A Recipe for Charge Density Prediction

Xiang Fu [1]   **Andrew Rosen** [2,3]   **Kyle Bystrom** [4]   **Rui Wang** [1]   **Albert Musaelian** [4]   **Boris Kozinsky** [4,5]   **Tess Smidt** [1]
Tommi Jaakkola [1]

Abstract

In density functional theory, the charge density is the core attribute of atomic systems from which all chemical properties can be derived. Machine learning methods are promising as a means of significantly accelerating charge density predictions, yet existing approaches either lack accuracy or scalability. We propose a recipe that can achieve both. In particular, we identify three key ingredients: (1) representing the charge density with atomic and virtual orbitals (spherical fields centered at atom/virtual coordinates); (2) using expressive and learnable orbital basis sets (basis function for the spherical fields); and (3) using a high-capacity equivariant neural network architecture. Our method achieves state-of-the-art accuracy while being more than an order of magnitude faster than existing methods. Furthermore, our method enables flexible efficiency–accuracy trade-offs by adjusting the model and/or basis set sizes.

1. Introduction

Density functional theory (DFT) is a computational quantum chemistry method that has enabled countless advancements in the chemical sciences by providing a tractable means to calculate the electronic structure of molecules and materials (Jain et al., 2016). The central concept in DFT is the charge density, a fundamental quantity from which all derivable ground-state physicochemical properties of a system, such as energy and forces, can, in principle, be derived. The most widely used Kohn–Sham formalism (Kohn & Sham, 1965) of DFT offers a reasonable balance between accuracy and computational efficiency among conventional DFT workflows. However, it still scales with a complexity of

roughly $O(N_e^3)$ where N_e is the number of electrons, rendering it computationally expensive and limiting its viability for both large-scale systems and long-timescale *ab initio* molecular dynamics simulations.

In DFT, the solution to the Kohn–Sham equations are reliant on an iterative calculation to identify the charge density that minimizes the potential energy functional for a given atomic configuration. This process, known as converging the self-consistent field, is the main computational expense within DFT. With a machine learning (ML) model that can effectively bypass the Kohn–Sham equations by accurately and efficiently predicting the charge density, the number of steps required to converge the ground-state electron density can be drastically reduced or potentially eliminated altogether by using the predicted charge density as the initial guess. If accurate enough, a machine-learned charge density could also be used to directly predict electronic structure properties, such as the band gap, band structure, and electronic density of states of a material. Furthermore, the charge density itself can provide an enormous amount of insight into a molecule or material. From the charge density, partial atomic charges, dipole moments, atomic spin densities, and effective bond orders can all be directly computed through one of several population analysis methods (Fonseca Guerra et al., 2004; Limas & Manz, 2018). For some materials discovery tasks, the charge density can also be a crucial descriptor depending on the application area (Shen et al., 2023; Yao et al., 2024). Therefore, efficient and accurate representations and ML models for charge density prediction are highly desirable as a means of accelerating the discovery of promising molecules and materials.

In machine learning workflows, the charge density is a volumetric, data-rich object, usually represented as voxels with a grid resolution of around 0.1 Å (Jørgensen & Bhowmik, 2022; Shen et al., 2022). This poses a challenge, as even relatively small molecules and materials can require hundreds of thousands to millions of grid points to represent the charge density at this (relatively coarse) resolution. At the same time, small deviations in the charge density that result from too coarse of a representation can have a substantial impact on the energy and other derivable properties. This need for both efficiency and accuracy creates a significant challenge for ML methods.

[1]Masachusetts Institute of Technology [2]UC Berkeley [3]Lawrence Berkeley National Laboratory [4]Harvard John A. Paulson School Of Engineering And Applied Sciences [5]Robert Bosch Research and Technology Center. Correspondence to: Xiang Fu <xiangfu@csail.mit.edu>.
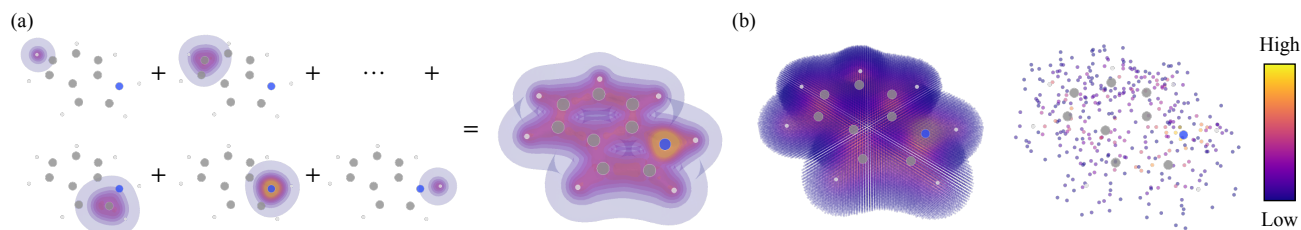
Figure 1. (a) Illustration of the orbital-based method for charge density representation for an example molecule (indole, C_8H_7N). The overall charge density is represented as a sum over spherical-harmonics-based atomic orbital basis functions (spherical fields) centered at each atom. (b) Left: Illustration of the probe-based method for charge density representation. The charge density is represented as a voxel where each grid point (probe node) represents a scalar density at that coordinate. The voxel for the example molecule is of size $108 \times 96 \times 40$. Grid points with very small charge densities (< 0.05) are not visualized. Right: For a probe-based machine learning prediction model, the voxel contains too many grid points to be processed simultaneously. Sampling of the voxel points is needed during training and inference. All charge densities use the same colormap scale at the right-most side of the figure. Atom color code: H (white), C (gray), N (blue). The charge density is from the QM9 charge density dataset (Jørgensen & Bhowmik, 2022).

Existing literature has mainly focused on two approaches for learning to predict the charge density. The first approach (orbital-based), illustrated in Figure 1 (a), is to predict atomic orbital basis set coefficients by regressing over coefficients extracted from DFT data (Fabrizio et al., 2019; Qiao et al., 2022; Rackers et al., 2023; Cheng & Peng, 2023; del Rio et al., 2023). The atomic orbital basis functions are based on the composition of radial functions and spherical harmonics. Under this scheme, the charge density is represented as a set of spherical fields centered around each atom. The real space charge density voxel can be constructed by overlaying the spherical fields and evaluating at each grid point. For orbital-based ML models, both the prediction of the basis set coefficients and the evaluation of the spherical fields are relatively scalable, making this approach efficient at inference time. However, this approach can suffer from sub-optimal accuracy due to the limited representation power of the chosen basis set. In particular, it is challenging for the atom-centered atomic orbitals to complex electronic structures between atoms.

The second approach (probe-based) (Gong et al., 2019; Jørgensen & Bhowmik, 2022; Koker et al., 2023; Pope & Jacobs, 2023), illustrated in Figure 1 (b), is to predict the charge density by inserting "probe nodes" at all grid coordinates of the charge density voxel and applying graph message passing between the atoms and these probe nodes. Finally, the scalar charge density at each grid coordinate is predicted through node-wise readout over the probe nodes. This approach, while expressive and accurate, is computationally expensive. To see why, recall that the number of grid points in the charge density voxel is usually very large for even a small atomic system. Conducting neural message passing over millions of nodes is both computationally and memory intensive. The large number of nodes usually requires sampling a subset of grid points from the charge density voxel (Figure 1 (b), right) in each training or

inference step (Jørgensen & Bhowmik, 2022).

This paper aims to address this accuracy–efficiency dilemma with a new recipe for building representations and ML models for charge density prediction. We identify three key ingredients:

1. We represent the charge density using an atomic orbital basis set (spherical fields centered at each atom) to leverage its efficiency and equivariant properties. Beyond orbitals placed at the atomic coordinates, we further introduce virtual orbitals to improve expressivity. In other words, we also place spherical fields centered at coordinates other than the atomic centers, while ensuring the placement algorithm is SE(3)-equivariant.

2. We use domain-informed and expressive basis sets. In particular, we construct an even-tempered Gaussian basis from an atomic orbital basis set. This allows us to smoothly control the expressivity of the atomic orbitals and enable flexible accuracy–efficiency trade-offs. We make the basis set exponents learnable to further improve expressivity.

3. We use a high-capacity equivariant neural network architecture (eSCN (Passaro & Zitnick, 2023)), which enables efficient training and inference with features of high tensor order for a large dataset.

We apply our recipe to the widely used QM9 charge density benchmark (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014; Jørgensen & Bhowmik, 2022). Our method outperforms existing state-of-the-art methods while being around $30\times$ faster. Furthermore, we can flexibly trade off accuracy and efficiency by adjusting the model/basis size; in doing so, we achieve up to $171\times$ efficiency compared to state-of-the-art methods with only a slight degradation in

Table 1. QM9 charge density prediction error and efficiency on the test set. Metrics for baseline models are from previous papers whenever possible and skipped (-) when unavailable. The metrics (↓ means lower the better, ↑ means higher the better) of the best-performing model are **bold**. The metrics are reported with corresponding standard errors when available. For SCDP models, K is the number of interaction layers in the eSCN backbone, L is the tensor order of the feature representation in the eSCN backbone, and β controls the expressiveness of the even-tempered Gaussian basis set. A higher K, higher L, or lower β indicates a more expressive model. eSCN + VO indicates that virtual orbitals are used. NMAE stands for normalized mean absolute error. Efficiency is measured by molecule per minute (mol. per min.).

	NMAE [%] ↓	NMAE, Split 2 [%] ↓	Mol. per min. [$\min^{-1}$] ↑
i-DeepDFT (Jørgensen & Bhowmik, 2022)	0.357 ± 0.001	-	-
e-DeepDFT (Jørgensen & Bhowmik, 2022)	0.284 ± 0.001	-	-
ChargE3Net (Koker et al., 2023)	0.196 ± 0.001	0.203 ± 0.003	3.95
InfGCN (Cheng & Peng, 2023)	0.869 ± 0.002	0.93	72.00
InfGCN, GTO only (Cheng & Peng, 2023)	-	3.72	-
GPWNO (Kim & Ahn, 2024)	-	0.73	-
SCDP models (Ours)			
eSCN, $K = 4, L = 3, \beta = 2.0$	0.504 ± 0.001	0.514 ± 0.003	675.47
eSCN, $K = 8, L = 6, \beta = 2.0$	0.434 ± 0.006	0.452 ± 0.017	567.19
eSCN, $K = 8, L = 6, \beta = 1.5$	0.381 ± 0.001	0.391 ± 0.002	442.25
eSCN + VO, $K = 8, L = 6, \beta = 2.0$	0.237 ± 0.001	0.250 ± 0.002	231.21
eSCN + VO, $K = 8, L = 6, \beta = 1.5$	0.206 ± 0.001	0.220 ± 0.002	177.14
eSCN + VO, $K = 8, L = 6, \beta = 1.3$	0.196 ± 0.001	0.209 ± 0.002	136.92
SCDP models fine-tuned with scaling factors (Ours)			
eSCN, $K = 4, L = 3, \beta = 2.0$	0.432 ± 0.001	0.438 ± 0.003	644.00
eSCN, $K = 8, L = 6, \beta = 2.0$	0.369 ± 0.007	0.386 ± 0.018	544.56
eSCN, $K = 8, L = 6, \beta = 1.5$	0.346 ± 0.001	0.354 ± 0.002	419.57
eSCN + VO, $K = 8, L = 6, \beta = 2.0$	0.207 ± 0.001	0.220 ± 0.002	221.19
eSCN + VO, $K = 8, L = 6, \beta = 1.5$	0.187 ± 0.001	0.200 ± 0.002	164.94
eSCN + VO, $K = 8, L = 6, \beta = 1.3$	$\mathbf{0.178 \pm 0.001}$	$\mathbf{0.191 \pm 0.002}$	125.29

accuracy. This tunability is valuable, as different applications, material classes, and available computing resources may require drastically different levels of accuracy in the charge density prediction. We conduct an ablation study to justify the significance of each proposed ingredient.

2. Results

We highlight the experimental results of our proposed method in the main paper and defer the related works and method sections to the appendix. Our experiments on the QM9 charge density benchmark aim to validate the effectiveness of our proposed recipe in both accuracy and efficiency. We refer to our method as **SCDP** models, which stands for **S**calable **C**harge **D**ensity **P**rediction models.

Dataset and metrics. The QM9 charge density dataset (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014; Jørgensen & Bhowmik, 2022) contains charge density calculations for 133,845 small organic molecules using the Vienna Ab initio Simulation Package (VASP). We adopt the original split, where 123,835, 50, and 10,000 data points are used for training, validation, and testing, respectively. There

are on average 18 atoms in each molecule and 666,462 grid points in each charge density voxel. The entire dataset takes 1.1 TB of disk space. Following previous works (Jørgensen & Bhowmik, 2022; Koker et al., 2023), we benchmark the prediction accuracy with the normalized mean absolute error, defined as:

$$\text{NMAE}(\hat{\rho}) = \frac{\int_{\mathbb{R}^3} |\rho(\boldsymbol{r}) - \hat{\rho}(\boldsymbol{r})| \, \mathrm{d}V}{\int_{\mathbb{R}^3} |\rho(\boldsymbol{r})| \, \mathrm{d}V}, \qquad (1)$$

where the integration is approximated by summing over the full charge density voxel. We benchmark the efficiency of different methods by the number of molecules predicted per minute (Mol. per min.) on a single NVIDIA A100-80GB-PCIe GPU for the QM9 test split.

Baseline Models. We compare SCDP to several previous works on the QM9 charge density prediction benchmark (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014; Jørgensen & Bhowmik, 2022). i-DeepDFT, e-DeepDFT (Jørgensen & Bhowmik, 2022), and ChargE3Net (Koker et al., 2023) are probe-based methods with different backbone architectures: i-DeepDFT
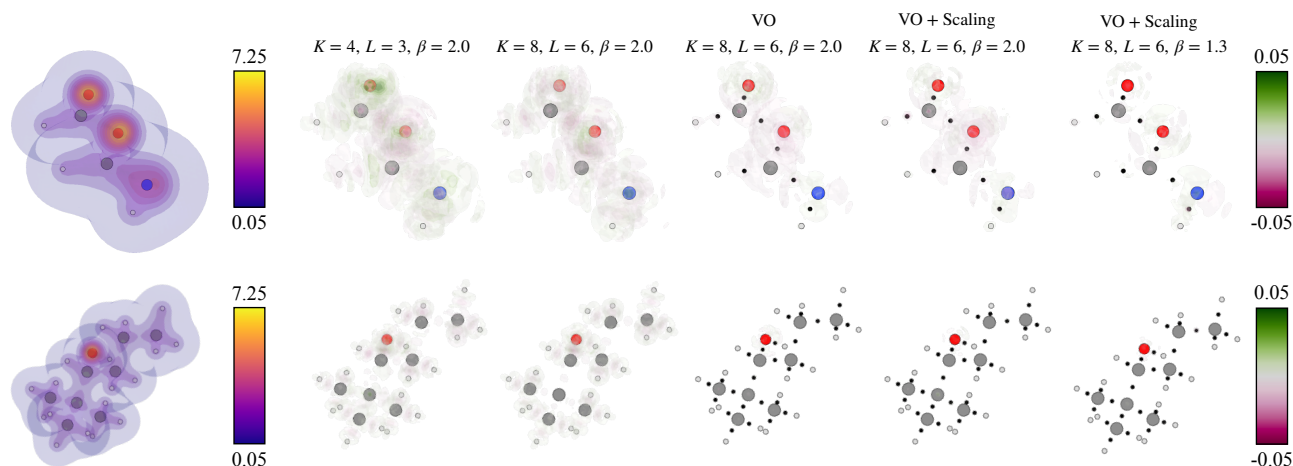
Figure 2. Visualization of the reference charge density and prediction errors for select SCDP models with two representative test molecules (top: $C_2H_3NO_2$ and bottom: $C_8H_{18}O$). The first column is the ground truth charge density with the corresponding color scale. The next five columns are prediction errors from various models which all use the same color scale in the rightmost for error magnitude. The prediction errors significantly reduce with larger model size, virtual orbitals, orbital exponent scaling, and a larger basis set. VO stands for virtual orbitals. Scaling stands for scaling factor fine-tuning. The virtual orbitals significantly reduce errors around chemical bonds. Atom color code: H (white), C (gray), N (blue), O (red), virtual nodes (small, black).
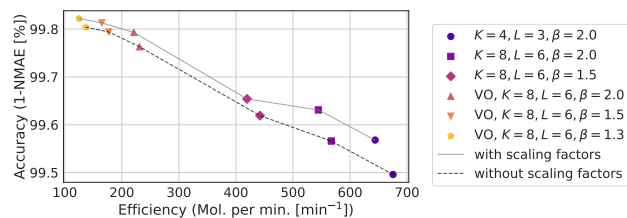

Figure 3. Efficiency–accuracy trade-off for SCDP models. The models with scaling factor fine-tuning form the Pareto front.

uses SchNet (Schütt et al., 2021), e-DeepDFT uses PaiNN (Schütt et al., 2021), while ChargE3Net uses higher-order equivariant features under the tensor field network framework (Thomas et al., 2018; Geiger & Smidt, 2022). InfGCN (Cheng & Peng, 2023) combines GTOs and a shallow network for probe-based inference. It also has a more efficient but less accurate GTO-only variant. GPWNO (Kim & Ahn, 2024) combines GTOs and plane-wave basis sets but still requires a large number (64,000) of probe nodes for constructing the plane wave prediction. NMAE results for i-DeepDFT, e-DeepDFT, and ChargE3Net are from (Koker et al., 2023). NMAE results for InfGC and GPWNO are also from the original papers, which uses a different test split from the default QM9 test split (last 1,600 molecules from the QM9 test split). We benchmark the efficiency of baseline models on our hardware when the source code and pretrained model are publicly available (ChargE3Net and InfGCN). We do not apply any modification to the original code but use optimized configurations for inference to better utilize our GPU: for ChargE3Net, we process 20,000 probes

in each batch instead of the default setting of 2,500 probes per batch, and for InfGCN, we process 40,000 probes in each batch with a batch size of 4.

A significant advance in both accuracy and efficiency. The metrics for all methods are presented in Table 1. We have a series of SCDP models with different model sizes, basis set sizes, as well as options on the inclusion of virtual orbitals and scaling factors. Our best-performing model uses the virtual orbitals described in Appendix B.1, an eSCN of 8 layers and feature representation of order $L = 6$, an even-tempered Gaussian basis with $\beta = 1.3$, and scaling factor fine-tuning. This model achieves an NMAE of 0.178 on the QM9 charge density test set, outperforming the state-of-the-art method ChargE3Net (Koker et al., 2023) — a probe-based method. While being more accurate, our best model also significantly outperforms ChargE3Net by $31.7\times$ in efficiency. Other configurations of our model with smaller model sizes, basis set sizes, and models without virtual orbitals can trade off accuracy for further gains in efficiency. The trade-off curves are visualized in Figure 3. Compared to a more efficient baseline model, InfGCN, all benchmarked configurations of our method are more efficient and significantly outperform in accuracy. These results convincingly demonstrate a significant advance in the accuracy–efficiency trade-off in ML methods for charge density prediction. Figure 5 in Appendix C shows the convergence of validation NMAE during pretraining and fine-tuning of SCDP models. More details on the hyperparameters for model construction and training are included in Appendix C, Table 3.

Ablation Analysis. We discuss the effectiveness of all in-

gredients through an ablation analysis of the performance of different SCDP models. Starting from the most lightweight model with $K = 4, L = 3, \beta = 2.0$ and no virtual orbitals, we first observe that increasing the model size to $K = 8, L = 6$ significantly improves the performance, reducing the NMAE from 0.504% to 0.434%. Next, we increase the basis set size by adjusting β from 2.0 to 1.5, which further reduces the NMAE to 0.381%. The introduction of the virtual orbitals renders a significant gain in accuracy by reducing the error from 0.434% to 0.237% for $\beta = 2.0$ and from 0.381% to 0.196% for $\beta = 1.5$. In particular, the charge density near chemical bonds is significantly more accurate after introducing the virtual orbitals, as visualized in Figure 2. On the other hand, for models with higher capacity, the improved accuracy comes at the cost of efficiency. As shown in Table 1 and Figure 3, higher capacity consistently improves performance while sacrificing efficiency. At the same time, all SCDP models remain highly efficient compared to baseline models. When the scaling factors are introduced, accuracy further improves at a slight cost on efficiency for all models. As shown in Figure 3, models with scaling factors from the Pareto front of all SCDP models benchmarked.

3. Discussion

Charge density is a fundamental quantity for atomic systems and is central to DFT. ML methods for charge density prediction are promising as a means of greatly accelerating DFT by circumventing the iterative procedure used to find the ground-state charge density given a set of atomic coordinates. In this paper, we propose a recipe that combines three ingredients: (1) virtual nodes; (2) expressive basis sets; and (3) high-capacity equivariant networks that collectively outperform state-of-the-art methods in accuracy while being more than an order of magnitude faster. Nevertheless, there are still many directions for further improving the performance of our proposed model. First, the simple heuristic of assigning virtual node coordinates to bond centers may not be optimal. With recent advances in auto-regressive (Daigavane et al., 2023) and diffusion-based (Hoogeboom et al., 2022; Xu et al., 2022) equivariant generative models for 3D atomic structures, learning to insert the virtual orbitals may be a promising avenue for optimizing the placement of virtual nodes, thus improving charge density prediction. Due to the "nearsighted" nature of electronic matter (Prodan & Kohn, 2005), an automated method for placing a higher density of virtual nodes near sites of chemical relevance may also be worthwhile to pursue. Second, we can use basis functions beyond Gaussian-type orbitals (e.g., Slater-type orbitals (Slater, 1930; Chong et al., 2004) or non-decay radial basis functions (Jing et al., 2024)) that may require fewer functions to achieve the same level of accuracy.

There are several limitations of the current paper that we aim to address in future work: (1) Despite substantial improvements in efficiency, the computational cost for training the current model is still significant: our best-performing model was pretrained for six days and fine-tuned for six days over four NVIDIA A100 GPUs for the QM9 charge density prediction task. The scaling factor fine-tuning stage requires a small learning rate, which prolongs training. The prediction model can benefit from resolving the training instability issues with the scaling factors as well as further improvement on the model architecture (Liao et al., 2023). (2) While our approach achieves state-of-the-art performance on the QM9 charge density prediction benchmark, its effectiveness in crystalline materials (Jain et al., 2013; Shen et al., 2022) has yet to be validated. The GTOs and the equivariant network can be applied to materials without modification. The bond-midpoint-based virtual node assignment for molecules can be generalized to crystals through a crystal graph construction algorithm, such as CrystalNN (Zimmermann & Jain, 2020). Alternatively, virtual nodes can be iteratively added to occupy void space inside the unit cell of the material using an algorithm based on the Voronoi diagram (Alexa et al., 2003). The virtual nodes are expected to play an even more important role in prediction accuracy — this is because the diverse atomic species in materials and their complex interactions induce even more complex charge density patterns. (3) To better validate the practical utility of the predicted charge density, evaluation on the reduction of self-consistent field calculations, or on recovering physical observables, such as energy and forces (Jørgensen & Bhowmik, 2022; Sunshine et al., 2023; Koker et al., 2023), will be highly valuable.

References

Alexa, M., Behr, J., Cohen-Or, D., Fleishman, S., Levin, D., and Silva, C. T. Computing and rendering point set surfaces. *IEEE Transactions on visualization and computer graphics*, 9(1):3–15, 2003.

Anderson, B., Hy, T. S., and Kondor, R. Cormorant: Covariant molecular neural networks. *Advances in neural information processing systems*, 32, 2019.

Bardo, R. D. and Ruedenberg, K. Even-tempered atomic orbitals. vi. optimal orbital exponents and optimal contractions of gaussian primitives for hydrogen, carbon, and oxygen in molecules. *The Journal of Chemical Physics*, 60(3):918–931, 1974.

Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csányi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems*, 35: 11423–11436, 2022.

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):2453, 2022.

Biewald, L. Experiment tracking with weights and biases, 2020. URL https://www.wandb.com/. Software available from wandb.com.

Bihani, V., Mannan, S., Pratiush, U., Du, T., Chen, Z., Miret, S., Micoulaut, M., Smedskjaer, M. M., Ranu, S., and Krishnan, N. A. EGraFFBench: evaluation of equivariant graph neural network force fields for atomistic simulations. *Digital Discovery*, 3(4):759–768, 2024.

Chanussot, L., Das, A., Goyal, S., Lavril, T., Shuaibi, M., Riviere, M., Tran, K., Heras-Domingo, J., Ho, C., Hu, W., et al. Open catalyst 2020 (oc20) dataset and community challenges. *Acs Catalysis*, 11(10):6059–6072, 2021.

Cheng, C. and Peng, J. Equivariant neural operator learning with graphon convolution. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=EjiA3uWpnc.

Chong, D. P., Van Lenthe, E., Van Gisbergen, S., and Baerends, E. J. Even-tempered slater-type orbitals revisited: From hydrogen to krypton. *Journal of computational chemistry*, 25(8):1030–1036, 2004.

Daigavane, A., Kim, S., Geiger, M., and Smidt, T. Symphony: Symmetry-equivariant point-centered spherical harmonics for molecule generation. *arXiv preprint arXiv:2311.16199*, 2023.

del Rio, B. G., Phan, B., and Ramprasad, R. A deep learning framework to emulate density functional theory. *npj Computational Materials*, 9(1):158, 2023.

Duval, A., Mathis, S. V., Joshi, C. K., Schmidt, V., Miret, S., Malliaros, F. D., Cohen, T., Liò, P., Bengio, Y., and Bronstein, M. A hitchhiker's guide to geometric GNNs for 3D atomic systems. *arXiv preprint arXiv:2312.07511*, 2023.

Eichkorn, K., Treutler, O., Öhm, H., Häser, M., and Ahlrichs, R. Auxiliary basis sets to approximate coulomb potentials. *Chemical physics letters*, 240(4):283–290, 1995.

Fabrizio, A., Grisafi, A., Meyer, B., Ceriotti, M., and Corminboeuf, C. Electron density learning of noncovalent systems. *Chemical science*, 10(41):9424–9432, 2019.

Focassio, B., Domina, M., Patil, U., Fazzio, A., and Sanvito, S. Linear jacobi-legendre expansion of the charge density for machine learning-accelerated electronic structure calculations. *npj Computational Materials*, 9(1):87, 2023.

Fonseca Guerra, C., Handgraaf, J.-W., Baerends, E. J., and Bickelhaupt, F. M. Voronoi deformation density (VDD) charges: Assessment of the Mulliken, Bader, Hirshfeld, Weinhold, and VDD methods for charge analysis. *Journal of computational chemistry*, 25(2):189–210, 2004.

Fu, X., Wu, Z., Wang, W., Xie, T., Keten, S., Gomez-Bombarelli, R., and Jaakkola, T. S. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL https://openreview.net/forum?id=A8pqQipwkt. Survey Certification.

Gasteiger, J., Becker, F., and Günnemann, S. Gemnet: Universal directional graph neural networks for molecules. *Advances in Neural Information Processing Systems*, 34: 6790–6802, 2021.

Geiger, M. and Smidt, T. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022.

Gong, S., Xie, T., Zhu, T., Wang, S., Fadel, E. R., Li, Y., and Grossman, J. C. Predicting charge density distribution of materials using a local-environment-based graph convolutional network. *Physical Review B*, 100(18):184103, 2019.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10. 1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

Hoogeboom, E., Satorras, V. G., Vignac, C., and Welling, M. Equivariant diffusion for molecule generation in 3D. In

International conference on machine learning, pp. 8867–8887. PMLR, 2022.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Inc., P. T. Collaborative data science, 2015. URL https://plot.ly.

Jain, A., Ong, S. P., Hautier, G., Chen, W., Richards, W. D., Dacek, S., Cholia, S., Gunter, D., Skinner, D., Ceder, G., et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

Jain, A., Shin, Y., and Persson, K. A. Computational predictions of energy materials using density functional theory. *Nature Reviews Materials*, 1(1):1–13, 2016.

Jing, B., Jaakkola, T. S., and Berger, B. Equivariant scalar fields for molecular docking with fast fourier transforms. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=BIveOmD1Nh.

Jørgensen, P. B. and Bhowmik, A. Equivariant graph neural networks for fast electron density estimation of molecules, liquids, and solids. *npj Computational Materials*, 8(1): 183, 2022.

Kim, S. and Ahn, S. Gaussian plane-wave neural operator for electron density estimation. *arXiv preprint arXiv:2402.04278*, 2024.

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

Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

Koker, T., Quigley, K., Taw, E., Tibbetts, K., and Li, L. Higher-order equivariant neural networks for charge density prediction in materials. *arXiv preprint arXiv:2312.05388*, 2023.

Larsen, A. H., Mortensen, J. J., Blomqvist, J., Castelli, I. E., Christensen, R., Dułak, M., Friis, J., Groves, M. N., Hammer, B., Hargus, C., et al. The atomic simulation environment—a python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.

Liao, Y.-L. and Smidt, T. Equiformer: Equivariant graph attention transformer for 3D atomistic graphs. *arXiv preprint arXiv:2206.11990*, 2022.

Liao, Y.-L., Wood, B., Das, A., and Smidt, T. EquiformerV2: Improved equivariant transformer for scaling to higher-degree representations. *arXiv preprint arXiv:2306.12059*, 2023.

Limas, N. G. and Manz, T. A. Introducing DDEC6 atomic population analysis: part 4. efficient parallel computation of net atomic charges, atomic spin moments, bond orders, and more. *RSC advances*, 8(5):2678–2707, 2018.

Nickolls, J., Buck, I., Garland, M., and Skadron, K. Scalable parallel programming with cuda: Is cuda the parallel programming model that application developers have been waiting for? *Queue*, 6(2):40–53, 2008.

Ong, S. P., Richards, W. D., Jain, A., Hautier, G., Kocher, M., Cholia, S., Gunter, D., Chevrier, V. L., Persson, K. A., and Ceder, G. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

Passaro, S. and Zitnick, C. L. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In *International Conference on Machine Learning*, pp. 27420–27438. PMLR, 2023.

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.

Pope, P. and Jacobs, D. Towards combinatorial generalization for catalysts: A kohn-sham charge-density approach. *Advances in Neural Information Processing Systems*, 36, 2023.

Pritchard, B. P., Altarawy, D., Didier, B., Gibson, T. D., and Windus, T. L. New basis set exchange: An open, up-to-date resource for the molecular sciences community. *Journal of chemical information and modeling*, 59(11): 4814–4820, 2019.

Prodan, E. and Kohn, W. Nearsightedness of electronic matter. *Proceedings of the National Academy of Sciences*, 102(33):11635–11638, 2005.

Qiao, Z., Christensen, A. S., Welborn, M., Manby, F. R., Anandkumar, A., and Miller III, T. F. Informing geometric deep learning with electronic interactions to accelerate quantum chemistry. *Proceedings of the National Academy of Sciences*, 119(31):e2205221119, 2022.

Rackers, J. A., Tecot, L., Geiger, M., and Smidt, T. E. A recipe for cracking the quantum scaling limit with machine learned electron densities. *Machine Learning: Science and Technology*, 4(1):015027, 2023.

Ramakrishnan, R., Dral, P. O., Rupp, M., and Von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Scientific data*, 1(1):1–7, 2014.

Ruddigkeit, L., Van Deursen, R., Blum, L. C., and Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *Journal of chemical information and modeling*, 52(11):2864–2875, 2012.

Schuchardt, K. L., Didier, B. T., Elsethagen, T., Sun, L., Gurumoorthi, V., Chase, J., Li, J., and Windus, T. L. Basis set exchange: a community database for computational sciences. *Journal of chemical information and modeling*, 47(3):1045–1052, 2007.

Schütt, K., Unke, O., and Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, pp. 9377–9388. PMLR, 2021.

Shen, J.-X., Munro, J. M., Horton, M. K., Huck, P., Dwaraknath, S., and Persson, K. A. A representation-independent electronic charge density database for crystalline materials. *Scientific data*, 9(1):661, 2022.

Shen, J.-X., Li, H. H., Rutt, A., Horton, M. K., and Persson, K. A. Topological graph-based analysis of solid-state ion migration. *npj Computational Materials*, 9(1):99, 2023.

Slater, J. C. Atomic shielding constants. *Physical review*, 36(1):57, 1930.

Sun, Q., Zhang, X., Banerjee, S., Bao, P., Barbry, M., Blunt, N. S., Bogdanov, N. A., Booth, G. H., Chen, J., Cui, Z.-H., et al. Recent developments in the PySCF program package. *The Journal of chemical physics*, 153(2), 2020.

Sunshine, E. M., Shuaibi, M., Ulissi, Z. W., and Kitchin, J. R. Chemical properties from graph neural network-predicted electron densities. *The Journal of Physical Chemistry C*, 127(48):23459–23466, 2023.

Thomas, N., Smidt, T., Kearnes, S., Yang, L., Li, L., Kohlhoff, K., and Riley, P. Tensor field networks: Rotation-and translation-equivariant neural networks for 3D point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Unke, O. T., Chmiela, S., Sauceda, H. E., Gastegger, M., Poltavsky, I., Schütt, K. T., Tkatchenko, A., and Müller, K.-R. Machine learning force fields. *Chemical Reviews*, 121(16):10142–10186, 2021.

Weigend, F. and Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7:3297, 2005. doi: 10.1039/ b508541a.

Weiler, M., Geiger, M., Welling, M., Boomsma, W., and Cohen, T. S. 3D steerable CNNs: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems*, 31, 2018.

Xu, M., Yu, L., Song, Y., Shi, C., Ermon, S., and Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation. *arXiv preprint arXiv:2203.02923*, 2022.

Yao, Z., Zhao, Y., Zhang, W., and Burton, L. A. Assessing the design rules of electrides. *Journal of Materials Chemistry C*, 2024.

Zimmermann, N. E. and Jain, A. Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity. *RSC advances*, 10(10):6063–6081, 2020.

Zitnick, L., Das, A., Kolluru, A., Lan, J., Shuaibi, M., Sriram, A., Ulissi, Z., and Wood, B. Spherical channels for modeling atomic interactions. *Advances in Neural Information Processing Systems*, 35:8054–8067, 2022.

A. Related Works

ML methods for charge density prediction. Orbital-based methods predict coefficients for the orbital basis set functions to recover the target charge density. Past works have explored Gaussian processes (Fabrizio et al., 2019) and graph neural networks (Qiao et al., 2022; Rackers et al., 2023; Cheng & Peng, 2023; del Rio et al., 2023) in small molecules, water, and materials systems. (Focassio et al., 2023) used Jacobi-Legendre expansion — a many-body extension of atomic orbitals — for representing and predicting the charge density. These approaches, while efficient, suffer from lower accuracy in benchmarks such as QM9 (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014; Jørgensen & Bhowmik, 2022) and the Materials Project (Jain et al., 2013; Shen et al., 2022) charge density datasets. Probe-based methods, on the other hand, predict the charge density by neural message passing between the atoms and probe nodes at all grid points. These methods (Gong et al., 2019; Jørgensen & Bhowmik, 2022; Koker et al., 2023; Pope & Jacobs, 2023) have shown superior accuracy in both molecules and materials but suffer from poor scalability, as they require neural processing of millions of probe nodes for molecule/material structures of tens of atoms. Recent works also explored a combination of atomic orbitals and probe-based methods (Cheng & Peng, 2023) or plane-wave basis sets (Kim & Ahn, 2024). However, both methods still require neural message passing with a large number of probe nodes, which limits their scalability. In the present work, we combine virtual nodes, even-tempered Gaussian basis, and trainable basis functions to greatly improve the expressivity of orbital basis functions.

Equivariant neural networks. Equivariant neural networks (Thomas et al., 2018; Anderson et al., 2019; Weiler et al., 2018; Gasteiger et al., 2021; Schütt et al., 2021; Geiger & Smidt, 2022; Liao & Smidt, 2022; Passaro & Zitnick, 2023) use equivariant representations and processing layers that can preserve roto-translational symmetries that are critical to atomistic modeling tasks. Equivariant models have shown advantages in ML potentials with respect to the accuracy, sample complexity, and molecular dynamics simulation capabilities (Unke et al., 2021; Batzner et al., 2022; Fu et al., 2023; Bihani et al., 2024) in addition to charge density prediction tasks (Jørgensen & Bhowmik, 2022; Koker et al., 2023). This is because atomic forces and charge densities are indeed SE(3)-equivariant with regard to the input atomic coordinates. In this work, we leverage recent advances in methods for building more expressive and scalable equivariant architectures (Passaro & Zitnick, 2023) to improve the accuracy and scalability of charge density prediction.

B. Methods

Our recipe for building ML charge density prediction capabilities involves two complementary aspects: the charge density representation and the prediction model.

B.1. Charge Density Representation

Gaussian-type orbitals (GTOs) are widely used as basis sets for representing electron configurations in quantum chemistry (Eichkorn et al., 1995). They are spherical Gaussian functions centered at atomic coordinates. For an atom i at coordinate r_i, a GTO basis function with exponent α, angular momentum quantum number (also called tensor order or degree) l, and magnetic quantum number m is given by the following expression:

$$\Phi_{\alpha,l,m,r_i}(r) \equiv R_l(r)Y_{l,m}\left(\frac{r-r_i}{r}\right)$$

$$= z_{\alpha,l}\exp(-\alpha r^2)r^l Y_{l,m}\left(\frac{r-r_i}{r}\right), \quad (2)$$

where $r = ||r - r_i||$ is the distance from a query coordinate r to the atom coordinate r_i and $Y_{l,m}$ are real spherical harmonics. $z_{\alpha,l}$ is a normalizing constant, such that $\int_{\mathbb{R}^3}||\Phi||_2^2\,dV = 1$. For the purpose of developing a machine learning model based on GTOs, we choose to represent the charge density, ρ, of an atomic system via a linear combination of many basis functions:

$$\rho(r) = \sum_i^N \sum_j^{N_b^i} \sum_{m=-l_{i,j}}^{l_{i,j}} c_{i,j,m}\Phi_{\alpha_{i,j},l_{i,j},m,r_i}(r), \quad (3)$$

where N is the number of atoms (including virtual ones when applicable), N_b^i is the number of l values (α values) for atom i. It should be noted that the charge density in Kohn–Sham DFT is not computed in this way; rather, Equation (3) is an artificial representation for the sake of training a machine learning model that is inspired by the orbital-like character of GTOs. The basis functions $\Phi_{\alpha_{i,j},l_{i,j},m,r_i}$ are chosen first as the basis set, with a fixed set of l and α values for each element. For example, the values for l and α for hydrogen in the def2-QZVPPD basis set (Weigend & Ahlrichs, 2005) are presented in Appendix C, Table 2. The number of basis functions for an atom i can be derived as $\sum_{j=1}^{N_b^i}(2\cdot l_{i,j}+1)$, because m can be an integer from $-l$ to l. A higher l value corresponds to a more complex angular part of the basis function and allows the corresponding spherical field to be more anisotropic. The number of orbital basis functions for elements H, C, N, O, and F of the def2-QZVPPD basis set (and its even-tempered variant, detailed later in this section) is included in Figure 4. Atoms with more complex electronic structures are often represented with more basis functions. The number of basis
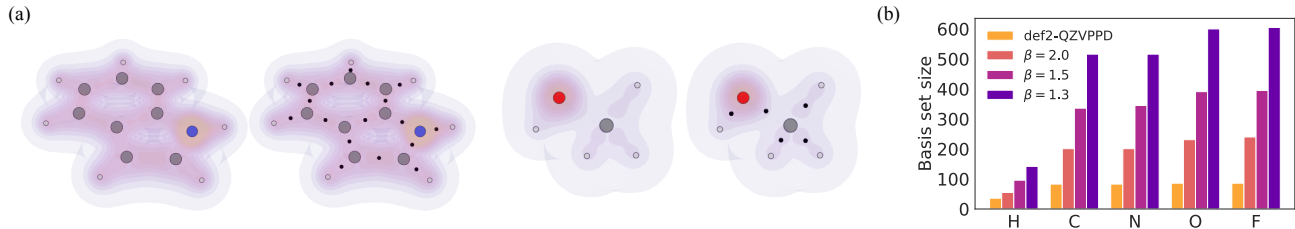
(a)

(b)

Figure 4. (a) Two example molecules (left: indole, C_8H_7N; right: methanol CH_3OH), before and after the bond-midpoint-based virtual coordinates (small black points) are inserted. Atom color code: H (white), C (gray), N (blue), O (red), virtual nodes (small, black). (b) The number of Gaussian-type orbital basis functions for selected elements in the `def2-QZVPPD` basis set and even-tempered Gaussian basis sets derived from it under different β, which controls the number of basis functions as described in Equation (4).

functions, ls, and α values are carefully chosen in existing basis sets such as `def2-QZVPPD`. We refer interested readers to the original papers (Weigend & Ahlrichs, 2005; Schuchardt et al., 2007; Pritchard et al., 2019) for more details regarding the construction of atomic orbital basis sets.

In training the machine learning model, after the basis set is determined, the coefficients $c_{i,j,m}$ are then fit such that Equation (3) best represent the charge density. GTOs have been studied in several previous works (Fabrizio et al., 2019; Rackers et al., 2023; Cheng & Peng, 2023) as a means of representing the charge density with promising results. However, their accuracy still bears significant room for improvement. We next introduce virtual orbitals, even-tempered Gaussian basis, and scaling factors for orbital exponents that greatly improve the expressive power of GTOs for charge density representation.

Virtual orbitals. The atom-centered spherical fields often struggle to capture non-local electronic structures, which induces representation errors. This limitation is effectively addressed with the introduction of virtual orbitals, which define sets of spherical fields located in a position other than the atomic centers. Due to the critical importance of chemical bonds in defining the overall electronic structure, we insert virtual nodes into the midpoint of all chemical bonds for a given molecule (illustrated in Figure 4 (a)). With this method, the coordinates to insert the virtual nodes are SE(3)-equivariant with regard to the input atom coordinates. Therefore, as long as the prediction of the basis set coefficients is SE(3)-equivariant, the overall charge density prediction will still be equivariant after the introduction of the virtual orbitals. We discuss potential extensions to virtual orbital assignments in Section 3. After the virtual nodes are created, one must decide which basis functions to use for the virtual orbitals. In this work, we use the basis functions of element O for the virtual nodes, which offers a balance in accuracy and efficiency based on preliminary experiments.

Even-tempered Gaussian basis. The number of basis func-

tions in existing basis sets, such as `def2-QZVPPD`, may be insufficient for representing complex charge densities. At the same time, expanding the number of basis functions requires care in choosing the values of l and α that improve expressivity effectively. As an example, the `def2-QZVPPD` basis set for hydrogen already contains basis functions with $l = 1$ and $\alpha = 2.292$. Extending this basis set with basis functions with $l = 1$ and $\alpha = 2.0$ will not significantly improve its expressivity because the spherical pattern will be similar to existing basis functions. A general methodology for controlling the basis set size is to use an even-tempered Gaussian basis set (Bardo & Ruedenberg, 1974). Based on a reference atomic orbital basis set (e.g., `def2-QZVPPD`), the even-tempered basis set constructs a series of GTOs with a set of angular momentum quantum numbers l determined by the atomic number and exponents α given by:

$$\alpha_k = \alpha \cdot \beta^k \quad \text{for } k = 0, 1, 2, \ldots, N_l. \quad (4)$$

For each spherical harmonics degree l, α and N_l are chosen such that the exponents in the reference atomic orbital basis set are well-covered[1]. β controls the number of basis functions — a smaller β creates a more expressive basis set with denser exponents. The use of an even-tempered basis set allows us to smoothly control the number of basis functions N_b^i effectively. Figure 4 (b) shows how the number of orbital basis functions for elements H, C, N, O, and F grows with a smaller β for the even-tempered Gaussian basis derived from the `def2-QZVPPD` basis set.

Scaling factors for orbital exponents. In existing orbital-based models (Fabrizio et al., 2019; Qiao et al., 2022; Rackers et al., 2023; Cheng & Peng, 2023; del Rio et al., 2023), while the coefficients for the basis functions are predicted by the ML model, the exponents are fixed for each atom type and not trainable. However, atoms in different local atomic environments can exhibit significantly different charge den-

---

[1]We adopt the implementation of PySCF (Sun et al., 2020) and refer interested readers to the original paper/code for more details on the construction of even-tempered Gaussian basis.

sity patterns around them, especially for the virtual orbitals that aim at capturing interatomic interactions. To further improve the expressivity of the basis set, we make the exponents trainable by learning a positive scaling factor $s > 0$, such that Equation (2) becomes:

$$\Phi_{\alpha,l,m,\boldsymbol{r}_i}(\boldsymbol{r}, s) = z_{\alpha,l,s} \exp(-s \cdot \alpha r^2) r^l Y_{l,m} \left(\frac{\boldsymbol{r} - \boldsymbol{r}_i}{r} \right). \tag{5}$$

The charge density is now represented with coefficients $c_{i,j,m}$ and scaling factors $s_{i,j}$ as:

$$\rho(\boldsymbol{r}) = \sum_i^N \sum_j^{N_b^i} \sum_{m=-l_{i,j}}^{l_{i,j}} c_{i,j,m} \Phi_{\alpha_{i,j},l_{i,j},m,\boldsymbol{r}_i}(\boldsymbol{r}, s_{i,j}), \tag{6}$$

where $z_{\alpha,l,s}$ is a normalizing constant such that $\int_{\mathbb{R}^3} ||\Phi||_2^2 \, dV = 1$. The introduction of the learnable scaling factors for the exponents significantly improves the expressive power of our charge density representation but is also prone to instability during training. We resolve the instability issue with a fine-tuning approach detailed in Appendix B.2.

B.2. Prediction Model

Using the atomic orbital basis set representation of charge density, the prediction model aims to predict the basis set coefficients $c_{i,j,m}$ and the scaling factors $s_{i,j}$ for each real and virtual node such that the predicted charge density matches the ground truth density obtained from DFT calculations. The model F takes as input the types $\mathcal{A} = \{\boldsymbol{a}_i \,|\, i = 1, \ldots, N\}$ and coordinates $\mathcal{R} = \{\boldsymbol{r}_i | i = 1, \ldots, N\}$ of all real and virtual nodes:

$$\{c_{i,j,m}, s_{i,j} | i = 1, \ldots, N; j = 1, \ldots, N_b^i;$$
$$m = -l_{i,j}, \ldots, l_{i,j}\} = F(\mathcal{A}, \mathcal{R}). \tag{7}$$

Backbone architecture. Our construction of the ML prediction model is motivated by the following:

- Charge density is SE(3)-equivariant with regard to the input atom coordinates. An equivariant model that can preserve this symmetry is desired. Concretely, the basis set coefficients $c_{i,j,m}$ are SE(3)-equivariant with regard to the input atom coordinates. The scaling factors $s_{i,j}$ are SE(3)-invariant with regard to the input atom coordinates.

- Charge density is data-rich and very sensitive to the local atomic environment. A high-capacity and expressive model is desired.

- Efficiency is key for general applications of charge density prediction. The model should be efficient while being expressive.

Based on these criteria, we consider equivariant model architectures and the balance between capacity and efficiency. For equivariant models, an important aspect of model expressivity is the representation of node and edge features in the form of irreducible representations (irreps) of SO(3)[2]: spherical harmonic coefficients. A higher degree of representation (L) is desired for building high-capacity models. Previous works have employed the PaiNN architecture (Schütt et al., 2021; Jørgensen & Bhowmik, 2022) that is based on Cartesian features (equivalent to $L = 1$) or architectures based on irreps of SO(3) and tensor products (Qiao et al., 2022; Koker et al., 2023; Cheng & Peng, 2023; Geiger & Smidt, 2022) for charge density prediction. However, these models suffer from limited expressivity or scalability. Cartesian features are limited in representing angular information ($L = 1$); meanwhile, the $O(L^6)$ complexity of tensor products limits the degree of representation that can be used while remaining computationally feasible.

In this work, we adopt the equivariant spherical channel network (eSCN) architecture (Passaro & Zitnick, 2023) as our model backbone. While using SE(3)-equivariant representations and processing layers, the convolution layers in eSCN reduce the SO(3) convolutions (Zitnick et al., 2022) or tensor products (Thomas et al., 2018; Batzner et al., 2022) to convolutions in SO(2) that are mathematically equivalent. It reduces the complexity of the convolution operation from $O(L^6)$ to $O(L^3)$. Further, the use of point-wise, spherical non-linearity in eSCN also distinguishes itself from e3nn-based equivariant models that only apply non-linearity to the scalar features in the irreps. In our experiments, we also find that eSCN outperforms alternative architectures, such as tensor field networks (Koker et al., 2023; Thomas et al., 2018) and MACE (Batatia et al., 2022). Using eSCN as the backbone architecture, we get the last-layer latent features $\boldsymbol{x}_i$ for all real/virtual nodes:

$$\{\boldsymbol{x}_i | i = 1, \ldots, N\} = \text{eSCN}(\mathcal{A}, \mathcal{R}). \tag{8}$$

Prediction layers. The features $\boldsymbol{x}_i$ are encoded using multi-channel spherical harmonic coefficients (irreps). Note that the prediction target, basis set coefficients $c_{i,j,m}$, are also encoded as multi-channel spherical harmonic coefficients. For example, for an eSCN with $L = 3$ and a latent dimension of 128, the last-layer latent atom features will be `128x0e + 128x1o + 128x2e + 128x3o`. For the (uncontracted) `def2-QZVPPD` basis set of hydrogen described in Table 2, its irreps are `7x0e + 4x1e + 2x2e + 1x3e` (even parity as charge density is reflection-invariant). The scaling factors $s_{i,j}$ are SE(3)-invariant and can be seen as scalar features of multi-channel irreps (`14x0e` for the

---

[2]We refer interested readers to (Geiger & Smidt, 2022) and (Duval et al., 2023) for more information on equivariant geometric neural networks.

`def2-QZVPPD` basis set of hydrogen). Therefore, we can make equivariant predictions of the basis set coefficients and invariant predictions of the scaling factors for each atom i through a fully connected tensor product layer over the atom features and additional processing:

$$\{c_{i,j,m}, \boldsymbol{h}_i | j = 1, \ldots, N_b^i; m = -l_{i,j}, \ldots, l_{i,j}\} = $$
$$\text{FullyConnectedTensorProduct}(\boldsymbol{x}_i, \boldsymbol{x}_i) \quad (9)$$

$$\{s_{i,j} | j = 1, \ldots, N_b^i\} = $$
$$C_1/(1 + \exp(-\text{Linear}(\boldsymbol{h}_i) + \ln C_2)) + C_3. \quad (10)$$

The basis set coefficients are directly obtained through the fully connected tensor product. The tensor product also produces scalar features $\boldsymbol{h}_i$ (`128x0e` for a 128-channel eSCN), which are used for predicting the scaling factors. The parameterization of Equation (10) allows the prediction to range from $(C_1, C_1 + C_3)$, and the scaling factors will be $C_1/C_2 + C_3$ when the linear network in Equation (10) is zero-initialized. By setting $C_1 = 1.5, C_2 = 2$, and $C_3 = 0.5$, we can limit the range of the scaling factors to be $(0.5, 2)$ (at most halve or double an exponent) and let initial scaling factors be 1 with a zero initialization of the linear layer in Equation (10).

With the predicted coefficients and scaling factors, the charge density prediction $\hat{\rho}$ can be obtained efficiently by evaluating Equation (6) at all grid coordinates of the charge density voxel. We train the model end-to-end with a mean-absolute error loss $\mathcal{L}$ over the charge density:

$$\mathcal{L} = \mathbb{E}_{\boldsymbol{r} \in \text{Data}} \left[|\rho(\boldsymbol{r}) - \hat{\rho}(\boldsymbol{r})| \right]. \quad (11)$$

Fine-tuning for scaling factor prediction. The scaling factors at the exponents lead to significant training instability when the network is trained from scratch. Therefore, we use a fine-tuning approach, where we first pre-train the model with fixed basis set exponents (an even-tempered Gaussian basis derived from `def2-QZVPPD`) and then fine-tune the prediction model with a small learning rate with the learning for scaling factors enabled. To achieve this, we zero-initialize the linear layer in Equation (10) and freeze its weights until the fine-tuning stage.

C. Supplementary Materials

Software. Basis-set-exchange-v0.9.1 (Schuchardt et al., 2007; Pritchard et al., 2019), PySCF-v2.5.0 (Sun et al., 2020), are used to build the orbital basis sets. PyTorch-v1.13.1 (Paszke et al., 2019) and CUDA-v11.6 (Nickolls et al., 2008) are used to build the SCDP models. The eSCN (Passaro & Zitnick, 2023) implementation is adopted from the Open Catalyst Project (Chanussot et al., 2021). We also acknowledge Numpy (Harris et al., 2020), ASE (Larsen et al., 2017), Pymatgen (Ong et al., 2013), wandb (Biewald, 2020), Matplotlib (Hunter, 2007), and Plotly (Inc., 2015).
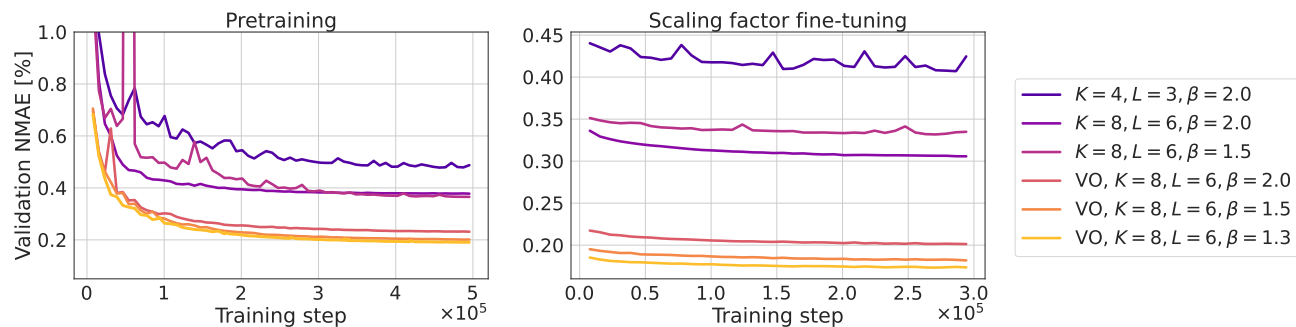
Figure 5. Convergence of validation NMAE during pretraining and finetuning.

Table 2. The (uncontracted) `def2-QZVPPD` basis set for H.

l	α
0	190.6916900
0	28.6055320
0	6.5095943
0	1.8412455
0	0.59853725
0	0.21397624
0	0.080316286
1	2.29200000
1	0.83800000
1	0.29200000
1	0.084063199228
2	2.06200000
2	0.66200000
3	1.39700000

Table 3. Hyperparameters for SCDP models. [1]The cutoff distance used for building the message passing graph. [2]The cutoff distance for computing the charge density using Equation (6). An orbital basis function only influences all grid coordinates within this distance.

Hyperparameter	Value
eSCN	
# interaction layers	$[4, 8]$
$L_{\max}$	$[3, 6]$
$m_{\max}$	2
sphere channels	128
hidden channels	256
edge channels	128
# sphere samples	128
radius cutoff[1]	6 Å
Basis set	
reference basis set	`def2-QZVPPD` (Weigend & Ahlrichs, 2005)
β	$[2.0, 1.5, 1.3]$
orbital inference cutoff[2]	5 Å
Training	
batch size	4
# grid point samples (training, without VO)	$100,000$
# grid point samples (validation/testing, without VO)	$200,000$
# grid point samples (training, with VO)	$60,000$
# grid point samples (validation/testing, with VO)	$120,000$
precision	32
gradient clipping	0.5
# training steps (pretraining)	$500,000$
# training steps (fine-tuning)	$300,000$
optimizer	Adam (Kingma & Ba, 2014)
Adam β_1	0.9
Adam β_2	0.999
Adam ϵ	1×10^{-8}
weight decay	0
initial learning rate (pretraining)	0.001
initial learning rate (fine-tuning)	2×10^{-5}
learning rate scheduler	exponential (LR = initial LR $\times 0.96^{\mathrm{step}/C}$)
terminal learning rate (pretraining)	1×10^{-5}
terminal learning rate (fine-tuning)	2×10^{-6}
Inference	
batch size (without VO)	8
batch size (with VO)	4
Max # grid points in a forward pass for Equation (6)	
eSCN, $K = 4, L = 3, \beta = 2.0$	$2,000,000$
eSCN, $K = 8, L = 6, \beta = 2.0$	$1,000,000$
eSCN, $K = 8, L = 6, \beta = 1.5$	$1,000,000$
eSCN + VO, $K = 8, L = 6, \beta = 2.0$	$600,000$
eSCN + VO, $K = 8, L = 6, \beta = 1.5$	$400,000$
eSCN + VO, $K = 8, L = 6, \beta = 1.3$	$400,000$