

JOINT MODELING OF SPATIAL AND TEMPORAL MULTISCALES IN MOLECULAR DYNAMICS

Anonymous authors

Paper under double-blind review

ABSTRACT

Molecular dynamics simulations are crucial for understanding complex physical, chemical, and biological processes at the atomic level. However, accurately capturing interactions across multiple spatial and temporal scales remains a significant challenge. We present a novel framework that jointly models spatial and temporal multiscale interactions in molecular dynamics. Our approach leverages Graph Fourier Transforms to decompose molecular structures into different spatial scales and employs Neural Ordinary Differential Equations to model the temporal dynamics at each scale. This unified framework explicitly links spatial structures with temporal evolution, enabling more accurate and comprehensive simulations of molecular systems. **We evaluate our model on the MD17 and alanine dipeptide datasets, demonstrating consistent performance improvements over state-of-the-art baselines across multiple molecules. Importantly, we introduce physics-based evaluation metrics, including bond lengths, bond angles, and dihedral angles, to assess the practical effectiveness of our model in preserving molecular geometry over long time horizons.** Ablation studies confirm the significant contributions of both spatial and temporal multiscale modeling components. Our method advances the simulation of complex molecular systems, potentially accelerating research in computational chemistry, drug discovery, and materials science.

1 INTRODUCTION

Molecular dynamics (MD) simulations are indispensable tools for investigating the behavior of molecular systems at the atomic level, offering profound insights into physical (Bear & Blaisten-Barojas, 1998), chemical (Wang et al., 2011), and biological (Salo-Ahen et al., 2020) processes. These simulations must capture interactions occurring across a wide range of spatial and temporal scales—from localized bond vibrations to long-range non-bonded interactions—posing significant computational challenges. Accurately modeling these multi-scale interactions requires sophisticated methods that can efficiently represent complex dependencies among atoms (Vakis et al., 2018).

An intrinsic link exists between the spatial scales of molecular interactions and their corresponding temporal dynamics. To illustrate this connection, consider the vibrational spectrum of a water molecule, a well-studied system in molecular spectroscopy. In Fourier-transform infrared (FTIR) spectroscopy (Griffiths, 1983), different absorption peaks correspond to distinct vibrational modes, each associated with specific spatial and temporal scales. As shown in Figure 1(a), high-frequency absorption peaks arise from localized O-H stretching vibrations, involving interactions over short spatial ranges and resulting in fast temporal dynamics. Conversely, low-frequency peaks correspond to collective motions like hydrogen bonding and bending modes, which involve longer spatial ranges and exhibit slower temporal dynamics. This spectrum exemplifies how localized spatial interactions correspond to faster temporal dynamics, while extended spatial interactions lead to slower dynamics.

Existing methods have attempted to address multi-scale interactions in MD simulations but often focus on either spatial or temporal scales independently. Coarse-grained techniques (Joshi & Deshmukh, 2021; Wang et al., 2019) reduce degrees of freedom to capture larger spatial scales but may lose important high-frequency details inherent in localized interactions. Neural operator-based methods like ITO (Schreiner et al., 2024), TimeWarp (Klein et al., 2024), and EGNO (Xu et al., 2024) are effective in extracting different temporal scales of signals but do not explicitly model the varying spatial scales that give rise to these temporal dynamics. Similarly, graph neural ODEs such

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

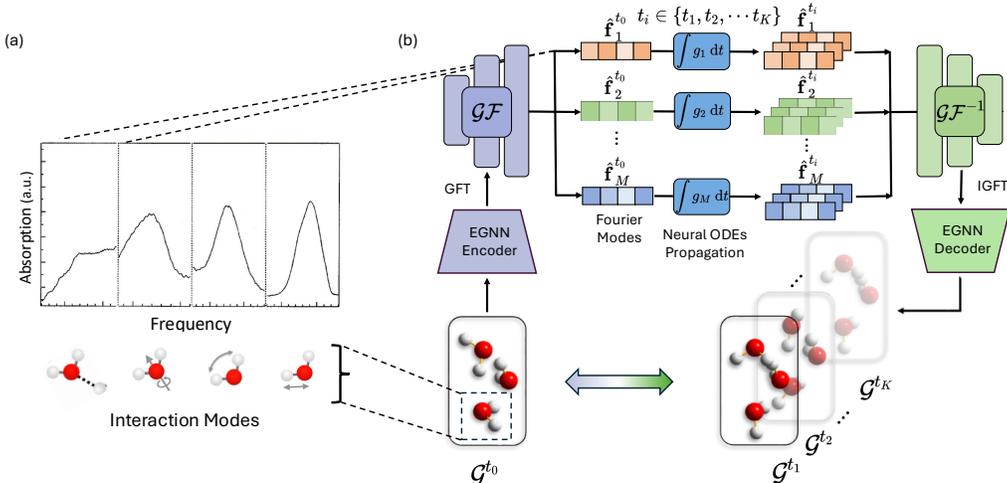


Figure 1: Overview of our framework for joint modeling of spatial and temporal multiscales in molecular dynamics. (a) FTIR spectrum of a water molecule, showing absorption peaks at different frequencies corresponding to various vibrational modes (Brubach et al., 2005). High-frequency peaks are associated with localized vibrations (e.g., O-H stretching), while low-frequency peaks correspond to global motions (e.g., bending modes). (b) Our model decomposes the molecular graph into different spatial scales using Graph Fourier Transform and models the temporal dynamics at each scale using Neural ODEs, capturing the correlation between spatial and temporal multiscale interactions.

as NDCN (Zang & Wang, 2020) and GG-ODE (Huang et al., 2023a) model temporal evolution in continuous time but often overlook the intrinsic connection between spatial structures and temporal dynamics.

The recognition of the aforementioned challenge naturally raises a pivotal question:

How to simultaneously model spatial and temporal multi-scale interactions in molecular dynamics for comprehensive and accurate simulations?

In this work, we propose a novel approach that explicitly models the joint spatial and temporal multiscale interactions in molecular systems. Inspired by molecular spectroscopy techniques like FTIR, which reveal the dual connotations of frequencies in both spatial and temporal domains, we leverage Graph Fourier Transformation to decompose molecular interactions into different spatial scales. This transformation captures the high-frequency (localized) and low-frequency (extended) spatial interactions, effectively encoding the spatial multiscale nature of the molecule.

To model the temporal dynamics associated with each spatial scale, we employ Neural Ordinary Differential Equations (Neural ODEs) (Chen et al., 2018). The adaptive stepping nature of Neural ODEs allows for flexible and efficient modeling of dynamics at different time scales, accommodating both fast and slow temporal evolutions. By propagating the decomposed spatial scales through time using Neural ODEs, we capture the temporal dynamics corresponding to each spatial scale. After temporal propagation, we reconstruct the molecular state in the physical space by an inverse Graph Fourier Transformation.

This joint modeling approach is motivated by the natural association between spatial and temporal scales in molecular systems: localized spatial interactions (high-frequency spatial scales) correspond to faster temporal dynamics, while extended spatial interactions (low-frequency spatial scales) correspond to slower temporal dynamics. By explicitly linking these scales, our method provides a comprehensive framework for capturing the complex interplay between spatial structures and temporal dynamics in MD simulations.

108 By integrating spatial and temporal multiscale modeling, our approach advances the simulation of
109 complex molecular systems. Extensive experiments on molecular dynamics datasets demonstrate
110 that our model outperforms baseline models, highlighting the effectiveness of capturing both spatial
111 and temporal multiscale interactions.

112 Our contributions can be summarized as follows:

- 114 1. **A New Perspective on Multiscale Modeling:** We provide a novel viewpoint by explic-
115 itly linking spatial and temporal scales in molecular dynamics, addressing the multiscale
116 challenge through joint modeling of spatial structures and temporal dynamics.
- 118 2. **A Novel Framework for Joint Multiscale Modeling:** We introduce a framework that
119 utilizes Graph Fourier Transformation to decompose molecular interactions into different
120 spatial scales and employs Neural ODEs to model the temporal dynamics at each scale.
121 The adaptive stepping of Neural ODEs allows for flexible modeling of dynamics at various
122 time scales.
- 123 3. **An Effective Model Architecture:** We design a new model architecture that implements
124 our proposed framework, effectively capturing and propagating different spatial scales
125 through time to model the multiscale nature of molecular systems. Extensive experiments
126 and ablations show the effectiveness of our architecture.

128 2 RELATED WORK

129 We review relevant works on multi-scale modeling in molecular dynamics, focusing on coarse-
130 grained methods, neural operator models, and graph neural ODEs.

131 2.1 COARSE-GRAINED MOLECULAR DYNAMICS

132 Coarse-grained molecular dynamics (CGMD) (Noid, 2013; Saunders & Voth, 2013) reduce com-
133 putational complexity by grouping atoms into larger units or “beads,” simplifying molecular repre-
134 sentations. This approach effectively captures large-scale molecular motions over extended periods,
135 facilitating the study of complex systems like protein folding and polymer dynamics. **F3low (Li
136 et al., 2024) extends CG modeling to the SE(3) manifold using flow-matching for autoregressive
137 sampling, enabling broader exploration of conformational space and improved insights into protein
138 folding. Two for One (Arts et al., 2023) uses score-based generative models to learn CG force fields
139 without force inputs, simplifying training and improving performance in protein simulations while
140 preserving dynamics.** However, CGMD primarily addresses spatial scaling and often sacrifices
141 fine-grained details inherent in localized interactions, such as bond vibrations and stretching. More-
142 over, it lacks explicit mechanisms for modeling temporal multi-scale dynamics, limiting its ability
143 to capture the interplay between fast and slow processes in molecular systems.

144 2.2 NEURAL OPERATOR MODELS

145 Neural operator models (Kovachki et al., 2023) have emerged as powerful tools for learning map-
146 pings between infinite-dimensional function spaces, showing promise in modeling complex dynamic
147 processes. **Conditional Diffusion Model (Hsu et al., 2024) introduces score dynamics (SD), which
148 learns accelerated evolution operators from molecular simulations using diffusion models, enabling
149 larger timesteps and faster simulations while maintaining accuracy.** In molecular dynamics, meth-
150 ods like Implicit Transfer Operator (ITO) (Schreiner et al., 2024), Timewarp (Klein et al., 2024),
151 and Equivariant Graph Neural Operator (EGNO) (Xu et al., 2024) aim to capture temporal dynam-
152 ics over multiple scales. ITO and Timewarp employ coarse-graining and time-stepping techniques
153 to accelerate simulations but focus primarily on temporal scaling, potentially losing critical spatial
154 details. EGNO models temporal evolution using neural operators with SE(3) equivariance but does
155 not explicitly handle spatial multi-scale interactions. These methods often treat spatial and temporal
156 scales separately, lacking a joint modeling framework that captures their intrinsic connection.

2.3 GRAPH NEURAL ODE MODELS

Graph Neural ODEs integrate Graph Neural Networks into continuous-time dynamics using Neural ODE frameworks (Chen et al., 2018), enabling the modeling of time-evolving graph-structured data (Zang & Wang, 2020; Huang et al., 2023a;b). While effective in capturing temporal evolution on graphs, these models typically rely on local message passing, which may not adequately represent multi-scale spatial interactions present in molecular systems. Moreover, they often do not explicitly link spatial scales with temporal dynamics, potentially overlooking the relationship between localized spatial interactions and fast temporal dynamics, as well as extended spatial interactions and slow dynamics.

3 PROBLEM DEFINITION

For the task of molecular dynamics, we aim to predict the future conformations of a molecular system based on its initial structure. Let $\mathcal{G}^{t_0} = (\mathcal{V}, \mathcal{E})$ represent the molecular graph at the initial time t_0 , where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ is the set of N atoms in the molecule, and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of optional edges representing chemical bonds or interactions between atoms.

The position of each atom $v_i \in \mathcal{V}$ at time t_0 is given by its 3D coordinates $\mathbf{x}_i^{t_0} \in \mathbb{R}^3$. Our objective is to predict the atomic positions $\mathbf{X}(t_k) := \{\mathbf{x}_1^{t_k}, \dots, \mathbf{x}_N^{t_k}\}$ at future time points t_1, t_2, \dots, t_K , where t_1, t_2, \dots, t_K are irregularly sampled on a continuous time scale. All data are parallel sequences from molecular systems of the same type, meaning the data contain conformations sampled from similar molecular systems.

4 METHODOLOGY

In this section, we introduce our framework for jointly modeling spatial and temporal multiscale interactions in molecular dynamics. Our approach integrates Equivariant Graph Neural Networks (EGNNs) (Satorras et al., 2021) with Graph Fourier Transforms (GFTs) and Neural Ordinary Differential Equations (Neural ODEs) (Chen et al., 2018).

4.1 EQUIVARIANT GRAPH NEURAL NETWORK ENCODER

We start by constructing a molecular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node $i \in \mathcal{V}$ represents an atom with features \mathbf{h}_i and position $\mathbf{x}_i \in \mathbb{R}^3$. The edges \mathcal{E} represent bonds or interactions between atoms.

Our EGNN encoder processes the molecular graph to produce updated node features and positions that are equivariant to rotations and translations.

Specifically, the EGNN updates node features and positions through message passing:

$$\mathbf{m}_{ij} = \phi_e(\mathbf{h}_i, \mathbf{h}_j, \mathbf{x}_i - \mathbf{x}_j), \quad (1)$$

$$\mathbf{h}'_i = \phi_h(\mathbf{h}_i, \sum_{j \in \mathcal{N}(i)} \mathbf{m}_{ij}), \quad (2)$$

$$\mathbf{x}'_i = \mathbf{x}_i + \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \psi(\mathbf{m}_{ij})(\mathbf{x}_i - \mathbf{x}_j), \quad (3)$$

where ϕ_e , ϕ_h , and ψ are learnable functions, and $\mathcal{N}(i)$ denotes the neighbors of node i .

In the above equations, \mathbf{h}_i are scalar latent features (rotationally invariant), and \mathbf{x}_i are vectorial features (rotationally equivariant). The updates are carefully designed to ensure that the updated positions \mathbf{x}'_i maintain translational and rotational equivariance, as the position updates depend only on relative positions $\mathbf{x}_i - \mathbf{x}_j$.

4.2 GRAPH FOURIER TRANSFORM

To capture multiscale spatial interactions, we apply the Graph Fourier Transform (GFT) to the node features and positions.

The GFT of the node features $\mathbf{H} \in \mathbb{R}^{N \times F}$ is given by:

$$\tilde{\mathbf{H}} = \mathbf{U}^\top \mathbf{H}, \quad (4)$$

where \mathbf{U} is the matrix of eigenvectors of the graph Laplacian \mathbf{L} .

Similarly, the GFT of the positions $\mathbf{X} \in \mathbb{R}^{N \times 3}$ is:

$$\tilde{\mathbf{X}} = \mathbf{U}^\top \mathbf{X}. \quad (5)$$

To preserve translational invariance, we note that the first eigenvector of the graph Laplacian corresponds to the zero eigenvalue and represents the DC component (uniform translation). By centering the positions \mathbf{X} before applying the GFT, i.e., $\mathbf{X}_c = \mathbf{X} - \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$, we ensure that the GFT captures only the variations in positions, which are invariant to translations.

4.3 NEURAL ODES FOR TEMPORAL DYNAMICS

Each frequency component from the GFT corresponds to a spatial scale. We model the temporal evolution of each component using Neural ODEs:

$$\frac{d\tilde{\mathbf{H}}_k(t)}{dt} = f_\theta(\tilde{\mathbf{H}}_k(t), t), \quad \forall k, \quad (6)$$

where $\tilde{\mathbf{H}}_k$ denotes the k -th frequency component, and f_θ is a neural network parameterized by θ .

Similarly, for positions:

$$\frac{d\tilde{\mathbf{X}}_k(t)}{dt} = g_\phi(\tilde{\mathbf{X}}_k(t), t), \quad \forall k. \quad (7)$$

Importantly, by integrating the Neural ODEs in the spectral domain and ensuring that the DC component is handled appropriately, we maintain translational equivariance in the reconstructed positions.

4.4 INVERSE GRAPH FOURIER TRANSFORM

After evolving the frequency components over time, we reconstruct the node features and positions using the Inverse Graph Fourier Transform (IGFT):

$$\mathbf{H}(t) = \mathbf{U} \tilde{\mathbf{H}}(t), \quad (8)$$

$$\mathbf{X}_c(t) = \mathbf{U} \tilde{\mathbf{X}}(t). \quad (9)$$

We then recover the absolute positions by adding back the mean position:

$$\mathbf{X}(t) = \mathbf{X}_c(t) + \mathbf{x}_{\text{mean}}, \quad (10)$$

where \mathbf{x}_{mean} is the mean position computed earlier. This ensures that the reconstructed positions maintain the correct translational behavior.

By carefully handling the mean and the DC component, we ensure that the entire process from EGNN to GFT, Neural ODEs, and IGFT preserves translational equivariance.

4.5 EQUIVARIANT GRAPH NEURAL NETWORK DECODER

Finally, we refine the reconstructed features and positions using an EGNN decoder, which has a similar structure to the encoder.

$$\mathbf{m}'_{ij} = \phi'_e(\mathbf{h}_i(t), \mathbf{h}_j(t), \mathbf{x}_i(t) - \mathbf{x}_j(t)), \quad (11)$$

$$\mathbf{h}_i'' = \phi'_h(\mathbf{h}_i(t), \sum_{j \in \mathcal{N}(i)} \mathbf{m}'_{ij}), \quad (12)$$

$$\mathbf{x}_i'' = \mathbf{x}_i(t) + \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \psi'(\mathbf{m}'_{ij})(\mathbf{x}_i(t) - \mathbf{x}_j(t)). \quad (13)$$

Again, the updates depend only on relative positions, ensuring that translational equivariance is maintained throughout the decoding process.

4.6 LOSS FUNCTION

To train the model, we use the Mean Squared Error (MSE) loss, which measures the difference between the predicted atomic positions and the ground truth positions at each predicted time point. Given that the goal is to predict the molecular conformations at time points t_1, t_2, \dots, t_K , the MSE is calculated as follows:

Let $\mathbf{x}_i^{t_j} \in \mathbb{R}^3$ be the ground truth 3D coordinates of atom i at time t_j , and $\tilde{\mathbf{x}}_i^{t_j} \in \mathbb{R}^3$ be the predicted coordinates for the same atom at time t_j . The MSE loss is defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{NK} \sum_{j=1}^K \sum_{i=1}^N \left\| \mathbf{x}_i^{t_j} - \tilde{\mathbf{x}}_i^{t_j} \right\|_2^2, \quad (14)$$

where N is the number of atoms and K is the number of time points.

This loss encourages the model to minimize the Euclidean distance between the predicted and actual atomic positions across all time steps, ensuring accurate trajectory prediction for the molecular system.

The MSE loss is applied at each time point, thus aligning the predicted future states with the true molecular dynamics trajectory. The model is trained by minimizing \mathcal{L}_{MSE} over all predicted time points.

5 EXPERIMENTS

In this section, we present a comprehensive evaluation of our proposed model on molecular dynamics datasets, comparing its performance against several baseline methods and conducting ablation studies to assess the contributions of different model components.

Our experiments are designed to address the following research questions:

- **RQ1:** Does jointly modeling spatial and temporal multiscale interactions improve prediction accuracy?
- **RQ2:** How do the different components of our model (the spatial multiscale modeling and the temporal multiscale modeling) contribute to its performance?
- **RQ3:** Does extracting different scales stabilize predictions over longer periods of time?
- **RQ4:** Does our model capture the latent correlation between specific spatial scales and specific temporal scales?

5.1 DATASET

We evaluate our model using the MD17 dataset (Chmiela et al., 2017), which contains molecular dynamics trajectories for eight small molecules: Aspirin, Benzene, Ethanol, Malonaldehyde, Naphthalene, Salicylic Acid, Toluene, and Uracil. The dataset provides atomic trajectories simulated

324 under quantum mechanical forces, capturing realistic molecular motions. We also evaluated our
325 model on the alanine dipeptide dataset (Schreiner et al., 2024), a standard benchmark for studying
326 conformational dynamics in proteins. Our task is to predict the future positions of atoms given the
327 initial state of the molecular system.

329 5.2 EXPERIMENTAL SETUP

330
331 For each molecule, we partition the trajectory data into training, validation, and test sets, using 500
332 samples for training, 2000 for validation, and 2000 for testing. The time scope ΔT for each piece
333 of data is set to 3000 simulation steps, providing a challenging prediction horizon.

334 A key aspect of our experimental setup is the use of **irregular timestep sampling**, in contrast to the
335 equi-timestep sampling used in some baseline models like EGNO, to better mimic the variable time
336 intervals in real-world physical systems. This setting tests the models’ ability to handle irregular
337 temporal data. Nevertheless, we also provide evaluations based on equi-timestep sampling in the
338 appendix for completeness.

339 Further details on the experimental setup, including hyperparameter choices and implementation
340 specifics, are also provided in the uploaded codebase to OpenReview. Detailed specification can
341 also be found in the appendix.

343 5.3 BASELINE MODELS

344
345 We compare our model against several state-of-the-art approaches:

- 346
347 • **NDCN** (Zang & Wang, 2020): A Graph Neural ODE model that integrates graph neural
348 networks into the ODE framework to learn continuous-time dynamics of networked sys-
349 tems.
- 350
351 • **EGNN** (Satorras et al., 2021): An Equivariant Graph Neural Network that models molec-
352 ular systems using 3D equivariant message passing but without explicit time propagation.
- 353
354 • **EGNO** (Xu et al., 2024): An Equivariant Graph Neural Operator that captures temporal
355 dynamics using neural operators with regular timesteps.
- 356
357 • **ITO** (Schreiner et al., 2024): An Implicit Time-stepping Operator that integrates differen-
358 tial equations into the learning process for temporal evolution.

359 These baselines represent a range of approaches for modeling molecular dynamics, including meth-
360 ods that focus on either spatial modeling (EGNN), or temporal modeling (NDCN, EGNO, ITO).

361 5.4 RESULTS AND ANALYSIS

362 To address **RQ1**, we evaluate the performance of our model and the baselines on the MD17 dataset.
363 The test Mean Squared Error (MSE) on the eight molecules for our model and the baseline methods
364 are summarized in Table 1. For completeness, we also provide results in equi-timestep setting in
365 Table 4 in the Appendix.

366
367
368 From Table 1, we observe that our model consistently outperforms the baseline methods across
369 all eight molecules under irregular timestep sampling. This demonstrates the effectiveness of our
370 approach in jointly modeling spatial and temporal multiscale interactions. The performance gains
371 are particularly significant for molecules like Benzene and Aspirin, where our model reduces the
372 MSE by substantial margins compared to the baselines.

373 The inferior performance of EGNO and ITO under irregular timestep sampling highlights their lim-
374 itations in handling variable time intervals due to their reliance on discrete Fourier transformations
375 in the temporal domain, which assume equi-timesteps. In contrast, EGNN, which, in our implemen-
376 tation, incorporates timestamp embeddings into a 3D GNN model, performs better than EGNO and
377 ITO in this setting but still falls short of our model. This underscores the advantage of our method
in modeling dynamics with irregular timesteps by explicitly capturing multiscale interactions.

Table 1: MSE ($\times 10^{-2} \text{ \AA}^2$) on MD17 dataset, timesteps sampled with **irregular** intervals. Upper part: comparison between the baselines and our model. Best results are marked in **bold**, and second-to-best results are underlined. Lower part: results for ablation models. Compared with the base model performance, deteriorated results are marked in **orange**, while improved results are marked in **green**.

	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
NDCN	29.75 \pm 0.02	70.13 \pm 0.98	10.05 \pm 0.02	42.28 \pm 0.07	2.30 \pm 0.00	3.43 \pm 0.05	12.33 \pm 0.00	2.39 \pm 0.00
EGNN	<u>9.09</u> \pm 0.10	<u>49.15</u> \pm 1.68	<u>4.46</u> \pm 0.01	<u>12.52</u> \pm 0.05	<u>0.40</u> \pm 0.02	<u>0.89</u> \pm 0.01	<u>8.98</u> \pm 0.09	<u>0.64</u> \pm 0.00
EGNO	10.60 \pm 0.01	52.53 \pm 2.40	4.52 \pm 0.06	12.89 \pm 0.06	0.46 \pm 0.01	1.07 \pm 0.00	9.31 \pm 0.10	0.67 \pm 0.01
ITO	12.74 \pm 0.10	57.84 \pm 0.86	7.23 \pm 0.00	19.53 \pm 0.01	1.77 \pm 0.01	2.53 \pm 0.03	9.96 \pm 0.04	1.71 \pm 0.15
Ours	8.85 \pm 0.02	40.86 \pm 0.98	4.41 \pm 0.06	12.49 \pm 0.00	0.40 \pm 0.01	0.87 \pm 0.01	8.63 \pm 0.04	0.62 \pm 0.02
w/o Fourier	8.79 \pm 0.06	46.23 \pm 3.16	4.42 \pm 0.01	12.50 \pm 0.03	0.46 \pm 0.00	0.95 \pm 0.07	8.70 \pm 0.01	0.61 \pm 0.01
w/o ODE	8.93 \pm 0.01	43.09 \pm 2.07	4.42 \pm 0.00	12.49 \pm 0.01	0.41 \pm 0.01	0.88 \pm 0.02	8.65 \pm 0.08	0.62 \pm 0.02

To further validate the effectiveness of our approach on more complex molecular dynamics scenarios, we extend our experiments to include the **alanine dipeptide** dataset, a standard benchmark in molecular dynamics simulations Schreiner et al. (2024). This dataset captures essential aspects of protein backbone dynamics and presents a greater challenge due to its conformational flexibility and rich torsional dynamics.

Experimental Results on Alanine Dipeptide *Units and Calculations:* Note that the alanine dipeptide dataset operates in nanometers (nm), while the MD17 dataset uses angstroms (\AA). When calculating bond lengths and bond angles, we consider all heavy atoms (excluding hydrogen atoms) to focus on the primary structural framework of the molecule.

The Mean Squared Error (MSE) results on the alanine dipeptide dataset are summarized in Table 2.

Table 2: Mean Squared Error (MSE) ($\times 10^{-3} \text{ nm}^2$) on the alanine dipeptide dataset. Best results are marked in **bold**.

Model	MSE ($\times 10^{-3} \text{ nm}^2$)
NDCN	12.27 \pm 0.19
ITO	26.95 \pm 0.19
EGNO	6.92 \pm 0.26
EGNN	5.67 \pm 0.08
Ours	5.24 \pm 0.19

From Table 2, we observe that our model achieves the lowest MSE among all compared methods, with approximately 7.5% reduction in error compared to EGNN. This indicates a significant improvement over the baselines and demonstrates the effectiveness of our approach on more complex molecular dynamics data.

In-depth Analysis of Molecular Structure Recovery To provide a more comprehensive evaluation, we analyzed the bond lengths and bond angles predicted by our model compared to the ground truth. Table 3 presents the Mean Absolute Errors (MAEs) and relative errors for bond lengths and bond angles.

Table 3: Mean Absolute Errors (MAEs) and relative errors for bond lengths and bond angles on the alanine dipeptide dataset. Best results are marked in **bold**.

Model	Bond Length MAE (nm)	Rela. Err. (%)	Bond Angle MAE ($^\circ$)	Rela. Err. (%)
EGNN	0.0209 \pm 0.0006	15.32 \pm 0.49	12.44 \pm 0.91	10.48 \pm 0.76
EGNO	0.0229 \pm 0.0018	16.75 \pm 1.23	10.54 \pm 0.11	8.89 \pm 0.11
Ours	0.0188 \pm 0.0022	13.74 \pm 1.66	10.47 \pm 1.03	8.80 \pm 0.89

Our model achieves the lowest MAEs and relative errors in both bond lengths and bond angles, demonstrating its superior ability to accurately recover internal molecular structures compared to the baselines.

Ramachandran Plot Analysis We further evaluate the ability of our model to capture torsional dynamics by analyzing the distribution of dihedral (torsion) angles ϕ (phi) and ψ (psi). Figure 2 presents the Ramachandran plots comparing the predicted and ground truth distributions.

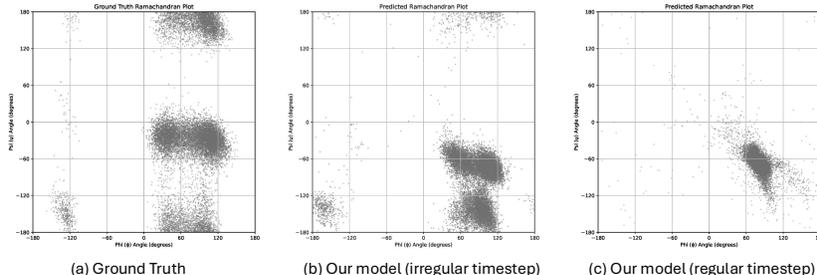


Figure 2: Ramachandran plots showing the distribution of dihedral angles ϕ and ψ for alanine dipeptide. (a) Ground truth distribution. (b) Distribution predicted by our model, trained on irregular timestep sampling setting; (c) trained on regular timestep sampling setting.

As shown in Figure 2, our model accurately captures the characteristic conformational regions of alanine dipeptide, closely matching the ground truth distributions. This indicates that our model effectively models the torsional dynamics, which are critical for understanding protein folding and conformational changes. We also observed that the model trained with irregular timestep sampling better recovers the characteristic patterns in the Ramachandran plot, indicating an improved ability to learn the molecule’s conformational space by incorporating the conformations at multiple time scales.

These additional experiments on alanine dipeptide further confirm the robustness and effectiveness of our approach in modeling complex molecular dynamics with irregular timesteps, outperforming existing state-of-the-art methods.

5.5 ABLATION STUDIES

To address **RQ2**, we conduct ablation studies to assess the contributions of the key components in our model: the Graph Fourier Transformation for spatial multiscale modeling and the Neural ODE for temporal multiscale modeling. We consider two ablated versions of our model:

- **w/o Fourier**: The model without the Graph Fourier Transformation. In this variant, we directly use the spatial features without decomposing them into different spatial scales.
- **w/o ODE**: The model without the Neural ODE component. In this variant, we replace the Neural ODE with a standard feed-forward neural network for temporal propagation.

The results of the ablation studies are presented in the lower part of Table 1. The results for the equi-timestep setting is provided in the lower part of Table 4 in the Appendix.

From the ablation results, we observe that removing either the Graph Fourier Transformation or the Neural ODE component leads to a degradation in performance on most molecules. For instance, without the Fourier component, the MSE on Benzene increases from 40.86 to 46.23, and without the ODE component, it increases to 43.09. This indicates that both the spatial multiscale modeling via Graph Fourier Transformation and the temporal multiscale modeling via Neural ODE contribute significantly to the overall performance of the model.

In some cases, the ablated models perform comparably to the full model (e.g., on Aspirin and Uracil). However, the full model consistently achieves the best overall performance across all molecules, confirming the importance of jointly modeling both spatial and temporal multiscales.

5.6 LONG-TERM PREDICTION STABILITY

To address **RQ3**, we evaluate the performance of our model and the baselines on longer-term predictions. We test the models on predicting atomic positions over extended time horizons of 1000, 2000, 3000, 4000, and 5000 simulation steps on two representative molecules: Benzene and Malonaldehyde. The results are shown in Figure 3.

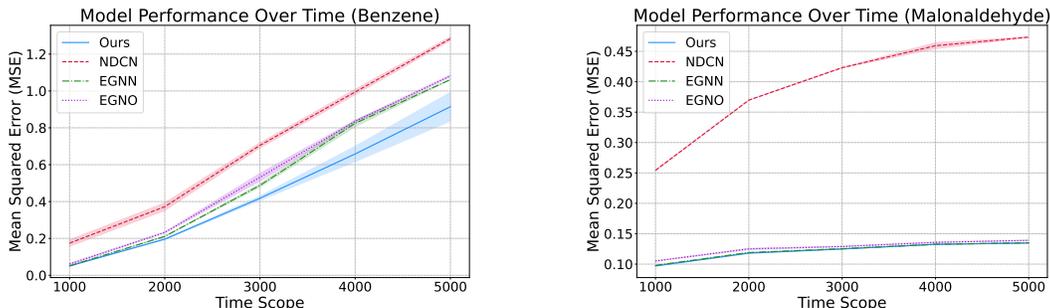


Figure 3: Long-term prediction MSE over different time horizons for Benzene and Malonaldehyde.

From Figure 3, we observe that our model maintains superior performance and demonstrates more stable predictions over longer periods compared to the baselines. The error growth rate of our model is less steep, indicating that it is better at capturing the dynamics that govern the system over extended time scales. This can be attributed to our model’s ability to identify and propagate the lower-frequency modes, which carry more global information and are crucial for long-term predictions. The baselines, lacking explicit multiscale modeling, tend to accumulate errors more rapidly over time.

5.7 CASE STUDY: CORRELATION BETWEEN SPATIAL AND TEMPORAL SCALES

To address **RQ4**, we conduct a case study on the Benzene molecule to investigate whether our model captures the latent correlation between specific spatial scales and temporal scales. We analyze the Fourier coefficients of different spatial modes over time to observe how they evolve.

As shown in Figure 4, the lower spatial modes (e.g., mode 0), which represent global, low-frequency spatial scales on the graph, also encode temporal signals of lower frequency. These modes change gradually over time, reflecting slower dynamics associated with extended spatial interactions. Conversely, higher spatial modes (e.g., mode 3), corresponding to more localized spatial interactions, has a significantly lower portion for the lower frequency temporal signals, exhibit higher-frequency temporal dynamics with rapid fluctuations. This demonstrates that our model effectively captures the inherent correlation between specific spatial scales and their associated temporal scales, validating our approach of jointly modeling spatial and temporal multiscale interactions.

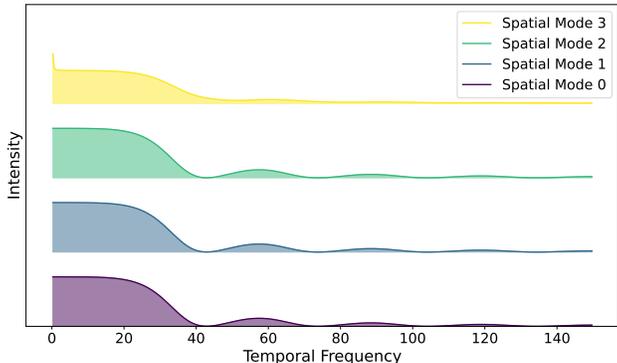


Figure 4: Fourier coefficients of different spatial modes over time for Benzene. Lower spatial modes (e.g., mode 0) correspond to a higher portion of lower-frequency temporal signals, while higher spatial modes (e.g., mode 3) exhibit higher-frequency temporal dynamics. (x axis unit: ps^{-1})

REFERENCES

- 540
541
542 Marloes Arts, Victor Garcia Satorras, Chin-Wei Huang, Daniel Zugner, Marco Federici, Cecilia
543 Clementi, Frank Noé, Robert Pinsler, and Rianne van den Berg. Two for one: Diffusion mod-
544 els and force fields for coarse-grained molecular dynamics. *Journal of Chemical Theory and*
545 *Computation*, 19(18):6151–6159, 2023.
- 546 Michael Bear and Estela Blaisten-Barojas. Atomic and ion collisions in solids and at surfaces:
547 Theory, simulation and applications. *American Scientist*, 86(2):188, 1998.
- 548
549 J-B Brubach, Al Mermet, A Filabozzi, A Gerschel, and P Roy. Signatures of the hydrogen bonding
550 in the infrared bands of water. *The Journal of chemical physics*, 122(18), 2005.
- 551
552 Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary
553 differential equations. *Advances in neural information processing systems*, 31, 2018.
- 554
555 Stefan Chmiela, Alexandre Tkatchenko, Huziel E Sauceda, Igor Poltavsky, Kristof T Schütt, and
556 Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields.
557 *Science advances*, 3(5):e1603015, 2017.
- 558
559 Peter R Griffiths. Fourier transform infrared spectrometry. *Science*, 222(4621):297–302, 1983.
- 560
561 Tim Hsu, Babak Sadigh, Vasily Bulatov, and Fei Zhou. Score dynamics: Scaling molecular dynam-
562 ics with picoseconds time steps via conditional diffusion model. *Journal of Chemical Theory and*
563 *Computation*, 20(6):2335–2348, 2024.
- 564
565 Zijie Huang, Yizhou Sun, and Wei Wang. Generalizing graph ode for learning complex system dy-
566 namics across environments. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge*
567 *Discovery and Data Mining*, pp. 798–809, 2023a.
- 568
569 Zijie Huang, Wanjia Zhao, Jingdong Gao, Ziniu Hu, Xiao Luo, Yadi Cao, Yuanzhou Chen, Yizhou
570 Sun, and Wei Wang. Tango: Time-reversal latent graphode for multi-agent dynamical systems.
571 *arXiv preprint arXiv:2310.06427*, 2023b.
- 572
573 Soumil Y Joshi and Sanket A Deshmukh. A review of advancements in coarse-grained molecular
574 dynamics simulations. *Molecular Simulation*, 47(10-11):786–803, 2021.
- 575
576 Leon Klein, Andrew Foong, Tor Fjelde, Bruno Mlodozieniec, Marc Brockschmidt, Sebastian
577 Nowozin, Frank Noé, and Ryota Tomioka. Timewarp: Transferable acceleration of molecular
578 dynamics by learning time-coarsened dynamics. *Advances in Neural Information Processing*
579 *Systems*, 36, 2024.
- 580
581 Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, An-
582 drew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces
583 with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.
- 584
585 Shaoning Li, Yusong Wang, Mingyu Li, Jian Zhang, Bin Shao, Nanning Zheng, and Jian Tang.
586 F³low: Frame-to-frame coarse-grained molecular dynamics with se(3) guided flow matching,
587 2024. URL <https://arxiv.org/abs/2405.00751>.
- 588
589 William George Noid. Perspective: Coarse-grained models for biomolecular systems. *The Journal*
590 *of chemical physics*, 139(9), 2013.
- 591
592 Outi MH Salo-Ahen, Ida Alanko, Rajendra Bhadane, Alexandre MJJ Bonvin, Rodrigo Vargas Hon-
593 orato, Shakhawath Hossain, André H Juffer, Aleksei Kabedev, Maija Lahtela-Kakkonen, An-
Anders Støttrup Larsen, et al. Molecular dynamics simulations in drug discovery and pharmaceutical
development. *Processes*, 9(1):71, 2020.
- Victor Garcia Satorras, Emiel Hoogetboom, and Max Welling. E (n) equivariant graph neural net-
works. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.
- Marissa G Saunders and Gregory A Voth. Coarse-graining methods for computational biology.
Annual review of biophysics, 42(1):73–93, 2013.

- 594 Mathias Schreiner, Ole Winther, and Simon Olsson. Implicit transfer operator learning: Multiple
595 time-resolution models for molecular dynamics. *Advances in Neural Information Processing*
596 *Systems*, 36, 2024.
- 597 Antonis I Vakis, Vladislav A Yastrebov, Julien Scheibert, Lucia Nicola, Daniele Dini, Clotilde Min-
598 fray, Andreas Almqvist, Marco Paggi, Seunghwan Lee, Georges Limbert, et al. Modeling and
599 simulation in tribology across scales: An overview. *Tribology International*, 125:169–199, 2018.
- 600 Jiang Wang, Simon Olsson, Christoph Wehmeyer, Adrià Pérez, Nicholas E Charron, Gianni De Fab-
601 rittiis, Frank Noé, and Cecilia Clementi. Machine learning of coarse-grained molecular dynamics
602 force fields. *ACS central science*, 5(5):755–767, 2019.
- 603 Quan-De Wang, Jing-Bo Wang, Juan-Qin Li, Ning-Xin Tan, and Xiang-Yuan Li. Reactive molecular
604 dynamics simulation and chemical kinetic modeling of pyrolysis and combustion of n-dodecane.
605 *Combustion and Flame*, 158(2):217–226, 2011.
- 606 Minkai Xu, Jiaqi Han, Aaron Lou, Jean Kossaifi, Arvind Ramanathan, Kamyar Azizzadenesheli,
607 Jure Leskovec, Stefano Ermon, and Anima Anandkumar. Equivariant graph neural operator for
608 modeling 3d dynamics. *arXiv preprint arXiv:2401.11037*, 2024.
- 609 Chengxi Zang and Fei Wang. Neural dynamics on complex networks. In *Proceedings of the 26th*
610 *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 892–902,
611 2020.

615 A ADDITIONAL EXPERIMENT RESULTS

616 Table 4: MSE ($\times 10^{-2}$) on MD17 dataset, timesteps sampled with **equal** intervals. Upper part:
617 comparison between the baselines and our model. Best results are marked in **bold**, and second-to-
618 best results are underlined. Lower part: results for ablation models. Compared with the base model
619 performance, deteriorated results are marked in **orange**, while improved results are marked in **green**.

	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
NDCN	31.73 \pm 0.40	56.21 \pm 0.30	10.74 \pm 0.02	46.55 \pm 0.28	2.25 \pm 0.01	3.58 \pm 0.11	13.92 \pm 0.02	2.38 \pm 0.00
EGNN	<u>9.24</u> \pm 0.07	57.85 \pm 2.70	<u>4.63</u> \pm 0.00	12.81 \pm 0.01	<u>0.38</u> \pm 0.01	0.85 \pm 0.00	<u>10.41</u> \pm 0.04	<u>0.56</u> \pm 0.02
EGNO	9.41 \pm 0.09	<u>55.13</u> \pm 3.21	<u>4.63</u> \pm 0.00	12.81 \pm 0.01	0.40 \pm 0.01	0.93 \pm 0.01	10.43 \pm 0.10	0.59 \pm 0.01
ITO	20.56 \pm 0.03	57.25 \pm 0.58	8.60 \pm 0.27	28.44 \pm 0.73	1.82 \pm 0.17	2.48 \pm 0.34	12.47 \pm 0.30	1.33 \pm 0.12
Ours	9.16 \pm 0.10	50.93 \pm 6.28	4.62 \pm 0.00	<u>12.83</u> \pm 0.00	0.35 \pm 0.01	<u>0.88</u> \pm 0.05	10.31 \pm 0.11	0.54 \pm 0.01
w/o Fourier	8.79 \pm 0.26	51.28 \pm 5.38	4.62 \pm 0.00	12.81 \pm 0.01	0.37 \pm 0.01	0.90 \pm 0.03	10.35 \pm 0.10	0.60 \pm 0.01
w/o ODE	9.19 \pm 0.15	55.13 \pm 2.73	4.63 \pm 0.01	12.80 \pm 0.00	0.37 \pm 0.02	0.86 \pm 0.05	10.32 \pm 0.04	0.54 \pm 0.02

631 B EXPERIMENT SETUP

632 In this section, we provide a detailed description of the experimental setup for our model, including
633 training procedures, model architecture, dataset inputs and outputs, hyperparameters, and an in-
634 depth examination of how the input features \mathbf{x} and \mathbf{h} are transformed throughout the model. This
635 information is based on the provided configuration file `config_md17_no.json` and the main
636 training script `main_md17_no.py`.

637 B.1 DATASETS

638 B.1.1 MD17 DATASET

- 639 • **Description:** The MD17 dataset consists of molecular dynamics simulations for various
640 small organic molecules. It provides atomic trajectories computed using *ab initio* methods,
641 capturing realistic molecular motions.
- 642 • **Molecule Used:** In this experiment, we focus on the **Alanine Dipeptide (ala2)** molecule,
643 specified by "mol": "ala2" in the configuration.

- 648 • **Data Directory:** The data is expected to be located in the directory specified by
649 "data_dir": "ala".
- 650 • **Data Split:**
 - 651 – **Training Set:** 500 samples (as specified by "max_training_samples": 500).
 - 652 – **Validation Set:** 2,000 samples.
 - 653 – **Test Set:** 2,000 samples.
- 654 • **Time Steps:** Each sample contains **8 time steps** ("num_timesteps": 8), capturing
655 the progression of the molecular state over time. The time interval between frames is
656 determined by "delta_frame": 3000.
- 657 • **Uneven Sampling:** Disabled ("uneven_sampling": false), indicating that
658 timesteps are sampled at regular intervals.

661 B.2 MODEL ARCHITECTURE

662
663 Our model aims to jointly model spatial and temporal multiscale interactions in molecular dynamics
664 simulations. The primary components of the model include:

665 B.2.1 1. ENCODER (EQUIVARIANT GRAPH NEURAL NETWORK)

- 666 • **Purpose:** Encode the initial molecular structure and properties into latent representations.
- 667 • **Input Features:**
 - 668 – **Atomic Positions** (\mathbf{x}): 3D coordinates of atoms at each time step.
 - 669 – **Atomic Features** (\mathbf{h}): Scalars representing properties like atom types or initial em-
670 beddings.
- 671 • **Process:** Utilize an **Equivariant Graph Neural Network (EGNN)** to encode spatial rela-
672 tionships while preserving equivariance under rotations and translations. Update \mathbf{h} and \mathbf{x}
673 based on neighboring information.

674 B.2.2 2. GRAPH FOURIER TRANSFORM (GFT)

- 675 • **Purpose:** Decompose the molecular graph into different spatial frequency components to
676 capture multiscale spatial interactions.
- 677 • **Process:** Apply GFT to the latent features \mathbf{h} and positions \mathbf{x} separately, transforming them
678 into the spectral domain. The number of modes is specified by "num_modes": 2,
679 indicating that we capture two spatial frequency components.

680 B.2.3 3. NEURAL ORDINARY DIFFERENTIAL EQUATIONS (NEURAL ODES)

- 681 • **Purpose:** Model the temporal evolution of each frequency component obtained from the
682 GFT.
- 683 • **Process:** For each frequency mode, we define a Neural ODE that models the continuous-
684 time dynamics. The ODE solver used is "solver": "dopri5", with relative and
685 absolute tolerances specified by "rtol": 1e-3 and "atol": 1e-4, respectively.
686 Temporal embeddings are incorporated via "time_emb_dim": 32.

687 B.2.4 4. INVERSE GRAPH FOURIER TRANSFORM (IGFT)

- 688 • **Purpose:** Transform the evolved frequency components back to the spatial domain to re-
689 construct the updated positions and features.

690 B.2.5 5. DECODER (EQUIVARIANT GRAPH NEURAL NETWORK)

- 691 • **Purpose:** Refine the reconstructed features and produce the final predicted atomic posi-
692 tions.
- 693 • **Process:** Apply an EGNN similar to the encoder to update \mathbf{h} and \mathbf{x} based on the evolved
694 features.

B.3 HYPERPARAMETERS

The hyperparameters used in the experiment are as follows (from `config_md17_no.json`):

- **Experiment Name:** "exp_name": "md17_exp"
- **Number of Time Steps:** "num_timesteps": 8
- **Batch Size:** "batch_size": 50
- **Epochs:** "epochs": 5000
- **Use CUDA:** "no_cuda": false
- **Random Seed:** "seed": 53
- **Learning Rate:** "lr": 0.0001
- **Hidden Feature Dimension:** "nf": 64
- **Model Type:** "model": "fourier"
- **Attention Mechanism:** "attention": 0
- **Number of Layers:** "n_layers": 5
- **Maximum Training Samples:** "max_training_samples": 500
- **Data Directory:** "data_dir": "ala"
- **Normalize Differences:** "norm_diff": false
- **Learnable Fourier Basis:** "learnable": false
- **Weight Decay:** "weight_decay": 1e-15
- **Use Tanh Activation:** "tanh": false
- **Delta Frame:** "delta_frame": 3000
- **Molecule:** "mol": "ala2"
- **Time Embedding Dimension:** "time_emb_dim": 32
- **Number of Modes:** "num_modes": 2
- **ODE Solver:** "solver": "dopri5"
- **Relative Tolerance:** "rtol": 1e-3
- **Absolute Tolerance:** "atol": 1e-4
- **Uneven Sampling:** "uneven_sampling": false
- **Fourier Basis:** "fourier_basis": "graph"
- **Disable Fourier Block:** "no_fourier": false
- **Disable ODE Block:** "no_ode": false

B.4 TRAINING PROCEDURE

- **Optimizer:** Adam optimizer with the specified learning rate and weight decay.
- **Learning Rate Scheduler:** StepLR scheduler with a step size and gamma (not specified in the config, but default values can be inferred from the code).
- **Loss Function:** Mean Squared Error (MSE) between predicted and true atomic positions.
- **Training Loop:**
 - Iterate over the specified number of epochs ("epochs": 5000).
 - For each batch, compute the loss and update model parameters via backpropagation.
 - Evaluate the model on the validation and test sets at regular intervals.

C BROADER IMPACT

In this paper, we introduced a novel framework for modeling molecular dynamics by jointly capturing spatial and temporal multiscale interactions. By leveraging the Graph Fourier Transform to decompose molecular structures into different spatial scales and employing Neural Ordinary Differential Equations to model the temporal dynamics at each scale, our approach explicitly links spatial structures with temporal evolution. Extensive experiments on the MD17 dataset demonstrated that our model consistently outperforms state-of-the-art baseline methods across multiple molecules, particularly under challenging conditions like irregular timestep sampling and long-term prediction horizons. Ablation studies confirmed the significant contributions of both the spatial multiscale modeling and the temporal multiscale modeling components. Our method provides a comprehensive and accurate representation of molecular behavior over time, advancing the simulation of complex molecular systems.

The proposed framework has the potential to significantly impact various domains by enhancing the accuracy and efficiency of molecular simulations. In computational chemistry and drug discovery, it can accelerate the understanding of molecular interactions and aid in the development of new therapeutics. In materials science, it can assist in designing novel materials with desired properties by accurately modeling molecular dynamics at different scales. Additionally, the method's computational efficiency may reduce resource requirements for large-scale simulations. However, it is important to consider potential limitations, such as the need for high-quality data and computational resources, as well as ethical considerations regarding the misuse of advanced simulation technologies. By addressing these factors, our work aims to contribute responsibly to the advancement of molecular science and technology.