Gaussian Plane-Wave Neural Operator For Electron Density Estimation

Seongsu Kim [1] **Sungsoo Ahn** [1]

Abstract

This work studies machine learning for electron density prediction, which is fundamental for understanding chemical systems and density functional theory (DFT) simulations. To this end, we introduce the Gaussian plane-wave neural operator (GPWNO), which operates in the infinite-dimensional functional space using the plane-wave and Gaussian-type orbital bases, widely recognized in the context of DFT. In particular, both high- and low-frequency components of the density can be effectively represented due to the complementary nature of the two bases. Extensive experiments on QM9, MD, and material project datasets demonstrate GPWNO's superior performance over ten baselines.

1. Introduction

The study of electron density is fundamental for understanding the properties and behaviors of chemical systems. Its estimation is a crucial component of Kohn-Sham density functional theory (Kohn & Sham, 1965, DFT) for simulations like ground-state energy estimation with applications to the development of new battery cathodes (Kondrakov et al., 2017), solar cell materials (Chang et al., 2021) or drug design (Adekoya et al., 2022). However, the computational complexity of the DFT is cubic in the system size and is often prohibitively expensive for real-world problems.

Machine learning has shown promise in predicting electron density with less computational complexity. Early works investigated methods based on kernel ridge regression (Grisafi et al., 2018; Bogojeski et al., 2018) and Gaussian process-based regression (Grisafi et al., 2018; Fabrizio et al., 2019). Recently, deep learning methods have been increasingly investigated due to their remarkable capability to capture complex relationships within data. Sinitskiy & Pande (2018) utilized a voxel-based 3D convolutional net with a U-Net architecture to predict density at a voxel level. Jørgensen & Bhowmik (2020) and Jørgensen & Bhowmik (2022) aimed to predict the electron density using graph neural networks, where the graph consists of interacting atom vertices and a special query point vertex for which the density is predicted.

More recently, Cheng & Peng (2023) tackled the problem with neural operator, an attractive paradigm that can map the given molecule into a function space of electron density whose output can be evaluated at an arbitrary point. In particular, they proposed to predict the density using atom-centered Gaussian-type orbital (Hehre et al., 1972, GTO).

In particular, exploration of the plane-wave (PW) basis, a cornerstone in DFT solvers like VASP (Hafner, 2008), Quantum Espresso (Giannozzi et al., 2017), and ABINIT (Gonze et al., 2009), remains absent within the domain of neural operators. The PW basis comprises periodic wavefunctions that meet the periodic boundary condition of the molecule. It is an appealing approach due to its ease of manipulation via fast Fourier transformation and its ability to effectively capture low-frequency signals such as those associated with long-range interactions, e.g., Coloumb potentials. The most similar work is the Fourier neural operator (Li et al., 2020a, FNO) baseline considered by Cheng & Peng (2023), which uses the Fourier basis consisting of periodic functions. The difference is how the Fourier basis does not respect the periodic boundary condition while the PW basis does.

Motivated by this lack of exploration, we aim to predict the electron density using neural operators that operate on the PW basis. However, using a moderately-sized PW basis, e.g., with 10^3 elements, alone cannot accurately express the electron density, which is in agreement with the poor performance of FNO reported by Cheng & Peng (2023). While DFT solvers overcome this by increasing the size, e.g., up to 8.9×10^7, the neural operators are not scalable to such high frequencies. Our focus is to enjoy the benefits of PW basis while keeping the basis size realistic for neural operators. Prior works on DFT, e.g., pseudopotential (Harrison, 1966; Bachelet et al., 1982), linearized augmented plane-wave (Koelling & Arbman, 1975), projector augmented wave (Blöchl, 1994), and mixed density fitting (Sun et al., 2017), hint that it is possible to reduce the required basis

[1]Pohang University of Science and Technology (POSTECH), Pohang, South Korea. Correspondence to: Seongsu Kim <seongsukim@postech.ac.kr>, Sungsoo Ahn <sungsoo.ahn@postech.ac.kr>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

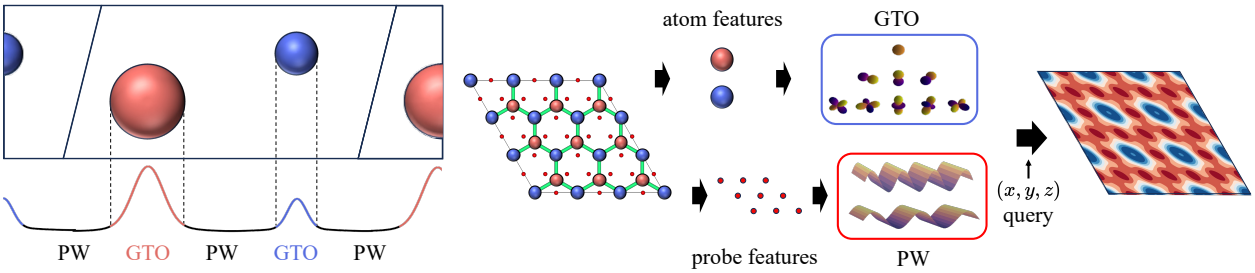


Figure 1: **Illustration of Gaussian plane wave neural operator (GPWNO). (left)** Our GPWNO decomposes the electron density estimation into two regions: high-frequency and fast-decaying regions for GTO basis and low-frequency and slow-decaying regions for the PW basis. **(right)** Our GPWNO makes two predictions based on the GTO and the PW bases. The PW-based prediction is from the lattice-based discretization via the probe nodes, while the GTO-based prediction is from the atom-wise message-passing scheme. Final output is evaluated at an arbitrary query point.

size by decoupling vicinity of atoms from the rest of the regions and mixing local orbitals with PWs.

Contribution. In this work, we propose Gaussian plane-wave neural operator (GPWNO) for estimating the electron density. Our main idea is to decompose the electron density into local atom-wise Gaussian-type orbital (GTO) and global plane-wave (PW) bases, where the bases complement each other for the expression of high- and low-frequency signals. In addition, our design choice allows prediction that is both SE(3)-equivariant and satisfies the periodic boundary condition. We provide an overview of GPWNO in Figure 1.

Our main component is the PW-based electron density prediction, which is based on PW neural operator layers with the convolution parameterized in the reciprocal space of the molecule, i.e., PW coefficients. To this end, we discretize the signal via probe nodes regularly placed along the lattice structure and apply discrete Fourier transformations to compute the convolution. Finally, the density at query nodes is evaluated via message passing. Our main idea on the lattice-based discretization, which allows convolution of the signals in the reciprocal space of molecules while satisfying the periodic boundary condition.

Our next component is the GTO-based electron density prediction, which can be defined as a neural operator in the graphon space (Cheng & Peng, 2023). We also introduce high-frequency masking scheme that allows the PW basis to concentrate on regions outside the vicinity of atoms. Our main idea here is to exploit the nature of GTO complementing the PW basis, which was inspired by DFT solvers like Gaussian plane-wave methods (Lippert et al., 1997), LAPW with local orbitals (Koelling & Arbman, 1975), and mixed density fitting (Sun et al., 2017).

We empirically verify the effectiveness of our methods on three datasets: QM9 (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014; Jørgensen & Bhowmik, 2022), MD (Brockherde et al., 2017; Bogojeski et al., 2020) datasets. We additionally benchmark our algorithm on the newly curated datasets from the material project (Jain et al., 2013; Shen et al., 2021, MP), categorized by seven different crystal family types, e.g., the crystal being triclinic and hexagonal. We show that our algorithm consistently outperforms the prior works by a large margin. We validate the effectiveness of our algorithmic components through an ablation study.

In summary, our contribution is twofold:

- We propose GPWNO for electron density prediction based on the mixture of GTO and PW bases which complement each other for the expression of high- and low-frequency components in the density. We design the framework to satisfy SE(3)-equivariance and periodic boundary conditions.
- We empirically validate our GPWNO for the QM9, MD, and newly curated MP datasets with comparison against ten baselines. We also empirically verify the effectiveness of each component of our algorithm through ablation studies.

2. Related Works

Machine learning for electron density estimation. Initially, researchers have resorted to statistical approaches such as kernel ridge regression (Grisafi et al., 2018; Bogojeski et al., 2018) and symmetry-adapted Gaussian processes (Grisafi et al., 2018; Fabrizio et al., 2019) for electron density estimation. Recent advance in deep learning have opened new avenues for modeling these functions. For example, Sinitskiy & Pande (2018) employed a voxel-based 3D U-Net convolutional network architecture, to predict electron densities at the voxel level.

Furthermore, Jørgensen & Bhowmik (2020) and Jørgensen & Bhowmik (2022) approached electron density prediction using neural message passing on graphs. Koker et al. (2023) advanced this field by utilizing higher-order equivariant features for more expressive representations. Finally, Cheng & Peng (2023) embraced the idea of coefficient learning

within the multicentric approximation framework and developed an equivariant graph neural network (GNN) based on tensor products for predicting density spectra.

Neural operator. Neural operators (Li et al., 2020b;a; Kovachki et al., 2023; Tran et al., 2023, NOs) form a paradigm to map data into infinite-dimensional function spaces where these spaces and the data share a common domain, e.g., from 3D vehicle geometries to pressure on car surfaces (Li et al., 2023), from molecular systems to electron densities (Cheng & Peng, 2023) and from seismic wavefield to probability to P- and S- wave arrivals (Sun et al., 2023). The input function for neural operators can be represented with any level of discretization, grid, resolution, or mesh. The output function, in turn, can be evaluated at any arbitrary point.

Initially, the graph neural operator (Li et al., 2020b, GNO) proposed to integrate kernel operation within graph structures, drawing similarities to the message-passing networks. Next, the Fourier neural operator (Li et al., 2020a, FNO) aims to parameterize and learn operators within the Fourier domain. This is done by discretizing the input domain and applying the fast Fourier transform (FFT). To improve the scalability of FNO, factorized-FNO (Tran et al., 2023, F-FNO), employs Fourier factorization, learning features in the Fourier space independently across each dimension. Finally, to overcome the limits of FNO being restricted to regular grids for using FFT operations, the geometry-informed neural operator (Li et al., 2023, GINO) maps irregular grids into a latent regular grid and then applies FFT.

3. Preliminary

3.1. Problem Definition

Here, we formulate the electron density prediction problem as learning a mapping from a three-dimensional Cartesian coordinate to a real number, given a molecule. We consider *periodic* molecules, where the periodicity is defined by the lattice structure. We also discuss extending our framework to aperiodic molecules.

To be specific, we represent a molecule of N atoms as $\mathcal{M} = (\mathcal{X}, \mathcal{A}, \boldsymbol{L})$ with atom positions $\mathcal{X}$, atomic features $\mathcal{A}$, and unit cell lattice $\boldsymbol{L}$. The atomic positions $\mathcal{X} = \{\boldsymbol{x}_u : u \in \mathcal{V}\}$ and the atom features $\mathcal{A} = \{\boldsymbol{z}_u : u \in \mathcal{V}\}$ where we use vertices $\mathcal{V}$ to index the atoms. Finally, we let the unit cell lattice $\boldsymbol{L} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]^\top$ describe the periodicity where each $\boldsymbol{a}_i$ represents a lattice vector.

Given the molecule $\mathcal{M} = (\mathcal{X}, \mathcal{A}, \boldsymbol{L})$, the infinite molecular representation $(\bar{\mathcal{X}}, \bar{\mathcal{A}})$ with atom positions $\bar{\mathcal{X}}$ and atom features $\bar{\mathcal{A}}$ is defined as follows:

$$\bar{\mathcal{X}} = \left\{ \boldsymbol{x}_{\boldsymbol{t},u} \middle| \boldsymbol{x}_{\boldsymbol{t},u} = \boldsymbol{x}_u + \sum_{d=1}^{3} t_d \boldsymbol{a}_d, \boldsymbol{t} \in \mathbb{Z}^3, u \in \mathcal{V} \right\}, \quad (1)$$

$$\bar{\mathcal{A}} = \{\boldsymbol{z}_{\boldsymbol{t},u} | \boldsymbol{z}_{\boldsymbol{t},u} = \boldsymbol{z}_u, \boldsymbol{t} \in \mathbb{Z}^3, u \in \mathcal{V}\}, \quad (2)$$

where $\boldsymbol{t} = [t_1, t_2, t_3] \in \mathbb{Z}^3$ indicates the change of unit cell over the infinitely repeating periodic patterns.

Consequently, the true electron density function $\rho : \mathbb{R}^3 \to \mathbb{R}$ also satisfies the following *periodic boundary condition*:

$$\rho(\boldsymbol{x}) = \rho\left(\boldsymbol{x} + \sum_{d=1}^{3} t_d \boldsymbol{a}_d\right), \quad (3)$$

where $\boldsymbol{t} = [t_1, t_2, t_3] \in \mathbb{Z}^3$ indicates the change of unit cell over the infinitely repeating periodic patterns.

Extension to aperiodic molecules. For aperiodic molecules, we follow the common practice of formulating them as periodic molecules with a sufficiently large unit cell lattice $\boldsymbol{L}$ with zero-padded boundaries. Examples include DFT solvers (Blöchl, 1995; Hafner, 2008; Ulian et al., 2013) and graph neural networks (Kosmala et al., 2023).

3.2. Plane-Wave Basis

The plane-wave (PW) basis is widely used in electronic structure computation of solid-state systems (Martin, 2004; Kittel, 2005) because it satisfies the periodic boundary condition by construction and is easily manipulated through fast Fourier transformations. We consider the truncated version of the (infinite) PW basis determined by a cutoff in the maximum frequency (or energy), as commonly employed by DFT solvers and Fourier neural operators.

Formally, given a lattice $\boldsymbol{L} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]^\top$, the wavevectors are constructed from reciprocal primitive vectors $(\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3)$ where $\boldsymbol{b}_1$ is defined by $(\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3)$ as:

$$\boldsymbol{b}_1 = 2\pi \frac{\boldsymbol{a}_2 \times \boldsymbol{a}_3}{\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)} \quad (4)$$

and $\boldsymbol{b}_2$ and $\boldsymbol{b}_3$ are defined similarly by $(\boldsymbol{a}_2, \boldsymbol{a}_3, \boldsymbol{a}_1)$ and $(\boldsymbol{a}_3, \boldsymbol{a}_1, \boldsymbol{a}_2)$, respectively. Note the dual relationship between the lattice vectors and the primitive vectors, e.g., $\boldsymbol{a}_1 \cdot \boldsymbol{b}_1 = 2\pi$ and $\boldsymbol{a}_1 \cdot \boldsymbol{b}_2 = 0$.

Finally, the PW basis element $\phi_{\boldsymbol{\lambda}}(\boldsymbol{x})$ indexed by a tuple of integers $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3]^\top \in \mathbb{Z}^3$ is defined as follows:

$$\phi_{\boldsymbol{\lambda}}(\boldsymbol{x}) = \exp\left(j \sum_{d=1}^{3} \lambda_d \boldsymbol{b}_d \cdot \boldsymbol{x}\right), \quad (5)$$

where j is the imaginary number. It should be noted that the plane wave basis element for any $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3]^\top \in \mathbb{Z}^3$ satisfies the periodic boundary condition, i.e. Equation (3). We also point out how the PW basis was originally designed for the wavefunctions to solve Schrödinger's equation (Kittel, 2005). However, in this work, we use them for parameterization of the electron density functions which is the square of the magnitude of the wavefunction.

3.3. Gaussian-Type Orbital Basis

Gaussian-type orbitals (Hehre et al., 1972, GTO) offer an alternative to the PW basis for the expression of electronic structures. We consider GTOs in the following form:

$$\phi_{nlm}(\boldsymbol{x}, \boldsymbol{x}_0) = C_{nlm} R_{nl}(\|\boldsymbol{r}\|) Y_{lm}\left(\frac{\boldsymbol{r}}{\|\boldsymbol{r}\|}\right), \tag{6}$$

where $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{x}_0$, $\boldsymbol{x}_0$ is center of the GTO, C_{nlm} is a normalization constant, $R_{nl}(\cdot)$ is the Gaussian radial basis function (RBF), and $Y_{lm}(\cdot)$ is the spherical harmonics (SH). The index of radial basis n and degree l of the spherical harmonics is a non-negative integer, and the magnetic order m is an integer satisfying $-l \leq m \leq l$. The GTO basis does not satisfy the periodic boundary condition Equation (3).

The use of Gaussian-type orbitals in quantum chemistry has become popular due to their mathematical simplicity and efficiency in electronic structure computation (Hehre et al., 1972). They allow straightforward integrations and often result in faster convergence in numerical simulations compared to the Slater-type orbitals (Gill, 1994).

4. Method

4.1. Overview

Here, we provide a high-level description of our Gaussian plane wave neural operator (GPWNO). Our main idea is to introduce the plane-wave basis set for effectively representing the electron density functions. However, since the electron density function consists of high-frequency elements near the atoms, we additionally introduce the GTO-based basis set for representing such elements. We depict the visualization of the overall architecture in Appendix A and we provide the complexity analysis of our model in Appendix G

Decomposition of prediction. We parameterize our electron density prediction $\rho_{\text{pred}} : \mathbb{R}^3 \to \mathbb{R}$ as follows:

$$\rho_{\text{pred}}(\boldsymbol{x}) = \rho_{\text{PW}}(\boldsymbol{x}) + \rho_{\text{GTO}}(\boldsymbol{x}), \tag{7}$$

where ρ_{PW} and ρ_{GTO} correspond to our prediction using the plane-wave basis set and the GTO-based basis set, respectively. For training, we use the mean-squared error:

$$\mathcal{L}_{\mathcal{M}} = \frac{1}{|\mathcal{D}_{\mathcal{M}}|} \sum_{\boldsymbol{x} \in \mathcal{D}_{\mathcal{M}}} \|\rho(\boldsymbol{x}) - \rho_{\text{pred}}(\boldsymbol{x})\|^2, \tag{8}$$

where we train over different molecules $\mathcal{M} \in \mathcal{D}$ and the locations $\boldsymbol{x} \in \mathcal{D}_{\mathcal{M}}$ is the dataset of locations where the electron density is known, e.g., computed by DFT simulations.

PBC implementation. Importantly, our parameterization guarantees the predictions to satisfy the periodic boundary conditions on PW and GTO-based prediction. Specifically, we utilized the crystal graph framework introduced by Xie & Grossman (2018), which emulates periodic boundary conditions common in computational physics, as outlined in Landau & Binder (2021). In simple terms, we consider each atom to have multiple coordinates, each translated by integer multiples of lattice vector, while maintaining identical node features during message passing.

Equivariance. Both the electron density function and our predictions are scalar fields that are SE(3)-equivariant. In particular, our PW-based prediction is a scalar field since the PW basis aligns with the lattice, which transforms along with the molecule and hence the PW coefficients are invariant.[1] The proof for the GTO-based prediction being a scalar field follows that of Cheng & Peng (2023).

Masking for decomposition. Moreover, the density prediction $\rho_{\text{PW}}(\boldsymbol{x})$, as characterized by the PW basis set, is designed to be zero within a specific cutoff radius for certain atoms. This approach aids in segmenting the target region for each basis and integrates the physical understanding that high-frequency electron density components are indicative of core (non-valence) electrons, which are typically located near an atom's center.

4.2. PW-based Prediction

Here, we further describe our PW-based prediction ρ of the electron density. To this end, following the Fourier neural operator (Li et al., 2020a, FNO), we let each layer of our network implicitly represent a signal updated in both the Fourier space and the Euclidean space. However, unlike the FNO approach, our signal is represented using the PW basis and the corresponding coefficients, which is a subset of Fourier series that satisfies the periodic boundary condition.

Signal initialization. To be specific, we initialize the first layer using composition of tensor field network (Thomas et al., 2018, TFN), grid point-wise embedding, and discrete Fourier transformation. We first obtain the atom-wise embedding of the molecule as $\mathbf{f}_u^{(0)} = \text{TFN}_u(\mathcal{M})$, where TFN_u denotes the TFN output associated with the atom $u \in \mathcal{V}$. We provide more details of our TFN architecture in Appendix B.

Lattice-based discretization. For point-wise embedding, we locate probe nodes $\mathcal{P}$ on a three-dimensional space inside the lattice $\mathbf{L} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]^\top$. Specifically, we uniformly discretized the space inside the lattice and located the probe node p at each point. i.e., a probe node $p \in \mathcal{P}$ is placed at

$$\boldsymbol{x}_p = \frac{d_1}{D}\boldsymbol{a}_1 + \frac{d_2}{D}\boldsymbol{a}_2 + \frac{d_3}{D}\boldsymbol{a}_3, \tag{9}$$

where $d_1, d_2, d_3 < D$ are non-negative integers and the total number of probe nodes is $M = D^3$. Note that this lattice generalizes the regular voxel grids with cubic cells.

[1] We provide proof for the aperiodic molecule in Appendix C

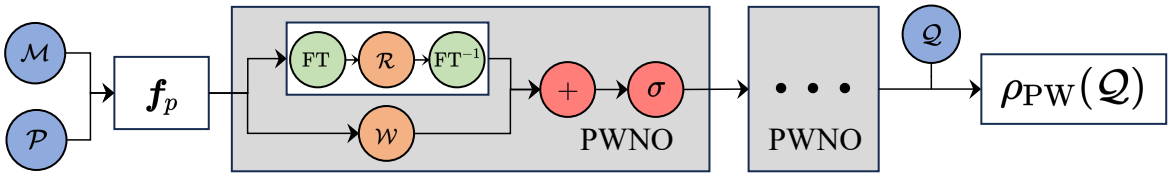


Figure 2: **Neural architecture for the PWNO layer.** Starting from a molecule $\mathcal{M}$, our framework makes a PW-based prediction ρ_{PW} for given query points $\mathcal{Q}$. The PWNO layer processes the discretized signal using convolution, which is expressed as applying FT, frequency-wise linear transformation $\mathcal{R}$ and inverse FT in sequence. Point-wise activation function σ is introduced for the non-linear transformation of the discretized signals.

Then, we aggregate messages from the atoms $u \in \mathcal{V}$ in vicinity of a probe point $p \in \mathcal{P}$:

$$\mathbf{f}_p^{(1)} = \mathrm{UPD} \circ \mathrm{AGG}(\{(r_{up}, \mathbf{f}_u^{(0)}) : u \in \mathcal{N}_{\mathcal{V}}(p)\}), \quad (10)$$

where $r_{up} = \|\boldsymbol{x}_p - \boldsymbol{x}_u\|$ and $\mathcal{N}_{\mathcal{V}}(p) \subseteq \mathcal{V}$ is the set of atoms within a certain radius for the probe node $p \in \mathcal{P}$. Furthermore, AGG and UPD are the aggregation and the update function typically used for graph neural networks. The grid embeddings $\{\mathbf{f}_p^{(1)} : p \in \mathcal{P}\}$ implicitly represent a signal $f^{(1)}$ evaluated at location $\boldsymbol{x}_p$, i.e., $\mathbf{f}_p^{(1)} = f^{(1)}(\boldsymbol{x}_p)$.

PW neural operator layer. To represent the interaction between the PW bases, we employ the PW neural operator layer (PWNO). PWNO iteratively updates the signal $f^{(h)} : \mathbb{R}^3 \to \mathbb{R}^C$ by a convolution operator using the PW basis where C is the channel dimension of signal, i.e., sequentially applying Fourier transformation, frequency-wise linear transformation, and inverse Fourier transformation:

$$f^{(h+1)} = \sigma(\mathcal{W} \cdot f^{(h)} + \mathcal{K} * f^{(h)}), \quad (11)$$

where $*$ is the convolution operator, $\mathcal{K}$ is the kernel, $\mathcal{W}$ is the point-wise linear operator, and σ is the point-wise non-linearities. i.e., GELU (Hendrycks & Gimpel, 2016). We implement the convolution operator via linear transformation in the Fourier space:

$$\mathcal{K} * f^{(h)} = \mathrm{FT}^{-1}(\mathcal{R} \cdot \mathrm{FT}(f^{(h)})), \quad (12)$$

where $\mathcal{R}$ is the frequency-wise linear transformation, FT is the Fourier transformation that maps point-wise representation into the PW basis coefficients, and FT^{-1} its inverse.

To implement the Fourier transformation in discrete space, we operate with point-wise representations $\{\mathbf{f}_p^{(h)} \in \mathbb{R}^C : p \in \mathcal{P}\}$ and frequency-wise representations $\{\hat{\mathbf{f}}_{\boldsymbol{\lambda}}^{(h)} \in \mathbb{R}^C : \boldsymbol{\lambda} \in \Lambda\}$ of the underlying signal $f^{(h)}$, e.g., $f^{(h)}(\boldsymbol{x}) = \sum_{\boldsymbol{\lambda} \in \Lambda} \hat{\mathbf{f}}_{\boldsymbol{\lambda}}^{(h)} \phi_{\boldsymbol{\lambda}}(\boldsymbol{x})$ where $\phi_{\boldsymbol{\lambda}}(\boldsymbol{x})$ is the PW with wavevector $\boldsymbol{\lambda}$. Our framework operates only on the wavevectors Λ of the PW basis, incorporating the periodicity constraints regarding the lattice structure Λ. We depicted the architecture of the PWNO layer in Figure 2.

Prediction network. At the final H-th layer, given a query point $\boldsymbol{x}_q$, we compute the output as follows:

$$\mathbf{f}_q^{(H)} = \mathrm{UPD} \circ \mathrm{AGG}(\{(r_{pq}, \mathbf{f}_p^{(H-1)}) : p \in \mathcal{N}_{\mathcal{P}}(q)\}), \quad (13)$$

where $r_{pq} = \|\boldsymbol{x}_q - \boldsymbol{x}_p\|$ and $\mathcal{N}_{\mathcal{P}}(q) \subseteq \mathcal{P}$ is the set of probe nodes within a certain radius for the query node q. Note that the query node p can be placed at any points in the continuous space unlike the probe node p which is regularly placed along the lattice. One can also process multiple queries in parallel without re-computing the final probe-wise features $\mathbf{f}_p^{(H-1)}$ for different queries.

High frequency masking. To promote the synergy between the PW and GTO-based predictions, we employ high frequency mask (HFM), which constrains the PW prediction to low frequency region that is distant from atoms. To this end, we propose to mask the PW-based prediction near the atom as follows:

$$a(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \min_{u \in \mathcal{V}} d(\boldsymbol{x}, \boldsymbol{x}_u) > r_{\text{mask}}, \\ 0 & \text{else.} \end{cases} \quad (14)$$

Note that our masking introduces discontinuities near the cutoff boundary, which we empirically observe to be ignorable.[2] This results in the following prediction:

$$\rho_{\mathrm{PW}}(\boldsymbol{x}_q) = a(\boldsymbol{x}) \mathbf{f}_q^{(H)},$$

where $\boldsymbol{x}_q$ is the input query point. We provide ablation studies on the effectiveness of HFM in Section 5.4.

4.3. GTO-based Prediction

Finally, we describe our prediction using the GTO basis set. Our philosophy mostly follows the multicentric expansion scheme of InfGCN (Cheng & Peng, 2023). Note that we can design the GTO-based prediction to satisfy the periodic boundary condition by introducing a periodical graph which is used in Xie & Grossman (2018).

[2] One could replace the masking function with an envelope function (Gasteiger et al., 2020b) for the continuous cutoff.

Table 1: **Evaluation of GPWNO for aperiodic materials.** We report the performance for QM9 and MD dataset in NMAE (%). The best number is highlighted in **bold**. The baseline results are from Cheng & Peng (2023). Each number is averaged over three runs. For brevity, we denote DimeNet and DeepDFT by DmNet and DDFT, respectively.

Dataset	LNO	FNO	GNO	DmNet++	DmNet	EGNN	DDFT2	DDFT	CNN	InfGCN	**GPWNO**
QM9	26.14	28.83	40.86	11.69	11.97	11.92	1.03	2.95	2.01	0.93	**0.73**
Ethanol	43.17	31.98	82.35	14.24	13.99	13.90	8.83	7.34	13.97	8.43	**4.00**
Benzene	38.82	20.05	82.46	14.34	14.48	13.49	5.49	6.61	11.98	5.11	**2.45**
Phenol	60.70	42.98	66.69	12.99	12.93	13.59	7.00	9.09	11.52	5.51	**2.68**
Resorcinol	35.07	26.06	58.75	12.01	12.04	12.61	6.95	8.18	11.07	5.95	**2.73**
Ethane	77.14	26.31	71.12	12.95	13.11	15.17	6.36	8.31	14.72	7.01	**3.67**
MDA	47.22	34.58	84.52	16.79	18.71	12.37	10.68	9.31	18.52	10.34	**5.32**

The GTO-based prediction ρ_{GTO} is constructed from the linear combination of GTO basis set as follows:

$$\rho_{\text{GTO}}(\boldsymbol{x}) = \sum_{u\in\mathcal{V}} \sum_{n=1}^{N} \sum_{l=1}^{L} \sum_{m=-l}^{l} \hat{f}_{u,nlm}\phi_{nlm}(\boldsymbol{x}, \boldsymbol{x}_u), \quad (15)$$

where $\phi_{nlm}(\boldsymbol{x}, \boldsymbol{x}_u)$ is the GTO as described in Equation (6) centered at each atom position $\boldsymbol{x}_u$ for $u \in \mathcal{V}$. Furthermore, $\hat{f}_{u,nlm} \in \mathbb{R}$ is the atom-wise coefficient term constructed from our neural operator given the molecule $\mathcal{M}$. Finally, N and L are the truncated maximum degrees for the radial basis and the spherical harmonics.

To be specific, the atom-wise coefficient terms are computed by TFN-based layer and its implementation (Thomas et al., 2018; Geiger & Smidt, 2022) to achieve the equivariant message passing. The spherical tensor feature $\hat{\mathbf{f}}^{(h)}_{u,nl} = \{\hat{f}^{(h)}_{u,nlm} : m \in \mathbb{Z}, -l \leq m \leq l\}$ is given by the following updates:

$$\hat{\mathbf{f}}^{(h+1)}_{u,nl} = \sum_{v\in\mathcal{N}_{\mathcal{V}}(u)} \sum_{k\geq 0} W^{(h)}_{nlk}(\boldsymbol{x}_v - \boldsymbol{x}_u)\hat{\mathbf{f}}^{(h)}_{v,nk}, \quad (16)$$

$$W^{(h)}_{nlk}(\boldsymbol{r}) = \sum_{J=\|k-l\|}^{k+1} \varphi^{J}_{nlk}(r) \sum_{m=-J}^{J} Y_{Jm}(\hat{r})\, Q^{lk}_{Jm}, \quad (17)$$

where $r = \|\boldsymbol{r}\|$, $\hat{r} = \frac{\boldsymbol{r}}{r}$, $\mathcal{N}_{\mathcal{V}}(u)$ is the set of neighbors of atom u, $Q^{lk}_{Jm} \in \mathbb{R}^{(2l+2)\times(2k+1)}$ is the Clebsh-Gordan matrix, and $\varphi^{J}_{nlk} : \mathbb{R}^{+} \to \mathbb{R}$ is a learnable scalar function.

5. Experiment

5.1. Experimental Setting

Baselines. The baseline methods of dataset include the GTO-based InfGCN (Cheng & Peng, 2023), voxel based model CNN (Çiçek et al., 2016), interpolation-based models, i.e., DeepDFT (Jørgensen & Bhowmik, 2020), DeepDFT2 (Jørgensen & Bhowmik, 2022), EGNN (Satorras et al., 2021), DimeNet (Gasteiger et al., 2020b), DimeNet++ (Gasteiger et al., 2020a), and neural operator based models, i.e., GNO (Li et al., 2020b), FNO (Li et al., 2020a), and LNO (Lu et al., 2019). We faithfully compare with the official numbers reported by the prior work (Cheng & Peng, 2023) when possible. We also outperform on another dataset, which we have not detailed here. We comprehensively compare in Table 8 of Appendix F. Our implementation of the GPWNO is published in the GitHub[3].

Evaluation metric. For evaluation, we follow prior works (Jørgensen & Bhowmik, 2020; Cheng & Peng, 2023) in using the normalized mean absolute error (NMAE):

$$\text{NMAE}(\rho, \rho_{\text{pred}}) = \frac{\int_{\mathbb{R}^3} \|\rho(\boldsymbol{x}) - \rho_{\text{pred}}(\boldsymbol{x})\| \mathrm{d}\boldsymbol{x}}{\int_{\mathbb{R}^3} \|\rho(\boldsymbol{x})\| \mathrm{d}\boldsymbol{x}}. \quad (18)$$

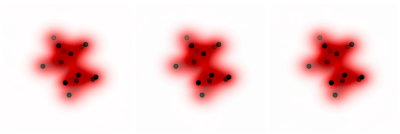
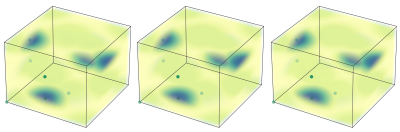
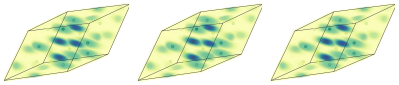
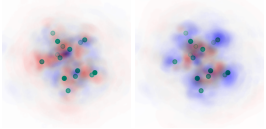
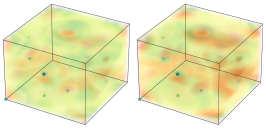
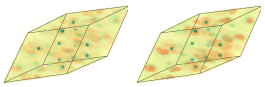
We provide more details regarding experiments setup, baselines, and evaluations in Appendix D.

5.2. Aperiodic Materials

Dataset. For aperiodic molecules, following the prior work (Cheng & Peng, 2023), we consider electron density prediction for the QM9 dataset (Ruddigkeit et al., 2012; Ramakrishnan et al., 2014) and the molecular dynamics (MD) dataset (Bogojeski et al., 2018). Specifically, the electron density dataset of the QM9 was provided by Jørgensen & Bhowmik (2022), and the MD was provided by Brockherde et al. (2017); Bogojeski et al. (2020). The MD dataset is further categorized into six smaller datasets corresponding to six types of molecules, i.e., ethanol, benzene, phenol, resorcinol, ethane, and malonaldehyde (MDA). Each smaller dataset contains snapshots of different geometries of a molecule. More details regarding the train and dataset are in the Appendix D.

Quantitative results. We present our experimental results in Table 1. The results demonstrate that our algorithm significantly outperforms existing baselines. In particular, in the QM9 and MD datasets, our approach substantially improves over the second-best model, i.e., InfGCN, by approximately reducing the NMAE by 20% in QM9 and 50% in MD. One

[3] https://github.com/seongsukim-ml/GPWNO

	QM9 ($O_2C_3H_4$)			Monoclinic ($Sr(P_3Pt_2)_2$)			Tetragonal($Eu(ZnGe)_2$)		
	GT	**GPWNO**	InfGCN	GT	**GPWNO**	InfGCN	GT	**GPWNO**	InfGCN
Pred.									
Error									
NMAE		**0.47**	0.59		**5.48**	5.78		**4.97**	5.28

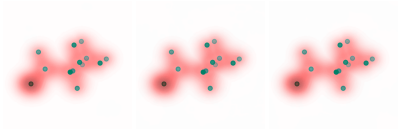
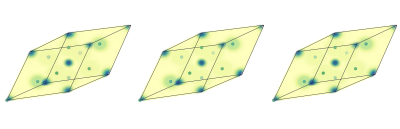
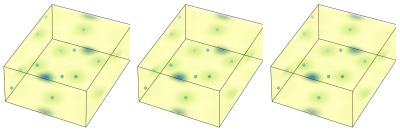
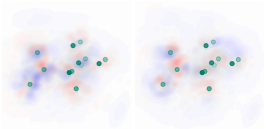
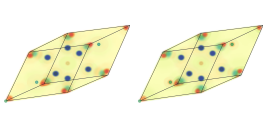
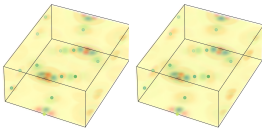
	QM9 (C_3ONH_5)			Triclinic ($Li_9Mn_2Co_5O_{16}$)			Orthorhombic (Dy_5Sn_{11})		
Pred.									
Error									
NMAE		**0.56**	0.60		**4.00**	5.89		**4.59**	5.15

Figure 3: **Visualization of ground truth (GT), density prediction, and error in NMAE (%).** For electron density prediction (Pred.) on QM9 and MP datasets, red and blue colors indicate higher electron density, respectively. For the error (Error), lighter colors indicates lower error.

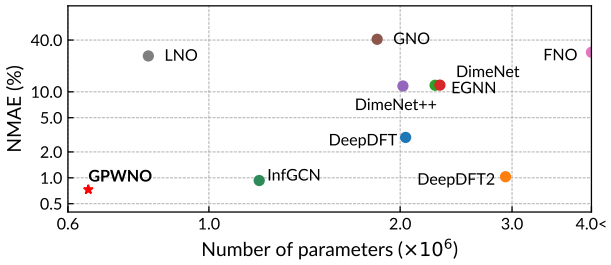


Figure 4: **Comparison between the baselines on QM9 (lower left is better).**

can also observe that FNO performs poorly despite its similarity to our GPWNO in using Fourier transformation as an intermediate layer. We hypothesize that the FNO cannot express the high-frequency details in the true electron density due to inefficient memory usage.

We also depict the tradeoff between the number of parameters and performance for the QM9 dataset in Figure 4, where one can observe how GPWNO outperforms the considered baselines despite using the smallest number of parameters. The corresponding numbers are reported in Appendix F.

Qualitative result. We additionally visualize the predicted densities for the QM9 dataset in Figure 3. Here, the performance of InfGCN is worse than GPWNO.

5.3. Periodic Systems

Dataset. To evaluate our model on periodic systems with diverse lattice structures, we newly conduct a benchmark from the materials project (Jain et al., 2013; Shen et al., 2021, MP). Precisely, we categorize the electron density for 117,535 molecules provided by Shen et al. (2021) into seven crystal families: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal, and cubic. Additionally, we conduct an extra dataset as a mixture of these crystal families. The size of the datasets ranges from 7,681 to 26,081. We use 1,000 or 500 samples for the test and validation, depending on the dataset size. We compare this dataset with the baseline results obtained from running the official repository codes on the newly curated datasets.

Quantitative results. We report the experimental results in Table 2. Here, one can observe how our algorithm significantly outperforms the baselines, regardless of the crystal family. We also note that DeepDFT and DeepDFT2 perform

Table 2: **Evaluation of GPWNO for periodic materials.** NMAE (%) in the MP dataset, categorized by the seven crystal family and their combinations. The best number is highlighted in **bold**. Each number is averaged over three runs.

Model	Mixed	Triclinic	Monoclinic	Orthorhombic	Tetragonal	Trigonal	Hexagonal	Cubic
DeepDFT	11.50	32.33	50.54	22.30	37.68	23.89	18.61	27.30
DeepDFT2	15.11	30.20	49.74	20.50	41.21	22.13	12.87	27.63
InfGCN	5.35	4.36	4.63	5.06	5.02	5.21	4.96	4.93
GPWNO	**4.84**	**3.92**	**4.44**	**4.73**	**4.52**	**4.92**	**4.55**	**4.32**

Table 3: **Ablation study on MD and MP.** Performance in NMAE (%) for the models using one or more of our algorithmic components, i.e., plane wave (PW), Gaussian-type orbital (GTO), and high-frequency masking (HFM), are reported. The best number is highlighted in **bold**. Each number is averaged over three runs.

Model			MD			MP		
PW	GTO	HFM	Benzene	Phenol	Resorcinol	Mixed	Triclinic	Cubic
✓	✗	✗	6.92	8.27	8.12	48.69	49.20	48.12
✗	✓	✗	2.90	3.24	2.96	5.35	4.36	4.93
✓	✓	✗	2.62	2.83	2.77	5.00	3.96	4.62
✓	✓	✓	**2.45**	**2.68**	**2.73**	**4.84**	**3.92**	**4.32**

poorly for the dataset, despite using the official repository.

Qualitative results. We provide visualization of the predicted densities on the MP dataset in Figure 3. Here, one can observe how the overall error and near atom region error both are reduced on GPWNO compared to InfGCN.

5.4. Ablation Study

We conduct extensive ablation studies on the MD and MP to verify the effectiveness of our method. We provide the results in Table 3 and Figure 5. In Table 3, we report the performance of our framework with and without each algorithmic component. In Figure 5, we check whether if introducing high-frequency components to the PW-only prediction eliminates the necessity of GTO basis. The detailed number of Figure 5 is reported in Table 9 of Appendix F.

Independently evaluating PW and GTO basis. In Table 3, one can observe how PW- and GTO-based predictions perform without additional algorithmic components. Intriguingly, one can observe how the PW performs poorly without the GTO basis. We hypothesize this to be since the PW basis cannot capture high frequency components in the true signal. This is additionally supported by Figure 5, where PW demonstrates reasonable performance for the MD dataset when increasing the size of the PW basis.

Synergy between PW and GTO basis. One can verify the complementary behavior of PW and GTO basis from Table 3 since combining the two predictions consistently improve the performance. This verifies our idea that employing a dual representation is more effective than relying on a single basis alone, demonstrating the synergistic advantages of utilizing both representations.

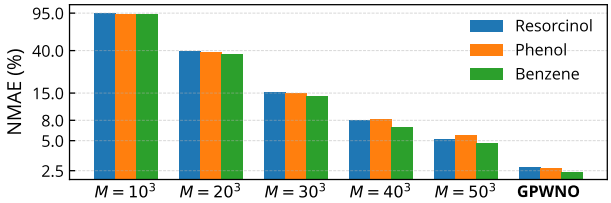


Figure 5: **Increasing number of probe nodes M for PW-only prediction and comparing with GPWNO.**

Effectiveness of HFM. When comparing our algorithm with and without HFM in Table 3, one can verify that HFM indeed enables the GTO-based prediction to concentrate on regions near atoms and PW-based prediction focus on regions far from atoms. Notably, HFM achieves this improvement in efficiency without introducing any additional model parameters.

6. Conclusion

In this work, we propose a new machine learning algorithm for electron density estimation. Through extensive experiments, we demonstrated the effectiveness of our model. While our experiments clearly demonstrate the superiority our GPWNO over the baselines, the evaluation metric is not fully aligned with our potential applications, e.g., errors in predicting the exchange-correlation energy. Since DFT computations required for this evaluation are expensive and rely on commercial licenses, we leave this for future work.

Impact Statement

Our framework can advance drug design and material discovery by accelerating molecular simulation. Such improvements often improve the quality of human life. However, they should also be approached with caution due to the potential misuse in the development of hazardous products.

Acknowledgement

This work was partly supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.RS-2019-II191906, Artificial Intelligence Graduate School Program(POSTECH)), the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2022R1C1C1013366), and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2022R1A6A1A0305295413).

We thank Seungjin Kang, Suhwan Song, Byunggook Na, Yongdeok Kim, Minjong Lee, Seonghyun Park, Hyosoon Jang, Juwon Hwang, Kiyoung Seong, and Dongyeop Woo for providing helpful feedbacks and suggestions in preparing the early version of the manuscript.

References

Adekoya, O. C., Adekoya, G. J., Sadiku, E. R., Hamam, Y., and Ray, S. S. Application of dft calculations in designing polymer-based drug delivery systems: An overview. *Pharmaceutics*, 14(9):1972, 2022. 1

Bachelet, G. B., Hamann, D. R., and Schlüter, M. Pseudopotentials that work: From h to pu. *Physical Review B*, 26(8):4199, 1982. 1

Bartlett, R. J. Many-body perturbation theory and coupled cluster theory for electron correlation in molecules. *Annual review of physical chemistry*, 32(1):359–401, 1981. 19

Blöchl, P. Electrostatic decoupling of periodic images of plane-wave-expanded densities and derived atomic point charges. *The Journal of chemical physics*, 103(17):7422–7428, 1995. 3

Blöchl, P. E. Projector augmented-wave method. *Physical review B*, 50(24):17953, 1994. 1

Bogojeski, M., Brockherde, F., Vogt-Maranto, L., Li, L., Tuckerman, M. E., Burke, K., and Müller, K.-R. Efficient prediction of 3d electron densities using machine learning. *arXiv preprint arXiv:1811.06255*, 2018. 1, 2, 6, 15

Bogojeski, M., Vogt-Maranto, L., Tuckerman, M. E., Müller, K.-R., and Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nature communications*, 11(1):5223, 2020. 2, 6, 15

Brockherde, F., Vogt, L., Li, L., Tuckerman, M. E., Burke, K., and Müller, K.-R. Bypassing the kohn-sham equations with machine learning. *Nature communications*, 8(1):872, 2017. 2, 6, 15

Chang, J., Jiang, L., Wang, G., Zhao, W., Huang, Y., and Chen, H. Lead-free perovskite compounds cssn 1- x ge x i 3- y br y explored for superior visible-light absorption. *Physical Chemistry Chemical Physics*, 23(26): 14449–14456, 2021. 1

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. 1, 2, 3, 4, 5, 6, 17, 18, 19

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pp. 424–432. Springer, 2016. 6

Fabrizio, A., Grisafi, A., Meyer, B., Ceriotti, M., and Corminboeuf, C. Electron density learning of non-covalent systems. *Chemical science*, 10(41):9424–9432, 2019. 1, 2

Gasteiger, J., Giri, S., Margraf, J. T., and Günnemann, S. Fast and uncertainty-aware directional message passing for non-equilibrium molecules. *arXiv preprint arXiv:2011.14115*, 2020a. 6

Gasteiger, J., Groß, J., and Günnemann, S. Directional message passing for molecular graphs. 2020b. URL https://openreview.net/forum?id=B1eWbxStPH. 5, 6

Geiger, M. and Smidt, T. e3nn: Euclidean neural networks. *arXiv preprint arXiv:2207.09453*, 2022. URL https://arxiv.org/abs/2207.09453. 6

Giannozzi, P., Andreussi, O., Brumme, T., Bunau, O., Nardelli, M. B., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Cococcioni, M., et al. Advanced capabilities for materials modelling with quantum espresso. *Journal of physics: Condensed matter*, 29(46):465901, 2017. 1

Gill, P. M. Molecular integrals over gaussian basis functions. In *Advances in quantum chemistry*, volume 25, pp. 141–205. Elsevier, 1994. 4

Gonze, X., Amadon, B., Anglade, P.-M., Beuken, J.-M., Bottin, F., Boulanger, P., Bruneval, F., Caliste, D., Caracas, R., Côté, M., et al. Abinit: First-principles approach to material and nanosystem properties. *Computer Physics Communications*, 180(12):2582–2615, 2009. 1

Grisafi, A., Fabrizio, A., Meyer, B., Wilkins, D. M., Corminboeuf, C., and Ceriotti, M. Transferable machine-learning model of the electron density. *ACS central science*, 5(1): 57–64, 2018. 1, 2

Hafner, J. Ab-initio simulations of materials using vasp: Density-functional theory and beyond. *Journal of computational chemistry*, 29(13):2044–2078, 2008. 1, 3

Harrison, W. *Pseudopotentials in the Theory of Metals*. Frontiers in Physics : a lecture note and reprint series. W.A. Benjamin, 1966. ISBN 9780805337310. URL https://books.google.co.kr/books?id=XV8sAAAAYAAJ. 1

Hehre, W. J., Ditchfield, R., and Pople, J. A. Self—consistent molecular orbital methods. xii. further extensions of gaussian—type basis sets for use in molecular orbital studies of organic molecules. *The Journal of Chemical Physics*, 56(5):2257–2261, 1972. 1, 4

Hendrycks, D. and Gimpel, K. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. 5

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. 2, 7

Jørgensen, P. B. and Bhowmik, A. Deepdft: Neural message passing network for accurate charge density prediction. *arXiv preprint arXiv:2011.03346*, 2020. 1, 2, 6, 15, 17, 19

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. 1, 2, 6, 15, 17, 19

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), 2022. doi: 10.1038/s41524-022-00863-y. 2, 6

Kittel, C. *Introduction to solid state physics*. John Wiley & sons, inc, 2005. 3

Koelling, D. and Arbman, G. Use of energy derivative of the radial solution in an augmented plane wave method: application to copper. *Journal of Physics F: Metal Physics*, 5(11):2041, 1975. 1, 2

Kohn, W. and Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965. 1, 19

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. 2

Kondrakov, A. O., Geßwein, H., Galdina, K., De Biasi, L., Meded, V., Filatova, E. O., Schumacher, G., Wenzel, W., Hartmann, P., Brezesinski, T., et al. Charge-transfer-induced lattice collapse in ni-rich ncm cathode materials during delithiation. *The journal of physical chemistry C*, 121(44):24381–24388, 2017. 1

Kosmala, A., Gasteiger, J., Gao, N., and Günnemann, S. Ewald-based long-range message passing for molecular graphs. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 17544–17563, 2023. 3

Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., and Anandkumar, A. Neural operator: Learning maps between function spaces, 2023. 3

Landau, D. and Binder, K. *A guide to Monte Carlo simulations in statistical physics*. Cambridge university press, 2021. 4

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020a. 1, 3, 4, 6

Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*, 2020b. 3, 6

Li, Z., Kovachki, N. B., Choy, C., Li, B., Kossaifi, J., Otta, S. P., Nabian, M. A., Stadler, M., Hundt, C., Azizzadenesheli, K., et al. Geometry-informed neural operator for large-scale 3d pdes. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 3

Lippert, B. G., PARRINELLO, J. H., and MICHELE. A hybrid gaussian and plane wave density functional scheme. *Molecular Physics*, 92(3):477–488, 1997. 2

Lu, L., Jin, P., and Karniadakis, G. E. Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *arXiv preprint arXiv:1910.03193*, 2019. 6

Mardirossian, N., McClain, J. D., and Chan, G. K. Lowering of the complexity of quantum chemistry methods by choice of representation. *The Journal of chemical physics*, 148(4), 2018. 19

Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511805769. 3

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. 2, 6

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. 2, 6

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. 6, 19

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. 17

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, 2021. 2, 7

Sinitskiy, A. V. and Pande, V. S. Deep neural network computes electron densities and energies of a large set of organic molecules faster than density functional theory (dft). *arXiv preprint arXiv:1809.02723*, 2018. 1, 2

Sun, H., Ross, Z. E., Zhu, W., and Azizzadenesheli, K. Phase neural operator for multi-station picking of seismic arrivals. *Geophysical Research Letters*, 50(24): e2023GL106434, 2023. 3

Sun, Q., Berkelbach, T. C., McClain, J. D., and Chan, G. K. Gaussian and plane-wave mixed density fitting for periodic systems. *The Journal of chemical physics*, 147(16), 2017. 1, 2

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. 4, 6

Tran, A., Mathews, A., Xie, L., and Ong, C. S. Factorized fourier neural operators. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tmIiMPl4IPa. 3, 17

Ulian, G., Tosoni, S., and Valdrè, G. Comparison between gaussian-type orbitals and plane wave ab initio density functional theory modeling of layer silicates: Talc [mg3si4o10 (oh) 2] as model system. *The Journal of chemical physics*, 139(20), 2013. 3

Wang, F. Q., Choudhary, K., Liu, Y., Hu, J., and Hu, M. Large scale dataset of real space electronic charge density of cubic inorganic materials from density functional theory (dft) calculations. *Scientific Data*, 9(1):59, 2022. 15

Xie, T. and Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.*, 120: 145301, Apr 2018. doi: 10.1103/PhysRevLett.120.145301. URL https://link.aps.org/doi/10.1103/PhysRevLett.120.145301. 4, 5

A. Overall Architecture

In Figure 6, we provide a detailed description of our model.

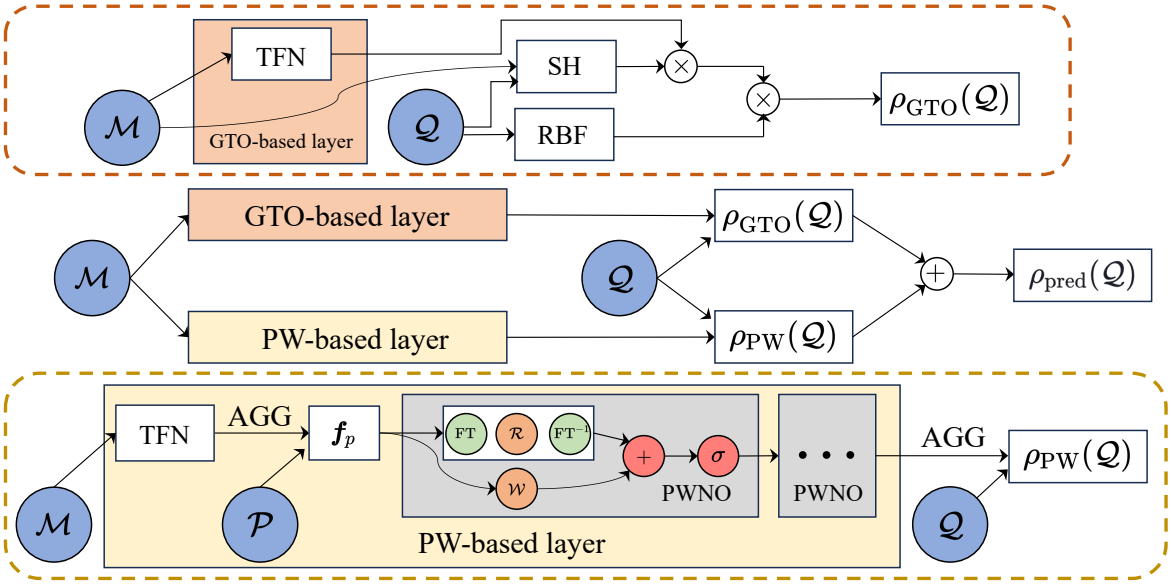


Figure 6: **Architecture of Gaussian plane-wave neural operator (GPWNO).** Starting from molecule $\mathcal{M} = (\mathcal{X}, \mathcal{A}, \boldsymbol{L})$ with the atomic positions $\mathcal{X}$, atomic features $\mathcal{A}$ and the lattice vectors $\boldsymbol{L}$, GPWNO aims to predict a scalar field ρ_{pred} for given query points $\mathcal{Q} \subset \mathbb{R}^3$. Notably, we decompose the target ρ_{pred} into two types of the bases: Gaussian-type orbital (GTO) basis and Plane wave (PW) basis. **(top)** GTO basis layer predict the ρ_{GTO} by the direct coefficients learning of spherical harmonics. **(bottom)** PW basis layer predict the signals of reciprocal space by using discretized voxel points $\mathcal{P}$ and the FNO layers, and predict the ρ_{PW} the message aggregation from voxel points $\mathcal{P}$.

B. Signal Initialization by Tensor Field Network

Here, we describe our TFN-based architecture for initializing a signal as an input to our PW neural operator. For a given molecule $\mathcal{M} = (\mathcal{X}, \mathcal{A}, \boldsymbol{L})$, we first initialize the atomic feature $\boldsymbol{f}_u^{(0)}$ of atom $u \in \mathcal{V}$ by the linear-embedding layer that takes input as the one-hot encoding of the atomic number of each atom. Then, we update the atomic feature with the TFN layer by treating $\boldsymbol{f}_u^{(0)} \in \mathbb{R}^C$ as a multi-channel rank-0 tensor feature of dimension C. i.e., $\boldsymbol{f}_u^{(0)} = \{f_{u,00}^1, \ldots, f_{u,00}^C\}$. Specifically, each embedding layer updates the spherical node feature $\boldsymbol{f}_{u,l}^{(h)} = \{\hat{f}_{u,lm}^{(h)} : m \in \mathbb{Z}, -l \leq m \leq l\}$ from the near atoms spherical node feature $\boldsymbol{f}_{v,l}^{(h)}$ by tensor product:

$$\hat{\mathbf{f}}_{u,l}^{(h+1)} = \sum_{v \in \mathcal{N}(u)} \sum_{k \geq 0} W_{lk}^{(h)}(\boldsymbol{x}_v - \boldsymbol{x}_u) \hat{\mathbf{f}}_{v,k}^{(h)}, \tag{19}$$

$$W_{lk}^{(h)}(\boldsymbol{r}) = \sum_{J=\|k-l\|}^{k+1} \varphi_{lk}^J(\|\boldsymbol{r}\|) \sum_{m=-J}^{J} Y_{Jm}\left(\frac{\boldsymbol{r}}{\|\boldsymbol{r}\|}\right) Q_{Jm}^{lk}. \tag{20}$$

Here, $\mathcal{N}(u)$ is the set of neighbors of atom u, $Q_{Jm}^{lk} \in \mathbb{R}^{(2l+2)\times(2k+1)}$ is the Clebsh-Gordan matrix, and $\varphi_{nlk}^J : \mathbb{R}^+ \to \mathbb{R}$ is a learnable scalar function. At the final embedding layer H, we set the output feature of the TFN as the multi-channel rank-0 tensor with dimension K. i.e., $\boldsymbol{f}_u^{(H)} \in \mathbb{R}^K$. Finally, we use this $\boldsymbol{f}_u^{(H)}$ as the input of the Equation (10).

C. Proof of Equivariance for the Lattice Construction on Aperiodic Molecule

We define the lattice $\boldsymbol{L}$ of the aperiodic molecule based on the principal component analysis (PCA). For a given molecule $\mathcal{M}$, let n is the number of the atoms and $\mathbf{X}$ is the atom positon matrix. i.e., $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]^\top \in \mathbb{R}^{n\times 3}$. Then, let $\boldsymbol{t} = \frac{1}{n}\sum_{u\in\mathcal{V}} \boldsymbol{x}_u \in \mathbb{R}^3$ be the centroid of the molecule $\mathbf{X}$ with n atoms and $\mathbf{C} = (\mathbf{X} - \mathbf{1}\boldsymbol{t}^\top)^\top(\mathbf{X} - \mathbf{1}^\top) \in \mathbb{R}^{3\times 3}$ be the covariance matrix computed after removing the centroid from $\mathbf{X}$. In the generic case, the eigenvalues of $\mathbf{C}$ satisfy $\lambda_1 < \lambda_2 < \lambda_3$. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ be the unit length corresponding eigenvectors. Then we define $\boldsymbol{L} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3]$ Note that eigenvectors are orthogonal to each other, so $\boldsymbol{L}$ can be used to the cubical lattice vector.

Next, we will show this $\boldsymbol{L}$ is SE(3) equivariant, meaning that the rotation of the data naturally transforms the lattice vector, i.e, $\boldsymbol{L}' = [\boldsymbol{v}_1\mathbf{R}^\top, \boldsymbol{v}_2\mathbf{R}^\top, \boldsymbol{v}_3\mathbf{R}^\top]$. Note that when we think about the direction of PCA eigenvector points out the maximal variance axis, it is natural that eigenvectors rotatated equivalently to the frame.

Let $(\mathbf{R}, \boldsymbol{k}) \in$ SO(3); then the position of molecule are transformed to $\mathbf{X}' = \mathbf{X}\mathbf{R} + \mathbf{1}\boldsymbol{k}^\top$. The centroid of the molecule is $\boldsymbol{t}' = \mathbf{R}^\top\boldsymbol{t} + \boldsymbol{k}$, and the covariance matrix becomes

$$
\begin{aligned}
\mathbf{C}' &= (\mathbf{X}' - \mathbf{1}\boldsymbol{t}'^\top)^\top(\mathbf{X}' - \mathbf{1}\boldsymbol{t}'^\top) \\
&= (\mathbf{X}\mathbf{R} + \mathbf{1}\boldsymbol{k}^\top - \mathbf{1}\boldsymbol{t}^\top\mathbf{R} - \mathbf{1}\boldsymbol{k}^\top)^\top(\mathbf{X}\mathbf{R} + \mathbf{1}\boldsymbol{k}^\top - \mathbf{1}\boldsymbol{t}^\top\mathbf{R} + \mathbf{1}\boldsymbol{k}^\top) \\
&= \mathbf{R}^\top(\mathbf{X} - \mathbf{1}\boldsymbol{t}^\top)^\top(\mathbf{X} - \mathbf{1}\boldsymbol{t}^\top)\mathbf{R} \\
&= \mathbf{R}^\top\mathbf{C}\mathbf{R}
\end{aligned}
$$

By the definition of the eigen decomposition, the eigenvector $\boldsymbol{v}$ of covaraince matrix $\mathbf{C}$ with eigenvalue λ satisfies $\mathbf{C}\boldsymbol{v} = \lambda\boldsymbol{v}$. Similary, the transformed covariance matrix $\mathbf{C}'$ has the new eigenvector $\boldsymbol{v}' = \mathbf{R}^\top\boldsymbol{v}$. We show that $\boldsymbol{v}'$ is the eigenvector of $\mathbf{C}'$ and its eigenvalue is λ:

$$
\begin{aligned}
\mathbf{C}'\boldsymbol{v}' &= (\mathbf{R}^\top\mathbf{C}\mathbf{R})(\mathbf{R}^\top\boldsymbol{v}) \\
&= \mathbf{R}^\top\mathbf{C}\boldsymbol{v} = \mathbf{R}^\top\lambda\boldsymbol{v} \\
&= \lambda\boldsymbol{v}'
\end{aligned}
$$

To this end, the lattice system of the rotated system is naturally rotated, that is $\boldsymbol{L}' = [\boldsymbol{v}_1', \boldsymbol{v}_2', \boldsymbol{v}_3'] = [\boldsymbol{v}_1\mathbf{R}^\top, \boldsymbol{v}_2\mathbf{R}^\top, \boldsymbol{v}_3\mathbf{R}^\top]$ as mentioned above.

D. Dataset and Hyperparameters

In this section, we provide more details about our dataset and train configuration. All the dataset contains the molecule graph of each species and its voxelized 3D electron charge density data on its pre-defined grid.

QM9. The dataset is provided by Jørgensen & Bhowmik (2022; 2020) The electron density of the dataset are calculated by VASP. To be specific, VASP with Perdew-Burke-Ernzerhof exchange correlation (PBE XC) functional and projector-augemented wave (PAW) is used.

MD. The dataset is provided by Brockherde et al. (2017); Bogojeski et al. (2018; 2020) and it is calculated by PBE XC functional with Quantum ESPRESSO using the PBE functional.

Cubic2. Cubic2 is not directly reported, since it shares the same materials structure with the cubic of the Materials Project. We report this to point out our model outperformed than previous works on this dataset, too. The data is provided by (Wang et al., 2022) and calculated by the VASP with the PAW. The performance on Cubic2 is in Table 8.

Materials project. This data is set of the periodic materials including cubic, hexagonal and so on. The detailed categorized condition is in Table 6.

Common statistics of the QM9, Cubic2 and MD are summarized in Table 4. MD dataset does not have a validation split, and the number of grids and grid lengths is for a single dimension so the number of voxels cubically with respect to it.

Table 4: **Statistics of the dataset.** We round up the number of grids and nodes to the nearest integer for readability.

Dataset	QM9	Cubic2	MD
train/val/test split	123835/50/1600	14421/1000/1000	1000(2000)/500(400)
min/mean/max #grid	40/88/160	32/94/448	20/20/20
min/mean/max #node	3/18/29	1/10/64	8/11/14
min/mean/max length (Bohr)	4.00/8.65/15.83	1.78/5.82/26.20	20/20/20
# node type	5	84	3

Additional statistics of the MP dataset are detailed in Table 5. The mixed dataset is composed of samples from each crystal family, with a distribution of 2000 for training, 150 for validation, and 150 for testing, resulting in a total of 14,000 training, 1,050 validation, and 1,050 testing samples. Note that the MP dataset may include non-cubical voxels, as its lattice structure is not confined to cubic forms. Differing from cubic structures, the generalized lattice types vary in length and grid dimensions along the $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$ directions of the lattice vector. Each number of the grid binning and lattice length are denoted by "# grid $\boldsymbol{a}_i$" and "length $\boldsymbol{a}_i$", respectivly. Accordingly, we provide statistics for each side, as well as the average binning of grid and length, which are represented as "# grid avg", and "length avg". These averages are calculated using the cube root of the product of the values along $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, and $\boldsymbol{a}_3$.

Table 5: **Dataset details of materials project.** We round up the number of grids to the nearest integer for readability. Each splitting of the numbers implies min, mean and max from the left to right.

	Triclinic	**Monoclinic**	**Orthorhombic**	**Tetragonal**	**Trigonal**	**Hexagonal**	**Cubic**
train/val/test split	12843/1000/1000	26081/1000/1000	22115/1000/1000	12031/1000/1000	9651/500/500	7681/500/500	15133/1000/1000
# atom type	83	86	84	85	84	87	88
# grid $\boldsymbol{a}_1$	36/103/1120	36/ 99/972	28/ 99/ 420	24/101/ 500	24/117/ 576	28/ 93/ 280	28/90/180
# grid $\boldsymbol{a}_2$	36/104/ 540	36/118/756	28/100/ 960	24/120/1120	20/112/ 540	28/ 98/ 280	28/90/192
# grid $\boldsymbol{a}_3$	36/182/1512	48/162/896	16/126/1120	36/155/1120	36/156/1500	16/125/1440	28/90/240
# grid avg.	47/115/ 540	57/121/387	30/105/ 420	38/119/ 384	40/123/ 384	33/ 99/ 178	28/90/184
Length $\boldsymbol{a}_1$ (Bohr)	2.36/ 6.67/ 74.23	2.30/ 6.44/65.40	1.78/6.46/28.11	1.61/ 6.60/32.62	1.52/7.64/38.79	1.71/6.01/18.30	1.78/5.85/12.10
Length $\boldsymbol{a}_2$ (Bohr)	2.36/ 6.76/ 34.99	2.36/ 7.66/50.18	1.78/6.57/61.13	1.61/ 7.85/74.88	1.28/7.34/36.35	1.71/6.33/18.30	1.78/5.85/12.80
Length $\boldsymbol{a}_3$ (Bohr)	2.49/11.95/101.39	2.94/10.58/59.56	0.98/8.20/75.07	2.34/10.13/74.43	2.46/10.17/100.46	0.98/8.16/95.06	1.78/5.86/15.92
Length avg. (Bohr)	3.10/ 8.46/ 35.94	3.78/ 8.23/29.96	2.07/7.08/41.50	2.60/8.19/28.01	2.68/8.38/35.60	2.22/6.84/33.73	1.78/5.85/12.23
# node	1/18/118	3/36/144	1/17/120	1/25/154	1/33/144	1/15/132	1/11/136

Table 6: **The condition of crystal familiy.** Here, each a, b, c represents the length of the corresponding lattice vector, i.e., $a = |\boldsymbol{a}_1|$, $b = |\boldsymbol{a}_2|$, and $c = |\boldsymbol{a}_3|$. Additionally, α, β, and γ denote the angles between $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$, $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, and $\boldsymbol{a}_2$ and $\boldsymbol{a}_1$, respectively.

Crystal Famlily	Edge	Angle
Triclinic	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma \neq 90^\circ$
Monoclinic	$a \neq b \neq c$	$\alpha = \gamma = 90^\circ, \beta \neq 120^\circ$
Orthorhombic	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Tetragonal	$a = b \neq c$	$\alpha = \beta = \gamma = 90^\circ$
Trigonal	$a = b = c$	$\alpha = \beta = \gamma \neq 90^\circ$
Hexagonal	$a = b \neq c$	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$
Cubic	$a = b = c$	$\alpha = \beta = \gamma = 90^\circ$

E. Model and Training Details

E.1. Model Details

We provide more details about the proposed GWPNO model and the baseline.

GPWNO.[4] Our implementation of GTO layer is inspired by the InfGCN. We used the spherical degree $l = 3$ or $l = 4$. The number of radio bases is settled $n = 16$ with the Gaussian parameters starting at 0.5 (Bohr) and ending at 5.0. We used 3 layers of the GTO layers and the number of the trainable parameter of GTO layer is 223K when $l = 3$ and 353K when $l = 4$.

In the implementation of the PW layer, we set the spherical degree $l = 3$ for embedding purposes. The number of probe nodes M was configured to either 20 or 40. To enhance the scalability of the PW layer, we incorporated the approach outlined by Tran et al. (2023). However, our empirical observations indicate that the performance of the model remains nearly unchanged with or without the use of Tran et al. (2023), provided that the number of probe nodes remains consistent. Regarding the trainable parameters in the PW neural operator layer, the count stands at 19.3K for $M = 20$ and increases to 47.0K when $M = 40$.

InfGCN.[5] InfGCN is a GTO-based model designed to learn the direct coefficients of the GTO basis. We used the official code implementation of InfGCN by Cheng & Peng (2023).

DeepDFT and **DeepDFT2.**[6] DeepDFT is a GNN based model that uses message passing to the query node where the charge density is predicted. DeepDFT2 uses PaiNN (Schütt et al., 2021) as the GNN architecture. We used the official code implementation of DeepDFT and DeepDFT2 by Jørgensen & Bhowmik (2020; 2022).

E.2. Training Details

In this section, we outline our model specifications and those of the baseline models. Additionally, we detail the hyperparameters related to training and testing utilized in our experiments. As previously mentioned, we make direct comparisons with the official results reported in prior work by Cheng & Peng (2023) whenever possible.

Table 7: **Hyperaparameter details**

Model	Dataset	`sp`	`cutoff`	`q.cutoff`	`p.cutoff`	M	`num.grad`	learning rate	batch size	learning rate decay	query points
GPWNO	QM9	4	3.0	1.5	0.75	40	80k	1e-3	32	0.5	1024
	MD	3	3.0	1.5	0.75	40	4k	5e-3			
	Cubic	4	5.0	2.5	1.25	20	10k	5e-3			
	MP	3	3.0	2.5	1.25	20	10k	5e-3			
InfGCN	MP	4	3.0	-	-	-	10k	5e-3	32	0.5	1024
DeepDFT DeepDFT2	MP	4	4.0	-	-	-	10k	3e-4	32	0.5	1024

We list the training specifications and major hyperparameters of our experiments in Table 7. Following the approach of Cheng & Peng (2023) and Jørgensen & Bhowmik (2020), we employ a sampling training scheme. In each training iteration, we sample multiple query points from the given data batch. For instance, we sample 1024 query points from a batch of 32 data points, resulting in a target size of (32×1024) query points during training. For testing, we evaluate all samples in the test dataset to avoid randomness. Minibatch inference does not impact the results, so we segment the full grid into small sections, evaluate the points sequentially as mini-batch, and compute the normalized mean absolute error (NMAE) across the entire inference dataset.

We utilized the ADAM optimizer, and the learning rate and its decay also reported in Table 7. In addition the number of the gradient step is also reported as `num.grad`.

Key hyperparameters in our model include the spherical harmonics (`sp`) order in the GTO layer and the radius cutoff distance for each message-passing operation. The `cutoff` parameter defines the neighborhood radius for the GTO layer. Additionally, the probe cutoff (`p.cutoff`) and query cutoff (`q.cutoff`) parameters specify the radius distances used in Equation (10) and Equation (13), respectively.

[4] https://github.com/seongsukim-ml/GPWNO
[5] https://github.com/ccr-cheng/infgcn-pytorch
[6] https://github.com/peterbjorgensen/DeepDFT

F. Additional Results

We provide the additional results that is not included in the main context for the space.

F.1. Additional Comparison

We provide the performance on another data set that is not reported in the main context. For QM9, we also evaluate the performance on the randomly rotated test set. We observe that our GPWNO also outperform than previous works on the rotated QM9. For Cubic2, GPWNO also achieve the state-of-the-art performance among the baseline.

Table 8: **The comparison of NMAE (%) reported in (Cheng & Peng, 2023).**

Dateset / Model		**Ours**	InfGCN	CNN	DeepDFT	DeepDFT2	EGNN	DimeNet	DimeNet++	GNO	FNO	LNO
QM9	rot	**4.68**	4.73	5.89	5.87	4.98	12.13	12.98	12.75	46.90	33.25	24.13
	unrot	**0.73**	0.93	2.01	2.95	1.03	11.92	11.97	11.69	40.86	28.83	26.14
MD	Ethanol	**4.00**	8.43	13.97	7.34	8.83	13.90	13.99	14.24	82.35	31.98	43.17
	Benzene	**2.45**	5.11	11.98	6.61	5.49	13.49	14.48	14.34	82.46	20.05	38.82
	Phenol	**2.68**	5.51	11.52	9.09	7.00	13.59	12.93	12.99	66.69	42.98	60.70
	Resorcinol	**2.73**	5.95	11.07	8.18	6.95	12.61	12.04	12.01	58.75	26.06	35.07
	Ethane	**3.67**	7.01	14.72	8.31	6.36	15.17	13.11	12.95	71.12	26.31	77.14
	Malonaldehyde	**5.32**	10.34	18.52	9.31	10.68	12.37	18.71	16.79	84.52	34.58	47.22
Cubic2		**7.69**	8.98	-	14.08	10.37	11.74	12.51	12.18	53.55	48.08	46.33

F.2. Numbers for Figure 4

The numbers for the performance of Figure 4 is reported in Table 9.

Table 9: **Numbers for Figure 4.**

# probes M	Benzene	Phenol	Resorcinol
10^3	93.81	94.06	96.11
20^3	36.71	38.41	39.70
30^3	14.04	15.18	15.36
40^3	6.92	8.27	8.12
50^3	**4.77**	**5.66**	**5.18**

G. Complexity Analysis

We provide the complexity of our model and the empirical results that demonstrates the analysis.

Time complexity. Our GPWNO scalses linearly with respect to the number of atoms, which is the same order as InfGCN. Our GPWNO scales log linearly with respect to the number of probe nodes due to the Fourier transform, but this is a hyperparameter that one can adjust to the given computational budget.

To be specific, we report the computational complexity of the considered algorithms as follows:

- InfGCN(Cheng & Peng, 2023), DDFT (Jørgensen & Bhowmik, 2020), DDFT2 (Jørgensen & Bhowmik, 2022), EGNN (Satorras et al., 2021): $O(d_{\text{atom}}N)$ where N and d_{atom} denote the number of atoms and maximum degree of the atom graph, respectively.
- Ours (GPWNO): $O(d_{\text{atom}}N + d_{\text{probe}}M + M \log M)$ where M and d_{probe} denotes number of probe nodes and maximum degree of the atom-to-probe node, respectively.

Here, the GPWNO complexities $O(d_{\text{atom}}N)$, $O(d_{\text{probe}}M)$, and $O(M \log M)$ are required for atom-wise message passing, atom-to-probe message passing, and Fourier operations, respectively. Since the degree d_{atom} depends on the cutoff radius, GPWNO scales linearly with the system size, i.e., number of atoms N. GPWNO scales log linearly with to number of probe nodes M, which is a hyperparameter that one can adjust to the given computational budget.

Empirical result. We empirically validate how the inference times of GPWNO scale linearly with the number of atoms N and log linearly with the number of probe nodes M. To vary N, we artificially increase the number of atoms N through supercell expansion, i.e., repeating the same cell to construct a larger supercell. Note that M is a hyperparameter we can tune without depending on the molecule. The results is depicted in Figure 7. These measurements were conducted using 100 forward passes for pre-defined 1000 query points on $\mathsf{SmIn_2Cu_9}$ (mp-1209644). $\mathsf{SmIn_2Cu_9}$ is sourced from the Materials Project database, consisting of 24 sites per single cell. We report the average and standard deviation over 5 trials.

Finally, our GPWNO is much scalable than the traditional DFT solvers, i.e., Linearized coupled cluster doubles method (Bartlett, 1981; Mardirossian et al., 2018) has $O(N^5)$ and Kohn-Sham density functional method (Kohn & Sham, 1965) has $O(N^3)$ complexity.

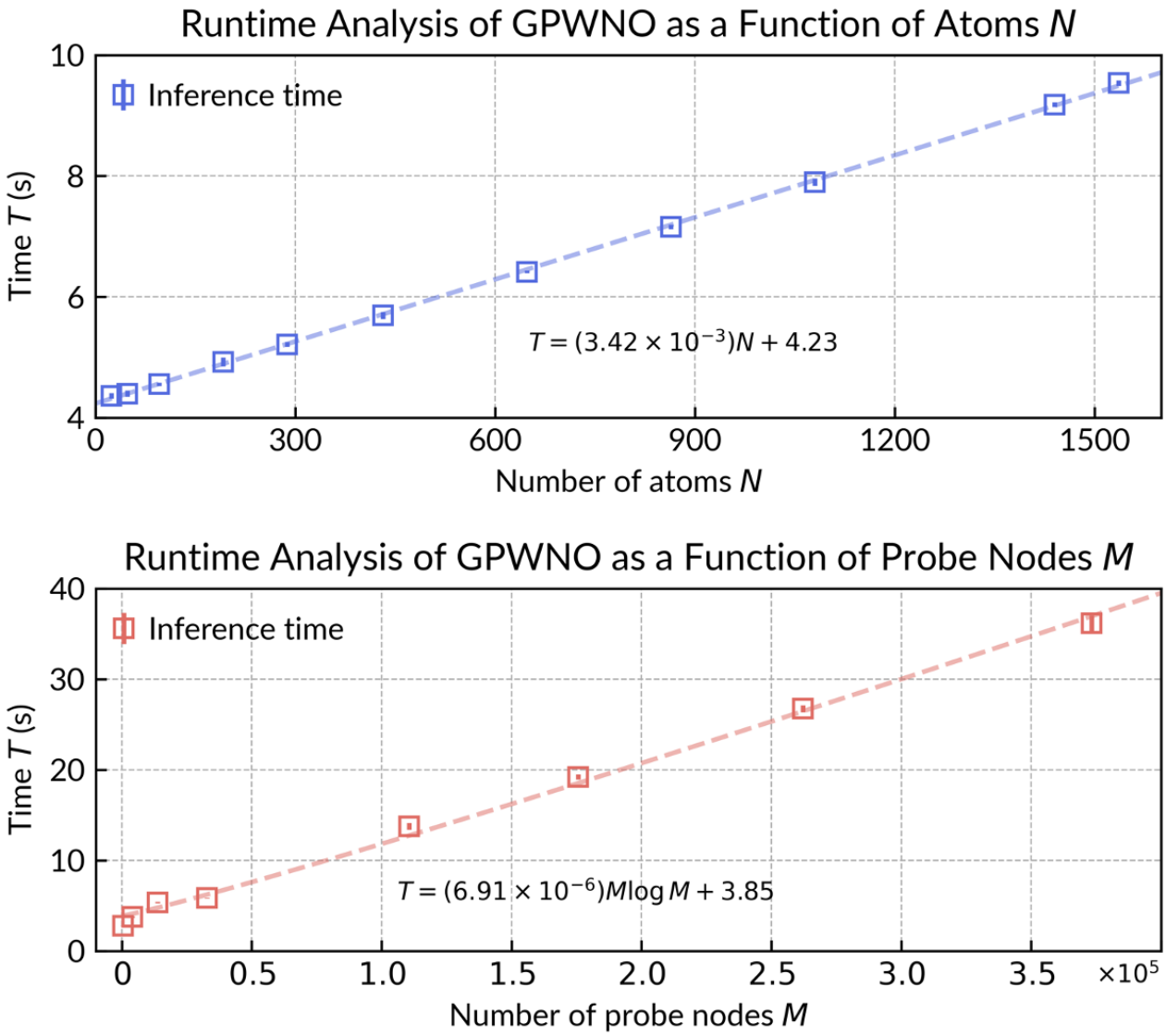


Figure 7: **Empirical validation of the time complexity by runtime analysis.** The inference times of GPWNO scale linearly with the number of atoms N and log linearly with the number of probe nodes M.