariant graph neural networks for particle physics tasks, 2023. URL https://arxiv.org/abs/2311.03094.

van der Ouderaa, T. F. A., Romero, D. W., and van der Wilk, M. Relaxing equivariance constraints with non-stationary continuous filters, 2022. URL https://arxiv.org/abs/2204.07178.

Villar, S., Hogg, D. W., Storey-Fisher, K., Yao, W., and Blum-Smith, B. Scalars are universal: Equivariant machine learning, structured like classical physics. *Advances in Neural Information Processing Systems*, 34:28848–28863, 2021.

Wang, R., Walters, R., and Yu, R. Approximately equivariant networks for imperfectly symmetric dynamics, 2022. URL https://arxiv.org/abs/2201.11969.

Weiler, M. and Cesa, G. General e (2)-equivariant steerable cnns. *Advances in neural information processing systems*, 32, 2019.

Weiler, M., Forré, P., Verlinde, E., and Welling, M. Coordinate independent convolutional networks–isometry and gauge equivariant convolutions on riemannian manifolds. *arXiv preprint arXiv:2106.06020*, 2021.

Xu, M., Powers, A., Dror, R., Ermon, S., and Leskovec, J. Geometric latent diffusion models for 3d molecule generation, 2023a. URL https://arxiv.org/abs/2305.01140.

Xu, M., Powers, A. S., Dror, R. O., Ermon, S., and Leskovec, J. Geometric latent diffusion models for 3d molecule generation. In *International Conference on Machine Learning*, 2023b.

A. Mathematical Prerequisites and Notations

Groups. A *group* is an algebraic structure defined by a set G and a binary operator $\cdot : G \times G \to G$, known as the *group product*. This structure $(G, \cdot)$ must satisfy four axioms: (1) *closure*, where $\forall_{h,g \in G} : h \cdot g \in G$; (2) the existence of an *identity* element $e \in G$ such that $\forall_{g \in G}, e \cdot g = g \cdot e = g$, (3) the existence of an *inverse* element, i.e. $\forall_{g \in G}$ there exists a $g^{-1} \in G$ such that $g^{-1} \cdot g = e$; and (4) *associativity*, where $\forall_{g,h,p \in G} : (g \cdot h) \cdot p{=}g \cdot (h \cdot p)$. Going forward, group product between two elements will be denoted as $g, g' \in G$ by juxtaposition, i.e., as $g\, g'$.

For Special Euclidean group $\mathrm{SE}(n)$, the group product between two roto-translations $g{=}(\mathbf{x}, \mathbf{R})$ and $g'{=}(\mathbf{x}', \mathbf{R}')$ is given by $(\mathbf{x}, \mathbf{R})\,(\mathbf{x}', \mathbf{R}'){=}(\mathbf{R}\mathbf{x}' + \mathbf{x}, \mathbf{R}\,\mathbf{R}')$, and its identity element is given by $e{=}(\mathbf{0}, \mathbf{I})$.

Homogeneous Spaces. A group can act on spaces other than itself via a *group action* $gT : G{\times}X \to X$, where X is the space on which G acts. For simplicity, the action of $g \in G$ on $x \in X$ is denoted as $g\,x$. Such a transformation is called a group action if it is homomorphic to G and its group product. That is, it follows the group structure: $(g\,g')\,x{=}g\,(g'\,x)\ \forall g, g' \in G, x \in X$, and $e\,x{=}x$. For example, consider the space of 3D positions $X = \mathbb{R}^3$, e.g., atomic coordinates, acted upon by the group $G{=}\mathrm{SE}(3)$. A position $\mathbf{p} \in \mathbb{R}^3$ is roto-translated by the action of an element $(\mathbf{x}, \mathbf{R}) \in SE(3)$ as $(\mathbf{x}, \mathbf{R})\,\mathbf{p}{=}\mathbf{R}\,\mathbf{p} + \mathbf{x}$.

A group action is termed *transitive* if every element $x \in X$ can be reached from an arbitrary origin $x_0 \in X$ through the action of some $g \in G$, i.e., $x{=}gx_0$. A space X equipped with a transitive action of G is called a *homogeneous space* of G. Finally, the *orbit* $G\,x := \{g\,x \mid g \in G\}$ of an element x under the action of a group G represents the set of all possible transformations of x by G. For homogeneous spaces, $X{=}G\,x_0$ for any arbitrary origin $x_0 \in X$.

Quotient spaces. The aforementioned space of 3D positions $X{=}\mathbb{R}^3$ serves as a homogeneous space of $G = \mathrm{SE}(3)$, as every element $\mathbf{p}$ can be reached by a roto-translation from $\mathbf{0}$, i.e., for every $\mathbf{p}$ there exists a $(\mathbf{x}, \mathbf{R})$ such that $\mathbf{p}{=}(\mathbf{x}, \mathbf{R})\,\mathbf{0}{=}\mathbf{R}\,\mathbf{0} + \mathbf{x}{=}\mathbf{x}$. Note that there are several elements in $SE(3)$ that transport the origin $\mathbf{0}$ to $\mathbf{p}$, as any action with a translation vector $\mathbf{x}{=}\mathbf{p}$ suffices regardless of the rotation $\mathbf{R}$. This is because any rotation $\mathbf{R}' \in \mathrm{SO}(3)$ leaves the origin unaltered.

We denote the set of all elements in G that leave an origin $x_0 \in X$ unaltered the *stabilizer subgroup* $\mathrm{Stab}_G(x_0)$. In subsequent analyses, the symbol H is used to denote the stabilizer subgroup of a chosen origin x_0 in a homogeneous space, i.e., $H{=}\mathrm{Stab}_G(x_0)$. We further denote the *left coset* of H in G as $g\,H := \{g\,h \mid h \in H\}$. In the example of positions $\mathbf{p} \in X{=}\mathbb{R}^3$ we concluded that we can associate a point $\mathbf{p}$ with many group elements $g \in SE(3)$ that satisfy $\mathbf{p}{=}g\,\mathbf{0}$. In general, letting g_x be any group element s.t. $x{=}g_x\,x_0$, then any group element in the left set $g_x\,H$ is also identified with the point $\mathbf{p}$. Hence, any $x \in X$ can be identified with a left coset $g_x H$ and vice versa.

Left cosets $g\,H$ then establish an *equivalence relation* $\sim$ among transformations in G. We say that two elements $g, g' \in G$ are equivalent, i.e., $g \sim g'$, if and only if $g\,x_0{=}g'\,x_0$. That is, if they belong to the same coset $g\,H$. The space of left cosets is commonly referred to as the *quotient space* G/H.

We consider *feature maps* $f : X \to \mathbb{R}^C$ as multi-channel signals over homogeneous spaces X. Here, we treat point clouds as sparse feature maps, e.g., sampled only at atomic positions. In the general continuous setting, we denote the space of feature maps over X with $\mathcal{X}$. Such feature maps undergo group transformations through *regular group representations* $\rho^{\mathcal{X}}(g) : \mathcal{X} \to \mathcal{X}$ parameterized by g, and which transform functions $f \in \mathcal{X}$ via $[\rho^{\mathcal{X}}(g)f](x){=}f(g^{-1}x)$.

Irreducible representations and spherical harmonics Given any representation, there are often orthogonal subspaces that do not interact with each other, making it possible to break our representation down into smaller pieces by restricting to these subspaces. Hence, it is useful to consider the representations that cannot be broken down. Such representations are terms irreducible representations or irreps. Given a group G, V a vector space, and $\rho : G \to GL(V)$ representation, a representation is irreducible if there is no nontrivial proper subspace $W \subset V$ such that $\rho|_W$ is a representation of G over space W . With each irrep there is an associated (harmonic) frequency l. The irreps of $SO(3)$ are given by the $(2l+1) \times (2l+1)$ dimensional rotation matrices called *Wigner-D matrices*. The central columns of these matrices comprise the set of $2l+1$ *spherical harmonics* $Y_m^{(l)} : \mathbb{S}^2 \to \mathbb{R}$, indexed by $m = -l, ..., l$.

B. Rapidash: A Regular Group Convolution Approach for Flexible Equivariant Modeling

The Rapidash architecture, employed throughout our empirical study, builds upon the efficient $\mathrm{SE}(3)$-equivariant regular group convolution framework operating on position-orientation space ($\mathbb{R}^3 \times S^2$) as introduced by PΘNITA (Bekkers et al., 2024). To facilitate a comprehensive investigation into the utility of equivariance and symmetry breaking, Rapidash extends this foundation by incorporating several key flexibilities: (i) versatile handling of various input and output geometric

quantities (e.g., scalars, vectors representing positions, normals, or pose information); (ii) enhanced scalability for large point clouds through multi-scale processing incorporating techniques like farthest point sampling; and (iii) convenient adaptability to different equivariance constraints, allowing for controlled comparisons between $\mathrm{SE}(n)$ and translation-only (T_n) equivariant models. Like PΘNITA, Rapidash primarily adopts the regular group convolution paradigm, distinguishing it from steerable G-CNNs or tensor field networks, although fundamental connections exist. This section elucidates the theoretical underpinnings of this approach.

B.1. Regular vs. Steerable Group Convolutions

Equivariant neural networks for $\mathrm{SE}(3)$ are often categorized into regular or steerable (tensor field) approaches (Weiler & Cesa, 2019).

- **Regular Group Convolutions:** These typically operate on multi-channel scalar fields defined over a group G or a homogeneous space $X \equiv G/H$ (like $\mathbb{R}^3 \times S^2 \equiv \mathrm{SE}(3)/\mathrm{SO}(2)$ in our case). Feature fields $f : X \to \mathbb{R}^C$ transform via the regular representation: $[\rho(g)f](x) = f(g^{-1}x)$. Convolutions are then a form of template matching of a kernel $k(g_y^{-1}x)$ with the input signal (Cohen & Welling, 2016; Bekkers, 2019). A key advantage is that point-wise nonlinearities can be applied directly to these scalar feature maps without breaking equivariance (cf. Bekkers et al., 2024, Appx. A.1).
- **Steerable Group Convolutions / Tensor Field Networks:** These operate on feature fields $f : \mathbb{R}^n \to V_\rho$ where the codomain V_ρ is a vector space carrying a representation ρ of $\mathrm{SO}(n)$ (often a sum of irreducible representations, irreps). Features transform via induced representations, affecting both the domain and codomain: $([\mathrm{Ind}_{SO(n)}^{SE(n)}\rho](g)f)(\mathbf{x}) = \rho(\mathbf{R})\, f(g^{-1}\mathbf{x})$ (Weiler et al., 2021). Kernels $k(\mathbf{x})$ must satisfy a steerability constraint $k(\mathbf{R}\mathbf{x}) = \rho_{out}(\mathbf{R})k(\mathbf{x})\rho_{in}(\mathbf{R}^{-1})$. While this allows for exact equivariance without discretizing the rotation group (if using irreps), applying nonlinearities typically requires specialized equivariant operations or transformations to a scalar basis, as standard element-wise activations on steerable features (vectors/tensors) can break equivariance (Weiler & Cesa, 2019).

B.2. Rapidash as a Regular Group Convolution and its Relation to Steerable Networks

Rapidash, like PΘNITA (Bekkers et al., 2024), processes feature fields $f : \mathbb{R}^3 \times S^2 \to \mathbb{R}^C$. That is, at each point $\mathbf{x} \in \mathbb{R}^3$, it maintains a scalar signal $f_\mathbf{x} : S^2 \to \mathbb{R}^C$ defined over the sphere of orientations S^2. This aligns with the regular group convolution paradigm, in which the convolution kernel acts as an $\mathbb{R}^3 \times S^2$, subject to a symmetry constraint due to the quotient space structure $\mathbb{R}^3 \times S^2 \equiv SE(3)/SO(2)$. Specifically, the $SE(3)$ group convolution over $\mathbb{R}^3 \times S^2$ is of the form

$$[Lf](\mathbf{x}, \mathbf{n}) = \iint k(\mathbf{R_n}^T(\mathbf{x}' - \mathbf{x}), \mathbf{R_n}^T\mathbf{n}')f(\mathbf{x}', \mathbf{n}')d\mathbf{x}'d\mathbf{n}', \tag{1}$$

with kernel constraint $\forall R_z \in SO(2) : k(R\mathbf{x}, R\mathbf{n}) = k(\mathbf{x}, \mathbf{n})$ with R_z a rotation around the z-axis. This symmetry constraint is automatically solved when conditioning message passing layers (such as convolution layers) on the invariants outlined in (Bekkers et al., 2024, Thm. 1). In terms of these invariants, the resulting discrete group convolution is given by

$$[Lf](\mathbf{x}_i, \mathbf{n}_i) = \sum_{j \in \mathcal{N}(i)} k(\mathbf{a}_{ij})f(\mathbf{x}_j, \mathbf{n}_j)\,, \tag{2}$$

with the invariant pair-wise attributes given by

$$\begin{pmatrix} a_{ij}^{(1)} \\ a_{ij}^{(2)} \\ a_{ij}^{(3)} \end{pmatrix} = \begin{pmatrix} \mathbf{n}_i^T(\mathbf{x}_j - \mathbf{x}_i) \\ \|\mathbf{n}_j \perp (\mathbf{x}_j - \mathbf{x}_i)\| \\ \mathbf{n}_i^T\mathbf{n}_i \end{pmatrix}\,, \tag{3}$$

with $\perp$ denoting part of the vector $\mathbf{n}_j$ orthogonal to $\mathbf{x}_j - \mathbf{x}_i$.

As detailed in (Bekkers et al., 2024, App A), the connection between regular and steerable convolutions is established through Fourier analysis on the group/homogeneous space (Kondor et al., 2018; Cesa et al., 2021). A scalar field $f_\mathbf{x}(\mathbf{n})$ over S^2 (as used in Rapidash/PΘNITA) can be decomposed into spherical harmonic coefficients (its Fourier transform) through a spherical Fourier transform. These coefficients for different spherical harmonic degrees correspond to features transforming under irreducible representations of $\mathrm{SO}(3)$, which are the building blocks of steerable tensor field networks.

Proposition B.1 (Equivalence via Fourier Transform)**.** *A regular group convolution operating on scalar fields $f : \mathbb{R}^n \times Y \to \mathbb{R}^C$ (where Y is S^{n-1} or $SO(n)$) can be equivalently implemented as a steerable group convolution operating on fields of Fourier coefficients $\hat{f} : \mathbb{R}^n \to V_\rho$ (where V_ρ is the space of combined irreducible representations) by performing point-wise Fourier transforms $\mathcal{F}_Y$ before the steerable convolution and inverse Fourier transforms $\mathcal{F}_Y^{-1}$ after.*

Remark B.2. This equivalence is discussed in depth by Bekkers et al. (2024, Appx. A.1), Brandstetter et al. (2022), and Cesa et al. (2021). Consequently, Rapidash could, in principle, be reformulated as a tensor field network by operating in the spherical harmonic (Fourier) domain.

B.3. Universal Approximation

The universal approximation capabilities of equivariant networks are crucial. For steerable tensor field networks, Dym & Maron (2020) proved universal approximation properties for equivariant graph neural networks. Building on such results, and the correspondence between regular and steerable views, it has been shown that message passing networks (which include architectures like Rapidash) conditioned on appropriate invariant attributes over position-orientation space ($\mathbb{R}^n \times S^{n-1}$) are equivariant universal approximators.

Corollary B.3 (Universal Approximation for Rapidash)**.** *Rapidash, as an instance of message passing networks over $\mathbb{R}^3 \times S^2$ with message functions conditioned on the bijective invariant attributes (derived in Bekkers et al. (2024, Thm. 1)), is an $\mathrm{SE}(3)$-equivariant universal approximator.*

Remark B.4. This follows from Bekkers et al. (2024, Cor. 1.1), which itself builds on universality results for steerable GNNs (Dym & Maron, 2020) and for invariant networks used to construct equivariant functions (Villar et al., 2021; Gasteiger et al., 2021). The key is that feature maps over $\mathbb{R}^3 \times S^2$ are sufficiently expressive.

B.4. Advantages of the Regular Group Convolution Viewpoint for Rapidash

Rapidash adopts the regular group convolution viewpoint, working with scalar signals on discretized spherical fibers ($f(\mathbf{x}, \mathbf{n}_k)$ where $\mathbf{n}_k$ are grid points on S^2). This offers practical advantages:

1. **Simplicity of Activation Functions:** Since the features $f(\mathbf{x}, \mathbf{n}_k)$ at each grid point are scalars (or vectors of scalars in the channel dimension), standard element-wise nonlinear activation functions (e.g., GELU, ReLU, SiLU) can be applied directly without breaking $\mathrm{SE}(3)$-equivariance. This is because the action of $g \in \mathrm{SE}(3)$ permutes these values on the fiber or spatially, but the activation acts on each scalar value independently. In contrast, steerable tensor field networks require specialized equivariant nonlinearities or norm-based activations on higher-order tensors to preserve equivariance (Weiler & Cesa, 2019), which can limit expressivity or introduce computational overhead.

2. **Computational Efficiency of Activations:** While steerable networks *can* apply scalar activations by first performing an inverse Fourier transform (to get scalar fields), applying the activation, and then a forward Fourier transform (back to irreps), this incurs significant computational cost at each nonlinearity (cf. Bekkers et al., 2024, Appx. A.1). By operating directly on scalar spherical signals, Rapidash avoids these repeated transformations. Previous work on steerable group convolutions has indeed found that element-wise activation functions applied to scalar fields (obtained via inverse Fourier transforms from steerable vector features) can be most effective (e.g., as implicitly done in some equivariant GNNs by taking norms or scalar products before activation, or as explicitly discussed for general steerable CNNs in Weiler & Cesa, 2019).

3. **Conceptual Simplicity:** The regular group convolution approach, involving template matching of kernels over signals on G/H, can be more intuitive and closer to standard CNN paradigms than navigating representation theory and Clebsch-Gordan tensor products often required for constructing steerable tensor field networks. Concepts like stride/sub-sampling and normalization layers readily transfer to this setting.

While discretizing the sphere S^2 introduces an approximation to full $\mathrm{SO}(3)$ equivariance (equivariance up to the grid resolution), empirical results, including those for PΘNITA (Bekkers et al., 2024) and our findings with Rapidash, demonstrate that this is not detrimental to achieving state-of-the-art performance and robust generalization.

B.5. Separable group convolutions

Regular group convolutions over the full space $SE(3)$ can be efficiently computed when the kernel is factorized via

$$k_{c'c}(\mathbf{x}, \mathbf{R}) = k_c^{\mathbb{R}^3}(\mathbf{x}) k_c^{\mathrm{SO}(3)}(\mathbf{R}) k_{c'c}^{(channel)},$$

with c, c' the row and column indices of the "channel mixing" matrix. Then the group convolution equation can be split into three steps that are each efficient to compute: a spatial interaction layer (message passing), a point-wise $\mathrm{SO}(3)$ convolution, and a point-wise linear layer (Knigge et al., 2022; Kuipers & Bekkers, 2023). It would result in the group convolutional counterpart of *depth-wise separable convolution* (Chollet, 2017), which separates convolution in two steps (spatial mixing and channel mixing). In particular, for our choice to the group convolution over $\mathbb{R}^3 \times S^2$ using the pair-wise invariants of (3) the kernel is parametrized as

$$k_{c'c}(\mathbf{a}_{ij}) = k_c^{\mathbb{R}^3}(a_{ij}^{(1)}, a_{ij}^{(2)}) k_c^{S^2}(a_{ij}^{(3)}) k_{c'c}^{(channel)}.$$

This form allows to split the convolution over several steps, which following (Bekkers et al., 2024) we adapt a ConvNext (Liu et al., 2022) as the main layer to parametrize `Rapidash`. Fig. **??** shows the steps performed in this block, relative a standard ConvNext block over position space only. Here, LN denotes layer norm, and GELU is used as activation function.

C. Sources of Equivariance Breaking

In the study of equivariant deep learning, it is important to distinguish between different notions of "breaking" symmetries. Some approaches, such as those explored by Lawrence et al. (2025a) or the pose-conditioning methods analyzed in Appendix D, utilize architectures that maintain specific (joint) equivariance properties to achieve "symmetry breaking" in the output (e.g., an output sample being less symmetric than the input, or overcoming the limitations of standard invariance). This section, in contrast, focuses on mechanisms by which the strict G-equivariance of a neural network architecture itself can be compromised or intentionally relaxed, leading to a model that no longer fully adheres to the mathematical definition of G-equivariance. We categorize these into *external* and *internal* sources of equivariance breaking.

C.1. External Equivariance Breaking

External equivariance breaking occurs when an inherently G-equivariant architecture loses its equivariance properties due to the way inputs are provided to the network. Consider a linear layer L designed to be G-equivariant (e.g., for $G = \mathrm{SE}(3)$). Let v be a vector input that transforms naturally under the group action, and define $x_g = g \cdot v$ as its transformed version, described in global coordinates. When these coordinates x_g are provided, for example, as a set of scalar triplets rather than as a geometric vector type that the layer is designed to process equivariantly, they may be treated independently by the network without regard to their collective transformation under G. In such cases, we have:

$$\text{if } x \neq x_g, \text{ then often } L(x_g) \neq g \cdot L(x) \text{ (or } L(x_g) \neq L(x) \text{ for invariant L)} \tag{4}$$

This means the network processes transformed inputs in a way that does not respect the group symmetry, thereby breaking the intended G-equivariance of the function L. In contrast, when inputs are specified as types that transform appropriately under the group action (e.g., as vectors for an $SE(3)$-equivariant layer that expects vector inputs), equivariance can be maintained:

$$L(g \cdot v) = g \cdot L(v) \quad \forall g \in G \tag{5}$$

Moreover, for features that are truly invariant under G (such as one-hot encodings of atom types in QM9, which are invariant to $SO(3)$ rotations), the group action is trivial:

$$g \cdot x_{inv} = x_{inv} \quad \Longrightarrow \quad L(g \cdot x_{inv}) = L(x_{inv}) \tag{6}$$

This ensures that processing these features does not break the network's overall desired invariance to group transformations of other, non-invariant inputs.

C.2. Internal Equivariance Breaking

Internal equivariance breaking refers to the deliberate relaxation or incorrect specification of equivariance constraints within the layers themselves. Even if inputs are provided correctly, the layer operations may not fully respect the symmetry group G. Recent works have explored various approaches to controlled relaxation, including (Wang et al., 2022):

- Basis decomposition methods that mix G-equivariant and non-G-equivariant components within a layer.
- Learnable deviations from strict G-equivariance, often controlled by regularization terms.
- Progressive relaxation or enforcement of G-equivariance constraints during the training process.

In some cases, an internally broken layer L_{broken} might be expressed as a combination of a strictly G-equivariant part and a non-G-equivariant part:

$$L_{\text{broken}}(X) = L_{\text{equiv}}(X) + \alpha L_{\text{non-equiv}}(X) \tag{7}$$

where α controls the degree of deviation from strict G-equivariance. Some approaches (e.g., (Wang et al., 2022)) implement schemes where α is annealed during training.

C.3. Interplay of Equivariance Breaking Mechanisms

In practice, both external and internal equivariance-breaking can occur, sometimes simultaneously, and their effects can interact. For example, a network might employ layers with relaxed internal constraints (e.g., non-stationary filters) while also processing geometric inputs (like coordinates) in a manner that externally breaks the intended global symmetry (e.g., treating them as independent scalar features). The overall adherence of the model to a specific group symmetry G then depends on the interplay of these factors. As noted by Petrache & Trivedi (2023), while aligning data symmetry with architectural symmetry (strict equivariance) is ideal, models with partial or approximate equivariance can still exhibit improved generalization compared to fully non-equivariant ones.

D. Analysis of Symmetry Breaking and Pose-Conditioning in Equivariant Models

D.1. Standard Invariance and Effective Pose Entropy

This section formalizes the notion that a standard $SO(3)$-invariant model, when viewed within the framework of a jointly invariant function $f(X, Z)$, operates as if the auxiliary pose variable Z carries no specific information, i.e., it corresponds to a maximum entropy distribution over poses.

Proposition D.1 (Standard $SO(3)$-Invariance as Maximum Effective Pose Entropy)**.** *Let $f(X, Z)$ be a function $f : \mathcal{X} \times SO(3) \to Y$ that is jointly $SO(3)$-invariant, meaning $f(gX, gZ) = f(X, Z)$ for all $g \in SO(3)$. If the output of $f(X, Z)$ is also required to be standard $SO(3)$-invariant with respect to X alone, defining a function $f_{inv}(X)$ such that $f_{inv}(gX) = f_{inv}(X)$ for all $g \in SO(3)$, then:*

1. *The function $f(X, Z)$ must be independent of the auxiliary pose variable Z. Specifically, $f(X, Z) = f(X, Id)$ for any $Z \in SO(3)$, where Id is the identity element in $SO(3)$.*
2. *Consequently, such a model $f(X, Z)$ (which produces $f_{inv}(X)$) behaves as if Z is drawn from an uninformative, maximum entropy distribution (e.g., the uniform distribution over $SO(3)$). The model cannot utilize any specific orientation information conveyed by Z.*

Proof Sketch. To demonstrate part 1, that $f(X, Z) = f(X, Id)$:

1. By the joint $SO(3)$-invariance of f, we have $f(X, Z) = f(Z^{-1}X, Z^{-1}Z) = f(Z^{-1}X, Id)$.
2. The condition that the output of $f(X, Z)$ is standard $SO(3)$-invariant with respect to X means that for any fixed second argument (like Id), the function $f(\cdot, Id)$ must be $SO(3)$-invariant in its first argument. That is, $f(gX, Id) = f(X, Id)$ for all $g \in SO(3)$.
3. Applying this standard $SO(3)$-invariance with $g = Z^{-1}$ to the expression $f(Z^{-1}X, Id)$, we get $f(Z^{-1}X, Id) = f(X, Id)$.
4. Combining steps 1 and 3: $f(X, Z) = f(Z^{-1}X, Id) = f(X, Id)$.

This establishes that $f(X, Z)$ is independent of Z.

For part 1, if $f(X, Z)$ is independent of Z, it cannot make use of any particular value of Z to alter its output. From an informational perspective, Z provides no specific information to the model. This behavior is equivalent to Z being drawn from a distribution that reflects maximal uncertainty about the pose, which for a compact group like $SO(3)$ is the uniform (Haar) measure, corresponding to maximum entropy. Thus, the model $f_{inv}(X)$ cannot utilize any specific orientation information that might be notionally carried by Z. □

D.2. Proof of Corollary 3.4 (Expressivity Gain from Low-Entropy Pose Information)

Corollary 3.4 states: *Let the optimal invariant mapping for a task, $f^*(X, R)$, depend non-trivially on canonical orientation R. Based on the distinct informational content of Z outlined above: (a) Standard invariant models $f_{inv}(X)$ (maximum effective pose entropy) lack the expressivity to represent f^*; and (b) Pose-conditioned models $f_{cond}(X, R)$ (provided with zero-entropy pose information) can represent f^*.*

Let $G = SO(3)$ be the group of orientations. The optimal mapping $f^* : \mathcal{X} \times G \to \mathcal{Y}$ (where $\mathcal{X}$ is the space for X and $\mathcal{Y}$ for the output) has two key properties:

1. **Joint Invariance for an Invariant Task:** $f^*(gX, gR) = f^*(X, R)$ for all $g \in G$.

2. **Non-trivial Dependence on R:** For any given $X \in \mathcal{X}$, there exist $R_1, R_2 \in G$ such that $R_1 \neq R_2$ but $f^*(X, R_1) \neq f^*(X, R_2)$. This implies that R is an essential input for determining the output of f^*, not merely a redundant pose of X that could be factored out by invariance.

Proof of 1: Standard invariant models $f_{inv}(X)$ lack the expressivity to represent f^*.

A standard G-invariant model $f_{inv}(X)$ is a function $f_{inv} : \mathcal{X} \to \mathcal{Y}$. By definition, its output depends solely on X and must satisfy $f_{inv}(gX) = f_{inv}(X)$ for all $g \in G$. For any specific input $X_0 \in \mathcal{X}$, $f_{inv}(X_0)$ produces a single, uniquely determined output value, say Y_{inv_0}.

The model $f_{inv}(X)$ does not take R as an explicit input. As established by the principle that standard invariant models operate with maximum effective entropy regarding an auxiliary pose variable (see Appendix D.1, Proposition D.1), $f_{inv}(X)$ cannot vary its output based on different values of R for a fixed X. Its output is fixed once X is fixed.

Now, consider the target function $f^*(X_0, R)$. According to property (2) of f^*, there exist distinct $R_1, R_2 \in G$ such that $f^*(X_0, R_1) = Y_1^*$ and $f^*(X_0, R_2) = Y_2^*$, where $Y_1^* \neq Y_2^*$.

If $f_{inv}(X)$ were to represent $f^*(X, R)$, then for the input X_0, it would need to output Y_1^* when the canonical orientation is R_1, and simultaneously output Y_2^* when the canonical orientation is R_2. However, $f_{inv}(X_0)$ can only produce its single output Y_{inv_0}. Since $Y_1^* \neq Y_2^*$, Y_{inv_0} cannot equal both. Therefore, $f_{inv}(X)$ cannot represent the function $f^*(X, R)$ due to its inability to differentiate its output based on R.

Proof of 2: Pose-conditioned models $f_{cond}(X, R)$ *can* represent f^*.

A pose-conditioned model $f_{cond}(X, R)$ is a function $f_{cond} : \mathcal{X} \times G \to \mathcal{Y}$. It explicitly takes both X and R as inputs. It is designed to satisfy the same joint G-invariance as the target: $f_{cond}(gX, gR) = f_{cond}(X, R)$.

The target function $f^*(X, R)$ is a specific function mapping from the domain $\mathcal{X} \times G$ to $\mathcal{Y}$ that adheres to this joint G-invariance. We assume that the class of models from which $f_{cond}(X, R)$ is drawn (e.g., sufficiently large neural networks architected to respect joint G-invariance) are universal approximators for continuous functions on $\mathcal{X} \times G$ that satisfy the given symmetry requirements.

Since $f_{cond}(X, R)$ takes R as an explicit input (representing zero-entropy pose information for that instance, as R is known and fixed), it has the architectural capacity to make its output depend on R. For a given X_0, the model $f_{cond}(X_0, R)$ can learn to produce different outputs for different values of R. Specifically, it can learn to output $f^*(X_0, R_1)$ when its input R is R_1, and $f^*(X_0, R_2)$ when its input R is R_2.

Given that $f^*(X, R)$ is a well-defined function of (X, R) satisfying the joint G-invariance, and $f_{cond}(X, R)$ is a universal approximator for such functions with access to both X and R, $f_{cond}(X, R)$ can represent $f^*(X, R)$.

D.3. Generalization Benefits of Joint Invariance in Pose-Conditioned Models

While App. D established the expressivity advantage of pose-conditioned invariant models $f_{cond}(X, R)$ over standard invariant models $f_{inv}(X)$, one might consider if a sufficiently complex non-equivariant model $f_{neq}(X, R)$, which also takes pose R as input but lacks structural symmetry constraints, could achieve similar performance. Assuming both $f_{cond}(X, R)$ (satisfying joint invariance $f(gX, gR) = f(X, R)$) and $f_{neq}(X, R)$ have sufficient capacity to represent the optimal invariant solution $f^*(X, R)$, we argue that the equivariant structure of f_{cond} provides superior generalization guarantees.

This argument leverages theoretical results on the generalization benefits derived from incorporating known symmetries, such as Theorem 6.1 presented by Lawrence et al. (2025a), which extends foundational work on equivariance and generalization (Elesedy & Zaidi, 2021).

Proposition D.2 (Generalization Advantage of Jointly Invariant Pose-Conditioned Models)**.** *Let X be drawn from a G-invariant distribution $\mathbb{P}(X)$, where $G = SO(3)$. Let R be the provided pose, treated as an auxiliary variable $Z = R$ with the equivariant conditional distribution $\mathbb{P}(Z|X) = \delta(Z - R)$. Consider two models predicting an invariant output Y using inputs (X, R):*

- *$f_{cond}(X, R)$: A model satisfying joint invariance, $f_{cond}(gX, gR) = f_{cond}(X, R)$.*
- *$f_{neq}(X, R)$: Any model using the same inputs (X, R) that does **not** satisfy joint invariance.*

Under suitable regularity conditions (as specified in Lawrence et al. (2025a), Thm 6.1) and for a suitable risk function $R(f) = \mathbb{E}[\mathcal{L}(f(X, R), Y)]$ (e.g., using L_2 loss), the expected risk of the non-equivariant model is greater than or equal to that of the jointly invariant model:

$$R(f_{neq}) \geq R(f_{cond})$$

The difference $R(f_{neq}) - R(f_{cond})$ represents a non-negative generalization gap attributable to the component of f_{neq} orthogonal to the space of jointly invariant functions.

Remark D.3. This proposition follows from applying Theorem 6.1 (Lawrence et al., 2025a) to our setting. The conditions are met: $\mathbb{P}(X)$ is G-invariant (typically assumed or achieved via augmentation), $\mathbb{P}(Z|X) = \delta(Z - R)$ is equivariant, and G acts freely on $Z = R \in SO(3)$.

Proposition D.2 provides formal support for preferring the pose-conditioned equivariant (jointly invariant) architecture $f_{cond}(X, R)$ over an unstructured, non-equivariant model $f_{neq}(X, R)$ that uses the same input information. By incorporating the known relationship between transformations of the input X and the pose R via the joint invariance constraint, the equivariant model leverages a useful inductive bias. This bias restricts the hypothesis space to functions consistent with the underlying geometry, thereby reducing variance and improving expected generalization performance compared to a less constrained model, even if both models possess sufficient capacity to represent the optimal solution.

E. Implementation details

We implemented our models using PyTorch (Paszke et al., 2019), utilizing PyTorch-Geometric's message passing and graph operations modules (Fey & Lenssen, 2019), and employed Weights and Biases for experiment tracking and logging. A pool of GPUs, including A100, A6000, A5000, and 1080 Ti, was utilized as computational units. To ensure consistent performance across experiments, computation times were carefully calibrated, maintaining GPU homogeneity throughout.

For all experiments, we use `Rapidash` with 7 layers with 0 fiber dimensions for $\mathbb{R}^3$ and 0 or 8 fiber dimensions for $\mathbb{R}^3 \times S^2$. The polynomial degree was set to 2. We used the Adam optimizer (Kingma & Ba, 2014), with a learning rate of $1e-4$, and with a CosineAnnealing learning rate schedule with a warm-up period of 20 epochs.

QM9 QM9 dataset (Ramakrishnan R., 2014) contains up to 9 heavy atoms and 29 atoms, including hydrogens. We use the train/val/test partitions introduced in Gilmer et al. (2017), which consists of 100K/18K/13K samples respectively for each partition.

Diffusion Model Unlike EDM (Hoogeboom et al., 2022b) that uses a DDPM-like diffusion model with a deterministic sampler, we use a stochastic sampler proposed in (Karras et al., 2022). We condition the diffusion model with feature scaling

and noise scaling and combine outputs with skip connections, allowing for faster sampling. The sampler used in this work implements a stochastic differential equation with a second-order connection.