DIRECT PREDICTION OF TENSORIAL PROPERTIES WITH EQUIVARIANT MESSAGE-PASSING: APPLICATIONS TO NONLINEAR OPTICS

Peter R. Miedaner
Department of Chemistry
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
miedaner@mit.edu

Kin Long Kelvin Lee
Intel Labs
Hillsboro, OR 97124, USA
kin.long.kelvin.lee@intel.com

Shiang Fang
Department of Physics
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
shiangfang913@gmail.com

Tess Smidt
Department of Electrical Engineering
and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
tsmidt@mit.edu

Keith Nelson
Department of Chemistry
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
kanelson@mit.edu

ABSTRACT

Accurate machine-learned property prediction enables data-driven design and discovery of a wide range of materials. While prediction of scalar quantum mechanical properties like energies have recently reached unprecedented levels of accuracy, prediction of higher-order polar tensors remains relatively difficult and uncommon, despite their ubiquity in fields such as nonlinear optics. The ability to perform accurate ML-based predictions of optical tensors could greatly expedite the discovery of nonlinear optical media. Here, we present a study on the performance of a simple equivariant message-passing neural network for the prediction of molecular hyperpolarizability tensors. Our key findings demonstrate the ability for a modest architecture to perform highly accurate, direct prediction of the full 27-element hyperpolarizability tensor, which we attribute to the network respecting the natural transformation properties of polar tensors, and also the ability of the network to recognize the global symmetries of the input molecules. To provide a mechanistic understanding of these results, we employ dimensionality reduction techniques on the learned equivariant representations to visualize and reason about their latent structure.

1 INTRODUCTION

Neural network architectures which are equivariant under the group of 3D rotations, translations, and reflections, $E(3)$,(Geiger & Smidt, 2022) have shown impressive performance in the prediction of quantum mechanical scalar properties (Liao & Smidt, 2023)and interatomic potentials (Batzner et al., 2022; Batatia et al., 2022). It remains to be seen how these networks perform in the *direct* prediction of nonscalar properties, such as those ubiquitously encountered in the context of nonlinear optics (NLO), a field that has contributed immensely to our understanding and control of microscopic dynamics in physics, chemistry, and biology. These tensors, called optical susceptibilities, determine the complete response of a molecule or crystal subject to an external electric field, $\vec{E}$

(Bloembergen, 1982):

$$\vec{P} = \chi^{(1)}\vec{E} + \underbrace{\underbrace{\chi^{(2)}}_{\text{Rank=3}}\underbrace{(\vec{E}\otimes\vec{E})}_{\text{Rank=2}}}_{\text{Rank=3-2}} + \dots \tag{1}$$

Where $\vec{P}$ is the polarization, and $\chi^{(n)}$ is the order-n susceptibility. The progress in our understanding of NLO processes, however, has historically been limited to the rate of discovery of efficient nonlinear optical media, which is often slow and founded on trial-and-error workflows. Although conventional computational chemistry such as density functional theory has been a supplementary source of data (Valdivia-Berroeta et al., 2022), the time complexity of these methods makes it difficult to compute these properties with sufficient accuracy *and* coverage over chemical space. Intelligent algorithms may help greatly expedite the discovery of NLO media but has remained largely unexplored to date due to a lack of available data and the difficulty in predicting these properties. Since NLO processes in gases are generally weak, these effects are primarily observed experimentally in the solid-state, meaning anisotropic effects can be dramatic. Thus, it is essential to know the full susceptibility tensor, rather than the rotationally-averaged scalar value. The full susceptibility tensors are always subject to two strict symmetry constraints: the first is equivariance, expressed as:

$$\chi_m^{l\ (n)}(R\mathbf{r}) = \sum_{m'} \chi_m^{l\ (n)} T_{m'}^l(\mathbf{r}) \tag{2}$$

Where $\chi_m^{l\ (n)}$ is the order-n optical susceptibility tensor in the spherical basis with degree l and order m, $D_{mm'}$ is a Wigner-D matrix, and R is an arbitrary 3-dimensional rotation matrix. The second constraint is stated by Neumann's principle, which enforces that the symmetry group of the molecule $\mathcal{G}_{\text{Crys}}$ is a subset of the group of symmetries of the property tensor $\mathcal{G}_T$ (Neumann, 1885; Alejo-Molina et al., 2014)

$$\mathcal{G}_{\text{Crys}} \subseteq \mathcal{G}_T \tag{3}$$

And thus all symmetry elements of the global molecular point group leave the property tensor unchanged, further reducing the number of independent values in the tensor. Depending on the optical frequencies and materials, there may be additional symmetry constraints (e.g. Kleinman Kleinman (1962)) which can also be enforced in these networks.

The second-order tensor, $\chi^{(2)}$, often called hyperpolarizability, is the lowest-order nonlinear susceptibility. It describes common three-wave mixing processes such as sum and difference frequency generation, second harmonic generation, and optical rectification, all of which form the bases of numerous important spectroscopic techniques (Boyd, 2008). Previous attempts to predict this property using machine-learning have been limited to very narrow-classes of materials and prediction of the *scalar* rotational average (Munn et al., 1994; Tuan-Anh & Zaleśny, 2020; Öberg et al., 2001; Ivonina et al., 2021).

In this work, we construct a simple equivariant message-passing neural network designed to directly predict the full second-order hyperpolarizability tensor. This model, even when randomly initialized, respects both the equivariant and Neumann symmetry constraints, leading to highly efficient training on a diverse molecular dataset and ultimately accurate predictions of full hyperpolarizability tensors. In addition, a recent work by Lee et al. (2024) used dimensionality reduction to show that many of the higher-degree internal representations do not form a structured representation space which correlates with lower prediction accuracy. While this was demonstrated for scalar property prediction, it remains an open question whether the same is true in the direct prediction of higher-order tensorial properties, where one might intuitively expect higher-order components to contribute significantly. Building on this work, we employ PHATE analysis (Moon et al., 2019) to visualize the learned internal representations at both the nodal and graph level. We find that while higher-order components exhibit detailed structure at the node level and partial organization at the graph level. We additionally discuss the implications and applications of our findings.

2 RESULTS

Atomic coordinates of ∼12,000 molecules from the QM9 molecular dataset (Blum & Reymond, 2009; Rupp et al., 2012) were sampled, and used as input for density functional theory (DFT) calculations of hyperpolarizablity tensors in the zero-frequency/D.C. limit, at the B3LYP/6-31G(d) level

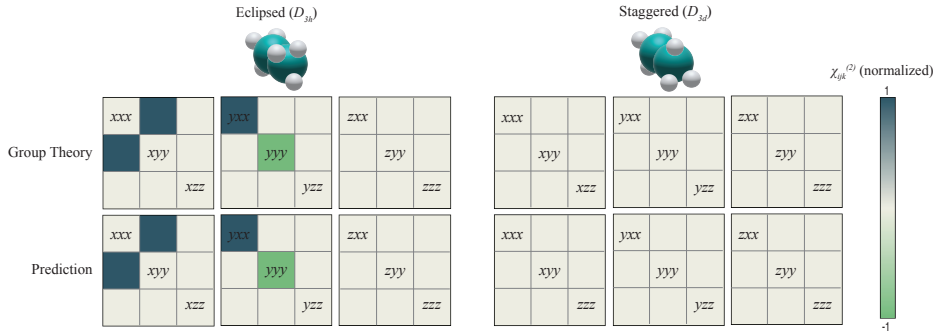

Figure 1: Toy model comparing the prediction symmetries of an untrained model of the full 27 element hyperpolarizability tensor in the Cartesian basis on the eclipsed (LEFT) and staggered (RIGHT) configurations of ethane.

of theory. In the D.C. limit, the hyperpolarizability tensor is fully symmetric with respect to permutations of its indices, which reduces the tensor from 27 to 10 independent values Kleinman (1962). In the spherical basis, three of these values come from the $l = 1$ (dipole) component, and the remaining seven from the $l = 3$ (octopole) component. This dataset is available at Ref.(Miedaner, 2025)

We implement a simplified equivariant message-passing architecture based on the NequIP convolutional layers, truncating the internal representations to order $l_{max} = 4$. The primary architecture difference is in the output step: rather than projecting to a scalar ($l = 0$) value, we implement an output head that reduces the full equivariant feature set to $l = 1$ and $l = 3$ components, which are then used to reconstruct the full hyperpolarizability tensor in the spherical or Cartesian bases. Figure 1 visualizes the power of this equivariant model architecture using ethane in both its eclipsed and staggered configurations as examples. While indistinguishable by their chemical composition, the conformations belong to two different point groups, namely D_{3h} and D_{3d} for the eclipsed and staggered configurations, respectively. The presence of an inversion center in D_{3d} requires all tensor elements to be vanishing, while group theory predicts four non-vanishing elements for the D_{3h} point group (Boyd, 2008). When the full hyperpolarizability tensors are reconstructed from an untrained, randomly initialized model, as shown in Figure 1, the output retains the correct symmetry of the tensor predicted from a group theoretical treatment, including identifying vanishing elements and preserving the relative signs and relationships of the nonvanishing elements. This places strong constraints on the space of learnable functions as a form of physical inductive bias, thereby eliminating the need for augmentation or many training samples.

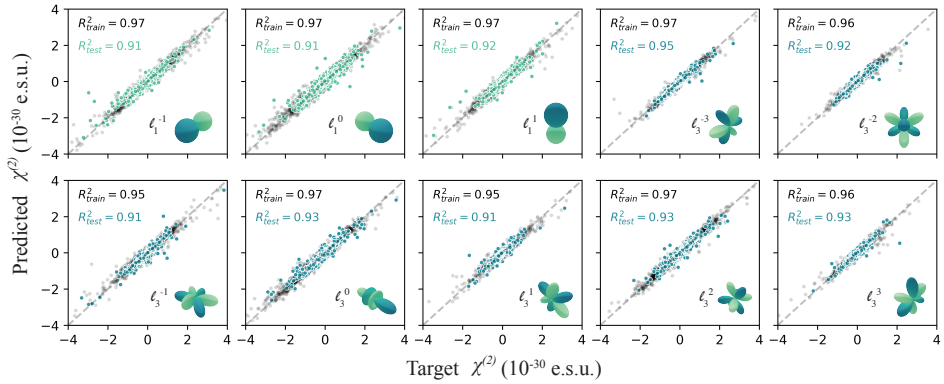

Figure 2: Prediction results for the training (black) and test (colored) data set for individual tensor elements in the spherical basis, and their corresponding R^2-values. The first three plots correspond to the 3 components from $l = 1$, and the last seven from the $l = 3$ component.

For model training, we split the ∼12,000 molecules uniformly into 80:10:10 training, testing, and validation ratios—data distributions with respect to hyperpolarizability values are shown in Appendix A, Figure 4. Model hyperparameters and loss functions are detailed in Table 1 of the appendix. The results for the ten hyperpolarizability elements in the spherical harmonic basis are shown in Figure 2. After just 30 training epochs, each component in the training split has near unity R^2 values. Similarly, prediction on the∼1000 test molecules showed $R^2 > 0.9$ for all 10 elements. The reserved validation set of ∼1000 molecules also has $R^2 > 0.9$ for all 10 elements, shown in Appendix B, Figure 6. The high R^2 values between training and test/validation sets demonstrates good generalization performance, and suggests the model is not overfitting to the training set.

To visualize the learned representations in the test dataset, we employ potential of heat diffusion for affinity-based transition embedding (PHATE) (Moon et al., 2019) Briefly, PHATE works by measuring local similarities and diffusing through the similarity distribution to identify long-range relationships. Unlike other popular data reduction techniques such as PCA(Greenacre et al., 2022), t-SNE(van der Maaten & Hinton, 2008), and UMAP (McInnes et al., 2018), PHATE is capable of preserving both local and global structure from the original manifold at lower dimensions; this allows us to reason about the latent structure based on absolute distances and the distribution of data points. We perform the analysis at node (atomic) and graph (molecular) levels (i.e. a pooled sum over the nodes in a graph) of the test split. We adopt the notation used in `e3nn` Geiger et al. (2022), and denote spherical harmonic degree and parity using notation $l = Lp$ where $L \in \{0, 1, 2, 3, 4\}$ for the spherical harmonic order and $p \in \{o, e\}$ for the parity.

The left panel in Figure 3 shows the PHATE projections for the $l = \{0e, 1o, 3o\}$ components at the node level before and after the interaction layers. The high degree of complexity in the structure and ordering of the node embeddings suggests that the higher-degree spherical harmonic components contribute significantly to the learned representation at the node/atomic level. The right panel in Figure 3 shows the PHATE analysis results at the graph level, after a summation of all the nodes. In all three components, the distributions show a degree of organization based on hyperpolarizability values, suggesting that all components are contributing at least weakly to the learned representation. Interestingly, the degree of fine structure for the nonscalar components, which was present at the node level, appear to be strongly washed out at the graph level. The scalar distribution, however, retains most of the fine structures that were present at the node level. While the results are shown for a small subset of steps and components, the trend was observed for all steps in the model and components with $l > 0$.

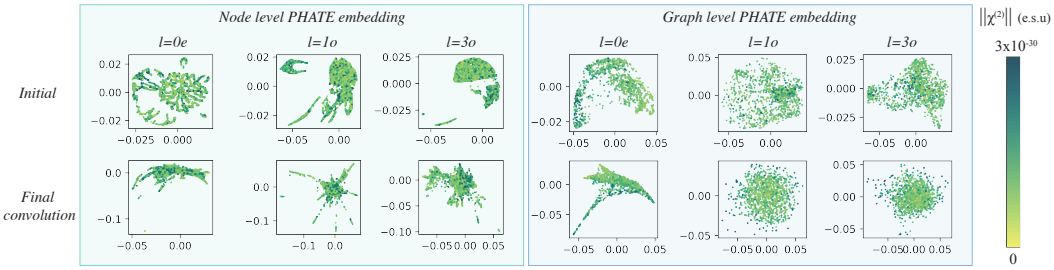

Figure 3: PHATE analysis results for the $l = \{0e, 1o, 3o\}$ components at the node (LEFT PANEL) and graph levels (RIGHT PANEL). Embeddings are shown before and after the interaction layers

To gain further insights, we perform a simple ablation experiment and train a model whose internal representations consist only of the internal representations of the hyperpolarizability tensor plus a scalar component (i.e. $l = \{0e, 1o, 3o\}$), which still respects the equivariance and Neumann symmetry constraints. The results, shown in the Appendix Figure 7, show a marked decrease in the prediction performance for all 10 components, but which is interestingly pronounced for components in the $l = 3$ representation.

3 Discussion and Conclusions

We have shown that using a simple equivariant network can lead to highly accurate prediction of hyperpolarizabilities on a diverse set of molecules. It was recently demonstrated that high-throughput DFT calculations of molecular hyperpolarizability can be an effective way to discovery novel sources of terahertz radiation(Valdivia-Berroeta et al., 2022). Replacing the computationally expensive DFT calculations with this machine learned predictor would greatly expand the accessible search space of THz generating molecules, and also open the possibility of inverse-designing molecular structures which would have high hyperpolarizabilities with a specified symmetry. While THz generating molecules is one area of interest, fast and accurate hyperpolarizability predictions can be used in the discovery and design of all second-order nonlinear media such as second harmonic generation and sum/difference frequency generation. While the results here demonstrate accurate prediction of molecular hyperpolarizability, the architecture of the neural network is agnostic to the phase of the material and could also be used to predict hyperpolarizabilities on crystals.

We also gained mechanistic insights into what aspects of the architecture lead to improved performance. Specifically, the PHATE analysis and ablation study suggest that higher degree spherical harmonic components are contributing to the graph level prediction, contrary to what is observed in the case of scalar prediction (Lee et al., 2024). In the ablation study, we find that some of the orders in the degree-3 component are affected much more strongly than others, which is a question of ongoing research. Most of the graph-level fine structure in the learned representation is localized in the scalar component. We speculate that this indicates higher-degree information is retained in the scalar component. Additionally, it suggests that a simple summation over the nodes may not be effective for high-degree components.

To summarize, we have provided a diverse dataset of molecular hyperpolarizability tensors, which we used to train a simple equivariant neural network. These networks produce highly accurate predictions of hyperpolarizability tensors. Our analysis suggests this prediction performance is largely related to the symmetry constraints which enforces the output tensor to reflect the equivariance and global symmetry of the input molecular structure, in addition to contributions of high-degree components to the learned representation. This work introduces a valuable tool in the discovery of the next generation of nonlinear optical media, which will be built upon in future works.

Acknowledgments

The authors acknowledge the MIT SuperCloud and Lincoln Laboratory Supercomputing Center for providing (HPC, database, consultation) resources that have contributed to the research results reported within this paper/report. P.R.M. and K.A.N. received support from the Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0019126. P.R.M received additional support from The MathWorks Inc.

References

Adalberto Alejo-Molina, Hendradi Hardhienata, and Kurt Hingerl. Simplified bond-hyperpolarizability model of second harmonic generation, group theory, and neumann’s principle. *J. Opt. Soc. Am. B*, 31(3):526–533, Mar 2014. doi: 10.1364/JOSAB.31.000526. URL https://opg.optica.org/josab/abstract.cfm?URI=josab-31-3-526.

Ilyes Batatia, David P Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/4a36c3c51af11ed9f34615b81edb5bbc-Paper-Conference.pdf.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13 (1):2453, 2022. doi: 10.1038/s41467-022-29939-5. URL https://doi.org/10.1038/s41467-022-29939-5.

Nicolaas Bloembergen. Nonlinear optics and spectroscopy. *Rev. Mod. Phys.*, 54:685–695, Jul 1982. doi: 10.1103/RevModPhys.54.685. URL https://link.aps.org/doi/10.1103/RevModPhys.54.685.

L. C. Blum and J.-L. Reymond. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. *J. Am. Chem. Soc.*, 131:8732, 2009.

Robert W. Boyd. *Nonlinear Optics, Third Edition*. Academic Press, Inc., USA, 3rd edition, 2008. ISBN 0123694701.

WA Falcon. Pytorch lightning, 2019. URL https://cir.nii.ac.jp/crid/1370013168774120069.

Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR 2019 Workshop on Representation Learning on Graphs and Manifolds*, 2019. URL https://arxiv.org/abs/1903.02428.

Mario Geiger and Tess Smidt. e3nn: Euclidean neural networks, 2022. URL https://arxiv.org/abs/2207.09453.

Mario Geiger, Tess Smidt, Alby M., Benjamin Kurt Miller, Wouter Boomsma, Bradley Dice, Kostiantyn Lapchevskyi, Maurice Weiler, Michał Tyszkiewicz, Simon Batzner, Dylan Madisetti, Martin Uhrin, Jes Frellsen, Nuri Jung, Sophia Sanborn, Mingjian Wen, Josh Rackers, Marcel Rød, and Michael Bailey. Euclidean neural networks: e3nn, April 2022. URL https://doi.org/10.5281/zenodo.6459381.

Michael Greenacre, Patrick J. F. Groenen, Trevor Hastie, Alfonso Iodice D'Enza, Angelos Markos, and Elena Tuzhilina. Principal component analysis. *Nature Reviews Methods Primers*, 2 (1):100, 2022. doi: 10.1038/s43586-022-00184-w. URL https://doi.org/10.1038/s43586-022-00184-w.

Mariia V. Ivonina, Yuuichi Orimoto, and Yuriko Aoki. Quantum chemistry–machine learning approach for predicting and elucidating molecular hyperpolarizability: Application to [2.2]paracyclophane-containing push–pull polymers. *The Journal of Chemical Physics*, 154(12): 124107, 03 2021. ISSN 0021-9606. doi: 10.1063/5.0040342. URL https://doi.org/10.1063/5.0040342.

D. A. Kleinman. Theory of second harmonic generation of light. *Phys. Rev.*, 128:1761–1775, Nov 1962. doi: 10.1103/PhysRev.128.1761. URL https://link.aps.org/doi/10.1103/PhysRev.128.1761.

Kin Long Kelvin Lee, Mikhail Galkin, and Santiago Miret. Deconstructing equivariant representations in molecular systems. In *AI for Accelerated Materials Design - NeurIPS 2024*, 2024. URL https://openreview.net/forum?id=pshyLoyzRn.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=KwmPfARgOTD.

Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018. doi: 10.21105/joss.00861. URL https://doi.org/10.21105/joss.00861.

Peter Miedaner. Molecular hyperpolarizability tensors, January 2025. URL https://doi.org/10.5281/zenodo.14780577.

Kevin R. Moon, David van Dijk, Zheng Wang, Scott Gigante, Daniel B. Burkhardt, William S. Chen, Kristina Yim, Antonia van den Elzen, Matthew J. Hirn, Ronald R. Coifman, Natalia B. Ivanova, Guy Wolf, and Smita Krishnaswamy. Visualizing structure and transitions in high-dimensional biological data. *Nature Biotechnology*, 37(12):1482–1492, 2019. doi: 10.1038/s41587-019-0336-3. URL https://doi.org/10.1038/s41587-019-0336-3.

R. W. Munn, N. S. Munn, and K. J. Hardie. Neural network analysis of molecular hyperpolarisabilities. *Advanced Materials for Optics and Electronics*, 4(1):19–26, 1994. doi: https://doi.org/10.1002/amo.860040104. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/amo.860040104.

F Neumann. Vorlesungen über die theorie der elastizität der festen körper und des lichtäthers, edited by oe meyer. *Leipzig: BG Teubner-Verlag*, 1885.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. *PyTorch: an imperative style, high-performance deep learning library*. Curran Associates Inc., Red Hook, NY, USA, 2019.

M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical Review Letters*, 108:058301, 2012.

Tran Tuan-Anh and Robert Zaleśny. Predictions of high-order electric properties of molecules: Can we benefit from machine learning? *ACS Omega*, 5(10):5318–5325, 2020. doi: 10.1021/acsomega.9b04339. URL https://doi.org/10.1021/acsomega.9b04339. PMID: 32201820.

Gabriel A. Valdivia-Berroeta, Zachary B. Zaccardi, Sydney K. F. Pettit, (Enoch) Sin-Hang Ho, Bruce Wayne Palmer, Matthew J. Lutz, Claire Rader, Brittan P. Hunter, Natalie K. Green, Connor Barlow, Coriantumr Z. Wayment, Daisy J. Ludlow, Paige Petersen, Stacey J. Smith, David J. Michaelis, and Jeremy A. Johnson. Data mining for terahertz generation crystals. *Advanced Materials*, 34(16):2107900, 2022. doi: https://doi.org/10.1002/adma.202107900. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/adma.202107900.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. URL http://jmlr.org/papers/v9/vandermaaten08a.html.

Kjell Öberg, Anders Berglund, Ulf Edlund, and Bertil Eliasson. Prediction of nonlinear optical responses of organic compounds. *Journal of Chemical Information and Computer Sciences*, 41(3):811–814, 2001. doi: 10.1021/ci000402q. URL https://doi.org/10.1021/ci000402q. PMID: 11410062.

A DATASET

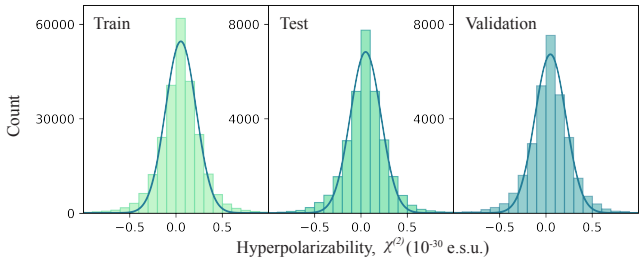

Figure 4: Distribution of hyperpolarizability tensor elements for the training, testing, and validation splits

Atomic coordinates for ∼12,000 molecules were randomly sampled from the QM9 database, and used as input for DFT calculations at the B3LYP/6-31G(d) level of theory. These calculations were performed using the `Polar` subroutine of the Gaussian 09 suite of electronic structure programs. This dataset offers a diverse range of small organic molecules and their hyperpolarizability tensors (Miedaner, 2025).

B MODEL

B.1 ARCHITECTURE AND TRAINING

Our simplified message-passing network was implemented with `e3nn` (Geiger et al., 2022), and closely resembles that used in Lee et al. (2024) which in turn was heavily inspired by NequIP (Batatia et al., 2022). An overview of the network architecture is given in Figure 5. The input to the

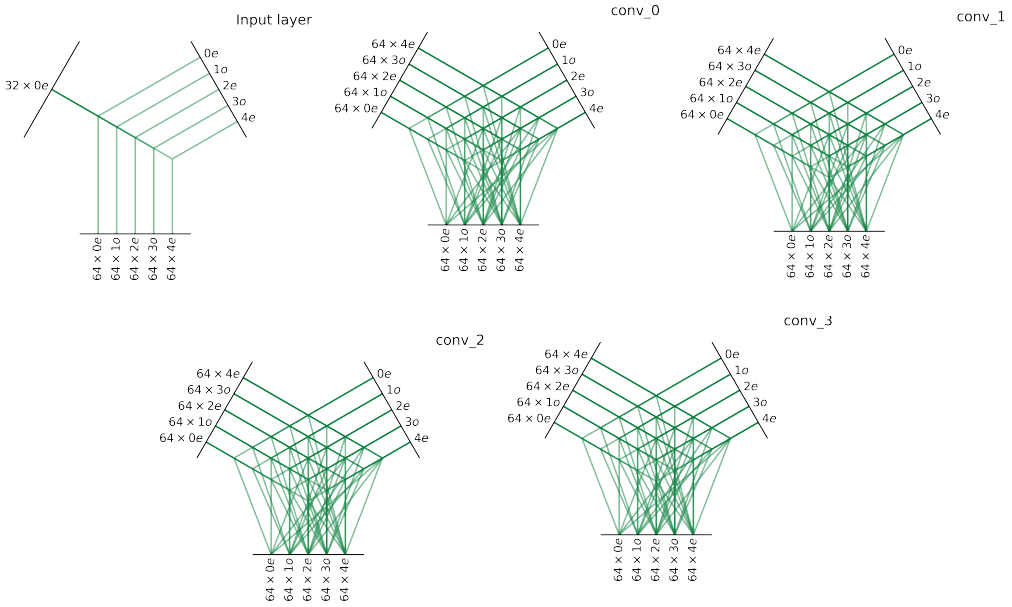

Figure 5: Number of dimensions for each irreducible representation present in the input and internal layers of the network used in this work.

network is a molecular graph containing nodes that correspond to atomic species and 3D coordinates; atomic numbers are mapped to scalar ($l = 0e$) vectors through an embedding table. Edges are formed between nodes via a radial cutoff value (see Table 1), and are subsequently used to compute pairwise distances that are projected onto a basis of radial Bessel functions and passed through a multi-layer perceptron comprising learnable weights. The internal layers consist of equivariant graph convolutions, which consists of fully connected tensor products between atomic and radial features. Features are truncated up to a cutoff of $l_{max} = 4$, i.e. the internal representations belong to the set $l = \{0e, 1o, 2e, 3o, 4e\}$. The final output step is a projection of the output from the final interaction layer to $l = 1o$ and $3o$ components.

The loss was computed individually for the $l = 1$ and $l = 3$ components, and the net loss function, $\mathcal{L}$ was defined as the average of the two values:

$$\mathcal{L} = \frac{\sum MSE_{l=1} + \sum MSE_{l=3}}{2} \tag{4}$$

Where MSE is mean-squared error. The model and training were implemented using PyTorch (Paszke et al., 2019), PyTorch Geometric (Fey & Lenssen, 2019), and PyTorch Lightning (Falcon, 2019) libraries. Training was performed using 2 GPUs and 30 total epochs. The hyperparameters used for model in this paper are given in Table 1.

B.2 VALIDATION DATASET RESULTS

Figure 6 shows the prediction results for the 10% of molecules in the validation split. An R^2 of > 0.9 is obtained for all elements of the tensor. The results are shown in the spherical basis, as described in the main text.

Table 1: Model hyperparameters

Hyperparameter	Value
Optimizer	AdamW
Learning rate	$1e-3$
Batch size	8
Internal l-values	$\{0e, 1o, 2e, 3o, 4e\}$
Degree normalization	3
Initial atom embedding dim.	32
Hidden dimensions	64
# of Bessel functions	10
Radial cutoff	4.0
# of interaction blocks	3

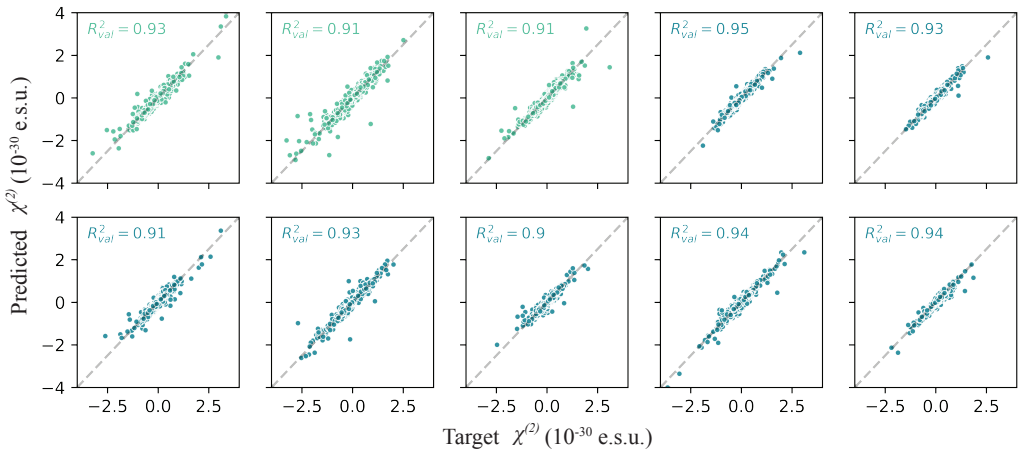

Figure 6: Prediction accuracy of trained model on validation split.

B.3 Ablation experiment

In order to further assess the contribution of higher degree spherical harmonic components to the learned representation, we perform a simple ablation experiment. We train a model with $l = \{0e, 1o, 3o\}$, which is the simplest model that will still enforce equivariance and Neumann's principle for a rank-3 tensor. Additionally, we increase the number of internal dimensions to keep the total number of trainable parameters approximately equal to the model used in the main text. The hyperparameters for the model used in the ablation experiment are shown in Table 2 The training and testing predictions are shown in Figure 7. There is a clear drop in the prediction performance for all ten components. Interestingly, the $3o$ components are affected much more strongly by the ablation as compared to the $1o$ components. These results further suggest the higher degree spherical harmonics contribute in the prediction of the hyperpolarizability tensor.

Table 2: Model hyperparameters

Hyperparameter	**Value**
Optimizer	AdamW
Learning rate	$1e-3$
Batch size	8
Internal l-values	$\{0e, 1o, 3o\}$
Degree normalization	3
Initial atom embedding dim.	64
Hidden dimensions	128
# of Bessel functions	10
Radial cutoff	4.0
# of interaction blocks	3

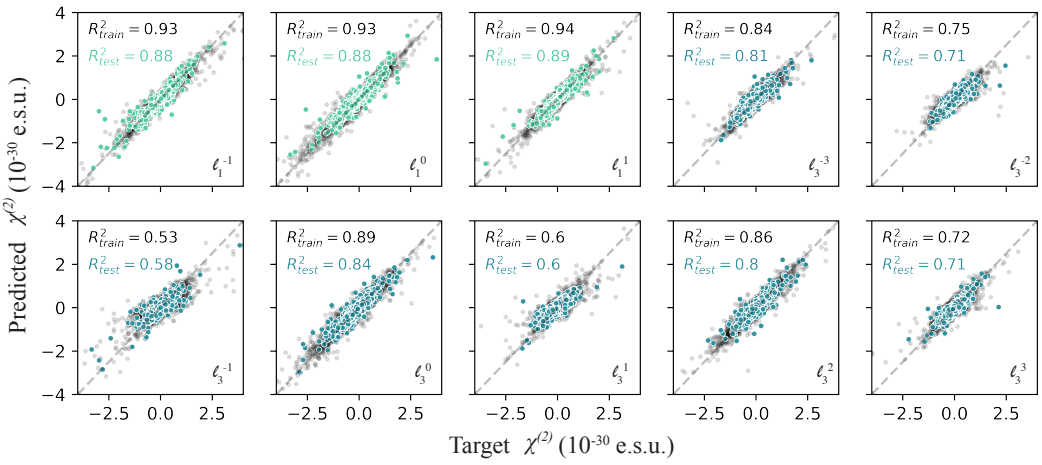

Figure 7: Prediction results for ablation experiment. A model trained on the same data in the main text but with $l = \{0e, 1o, 3o\}$.