DoMiNO: Down-scaling Molecular Dynamics with Neural Graph Ordinary Differential Equations

Fang Sun, Zijie Huang, Yadi Cao, Xiao Luo, Wei Wang, Yizhou Sun
Computer Science Department, UCLA
fts@cs.ucla.edu

Abstract

Molecular dynamics (MD) simulations are crucial for understanding and predicting the behavior of molecular systems in biology and chemistry. However, their wide adoption is hindered by two main challenges: (1) *computational cost*, because fine-grained simulations often require millions of small timesteps, and (2) *lack of flexibility*, as existing machine-learning-based surrogates typically operate at either a single small or single large timestep. These approaches either accumulate significant roll-out errors or lose the ability to produce fine-grained results if the timestep is large. To address these issues, we propose **DoMiNO**: Down-scaling Molecular Dynamics with Neural Graph Ordinary Differential Equations, a hierarchical framework that models multi-scale dynamics. Specifically, DoMiNO performs down-scaling by progressively up-sampling [1] the trajectory across multiple temporal resolutions, equipping each level with a Neural Graph ODE to capture that scale's dominant behavior. At inference, DoMiNO flexibly combines different timestep sizes to predict both short- and long-range dynamics with high fidelity. Empirical results on challenging MD benchmarks—ranging from small molecules to proteins—demonstrate the method's long-term stability, flexibility, and accuracy. Our implementation is available at https://github.com/FrancoTSolis/domino-code.

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), enabling insights into atomic-scale interactions over time. Yet, modeling these interactions for extended timescales is extremely **computationally intensive**. A typical timestep in MD might be on the order of a femtosecond (10^{-15} s) (Hollingsworth & Dror, 2018), while biologically relevant phenomenasuch as protein conformational changes can unfold on the nanosecond (10^{-9} s) to microsecond (10^{-6} s) range (McGeagh et al., 2011). Consequently, a full-scale MD simulation needs to traverse millions of steps, which can be prohibitively expensive.

Compounding this challenge is the **multi-scale** nature of molecular trajectories. To illustrate, imagine an molecular system [2] with fast motions (atoms vibrating around their equilibrium positions) occur on very short timescales, while slow, more structured motions (e.g., a protein's domain rearrangement) transpire less frequently but drive large-scale conformational changes. For computational efficiency, one might opt for larger time increments, but that risks oversmoothing the fine, fast dynamics. Conversely, smaller increments yield fine-grained resolution but at a 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), for example, 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

[1] Down-scaling refers to changing a larger, global timescale to a smaller, more local time scale. Meanwhile, up-sampling refers to progressively rebuilding finer details to reconstruct the high-resolution trajectory.

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

ODEs typically lack explicit mechanisms for capturing the inherent fast and slow 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 down-scaling 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 an up-sampling process, progressively adding fine-scale details to the coarse dynamics captured at lower levels. Only a limited number of points need to be sampled per bracket, significantly reducing the computational cost while maintaining accuracy. Our contributions can be summarized as follows:

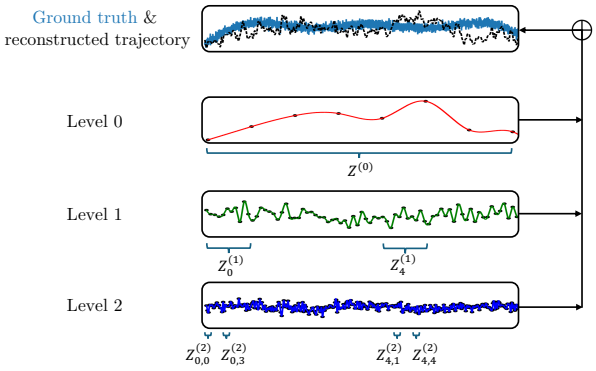

Figure 1: **Hierarchical reconstruction of molecular dynamics trajectories using DoMiNO.** The model progressively refines the trajectory through a multi-scale Neural Graph ODE framework. Level 0 captures slow, large-scale motions, while finer levels (Level 1, Level 2) introduce high-frequency details. This process can be seen as up-sampling, where fine-grained atomic dynamics are reconstructed from coarse representations. By selectively sampling a limited number of points at each level, the model significantly improves computational efficiency while maintaining accuracy.

- *New perspective.* We articulate the fast-slow dichotomy in MD simulations and develop a hierarchical down-scaling view to accommodate both modes.
- *Novel architecture.* We propose Neural Graph ODE modules operating at each level of detail, tied together by learned transitions that allow flexible timing predictions.
- *Good performance.* We validate DoMiNO on a broad range of molecular systems-small molecules to proteins-and demonstrate significant gains in long-term stability and accuracy compared to state-of-the-art baselines.

2 Related Work

Our approach builds upon a rich set of techniques for modeling dynamic systems and leverages insights from diverse computational frameworks.

2.1 Neural ODE-based Models

Neural ODE-based models have advanced continuous-time dynamic modeling by learning latent representations that evolve via differential equations. For example, LG-ODE (Huang et al., 2020) employs a latent ODE framework with graph neural networks to learn from irregularly sampled data Huang et al. (2020), while CG-ODE (Huang et al., 2021), HOPE (Luo et al., 2023b), and PG-ODE (Luo et al., 2023a) extend these ideas to capture multi-agent interactions and higher-order dynamics. Despite their promise, these models often struggle with long-term dependencies and incur high computational costs.

2.2 Wavelet-based Models

Wavelet transform (Zhang & Zhang, 2019) decomposes time series into multiple frequency components, enabling robust handling of non-stationary dynamics. Methods such as W-Transformer (Sasal et al., 2022) and Wavelet-ARIMA (Kriechbaumer et al., 2014) use these decompositions to enhance long-range dependency capture and reduce noise. However, the fixed frequency decomposition of wavelets can limit flexibility when modeling complex nonlinear dynamics.

2.3 GENERATIVE MD APPROACHES

Generative approaches in molecular dynamics, including normalizing flow and diffusion-based models, have been developed to enable large temporal leaps in simulations. Methods like Timewarp (Klein et al., 2024) and ITO (Schreiner et al., 2023) demonstrate significant speed-ups and improved sampling. Yet, these approaches can suffer from reduced interpretability and long-term stability, challenges our hierarchical, continuous-time framework seeks to overcome.

3 PROBLEM FORMULATION AND PRELIMINARIES

In this section, we formalize the MD prediction task and introduce the continuous-time framework that underlies our approach.

3.1 MOLECULAR DYNAMICS AS A GRAPH PROBLEM

Let a molecular system be represented by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each vertex $v_i \in \mathcal{V}$ corresponds to an atom and edges $\langle i, j \rangle \in \mathcal{E}$ capture atomic interactions (e.g., bonds, long-range interactions). Each atom v_i has feature $\mathbf{x}_i^t$ (e.g., velocity, charge) and a 3D position $\mathbf{r}_i(t)$. The MD prediction task is to learn a function

$$f : \left(\mathbf{X}^0, \mathcal{G}, T\right) \mapsto \widehat{\mathbf{R}}, \tag{1}$$

which generates the trajectory $\widehat{\mathbf{R}} = \{\hat{\mathbf{r}}_i(t)\}_{i=1}^N$ over time horizon T.

3.2 GRAPH NEURAL ODES

To continuously model atomic dynamics, we utilize Neural ODEs (Chen et al., 2018) in a graph-aware manner. Formally, let $\mathbf{z}_i(t)$ be a latent representation for atom i. We define:

$$\frac{d\mathbf{z}_i}{dt} = f_\theta \left(\mathbf{z}_i(t), \{\mathbf{z}_j(t)\}_{j \in \mathcal{N}(i)}\right), \tag{2}$$

where $\mathcal{N}(i)$ denotes neighbors of i. By integrating from t_0 to t :

$$\mathbf{z}_i(t) = \mathbf{z}_i\left(t_0\right) + \int_{t_0}^{t} f_\theta\left(\mathbf{z}_i(\tau)\right) d\tau, \tag{3}$$

we can flexibly query any time t without discrete autoregressive rollouts at every step.

4 METHOD

We propose a hierarchical model for molecular dynamics prediction that jointly learns continuous latent dynamics at multiple temporal scales. The model consists of three components: a PaiNN-based encoder that extracts initial latent states from observed trajectories; a hierarchical set of Neural Graph ODEs that progressively refine these latent representations through temporal down-scaling; and a decoder that fuses multi-level latent features to reconstruct atom trajectories.

4.1 ARCHITECTURE OVERVIEW

At the coarsest level (Level 0), the encoder $\sigma_{\text{enc}}^{(0)}$ processes the observed trajectory O and produces an initial latent state $z_0^{(0)}$. This state is evolved using a Neural ODE module $f_\theta^{(0)}$ via:

$$z_t^{(0)} = \text{ODESolve}\big(z_0^{(0)}, f_\theta^{(0)}, t\big). \tag{4}$$

For higher levels ($i > 0$), a projector $\sigma_{\text{proj}}^{(i)}$ initializes the latent state $z_0^{(i)}$ by refining the corresponding state $z_{t_j}^{(i-1)}$ from the previous level:

$$z_0^{(i)} = \sigma_{\text{proj}}^{(i)}\big(z_{t_j}^{(i-1)}\big), \tag{5}$$

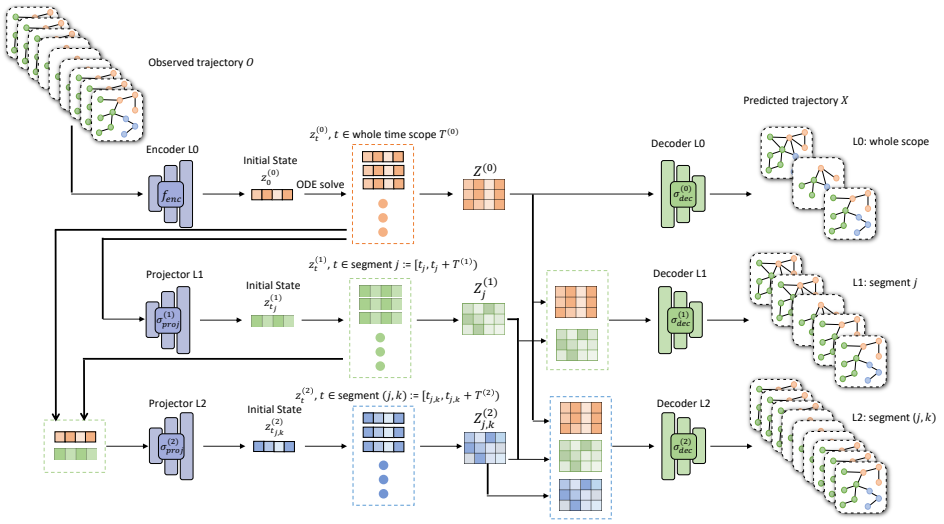

Figure 2: **Hierarchical architecture of DoMiNO.** The model decomposes molecular dynamics trajectories into multiple temporal scales using a sequence of Neural Graph ODEs. The lower levels (coarse resolution) capture large-scale, slow dynamics, while higher levels (fine resolution) refine these representations by incorporating high-frequency details. The final prediction is obtained by concatenating latent representations from all levels, ensuring both computational efficiency and accuracy in multi-scale trajectory reconstruction.

which is then evolved by $f_\theta^{(i)}$ to obtain $z_t^{(i)}$. Finally, the decoder σ_{dec} concatenates latent states across levels and reconstructs the predicted trajectory:

$$\hat{X} = \sigma_{\text{dec}}\Big(\text{concat}\big(z_t^{(0)}, z_t^{(1)}, \ldots, z_t^{(N)}\big)\Big). \tag{6}$$

This hierarchical structure enables efficient down-scaling and accurate multi-scale modeling. Figure 2 illustrates this architecture, where lower-level ODEs initialize and guide the evolution of higher-level ODEs, refining the prediction by incorporating progressively finer details.

4.2 Encoder: PaiNN Equivariant GNN

We use a PaiNN-based encoder to convert the observed trajectory into initial latent representations that capture both the geometric structure and temporal dynamics of the molecular system. For each atom i with 3D coordinates $\mathbf{R}_i$ and feature vector $\mathbf{x}_i$, we first compute relative distances for message passing:

$$\mathbf{d}_{ij} = \mathbf{R}_j - \mathbf{R}_i. \tag{7}$$

Then, at each GNN layer l, the node representation is updated via:

$$\mathbf{h}_i^{(l+1)} = \mathbf{h}_i^{(l)} + \sum_{j \in \mathcal{N}(i)} \phi_m\Big(\mathbf{h}_i^{(l)}, \mathbf{h}_j^{(l)}, \mathbf{d}_{ij}\Big), \tag{8}$$

where ϕ_m is a learnable message function. This message passing ensures equivariance to rotations and translations. In parallel, a temporal self-attention mechanism aggregates the sequence of node embeddings for each atom. Specifically, for atom i we define a global sequence vector:

$$\boldsymbol{a}_i = \tanh\Big(\Big(\frac{1}{N}\sum_t \hat{\boldsymbol{h}}_i^t\Big)\boldsymbol{W}_a\Big), \tag{9}$$

and obtain the fixed-dimensional representation:

$$\boldsymbol{u}_i = \frac{1}{N}\sum_t \sigma\Big(\boldsymbol{a}_i^T \hat{\boldsymbol{h}}_i^t\Big)\hat{\boldsymbol{h}}_i^t, \tag{10}$$

where $\hat{\boldsymbol{h}}_i^t = \sigma\big(\boldsymbol{W}_t[\boldsymbol{h}_i^t \,\|\, \Delta t]\big) + \text{TE}(\Delta t)$ and $\Delta t = t - t_{\text{start}}$. The final encoder outputs a factorized posterior over initial latent states:

$$q_\phi\big(\boldsymbol{Z}^0 | o_1, o_2, \ldots, o_N\big) = \prod_{i=1}^{N} q_\phi\big(\boldsymbol{z}_i^0 | o_1, o_2, \ldots, o_N\big). \tag{11}$$

4.3 HIERARCHICAL DOWN-SCALING ODES

Our model incorporates multiple levels of Neural ODEs to refine the latent dynamics. At Level 0, the latent state $z_0^{(0)}$ is evolved over time using:

$$z_t^{(0)} = \text{ODESolve}\big(z_0^{(0)}, f_\theta^{(0)}, t\big). \tag{12}$$

For each higher level $i > 0$, we initialize the state by projecting the corresponding latent state from the previous level:

$$z_0^{(i)} = \sigma_{\text{proj}}^{(i)}\big(z_{t_j}^{(i-1)}\big), \tag{13}$$

and then evolve it via:

$$z_t^{(i)} = \text{ODESolve}\big(z_0^{(i)}, f_\theta^{(i)}, t\big). \tag{14}$$

Additionally, for each level i we refine the representation within each segment by applying:

$$z_{t_{i,T_i}}^{(i)} = \sigma_{\text{proj}}^{(i-1)}\big(z_{t_{i-1,j}}^{(i-1)}\big) \tag{15}$$

and integrating:

$$z_0^{(i)} = z_0^{(i)} + \int_{t=t_{i-1,j}}^{t_{i,T_i}} g_i^{(i)}\Big(z_1^{(i)}, z_2^{(i)}, \ldots, z_N^{(i)}\Big)\, dt. \tag{16}$$

This nested interval formulation enables efficient down-scaling without traversing all fine-grained time points.

4.4 DECODER: MULTI-LEVEL LATENT REPRESENTATION CONCATENATION

To recover the predicted trajectory at a desired time t, we concatenate latent representations from all levels:

$$\boldsymbol{z}_{\text{concat}}^t = \Big[\boldsymbol{z}_t^{(0)},\, \boldsymbol{z}_t^{(1)},\, \ldots,\, \boldsymbol{z}_t^{(N)}\Big]. \tag{17}$$

The decoder σ_{dec} then maps this fused representation back to 3D coordinates:

$$\boldsymbol{x}_i^t = \sigma_{\text{dec}}\Big(\boldsymbol{z}_{\text{concat}}^t\Big). \tag{18}$$

This multi-level concatenation enriches the information available for accurate reconstruction of molecular trajectories.

5 EXPERIMENTS

In this section, we evaluate our model's performance on various datasets and compare it with representative baselines from different categories. We also conduct an ablation study to investigate the impact of the number of latent encoding levels on our model's performance.

5.1 DATASETS

We evaluate our model using two types of datasets: small molecules and proteins. The small molecules dataset includes the Lennard-Jones (LJ) system, TIP3P water, and TIP4P water. The protein dataset consists of alanine dipeptide (ALA2) (Schreiner et al., 2023). Appendix A provides more details as to the configuration and generation of the datasets. Given an observed sequence of 2000 time steps, the task is to extrapolate the trajectory for arbitrarily sampled points within the subsequent 8000 time steps.

5.2 BASELINES

- **Wavelet ARIMA** (Kriechbaumer et al., 2014): A multiscale statistical model that leverages wavelet decomposition to capture dynamics across different frequency components.
- **DESCINet** (Silva et al., 2023): A state-of-the-art hierarchical deep convolutional neural network designed for long time series forecasting.
- **ITO** (Schreiner et al., 2023): A generative model for molecular dynamics that employs denoising diffusion probabilistic models with SE(3) equivariant architectures.
- **LG-ODE** (Huang et al., 2020): A neural-ODE based framework for learning continuous multi-agent system dynamics from irregularly-sampled partial observations, incorporating graph structure.

5.3 EVALUATION

We evaluate our model's performance using two strategies: full domain evaluation and selected time domain evaluation. These strategies assess our model's ability to predict accurately and robustly on long-sequence molecular dynamics data.

Table 1: Mean Squared Error (MSE) on the four datasets. Best results are marked in bold.

Model	LJ	TIP3P	TIP4P	ALA2
Wavelet ARIMA	0.1965	0.1551	0.1556	0.0323
DESCINet	0.2099±0.0000	0.2034±0.0000	0.2085±0.0000	0.0344±0.0000
ITO 1-step	0.2813±0.0000	0.1838±0.0000	0.1832±0.0000	0.1756±0.0001
ITO rollout	0.5422±0.0002	0.4920±0.0008	0.5177±0.0014	0.7506±0.0029
LG-ODE	0.1859±0.0012	0.1520±0.0001	0.1511±0.0001	0.0447±0.0009
Ours	**0.1786±0.0012**	**0.1513±0.0001**	**0.1503±0.0001**	**0.0225±0.0034**

5.3.1 FULL DOMAIN EVALUATION

We evaluate the extrapolation performance of our model across the entire time domain by sampling multiple time points. Notably, Wavelet ARIMA is a deterministic, multiscale statistical model that produces results without error bars, while the other methods exhibit variance across different runs. Table 1 presents the Mean Squared Error (MSE) on the four datasets, clearly demonstrating that our model consistently outperforms the baselines in general molecular dynamics trajectory prediction.

5.3.2 SELECTED TIME DOMAIN EVALUATION

To confirm that our model's superior performance persists on longer sequences, we evaluate the models on selected time domains. As shown in Tables 3–5 in Appendix B, our model maintains consistently lower MSE at extended time horizons, demonstrating its robustness in handling long-sequence data. This underscores DoMiNO's ability to capture both short- and long-range molecular dynamics more effectively than the baselines.

5.4 ABLATION STUDY

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 illustrated in Figure 3, the full model consistently achieves the lowest Mean Squared Error (MSE) across all datasets, demonstrating the effectiveness of our hierarchical down-scaling framework in capturing multi-scale molecular dynamics.

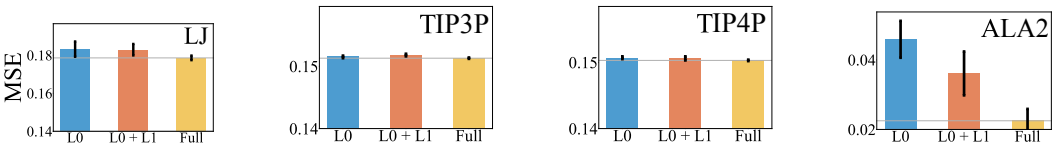

Figure 3: **Ablation study on hierarchical latent encoding.** The results show that incorporating multiple encoding levels leads to a consistent reduction in error, with the full hierarchical model (L0 + L1 + L2) achieving the best performance.

6 CONCLUSION

We introduced DoMiNO, a novel hierarchical framework that models molecular dynamics at multiple temporal scales using Neural Graph Ordinary Differential Equations. By progressively up-sampling input sequences and employing scale-specific ODE modules, DoMiNO effectively captures both fast and slow dynamics. Extensive experiments demonstrate its enhanced long-term prediction accuracy and computational efficiency, offering a promising direction for robust MD simulations.

7 ACKNOWLEDGMENTS

This work was partially supported by NSF 2211557, NSF 2119643, NSF 2303037, NSF 2312501, SRC JUMP 2.0 Center, Amazon Research Awards, and Snapchat Gifts.

REFERENCES

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

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

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.

Tom Darden, Darrin York, Lee Pedersen, et al. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *Journal of chemical physics*, 98:10089–10089, 1993.

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.

Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS computational biology*, 13(7):e1005659, 2017.

Denis J Evans and Brad Lee Holian. The nose-hoover thermostat. *Journal of Chemical Physics*, 83 (8):4069–4074, 1985.

Matt J Harvey, Giovanni Giupponi, and G De Fabritiis. Acemd: accelerating biomolecular dynamics in the microsecond time scale. *Journal of chemical theory and computation*, 5(6):1632–1639, 2009.

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. Coupled graph ode for learning interacting system dynamics. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 705–715, 2021.

Leon Klein, Andrew Foong, Tor Fjelde, Bruno Mlodozeniec, Marc Brockschmidt, Sebastian Nowozin, Frank Noé, and Ryota Tomioka. Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics. *Advances in Neural Information Processing Systems*, 36, 2024.

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.

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.

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

Xiao Luo, Yiyang Gu, Huiyu Jiang, Jinsheng Huang, Wei Ju, Ming Zhang, and Yizhou Sun. Graph ode with factorized prototypes for modeling complicated interacting dynamics. *arXiv preprint arXiv:2311.06554*, 2023a.

Xiao Luo, Jingyang Yuan, Zijie Huang, Huiyu Jiang, Yifang Qin, Wei Ju, Ming Zhang, and Yizhou Sun. Hope: High-order graph ode for modeling interacting dynamics. In *International Conference on Machine Learning*, pp. 23124–23139. PMLR, 2023b.

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.

M Germana Paterlini and David M Ferguson. Constant temperature simulations using the langevin equation with velocity verlet integration. *Chemical Physics*, 236(1-3):243–252, 1998.

Lena Sasal, Tanujit Chakraborty, and Abdenour Hadid. W-transformers: a wavelet-based transformer framework for univariate time series forecasting. In *2022 21st IEEE international conference on machine learning and applications (ICMLA)*, pp. 671–676. IEEE, 2022.

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

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.

Dengsheng Zhang and Dengsheng Zhang. Wavelet transform. *Fundamentals of image data mining: Analysis, Features, Classification and Retrieval*, pp. 35–44, 2019.

A DATA AND TRAINING CONFIGURATIONS

A.1 DATA GENERATION

The classical molecular dynamics simulations for the Lennard-Jones (LJ), TIP3P, and TIP4P-Ew systems were performed using OpenMM (Eastman et al., 2017). In these simulations, the systems are composed of either uniform single-atom setups (for LJ) or single-molecule configurations (for TIP3P and TIP4P-Ew) and are contained in simulation boxes of varying sizes. Periodic boundary conditions are applied in all directions, and a cut-off distance of 10 Å is used to manage interatomic interactions.

Each simulation begins with particles initialized at random positions and velocities in a static state. After a brief transient period, the systems reach equilibrium under constant volume and temperature (NVT), ensuring a canonical ensemble. A velocity Verlet integrator (Paterlini & Ferguson, 1998) is employed with a time step of 2.0 femtoseconds, while a Nosé–Hoover chain thermostat (Evans & Holian, 1985) maintains a constant temperature using a collision frequency of 1.0 per picosecond and a chain length of 10.

For each configuration, the simulation is run for 500,000 steps, with the state of the system recorded every 50 steps. This configuration results in an effective sequence length of 10,000 time points and an effective time step of 100 femtoseconds between stored states.

For the ALA2 dataset, the simulation parameters differ. The simulation is performed using the ACEMD (Harvey et al., 2009) code with the AMBER ff-99SB-ILDN forcefield and a Langevin integrator. The time step is set to 2 femtoseconds, and the simulation runs for 250 nanoseconds. Frames are saved every 1 picosecond, yielding an effective step size of 1000 femtoseconds. The simulation is conducted at 300 K in a periodic box with a volume of approximately $(2.3222\ \text{nm})^3$. The system is solvated with 651 TIP3P water molecules, and electrostatics are managed using Particle Mesh Ewald (PME) (Darden et al., 1993) with a real-space cutoff of 0.9 nm, a grid spacing of 0.1 nm, and updates every two time steps. All bonds between hydrogens and heavy atoms are constrained.

Table 2: Dataset Statistics

Dataset	Number of Atoms	Sequence Length	Step Size (fs)
LJ	258	10000	100
TIP3P	774	10000	100
TIP4P	753	10000	100
ALA2	22	10000	1000

A.2 TRAINING OBJECTIVE FUNCTION

We train the encoder, hierarchical ODEs, and decoder jointly by maximizing the evidence lower bound (ELBO):

$$\mathrm{ELBO}(\theta, \phi) = \mathbb{E}_{\boldsymbol{Z}^0 \sim q_\phi\left(\boldsymbol{Z}^0 | o_1, \ldots, o_N\right)} \Big[\log p_\theta(o_1, \ldots, o_N)\Big] - \mathrm{KL}\Big[q_\phi\big(\boldsymbol{Z}^0 | o_1, \ldots, o_N\big) \,\|\, p\big(\boldsymbol{Z}^0\big)\Big], \tag{19}$$

where q_ϕ denotes the approximate posterior and p_θ is the generative model defined by our ODEs.

B MORE EXPERIMENT RESULTS

Table 3: Mean Squared Error (MSE) on the LJ dataset for different terminate timesteps. Best results are marked in bold.

Terminate timestep	80-100	180-200	980-1000	1980-2000	2980-3000	4980-5000	7980-8000
Wavelet ARIMA	0.1881	0.2032	0.1900	0.1956	0.2087	0.1928	0.1969
DESCINet	0.2204±0.0000	0.2197±0.0000	0.2015±0.0000	0.2054±0.0000	0.2252±0.0000	0.2092±0.0000	0.2139±0.0000
ITO 1 step	0.2806±0.0002	0.2807±0.0002	0.2810±0.0001	0.2801±0.0002	0.2813±0.0001	0.2812±0.0002	0.2843±0.0001
ITO rollout	0.5133±0.0007	0.5314±0.0004	0.5012±0.0003	0.5255±0.0007	0.5178±0.0002	0.5032±0.0002	0.5437±0.0002
LG-ODE	0.1989±0.0003	0.1927±0.0003	0.1750±0.0002	0.1817±0.0002	0.1974±0.0002	0.1840±0.0008	0.1916±0.0024
Ours	**0.1982±0.0001**	**0.1923±0.0000**	**0.1745±0.0001**	**0.1808±0.0002**	**0.1954±0.0001**	**0.1799±0.0001**	**0.1840±0.0002**

Table 4: Mean Squared Error (MSE) on the TIP3P dataset for different terminate timesteps. Best results are marked in bold.

Terminate timestep	80-100	180-200	980-1000	1980-2000	2980-3000	4980-5000	7980-8000
Wavelet ARIMA	0.1554	0.1575	0.1538	0.1547	0.1546	0.1560	0.1560
DESCINet	0.2027±0.0001	0.2029±0.0000	0.2025±0.0001	0.2014±0.0001	0.2096±0.0001	0.2013±0.0000	0.2040±0.0000
ITO 1 step	0.1840±0.0001	0.1841±0.0001	0.1837±0.0001	0.1840±0.0001	0.1836±0.0001	0.1838±0.0001	0.1837±0.0001
ITO rollout	0.4778±0.0021	0.4846±0.0013	0.4924±0.0038	0.4924±0.0020	0.4897±0.0015	0.5031±0.0028	0.5025±0.0021
LG-ODE	0.1523±0.0001	0.1519±0.0001	0.1518±0.0003	0.1530±0.0004	0.1525±0.0002	0.1528±0.0003	0.1536±0.0004
Ours	**0.1516±0.0000**	**0.1515±0.0000**	**0.1512±0.0000**	**0.1520±0.0000**	**0.1519±0.0000**	**0.1517±0.0000**	**0.1520±0.0001**

Table 5: Mean Squared Error (MSE) on the TIP4P dataset for different terminate timesteps. Best results are marked in bold.

Terminate timestep	80-100	180-200	980-1000	1980-2000	2980-3000	4980-5000	7980-8000
Wavelet ARIMA	0.1520	0.1599	0.1579	0.1557	0.1548	0.1555	0.1557
DESCINet	0.2058±0.0003	0.2086±0.0001	0.2106±0.0001	0.2109±0.0001	0.2070±0.0000	0.2025±0.0000	0.2066±0.0001
ITO 1 step	0.1833±0.0000	0.1830±0.0000	0.1832±0.0000	0.1831±0.0000	0.1833±0.0000	0.1832±0.0000	0.1833±0.0000
ITO rollout	0.4929±0.0075	0.5123±0.0057	0.5012±0.0067	0.5256±0.0063	0.5178±0.0065	0.5032±0.0063	0.5437±0.0069
LG-ODE	0.1521±0.0000	0.1507±0.0000	**0.1512±0.0001**	0.1510±0.0002	0.1508±0.0001	0.1515±0.0003	0.1530±0.0005
Ours	**0.1502±0.0000**	**0.1506±0.0000**	0.1513±0.0000	**0.1501±0.0000**	**0.1500±0.0000**	**0.1503±0.0000**	**0.1511±0.0000**