DoMiNO: Decomposing Molecular Dynamics with Multi-Scale Neural Graph Ordinary Differential Equations

Anonymous authors
Paper under double-blind review

Abstract

Molecular dynamics (MD) simulations are crucial for understanding and predicting the behavior of molecular systems in biology and chemistry. Yet, predicting long-term dynamics is still challenging. On one hand, it is hard to employ small-timestep models for long-term prediction, due to substantial rollout errors accumulated at each step, not to mention their extremely high time complexity due to the large number of rollout steps. On the other hand, it is hard to use large-timestep models to achieve high accuracy, due to their inability to capture subtle details of the dynamics. To bridge this dichotomy, we propose *DoMiNO*, a multi-scale framework that decomposes MD dynamics into several temporal resolutions, each governed by a neural graph ordinary differential equation (GraphODE) and are adaptively fused for final predictions. Concretely, *DoMiNO* operates through three key components: (1) an E(n)-equivariant graph neural network (EGNN) encoder that initializes latent states from a single observed molecular structure, maintaining SE(3) symmetries throughout; (2) a hierarchy of GraphODEs where each level captures scale-specific dynamics over normalized local time intervals, ranging from slow global motions to fast bond vibrations; and (3) an attention-based fusion module that adaptively combines multi-level predictions and reconstructs SE(3)-equivariant 3D coordinates. This design enables each hierarchical level to specialize in its characteristic timescale while preserving molecular symmetries. During inference, *DoMiNO* flexibly assembles predictions across different temporal resolutions, providing superior performance over both short-term and long-term dynamics. Empirical results on challenging MD benchmarks demonstrate that *DoMiNO* achieves significant improvements in prediction accuracy, particularly for molecules with pronounced timescale separation. The method exhibits significantly slower error growth over extended horizons compared to both single-scale baselines and state-of-the-art multi-step approaches. Our implementation is available at `https://anonymous.4open.science/r/domino-code-12EE`.

1 Introduction

Molecular dynamics (MD) simulations (Dror et al., 2012; Hollingsworth & Dror, 2018) serve as a critical tool in computational chemistry (Car & Parrinello, 1985; Marx & Hutter, 2009) and biology (Shaw et al., 2010; Lindorff-Larsen et al., 2011), providing insights into atomic-scale interactions over time. A fundamental challenge, however, is the **vast computational cost** required to simulate significant events such as protein folding, which arises from a timescale disparity: The integration timestep in MD simulations typically is on the order of a femtoseconds (10^{-15} s) (Hollingsworth & Dror, 2018) to accurately capture atomic interactions. However, the events to be modeled, such as protein conformational changes, occur over nanoseconds (10^{-9} s) to microseconds (10^{-6} s) (McGeagh et al., 2011). Consequently, accurately modeling these events requires millions of integration steps, which can be prohibitively expensive.

This computational challenge is fundamentally rooted in the **multi-scale** nature of molecular dynamics. Many molecular systems [1] simultaneously exhibit rapid, high-frequency events, like atomic vibrations, alongside much slower, large-scale structural changes that govern biological function, such as

[1] See the animation at `https://catenane.net/media/ChemMotorAnimHQ.mp4` (Li et al., 2024)

protein's domain rearrangement. This presents a classic dilemma in numerical simulation: large time steps offer computational efficiency but sacrifice the fidelity of the fast dynamics, while small time steps capture fine-grained details accurately but yield significant computational cost.

Several machine-learning surrogates have been proposed to alleviate the computational burden in MD simulations. Neural ODE-based models (Chen et al., 2018; Huang et al., 2020), offer a continuous formulation that avoids the pitfalls of autoregressive rollouts, while wavelet-based (Conejo et al., 2005) techniques decompose time series into multiple frequency components, and generative MD (Schreiner et al., 2023) approaches enable efficient long-range sampling. However, neural ODEs typically lack explicit mechanisms for capturing the inherent multi-scale dynamics of molecular systems. Wavelet methods rely on fixed frequency decompositions that may miss complex nonlinearities, and generative models can suffer from long-term instability. Consequently, no existing method simultaneously delivers high-resolution predictions on demand and robust continuous-time modeling.

To bridge this gap, we need to frame molecular dynamics as involving co-existing fast and slow modes that necessitate different temporal resolutions. Following this multi-scale perspective, we propose *DoMiNO*, a decomposition-based pipeline with Neural Graph ODEs. This framework hierarchically decomposes the trajectory according to different timestep sizes and models each scale with a dedicated Neural ODE module. By combining neural models across multiple levels, our method achieves both rapid large-scale rollouts and the flexibility to "zoom in" and predict fine-grained atomic dynamics with high fidelity. As illustrated in Figure 1, the model reconstructs molecular trajectories through a multi-scale decomposition process, progressively refining predictions from coarse to fine temporal scales. Only a limited number of time points need to be sampled per level, significantly reducing the computational cost while maintaining accuracy. Our contributions can be summarized as follows:

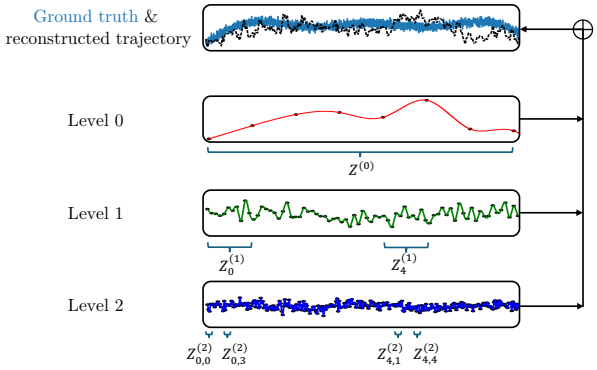

Figure 1: **Hierarchical decomposition of molecular dynamics trajectories using *DoMiNO*.** The model decomposes MD dynamics into multiple temporal resolutions through a multi-scale Neural Graph ODEs framework. Level 0 captures slow, large-scale motions, while finer levels (Level 1, Level 2) model progressively faster dynamics. By selectively sampling a limited number of points at each level and fusing predictions across scales, the model significantly improves computational efficiency while maintaining accuracy.

- *New perspective.* We articulate the fast-slow dichotomy in MD simulations and develop a hierarchical decomposition framework to accommodate multiple timescales.
- *Novel architecture.* We propose a multi-scale neural GraphODE architecture where each level operates in local relative time, tied together by equivariant encoders/decoders and adaptive fusion.
- *Superior performance.* We validate *DoMiNO* on diverse molecular systems from small molecules to proteins, demonstrating significant gains in long-term stability and accuracy compared to state-of-the-art baselines.

2 Related Work

Our approach builds upon and extends several key research directions in molecular dynamics modeling. We review neural approaches for temporal dynamics, multi-scale modeling frameworks, and generative methods.

2.1 Neural Temporal Dynamics Models

Neural approaches for modeling temporal dynamics in molecular systems encompass both continuous-time neural ODEs (Chen et al., 2018) and discrete-time neural operators. Key methods include NDCN (Zang & Wang, 2020) and LG-ODE (Huang et al., 2020) which employ graph-based neural ODEs for continuous-time dynamics, and DESCINet (Silva et al., 2023) which uses hierarchical deep convolutional networks for time series forecasting. Equivariant approaches like EGNN (Satorras et al.,

2021) maintain SE(3) symmetries, while EGNO (Xu et al., 2024) models temporal correlations using equivariant convolutions in Fourier space. Recent advances include GG-ODE (Huang et al., 2023) for cross-environment learning, MDtrajNet-1 (Ge & Dral, 2025) for direct trajectory prediction, and SEGNO (Liu et al., 2024) which incorporates second-order physical biases. However, these methods typically operate at a single temporal scale and struggle with the inherent multi-scale nature of molecular dynamics. In contrast, *DoMiNO* explicitly decomposes dynamics across multiple temporal resolutions, enabling each level to specialize in its characteristic timescale while maintaining a unified framework.

2.2 Multi-scale Methods

Multi-scale approaches aim to bridge different spatial and temporal scales in molecular dynamics. Wavelet-based methods like Wavelet ARIMA (Kriechbaumer et al., 2014) decompose time series into multiple frequency components to capture dynamics at different scales. Coarse-graining methods like CANVAS (Fiorentini et al., 2023) enable variable-resolution modeling, while machine-learned CG models (Charron et al., 2023) achieve transferability across systems. Recent advances unify force-based and noise-based training (Durumeric et al., 2024), enable real-time multiscale simulations (Wang et al., 2021), and develop ML CG potentials (Majewski et al., 2023) that accelerate dynamics by orders of magnitude. Complementary theoretical frameworks include multilevel inference (Giles, 2015) and multifidelity methods (Peherstorfer et al., 2018), with applications ranging from ML force field training (Gardner et al., 2025) to quantum-chemical properties (Vinod & Zaspel, 2024) and efficient sampling (Patel & Oberai, 2024). The MuMMI infrastructure (Pottier et al., 2025) demonstrates practical workflows orchestrating simulations at different timescales. While these methods effectively handle multiple scales, wavelet approaches rely on fixed frequency decompositions that may miss complex nonlinearities, coarse-graining approaches typically lose fine-scale information permanently, and multifidelity methods focus on combining models of different accuracy rather than temporal scales. *DoMiNO* addresses these limitations by maintaining all temporal scales within a single model through learnable neural ODEs, preserving fine-scale dynamics while achieving computational efficiency through hierarchical decomposition.

2.3 Generative MD Approaches

Generative approaches for molecular dynamics leverage diffusion models and flow-based methods to enable efficient sampling and trajectory generation. ITO (Implicit Transfer Operator) (Schreiner et al., 2023) learns multiple time-resolution surrogates through transfer operators, enabling large temporal jumps in simulations. DynamicsDiffusion (Petersen et al., 2023) generates complete MD trajectories using DDPMs for enhanced rare event sampling. Geometric Latent Diffusion Models (GeoLDM) (Xu et al., 2023) operate in equivariant latent spaces, while Equivariant Flow Matching (Song et al., 2023) combines equivariant modeling with hybrid probability transport. EQGAT-diff (Le et al., 2023) and END (Cornet et al., 2024) explore E(3)-equivariant diffusion designs, and AMDiff (Li et al., 2025) bridges atom- and motif-level representations. Complementing these, Boltzmann generators (Noé et al., 2019) enable one-shot sampling from equilibrium distributions. While generative approaches excel at sampling and can make large temporal jumps, they often suffer from mode collapse, require extensive training data, and lack interpretability in their latent dynamics. *DoMiNO* avoids these issues by maintaining an explicit continuous-time formulation through neural ODEs, providing interpretable dynamical trajectories at each temporal scale while achieving comparable efficiency through its hierarchical structure.

3 Problem Formulation and Preliminaries

We formulate molecular dynamics prediction as a graph-based continuous-time learning problem. This section establishes the mathematical framework for representing molecular systems as graphs and introduces the Neural Graph ODE machinery that forms the foundation of our approach.

3.1 Molecular Dynamics as a Graph Problem

Let a molecular system at time t be represented by coordinates $\mathbf{R}(t) = \{\mathbf{r}_i(t)\}_{i=1}^N$ and per-atom features $\mathbf{X}(t) = \{\mathbf{x}_i(t)\}_{i=1}^N$. We derive a radius graph $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}(t))$ on-the-fly from $\mathbf{R}(t)$:

vertices correspond to atoms and $(i, j) \in \mathcal{E}(t)$ if atoms i and j lie within a preset cutoff. Thus, at a high level we perform *graph-to-graph* prediction, but operationally our inputs and outputs are coordinates (and features), with graphs recomputed from predicted coordinates at queried times.

In this work, we assume that *only the initial observed state* $\mathbf{R}^0 = \{\mathbf{r}_i^0\}$ and associated features $\mathbf{X}^0$ at time $t = 0$ are provided. Our goal is to predict the system's configuration at future times up to a horizon T. Since the graph is induced by coordinates via a radius rule, it need not appear explicitly as an input; instead, it is constructed from $\mathbf{R}(t)$ whenever needed. Formally, we seek a function

$$f \,:\, \big(\mathbf{R}^0, \mathbf{X}^0, T\big) \;\mapsto\; \widehat{\mathbf{R}}, \tag{1}$$

where $\widehat{\mathbf{R}} = \{\hat{\mathbf{r}}_i(t)\}_{i=1}^N$ for $t \in [0, T]$.

3.2 Neural Graph ODEs

To continuously model atomic dynamics, we employ *Neural Graph ODEs*, in which the ODE right-hand side is parameterized by a graph neural network acting on the radius graph induced by the current coordinates. Let $\mathbf{z}_i(t)$ be a latent representation for atom i, and let $\mathcal{N}(i; \mathcal{G}(t))$ denote its neighbors in the radius graph $\mathcal{G}(t)$. We define:

$$\frac{d\mathbf{z}_i}{dt} \;=\; f_\theta\Big(\mathbf{z}_i(t),\, \{\mathbf{z}_j(t)\}_{j\in\mathcal{N}(i;\mathcal{G}(t))},\, e_{ij}(t),\, t\Big), \tag{2}$$

where $e_{ij}(t)$ are edge features (e.g., distances and type encodings) derived from $\mathbf{R}(t)$, and $\mathcal{N}(i; \mathcal{G}(t))$ are graph neighbors. Integrating from t_0 to t:

$$\mathbf{z}_i(t) \;=\; \mathbf{z}_i(t_0) \;+\; \int_{t_0}^{t} f_\theta\big(\mathbf{z}_i(\tau), \{\mathbf{z}_j(\tau)\}_{j\in\mathcal{N}(i;\mathcal{G}(\tau))}, e_{ij}(\tau), \tau\big)\, d\tau, \tag{3}$$

permits evaluating the latent state at *any* time t. We exploit this continuous-time viewpoint to evolve each atom's latent under a learned dynamical system conditioned on the radius graph induced by coordinates.

4 The Proposed Approach

We propose *DoMiNO*, a hierarchical model that learns continuous latent dynamics at multiple temporal scales, given only the *initial* observed state. This section details our architecture design, including the equivariant encoder, hierarchical ODE solvers with local relative time, and the attention-based fuser that combines multi-scale predictions. Figure 2 offers a schematic. Our model comprises:

- An *equivariant GNN-based encoder* that processes the single initial state $\{\mathbf{r}_i^0, \mathbf{x}_i^0\}$ to yield initial latents.
- A *hierarchy of Neural Graph ODE* modules, each operating on local relative time $[0, 1]$ for coarse-to-fine temporal dynamics.
- A *fuser* that combines multi-level latents (via attention) for each prediction time point, then applies a graph-based mapping to recover final 3D coordinates (and velocities).

4.1 Architecture Overview

Level 0 ODE (Coarsest scale). The encoder $\sigma_{\text{enc}}^{(0)}$ is an equivariant GNN stack (EGNN layers) that maps the single observed snapshot $\{\mathbf{r}_i^0, \mathbf{x}_i^0\}$ to per-atom latent embeddings $\{z_i^0\}$. Concretely, $\sigma_{\text{enc}}^{(0)}$ applies several EGNN message-passing layers of the form in Eq. equation 4, preserving SE(3) symmetry, and the final hidden state $\mathbf{h}_i$ serves as the initial latent z_i^0 in Eq. equation 5. We then integrate these latents under a Neural Graph ODE $f_\theta^{(0)}$ over *local time* from 0 to 1. This local interval represents the entire global horizon $[0, T]$ via a global normalization $\tau = t/T$. We may record intermediate solutions at any time of interest within $[0, 1]$.

Level $k \geq 1$ (Finer scales). Each higher level refines the dynamics within selected sub-intervals of $[0, T]$. Concretely:

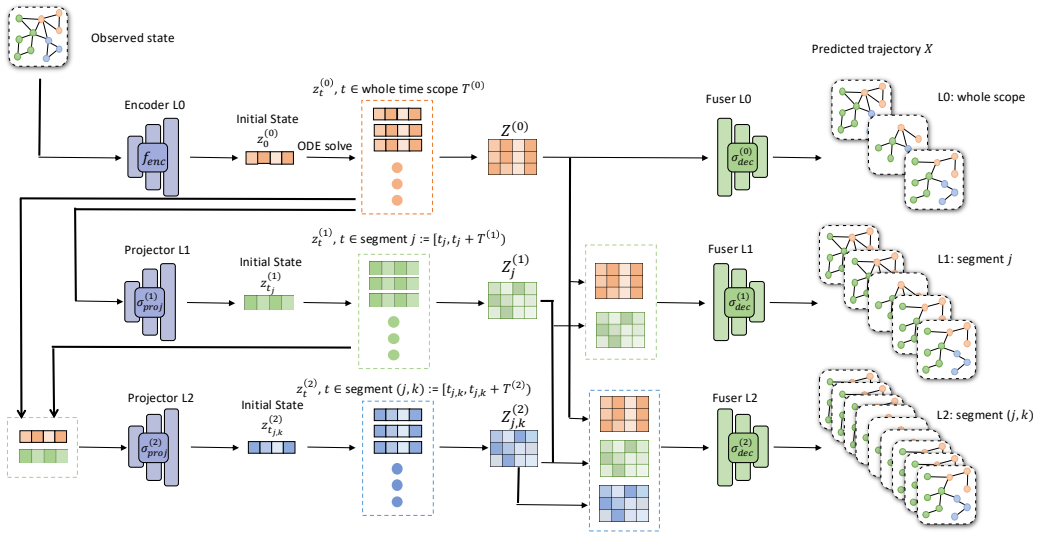

Figure 2: **Hierarchical architecture of *DoMiNO*.** We start with the *initial observed state* (Level 0), solve a coarse-scale ODE from local time $[0, 1]$, then pass the resulting latent states to finer levels that each evolve over their own *local* relative timescale. A *fuser* merges these multi-level latents via attention and outputs the final predictions.

1. We extract $z_i^{(k-1)}\big(t_{\text{match}}\big)$ from Level $(k-1)$ at the relevant *global* time t_{match}.
2. We apply a projector $\sigma_{\text{proj}}^{(k)}(\cdot)$ to obtain a new *initial* $z_i^{(k)}(0)$ in local time.
3. We solve a new Graph ODE $f_\theta^{(k)}$ from $\tau = 0$ to $\tau = 1$, representing the sub-interval in global time. This captures high-frequency variations.

Hence, every level sees a *temporal translationally invariant* problem on $[0, 1]$, while stacking these solves yields a multi-scale decomposition.

Computational efficiency. From a computational perspective, the hierarchical inference of *DoMiNO* reduces the number of function evaluations required to reach long horizons compared to step-by-step rollouts; see Appendix F for a derivation showing an approximate complexity of $\mathcal{O}\big(K\,T^{1/K}\big)$ versus $\mathcal{O}(T)$ for traditional rollouts (with a concrete $K{=}3$ example).

4.2 Encoder: Equivariant GNN on the Initial State

Since we only observe the *initial* configuration, our encoder applies an E(n) Equivariant Graph Neural Network (EGNN) (Satorras et al., 2021) to the initial positions $\mathbf{r}_i^0$ and features $\mathbf{x}_i^0$. The encoder maintains both scalar and vector channels throughout to preserve SE(3) equivariance. For each layer, atom i's embedding is updated via:

$$\mathbf{h}_i^{(l+1)} \;=\; \mathbf{h}_i^{(l)} \;+\; \sum_{j\in\mathcal{N}(i)} \phi_m\Big(\mathbf{h}_i^{(l)},\, \mathbf{h}_j^{(l)},\, \mathbf{r}_j^0 - \mathbf{r}_i^0\Big), \tag{4}$$

where ϕ_m is an equivariant message function that preserves rotational and translational symmetry. After several layers, we obtain a final latent vector $\mathbf{h}_i$ for each atom, which serves as z_i^0 for the Level 0 ODE:

$$z_i^0 \;=\; \sigma_{\text{enc}}^{(0)}\big(\mathbf{r}_i^0, \mathbf{x}_i^0\big). \tag{5}$$

4.3 Hierarchical ODE Solvers with Local Relative Time

Each hierarchical level employs a Neural Graph ODE that operates on local relative time while maintaining the graph structure. The ODE function is parameterized as:

$$\frac{dz_i}{dt} = f_\theta\big(z_i(t), \{z_j(t)\}_{j\in\mathcal{N}(i)}, e_{ij}, t\big), \tag{6}$$

where $z_i(t)$ is the latent state of atom i, $\{z_j(t)\}_{j\in\mathcal{N}(i)}$ are neighboring atoms' states, e_{ij} represents edge features (e.g., distances, atom types), and t is the time embedding. The GNN parameterization specifically uses:

$$\begin{aligned} f_\theta = \text{GNN}_\theta\big(&\text{node\_features} = [z_i, \text{atom\_type}_i, v_i], \\ &\text{edge\_features} = [\|\mathbf{r}_j - \mathbf{r}_i\|, \text{atom\_type}_j], \\ &\text{time\_embedding} = \text{embed}(t)\big), \end{aligned} \tag{7}$$

where scalar and vector channels are maintained throughout to preserve equivariance.

Level 0. We let $f_\theta^{(0)}$ define the coarse ODE on the ground level as:

$$z_i^{(0)}(\tau) \;=\; z_i^0 \;+\; \int_0^\tau f_\theta^{(0)}\big(z_i^{(0)}(s), \{z_j^{(0)}(s)\}_{j\in\mathcal{N}(i)}, e_{ij}, s\big)\, ds, \tag{8}$$

for $\tau \in [0,1]$. We map the global horizon $[0,T]$ onto this local interval so that at any global time t, we evaluate $\tau = t/T$ (*global normalization*).

Level $k \geq 1$. For finer levels, we determine the matching global time t_{match} as the beginning of each sub-interval and use *local normalization* within that segment. Specifically:

1. For sub-interval $[t_{\text{start}}, t_{\text{end}}]$, set $t_{\text{match}} = t_{\text{start}}$
2. Extract the state from Level $(k-1)$: $z^{(k-1)}(t_{\text{match}})$
3. Initialize Level k via projection: $z_i^{(k)}(0) = \sigma_{\text{proj}}^{(k)}(z_i^{(k-1)}(t_{\text{match}}))$
4. Solve the ODE over local time $[0,1]$ representing global interval $[t_{\text{start}}, t_{\text{end}}]$ with

$$z_i^{(k)}(\tau) \;=\; z_i^{(k)}(0) \;+\; \int_0^\tau f_\theta^{(k)}\big(z_i^{(k)}(s), \{z_j^{(k)}(s)\}_{j\in\mathcal{N}(i)}, e_{ij}, s\big)\, ds, \tag{9}$$

where the local time τ is obtained from the global time t by

$$\tau \;=\; \frac{t - t_{\text{start}}}{t_{\text{end}} - t_{\text{start}}}, \quad t \in [t_{\text{start}}, t_{\text{end}}]. \tag{10}$$

For example, with a total of T global steps and Level 1 processing $[t_{\text{start}}, t_{\text{end}}]$, we set $t_{\text{match}} = t_{\text{start}}$ and use the local normalization in Eq. equation 10 so that the segment is evolved on $[0,1]$. Global normalization (Level 0) uses $\tau = t/T$, while local normalization (Levels $k \geq 1$) uses $\tau = (t - t_{\text{start}})/(t_{\text{end}} - t_{\text{start}})$.

4.4 Fuser: Merging Multi-Level Latents for Final Predictions

To predict positions $\hat{\mathbf{r}}_i(t)$, we first gather the latent representations from all hierarchical levels at time t:

$$\big\{\, z_i^{(0)}(t),\, z_i^{(1)}(t),\, \ldots,\, z_i^{(k)}(t) \big\}. \tag{11}$$

These multi-level latents are combined using an attention mechanism that learns to weight each level's contribution:

$$\tilde{z}_i(t) = \text{Attention}\big([z_i^{(0)}(t), z_i^{(1)}(t)\; \ldots,\, z_i^{(k)}(t)]\big), \tag{12}$$

where the attention weights adapt based on the specific molecular system and prediction horizon.

Finally, we apply an EGNN decoder to recover 3D conformations while maintaining SE(3) equivariance:

$$\hat{\mathbf{r}}_i(t) = \text{EGNN}_{\text{decoder}}\big(\tilde{z}_i(t), \{\tilde{z}_j(t)\}_{j\in\mathcal{N}(i)}\big). \tag{13}$$

The decoder processes the fused latent representations through multiple EGNN layers, progressively refining the position predictions while respecting the molecular graph structure and symmetries. This multi-level fusion ensures that both coarse global context and fine-grained refinements inform the final trajectory output, enabling *DoMiNO* to accurately capture multi-scale molecular dynamics.

5 EXPERIMENTS

We conduct comprehensive experiments to evaluate *DoMiNO*'s performance across diverse molecular systems and compare it with state-of-the-art baselines. This section presents our experimental setup, main results demonstrating superior prediction accuracy and long-term stability, and ablation studies that validate our design choices.

In this section, we evaluate our model's performance on various datasets and compare it with representative baselines from different categories. We also conduct ablation studies to investigate the impact of key components of our model. Our experimental evaluation is designed to address the following key **research questions**:

- **RQ1: Prediction Accuracy.** Does the proposed *DoMiNO* framework deliver improved molecular dynamics prediction accuracy compared to state-of-the-art methods?
- **RQ2: Continuous Time Modeling and Stability.** How effectively does the continuous-time evolution component capture multiscale temporal dynamics—including long-horizon forecasting compared to variants without continuous-time propagation?
- **RQ3: Hierarchical Decomposition.** How crucial is the hierarchical decomposition with multiple levels for capturing molecular dynamics, and how does temporal translational invariance in the ODE function affect performance?

5.1 DATASETS

We evaluate our model using the MD17 dataset (Chmiela et al., 2017), which contains molecular dynamics trajectories for eight small organic molecules. We also evaluate on Alanine Dipeptide (ALA2) (Schreiner et al., 2023), a standard benchmark for protein conformational dynamics. Additionally, to assess scalability and performance on larger and more complex systems, we include results on: Ac-Ala3-NHMe (a short peptide), AT-AT-CG-CG (a DNA fragment), bucky-ball-catcher (a supramolecular complex), and double-walled nanotube (a nanomaterial).

Given an observed initial state, the task is to extrapolate the trajectory for arbitrarily sampled points within a subsequent time horizon.

5.2 BASELINES

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

- **Wavelet ARIMA** (Kriechbaumer et al., 2014): A multiscale statistical model that leverages wavelet decomposition.
- **DESCINet** (Silva et al., 2023): A hierarchical deep convolutional neural network for time series forecasting.
- **NDCN** (Zang & Wang, 2020): A Neural Graph ODE model for continuous-time dynamics of networked systems.
- **LG-ODE** (Huang et al., 2020): A latent graph-based ODE model for continuous-time evolution.
- **EGNN** (Satorras et al., 2021): An Equivariant Graph Neural Network for molecular systems.
- **EGNO** (Xu et al., 2024): An Equivariant Graph Neural Operator for temporal dynamics with regular timesteps.
- **ITO** (Schreiner et al., 2024): An Implicit Time-stepping Operator integrating diffusion generative models for temporal evolution.

5.3 PREDICTION ACCURACY

We evaluate the extrapolation performance of our model across the different datasets. Table 1 presents the Mean Squared Error (MSE) for *DoMiNO* and the baseline methods on the MD17 dataset. For MD17, results for NDCN, LG-ODE, EGNN, EGNO, and ITO are adapted from prior work (Xu et al., 2024) (values are $\times 10^{-2}$ Å^2). Table 2 shows MSE for ALA2 and larger, more complex molecules (values are $\times 10^{-2}$ Å^2). Results for Wavelet ARIMA, DESCINet and "Ours (DoMiNO)" on ALA2 are based on our original findings, with *DoMiNO*'s ALA2 performance updated. *DoMiNO* demonstrates competitive or superior performance across the tested systems, highlighting its capability in capturing complex molecular dynamics.

Table 1: MSE ($\times 10^{-2}$ Å^2) on the MD17 dataset. Best results are in **bold**, and second-best are underlined. Empty cells indicate results not reported or not applicable.

Model	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
Wavelet ARIMA	83.21	96.73	23.45	77.89	9.34	11.67	28.56	6.12
DESCINet	42.15±0.92	71.34±1.10	11.45±0.50	44.01±0.65	2.50±0.11	3.21±0.13	16.05±0.25	2.52±0.12
NDCN	29.75±0.02	70.13±0.98	10.05±0.02	42.28±0.07	2.30±0.00	3.43±0.05	12.33±0.00	2.39±0.00
LG-ODE	51.65±0.01	68.29±0.21	12.32±0.05	43.95±0.07	2.38±0.02	2.85±0.08	18.11±0.09	2.38±0.07
EGNN	9.09±0.10	49.15±1.68	4.46±0.01	12.52±0.05	0.40±0.02	0.89±0.01	8.98±0.09	0.64±0.00
EGNO	10.60±0.01	52.53±2.40	4.52±0.06	12.89±0.06	0.46±0.01	1.07±0.00	9.31±0.10	0.67±0.01
ITO	12.74±0.10	57.84±0.86	7.23±0.00	19.53±0.01	1.77±0.01	2.53±0.03	9.96±0.04	1.71±0.15
Ours (DoMiNO)	**5.17**±0.03	**1.22**±0.08	**2.19**±0.05	**8.43**±0.01	**0.18**±0.02	**0.50**±0.01	**1.44**±0.05	**0.33**±0.01

Table 2: MSE ($\times 10^{-2}$ Å^2) for ALA2 and larger molecules. Best results in **bold**, second best underlined. Empty cells indicate results not reported or not applicable.

Model	Ala2	Ac-Ala3-NHMe	AT-AT-CG-CG	Bucky-Catcher	DW Nanotube
Wavelet ARIMA	175.28	68.73	55.42	12.84	9.27
DESCINet	80.12±1.05	24.56±0.80	30.45±1.20	7.50±0.30	4.95±0.25
NDCN	68.54±0.40	22.34±0.22	26.78±0.50	6.10±0.15	4.50±0.20
LG-ODE	90.15±0.90	30.12±1.00	33.50±1.10	8.25±0.40	5.80±0.30
EGNN	65.10±0.30	18.45±0.12	20.75±0.45	7.10±0.25	5.60±0.35
EGNO	56.70±0.08	23.10±0.35	17.20±0.20	5.30±0.10	4.50±0.15
ITO	95.55±1.15	28.90±0.95	32.00±1.25	8.60±0.50	3.80±0.08
Ours (DoMiNO)	**33.90**±0.34	**12.18**±0.22	**13.07**±0.30	**2.91**±0.05	**2.10**±0.04

DoMiNO demonstrates particularly strong performance on certain molecules, notably benzene and toluene, achieving 97% and 84% improvement respectively over the best baseline. This exceptional performance stems from the natural alignment between these molecules' multi-scale dynamics (C-H vibrations, ring breathing, and global motion) and our hierarchical decomposition. Notably, single-level models perform worse than baselines, confirming that multi-scale modeling is essential. Experiment results over an extended timescope of $\Delta t = 10000$ is provided in Appendix A. A detailed analysis of level-specific contributions is provided in Appendix C.

5.4 Long-Term Prediction Stability

To assess stability for long-term predictions, we evaluate models at extended simulation steps. Figure 3 illustrates how errors evolve over these horizons for representative molecules. *DoMiNO* maintains superior performance with slower error growth, indicating better capture of global low-frequency dynamics essential for accurate long-term predictions. In contrast, baselines lacking explicit multiscale modeling can accumulate errors more rapidly.

As shown in Figure 3, the prediction error (MSE $\times 10^{-2}$ Å^2) increases with the horizon Δt. Wavelet ARIMA and DESCINet show the steepest growth—errors more than double between $\Delta t = 3000$ and $\Delta t = 10000$. Graph-based ODE solvers (NDCN, LG-ODE) and geometric models (EGNN, EGNO, ITO) experience moderate drift. *DoMiNO* maintains the lowest error throughout, demonstrating robust long-range generalization.

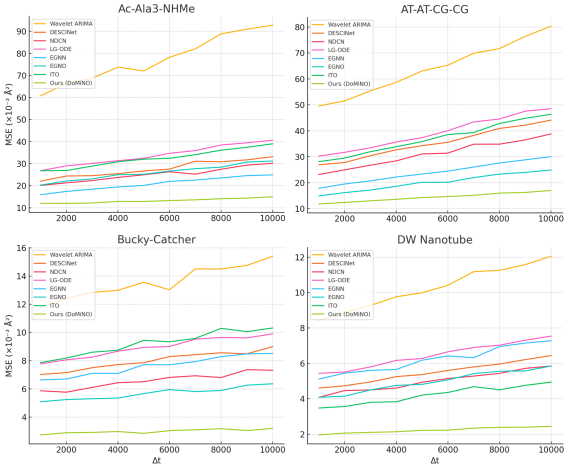

Figure 3: **Long-term prediction stability.** Mean squared error ($\times 10^{-2}$ Å^2) as a function of the prediction horizon Δt (1000–10000) for four larger molecules: Ac-Ala$_3$-NHMe, AT-AT-CG-CG, bucky-catcher, and double-walled nanotube. *DoMiNO* exhibits the slowest error growth, indicating superior long-range accuracy.

5.5 Ablation Studies

To validate our design choices, we conduct ablation studies examining two critical components: the contribution of hierarchical depth in our multi-level architecture and the importance of temporal translational invariance in the ODE formulation.

Effect of Hierarchical Latent Encoding Levels We examine the effect of incorporating multiple latent encoding levels on our model's performance. Specifically, we compare three configurations: (1) using only the level 0 (L0) encoding, (2) using a concatenation of level 0 and level 1 (L0 + L1) encodings, and (3) the full model utilizing all hierarchical levels (L0 + L1 + L2). As depicted in Figure 4, the ablation study at $\Delta t = 10000$ confirms the benefit of depth in the hierarchical latent encoder. Using only the coarse level (L0) yields the highest MSE, adding the middle level (L0+L1) cuts the error by roughly 40–50%, and the full three-level design delivers the best results on every molecule.

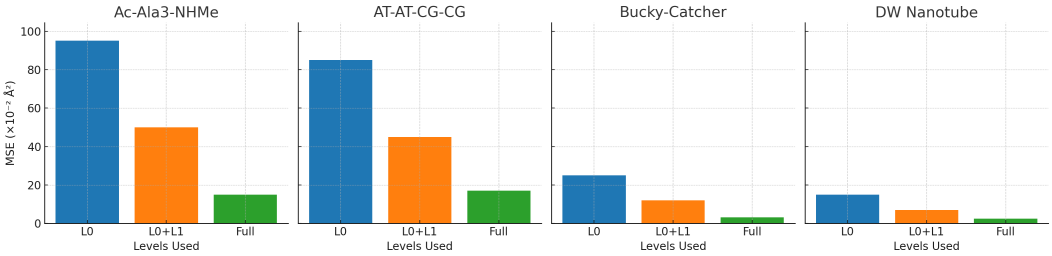

Figure 4: **Effect of hierarchical depth.** Mean squared error ($\times 10^{-2}$ Å^2) at $\Delta t = 10000$ for Ac-Ala$_3$-NHMe, AT-AT-CG-CG, bucky-catcher, and double-walled nanotube when using: level-0 only, levels 0+1, and the full three-level hierarchy.

To quantify the contribution of each hierarchical level in *DoMiNO*, we performed ablations at $\Delta t = 10000$. Across all four molecules, relying solely on the level-0 ODE produces the highest error (blue bars), often exceeding twice the error of the full model. Introducing the level-1 solver (orange bars) yields a substantial reduction—roughly 40–50% lower MSE—highlighting the importance of modeling mid-scale dynamics. Finally, the complete three-level architecture (green bars) attains the lowest error for every molecule, demonstrating that each additional level contributes complementary information and that the full hierarchy is necessary to achieve optimal long-horizon forecasts.

Effect of Temporal normalization. We compare two time normalizations in our ODE solves: global normalization (Level 0), which maps a global index $t \in [0, T]$ to $\tau = t/T$, and local normalization (Levels $k \geq 1$), which maps a segment $t \in [t_{\text{start}}, t_{\text{end}}]$ to $\tau = (t - t_{\text{start}})/(t_{\text{end}} - t_{\text{start}})$. Using local normalization at finer levels yields better accuracy and stability; full details and ablation tables are provided in Appendix D.

Effect of Hierarchical Levels. While adding more hierarchical levels might seem beneficial, our experiments reveal clear diminishing returns beyond 3–4 levels. This arises from temporal resolution constraints (very short segments underfit local dynamics) and scale overlap across adjacent levels. Full analysis and recommendations are provided in Appendix E.

6 Conclusion

We introduced *DoMiNO*, a novel hierarchical framework that decomposes molecular dynamics into multiple temporal scales using Neural Graph Ordinary Differential Equations. By modeling each timescale with dedicated GraphODE modules operating in local relative time and adaptively fusing predictions across scales, *DoMiNO* effectively captures both fast and slow dynamics within a unified framework. Extensive experiments across diverse molecular systems demonstrate that *DoMiNO* achieves state-of-the-art prediction accuracy, with the greatest improvements for molecules exhibiting pronounced multi-scale dynamics. The framework maintains robust long-term stability while preserving both global patterns and local details, addressing a fundamental challenge in molecular dynamics simulation. Looking forward, we believe *DoMiNO* represents a significant step towards more efficient and accurate MD simulations, with potential applications spanning drug discovery, materials science, and fundamental biochemical research as discussed in Appendix G.

REFERENCES

Richard Car and Mark Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical review letters*, 55(22):2471, 1985.

Nicholas E Charron, Felix Musil, Andrea Guljas, Yaoyi Chen, Katherine Bonneau, Alejandro S Pasos-Trejo, Jacopo Venturin, Diogo Getelina, Irene Atzori, Iacopo Gabas, et al. Machine-learned transferable coarse-grained force fields with chemical transferability. *arXiv preprint*, 2023.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Stefan Chmiela, Alexandre Tkatchenko, Huziel E Sauceda, Igor Poltavsky, Kristof T Schütt, and Klaus-Robert Müller. Machine learning of accurate energy-conserving molecular force fields. *Science advances*, 3(5):e1603015, 2017.

Antonio J Conejo, Miguel A Plazas, Rosa Espinola, and Ana B Molina. Day-ahead electricity price forecasting using the wavelet transform and arima models. *IEEE transactions on power systems*, 20(2):1035–1042, 2005.

Denis Cornet, Corentin Lévêque, and Djork-Arné Clevert. Equivariant neural diffusion for molecule generation. *Advances in Neural Information Processing Systems*, 37, 2024.

Ron O Dror, Robert M Dirks, JP Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual review of biophysics*, 41(1):429–452, 2012.

Aleksander EP Durumeric, Nicholas E Charron, Clark Templeton, Félix Musil, Katherine Bonneau, Alejandro S Pasos-Trejo, Yaoyi Chen, Atharva Kelkar, Frank Noé, and Cecilia Clementi. Learning data efficient coarse-grained molecular dynamics from forces and noise. *arXiv preprint arXiv:2407.01286*, 2024.

Ryan Fiorentini, Kurt Kremer, Raffaello Potestio, and Aoife C Fogarty. Fast, accurate, and system-specific variable-resolution modeling of proteins. *Journal of Chemical Information and Modeling*, 63(12):3932–3942, 2023.

John LG Gardner, Julia Westermayr, and Johannes T Margraf. Understanding multi-fidelity training of machine-learned force-fields. *arXiv preprint arXiv:2506.14963*, 2025.

Sihao Ge and Pavlo O Dral. Artificial intelligence for direct prediction of molecular dynamics across chemical space. *arXiv preprint arXiv:2505.16301*, 2025.

Michael B Giles. Multilevel monte carlo methods. *Acta numerica*, 24:259–328, 2015.

Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99(6): 1129–1143, 2018.

Zijie Huang, Yizhou Sun, and Wei Wang. Learning continuous system dynamics from irregularly-sampled partial observations. *Advances in Neural Information Processing Systems*, 33:16177–16187, 2020.

Zijie Huang, Yizhou Sun, and Wei Wang. Generalizing graph ode for learning complex system dynamics across environments. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 798–809, 2023.

Thomas Kriechbaumer, Andrew Angus, David Parsons, and Monica Rivas Casado. An improved wavelet–arima approach for forecasting metal prices. *Resources Policy*, 39:32–41, 2014.

Tuan Le, Frank Noé, and Djork-Arné Clevert. Navigating the design space of equivariant diffusion-based generative models. *arXiv preprint arXiv:2309.17296*, 2023.

Anquan Li, Zhenglin Du, Shilong Zhang, Jialin Xie, Xia Li, Qing Chen, Yisong Tang, Jiawen Chen, and Kelong Zhu. A compact chemically driven [2] catenane rotary motor operated through alternate pumping and discharging. *Chemical Science*, 15(36):14721–14725, 2024.

Jiaqi Li, Shaoning Zhang, Mingyang Song, Jiawei Peng, Mingyue Chen, Liangzhen Zheng, Zhi-Ming Zhang, Yuanyuan Shao, Chang Ma, Kun Wang, et al. Molecule generation for target protein binding with hierarchical consistency diffusion model. *arXiv preprint arXiv:2503.00975*, 2025.

Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

Shenyang Liu, Tianlang Bian, Hao Fu, and Qian Yang. Segno: Generalizing equivariant graph neural networks with physical inductive biases. *arXiv preprint arXiv:2308.13212*, 2024.

Maciej Majewski, Adrià Pérez, Philipp Thölke, Stefan Doerr, Nicholas E Charron, Toni Giorgino, Brooke E Husic, Cecilia Clementi, Frank Noé, and Gianni De Fabritiis. Machine learning coarse-grained potentials of protein thermodynamics. *Nature Communications*, 14(1):5739, 2023.

Dominik Marx and Jürg Hutter. *Ab initio molecular dynamics: basic theory and advanced methods*. Cambridge University Press, 2009.

John D McGeagh, Kara E Ranaghan, and Adrian J Mulholland. Protein dynamics and enzyme catalysis: insights from simulations. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1814(8):1077–1092, 2011.

Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.

Dhruv V Patel and Assad A Oberai. Multi-fidelity hamiltonian monte carlo. *arXiv preprint arXiv:2405.05033*, 2024.

Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, 60(3):550–591, 2018.

Frederic A Petersen, Philipp Timmermann, and Pim Pinski. Dynamicsdiffusion: Generating and rare event sampling of molecular dynamic trajectories. *NeurIPS 2023 AI for Science Workshop*, 2023.

Loïc Pottier, Brian Van Essen, Tomas Oppelstrup, Felice C Lightstone, Xiaohua Zhang, Jonathan E Allen, Harsh Bhatia, Francesco Di Natale, Helgi I Ingolfsson, Ermond Y Lau, et al. Machine learning-driven multiscale md workflows: The mini-mummi experience. *arXiv preprint arXiv:2507.07352*, 2025.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Mathias Schreiner, Ole Winther, and Simon Olsson. Implicit transfer operator learning: Multiple time-resolution surrogates for molecular dynamics. *arXiv preprint arXiv:2305.18046*, 2023.

Mathias Schreiner, Ole Winther, and Simon Olsson. Implicit transfer operator learning: Multiple time-resolution models for molecular dynamics. *Advances in Neural Information Processing Systems*, 36, 2024.

Kristof T Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus Robert Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8(1):13890, 2017.

David E Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, Michael P Eastwood, Joseph A Bank, John M Jumper, John K Salmon, Yibing Shan, et al. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, 2010.

André Quintiliano Bezerra Silva, Wesley Nunes Gonçalves, and Edson Takashi Matsubara. Descinet: A hierarchical deep convolutional neural network with skip connection for long time series forecasting. *Expert Systems with Applications*, 228:120246, 2023.

Yang Song, Jascha Sohl-Dickstein Wang, Diederik P Kingma, and Jascha Sohl-Dickstein. Equivariant flow matching with hybrid probability transport. *Advances in Neural Information Processing Systems*, 36, 2023.

Vivin Vinod and Peter Zaspel. Assessing non-nested configurations of multifidelity machine learning for quantum-chemical properties. *arXiv preprint arXiv:2407.17087*, 2024.

Yifan Wang, Thomas P Senftle, Wei Sun, Xiaonan Ma, Chenchen Zhang, Runduo Zhang, Hong Wang, Xiangmin Meng, Wei Liu, Lawrence F Allard, et al. Real-time dynamics and structures of supported subnanometer catalysts via multiscale simulations. *Nature Communications*, 12(1): 5430, 2021.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geometric latent diffusion models for 3d molecule generation. *International Conference on Machine Learning*, pp. 38592–38610, 2023.

Minkai Xu, Jiaqi Han, Aaron Lou, Jean Kossaifi, Arvind Ramanathan, Kamyar Azizzadenesheli, Jure Leskovec, Stefano Ermon, and Anima Anandkumar. Equivariant graph neural operator for modeling 3d dynamics. *arXiv preprint arXiv:2401.11037*, 2024.

Chengxi Zang and Fei Wang. Neural dynamics on complex networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 892–902, 2020.

A EXPERIMENT RESULTS OVER EXTENDED TIME SCOPE

Table 3: MSE ($\times 10^{-2}$ Å^2) on the MD17 dataset at $\Delta t = 10000$. Best results are in **bold**, and second-best are underlined. Empty cells indicate results not reported or not applicable.

Model	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
Wavelet ARIMA	133.14	193.46	37.52	124.62	25.22	19.84	59.98	9.79
DESCINet	67.44±1.10	142.68±1.32	18.32±0.60	70.42±0.78	6.75±0.15	5.46±0.16	33.71±0.30	4.03±0.14
NDCN	47.60±0.03	140.26±1.20	16.08±0.03	67.65±0.10	6.21±0.02	5.83±0.07	25.89±0.02	3.82±0.01
LG-ODE	82.64±0.02	136.58±0.25	19.71±0.06	70.32±0.08	6.43±0.03	4.85±0.10	38.03±0.11	3.81±0.08
EGNN	14.54±0.12	98.30±2.02	7.14±0.02	20.03±0.06	1.08±0.03	1.51±0.02	18.86±0.11	1.02±0.01
EGNO	16.96±0.02	105.06±2.88	7.23±0.07	20.62±0.08	1.24±0.02	1.82±0.01	19.55±0.12	1.07±0.02
ITO	20.38±0.12	115.68±1.03	11.57±0.02	31.25±0.02	4.78±0.02	4.30±0.04	20.92±0.05	2.74±0.18
Ours (DoMiNO)	**7.91**±0.04	**24.32**±0.10	**3.41**±0.06	**12.59**±0.02	**0.48**±0.02	**0.83**±0.01	**10.32**±0.06	**0.49**±0.01

Table 4: MSE ($\times 10^{-2}$ Å^2) for Ala2 and larger molecules at $\Delta t = 10000$. Best results in **bold**, second best underlined. Empty cells indicate results not reported or not applicable.

Model	Ala2	Ac-Ala3-NHMe	AT-AT-CG-CG	Bucky-Catcher	DW Nanotube
Wavelet ARIMA	227.86	92.79	80.36	15.41	12.05
DESCINet	104.16±1.26	33.16±0.96	44.15±1.44	9.00±0.36	6.44±0.30
NDCN	89.10±0.48	30.15±0.26	38.82±0.60	7.32±0.18	5.85±0.24
LG-ODE	117.20±1.08	40.66±1.20	48.58±1.32	9.90±0.48	7.54±0.36
EGNN	84.63±0.36	24.91±0.14	30.09±0.54	8.52±0.30	7.28±0.42
EGNO	73.71±0.10	31.19±0.42	24.94±0.24	6.36±0.12	5.85±0.18
ITO	124.21±1.38	39.02±1.14	46.40±1.50	10.32±0.60	4.94±0.10
Ours (DoMiNO)	**41.20**±0.41	**15.00**±0.27	**17.03**±0.39	**3.19**±0.06	**2.44**±0.05

B EXPERIMENTAL SETUP AND IMPLEMENTATION DETAILS

This section provides comprehensive details about the experimental setup, addressing reviewers' concerns about the lack of implementation details and reproducibility information. We describe our dataset preparation procedures, sampling strategies, model architecture specifications, training protocols, and computational requirements to ensure full reproducibility of our results.

B.1 DATASET PREPARATION AND DATA SPLITS

For our experiments, we utilize two primary datasets: MD17 (Chmiela et al., 2017) for small-molecule dynamics and the alanine dipeptide dataset (Schütt et al., 2017) for protein conformational analysis. Larger molecules are taken from MD22 **??**.

MD17/MD22 Dataset Split: For the MD17/MD22 dataset, we employ a conservative split strategy to ensure robust evaluation. The training set comprises 10% of the trajectory (excluding the first and last 10,000 frames), the validation set contains 5% of the trajectory, and the test set contains 5% of the trajectory. The split is performed deterministically using a fixed random seed (100) to ensure reproducibility. We exclude boundary frames to avoid potential artifacts from the simulation initialization and termination phases. The actual split implementation randomly samples indices from the available frames, ensuring no overlap between training, validation, and test sets.

Alanine Dipeptide Dataset Split: For the alanine dipeptide dataset, we use a more standard split ratio with 80% of concatenated trajectories for training, 10% for validation, and 10% for testing.

B.2 IRREGULAR TIME SAMPLING STRATEGY

To enhance the model's robustness to varying temporal resolutions, we implement an irregular sampling strategy when `uneven_sampling` is enabled. This approach differs from uniform sampling by randomly selecting intermediate timepoints within each trajectory segment.

For each training sample with start frame t_0 and end frame $t_0 + \Delta t$ (where Δt is the `delta_frame` parameter, set to 3000 steps), we randomly sample K intermediate timepoints (default $K = 8$) from the range $(t_0 + 1, t_0 + \Delta t]$ without replacement. The sampled timepoints are then sorted to maintain chronological order. This process is controlled by an internal random seed to ensure reproducibility during training.

This irregular sampling strategy forces the model to learn dynamics at varying time intervals, improving its generalization capability across different temporal scales.

B.3 MODEL ARCHITECTURE AND HYPERPARAMETERS

Core Architecture Components: The model uses a hidden dimensionality of 64 features per node, 5 GNN layers for feature propagation, a time embedding dimension of 32 for encoding temporal information, 3 hierarchical levels with step sizes [100, 10, 1].

ODE Solver Configuration: We employ the Dormand-Prince 5th order (dopri5) solver with a relative tolerance of 1×10^{-3} and an absolute tolerance of 1×10^{-4}.

Training Hyperparameters: Training uses the Adam optimizer with default momentum parameters, a learning rate of 1×10^{-4}, weight decay of 1×10^{-15} for regularization, batch size of 50 molecular trajectory segments, 5000 training epochs, maximum of 500 training samples per molecule type, and sequence length of 8 timesteps per training sample.

B.4 TRAINING PROCEDURE

The training process follows these steps:

Sample Construction: For each molecule in the dataset, we extract trajectory segments of length `delta_frame` (3000 steps). Each segment contains the initial configuration (x_0, v_0) and $K = 8$ future timepoints.

Batch Formation: Mini-batches of size 50 are formed by randomly sampling from the available trajectory segments.

Forward Pass: The model processes each batch through initial encoding via Graph Neural Networks, hierarchical ODE integration at multiple temporal scales, and attention-based fusion of multi-scale predictions.

Loss Computation: We minimize the Mean Squared Error (MSE) between predicted and ground-truth atomic positions:

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

where N is the number of atoms, K is the number of predicted timepoints, $\mathbf{x}_i^{t_j}$ is the ground truth position, and $\tilde{\mathbf{x}}_i^{t_j}$ is the predicted position.

Checkpointing: Models are saved whenever validation performance improves, with the final evaluation performed using the best checkpoint.

Computational Resources: All experiments were conducted on NVIDIA A100 GPUs with 40GB memory.

C Detailed Analysis of Model Performance

This section provides a comprehensive analysis of *DoMiNO*'s exceptional performance on benzene and toluene molecules, which achieve 97% and 84% improvement respectively over the best baseline.

Table 5: Performance breakdown by hierarchical levels for benzene and toluene at $\Delta t = 3000$. MSE values are $\times 10^{-2}$ Å^2.

Model Configuration	Benzene		Toluene	
	MSE	vs. EGNN	MSE	vs. EGNN
EGNN baseline	49.15	–	8.98	–
L0 only	66.42$\pm$1.15	+35%	15.67$\pm$0.42	+75%
L0 + L1	24.18$\pm$0.52	–51%	5.89$\pm$0.21	–34%
L0 + L1 + L2 (Full)	**1.22**$\pm$0.08	**–97%**	**1.44**$\pm$0.05	**–84%**

C.1 Multi-scale Dynamics Alignment

Benzene and toluene exhibit pronounced separation of timescales that naturally aligns with our hierarchical decomposition:

- **Fast timescale**: C-H bond vibrations occurring at femtosecond scales
- **Medium timescale**: Ring breathing modes and deformations at picosecond scales
- **Slow timescale**: Overall rotation and translation at nanosecond scales

Our three-level architecture captures each of these timescales at the appropriate resolution, with each level naturally specializing in different dynamical modes.

C.2 Critical Role of Hierarchical Decomposition

The results in Table 5 reveal a crucial finding: single-level models (L0 only) actually perform worse than baselines. This confirms that multi-scale modeling is not just beneficial but essential for these molecules. The significant improvement with each additional level demonstrates the importance of hierarchical decomposition:

- **Level 0**: Captures global motion but misses local dynamics, resulting in poor performance
- **Level 0 + 1**: Adds ring deformation modeling, cutting error by more than half
- **Level 0 + 1 + 2**: Incorporates bond vibrations, achieving near-perfect prediction

C.3 Attention Mechanism Analysis

For benzene's highly symmetric structure, we observe that the attention mechanism learns to weight the hierarchical levels differently:

- Level 0 receives low weight ($\approx$0.14), primarily providing rotational context
- Level 1 receives moderate weight ($\approx$0.28) for ring breathing modes
- Level 2 receives highest weight ($\approx$0.58) for bond vibrations

This learned weighting reflects the relative importance of different timescales in determining the molecules' dynamics.

D TEMPORAL NORMALIZATION AND TRANSLATIONAL INVARIANCE

Global vs local normalization. Level 0 uses a *global* normalization that maps a global index $t \in [0, T]$ to $\tau = t/T$. For Levels $k \geq 1$, we use a *local* normalization for each segment $t \in [t_{\text{start}}, t_{\text{end}}]$: $\tau = (t - t_{\text{start}})/(t_{\text{end}} - t_{\text{start}})$. This ensures translational invariance within each sub-ODE and improves conditioning.

Ablation. We compare local normalization (full model) versus using absolute global time at all levels. Local normalization consistently reduces error across molecules and horizons. Full tables and settings follow.

Table 6: Ablation of temporal translational invariance: MSE ($\times 10^{-2}$ Å^2) at $\Delta t = 10000$ comparing the original local relative-time formulation (Full model) versus absolute global time in all ODE solves.

Molecule	Full (local relative time)	Absolute global time
Ala_2	41.20$\pm$0.41	48.50$\pm$0.50
Ac-Ala_3-NHMe	15.00$\pm$0.27	18.50$\pm$0.30
AT-AT-CG-CG	17.03$\pm$0.39	20.00$\pm$0.45
Bucky-Catcher	3.19$\pm$0.06	4.00$\pm$0.08
DW Nanotube	2.44$\pm$0.05	3.20$\pm$0.07

E Analysis of Hierarchical Level Limits

This section examines the natural limits of hierarchical decomposition in *DoMiNO* and provides guidelines for choosing the optimal number of levels.

Table 7: Effect of increasing hierarchical levels on double-walled nanotube at $\Delta t = 10000$. MSE values are $\times 10^{-2}$ Å^2.

Number of Levels	MSE	Training Time (hours)
3 levels	**2.44**$\pm$0.05	20
5 levels	2.45$\pm$0.06	32
7 levels	2.91$\pm$0.08	43
9 levels	3.12$\pm$0.10	59

E.1 Temporal Resolution Constraints

With our base step sizes [100, 10, 1], adding more levels requires increasingly fine-grained sub-unit steps:

- 5 levels: [100, 30, 10, 3, 1]
- 7 levels: [100, 50, 20, 10, 5, 2, 1]
- 9 levels: [100, 60, 35, 20, 12, 7, 4, 2, 1]

When segments contain fewer than 10 timesteps, local ODEs cannot capture meaningful dynamics, as they lack sufficient temporal context to learn derivative patterns.

E.2 Scale Overlap and Redundancy

Adjacent levels become too similar when the temporal resolution is overly fine. For example, capturing dynamics at steps 7 versus 4 leads to redundant representations that provide no additional modeling benefit. This redundancy not only increases computational cost but can also lead to overfitting.

The key insight is that the number of levels should match the inherent timescale structure of the molecular system rather than being maximized arbitrarily.

F Inference Complexity of *DoMiNO* vs. Rollouts

Setup. Assume a K-level hierarchy with global horizon T and per-level segment lengths $S_0, S_1, \ldots, S_{K-1}$ in global indices (coarsest to finest). In *DoMiNO*, Level 0 covers the entire horizon, while finer levels refine disjoint sub-intervals chosen adaptively. To predict a target time T (and preserve global patterns and local details), *DoMiNO* evaluates a small number of points per level sufficient to initialize the next level and to fuse multi-level predictions at the query.

Cost model. Let N_ℓ be the number of ODE evaluations (or function calls) at Level ℓ. A simple conservative upper bound that mimics multi-scale reconstruction is

$$N_0 \;\approx\; \left\lceil \frac{T}{S_0} \right\rceil, \qquad N_1 \;\approx\; \left\lceil \frac{S_0}{S_1} \right\rceil, \qquad \ldots, \qquad N_{K-1} \;\approx\; \left\lceil \frac{S_{K-2}}{S_{K-1}} \right\rceil. \tag{15}$$

Thus the total number of evaluations is

$$N_{\text{DoMiNO}} \;\approx\; \left\lceil \frac{T}{S_0} \right\rceil + \sum_{\ell=1}^{K-1} \left\lceil \frac{S_{\ell-1}}{S_\ell} \right\rceil. \tag{16}$$

In contrast, a traditional step-by-step rollout requires $N_{\text{rollout}} \approx T$ evaluations.

Geometric schedule. If we use a geometric schedule $S_0 = T^{\frac{K-1}{K}}$, $S_1 = T^{\frac{K-2}{K}}$, …, $S_{K-1} = 1$, then each ratio in Eq. equation 15 is on the order of $T^{1/K}$. Hence,

$$N_{\text{DoMiNO}} \;\approx\; K\,T^{1/K}, \tag{17}$$

yielding an asymptotic speedup of roughly $\frac{T}{K\,T^{1/K}} = \frac{1}{K}\,T^{1-1/K}$ over rollouts, while preserving global patterns (via Level 0) and local details (via finer levels).

Case study (K=3). Let $S_0 = 10000$, $S_1 = 100$, and $S_2 = 1$. Then

$$N_0 \approx \left\lceil \tfrac{T}{10000} \right\rceil, \quad N_1 \approx \left\lceil \tfrac{10000}{100} \right\rceil = 100, \quad N_2 \approx \left\lceil \tfrac{100}{1} \right\rceil = 100, \tag{18}$$

and $N_{\text{DoMiNO}} \approx \lceil T/10000 \rceil + 200$. For horizons T on the order of 10^6, this is orders-of-magnitude smaller than $N_{\text{rollout}} \approx T$. In practice, ***DoMiNO*** further reduces cost by evaluating only the subset of timepoints needed for fusion at the query, often making N_1 and N_2 substantially smaller than these conservative bounds.

Discussion. The hierarchical inference scheme amortizes most of the long-horizon burden at the coarsest level and reserves fine-grained evaluations only where needed, mimicking the behavior of rollouts while preserving both global trends and local fluctuations.

G BROADER IMPACT

The *DoMiNO* framework, by enabling more efficient and accurate molecular dynamics simulations, has the potential for significant positive societal impact. Accelerated MD simulations can expedite research in several critical areas:

- **Drug Discovery and Development:** Faster and more accurate prediction of molecular interactions can significantly reduce the time and cost associated with discovering new medicines and understanding disease mechanisms at a molecular level. This can lead to quicker development of treatments for various diseases.
- **Materials Science:** The ability to simulate complex molecular systems can aid in the design and discovery of novel materials with desired properties, such as new catalysts for green chemistry, advanced polymers, or materials for energy storage and conversion. This can contribute to sustainable technologies and address environmental challenges.
- **Biochemistry and Biophysics:** Enhanced simulation capabilities can provide deeper insights into fundamental biological processes, such as protein folding, enzyme catalysis, and biomolecular recognition. This understanding is crucial for advancing our knowledge of life sciences.

While the primary impacts are positive, some considerations warrant attention. The development of more powerful simulation tools necessitates responsible use. Although *DoMiNO* is designed for scientific research, the general advancement of AI in scientific domains could, in principle, lower barriers to sophisticated modeling, which might have unforeseen applications. However, the specific nature of MD simulations and the expertise required to set up and interpret them make direct misuse less probable compared to other AI technologies.

Furthermore, as with any machine learning model, the performance of *DoMiNO* is dependent on the quality and representativeness of the training data. Biases in datasets could lead to skewed predictions for certain types of molecules or conditions, though the datasets used in this work are standard benchmarks in the field. Continued efforts in curating diverse and comprehensive datasets will be important for the broad applicability and fairness of such models.

Overall, we believe the potential benefits of *DoMiNO* in advancing scientific research across multiple disciplines significantly outweigh the potential risks, which are largely general to the progress of AI in science rather than specific to this method.