UCLA Samueli School of Engineering

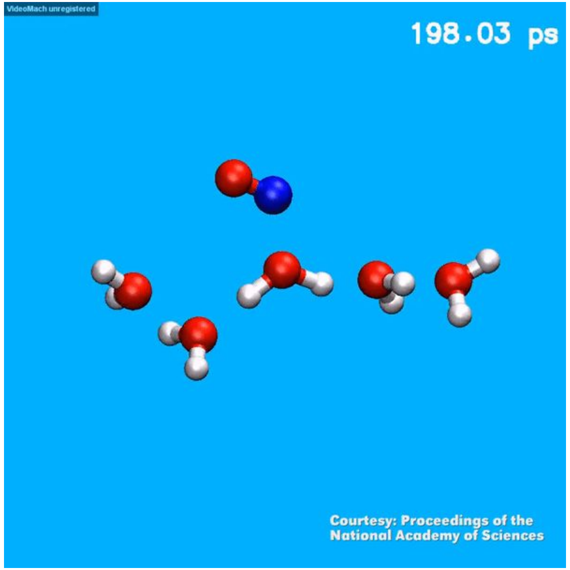
ICLR

Machine Learning Multiscale Processes

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

Authors: Fang Sun, Zijie Huang, Yadi Cao, Xiao Luo, Wei Wang, Yizhou Sun
Presenter: Prof. Yizhou Sun
Computer Science Department, UCLA
Date: Apr. 27, 2025

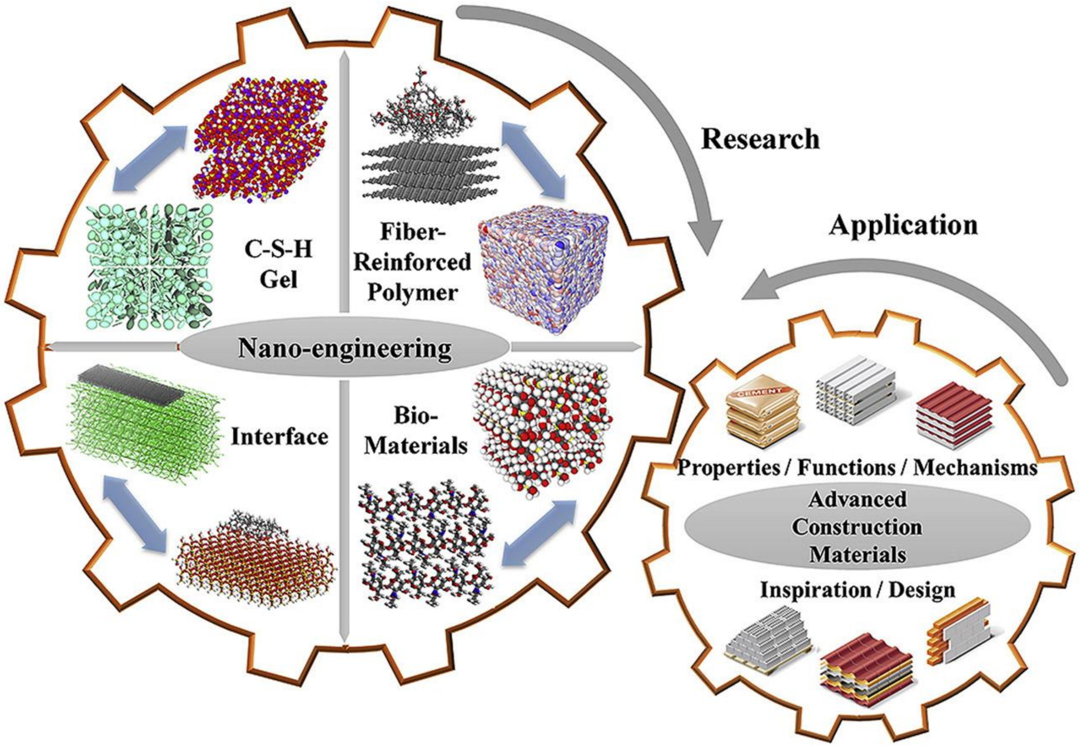
Molecular Dynamics (MD) Prediction



- A computer simulation method for analyzing the physical movements of atoms and molecules.
- Times-series task formulation: **Predict the atom trajectories $r_i(t), i = 1 \ldots N$ for any desired time point t.**

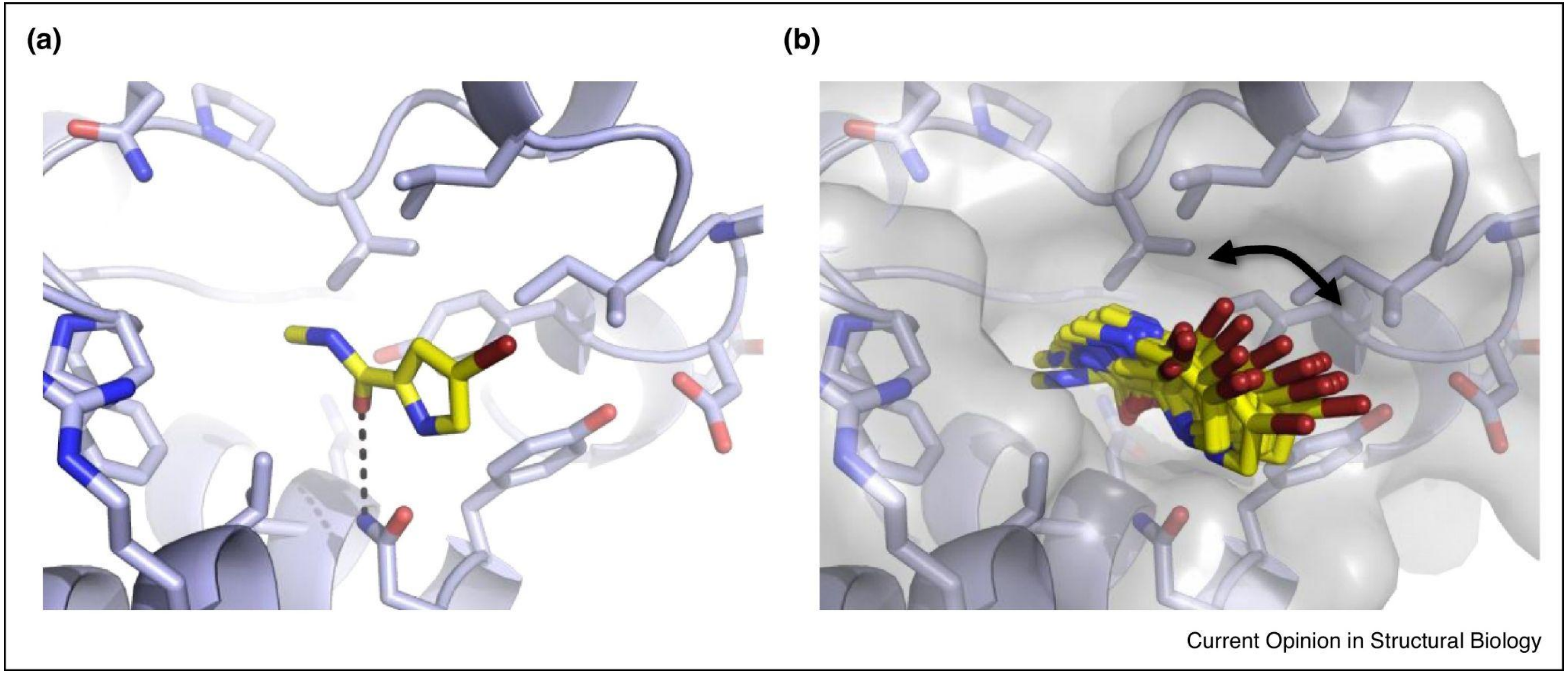
UCLA Samueli School of Engineering

Why MD prediction? — Material Science



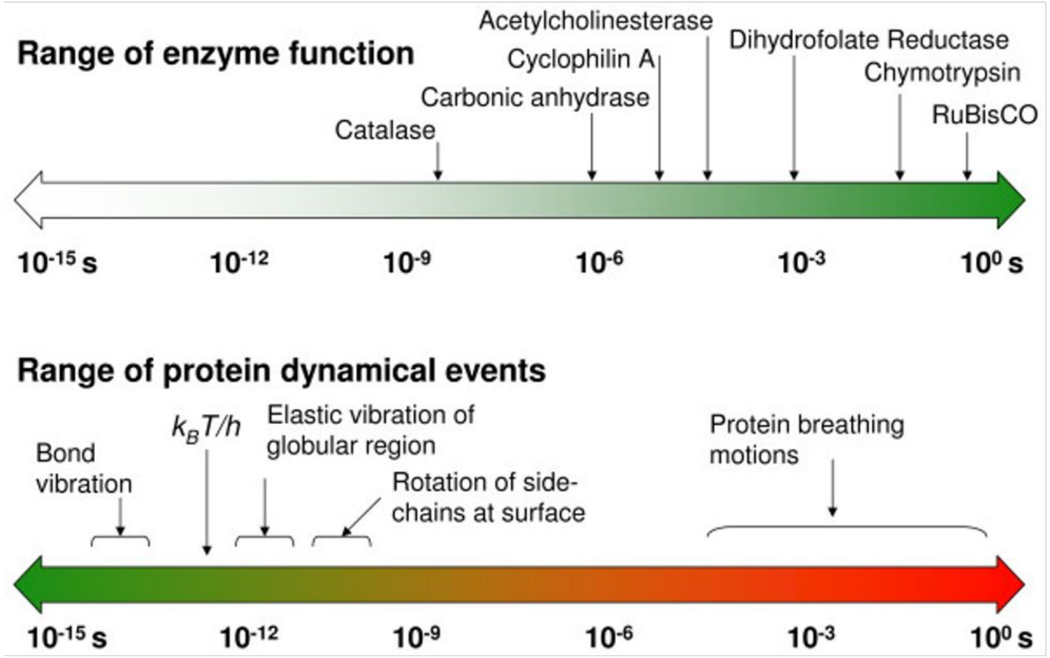
https://www.sciencedirect.com/science/article/pii/S1359836817336077

Why MD prediction? — Drug Design



https://www.sciencedirect.com/science/article/pii/S0959440X17301100

From the multi time-scale perspective



[Agarwal, 2006]

- Protein folding: $10^{-3}\ s$
- Enzyme Reaction: $10^{-8} \sim 10^{-7}\ s$
- Diffusion Limit: $10^{-9}\ s$
- MD step-size: $10^{-15}\ s = 1\ fs$

⇒ Meaningful MD simulations need to propagate through 10^6 steps!

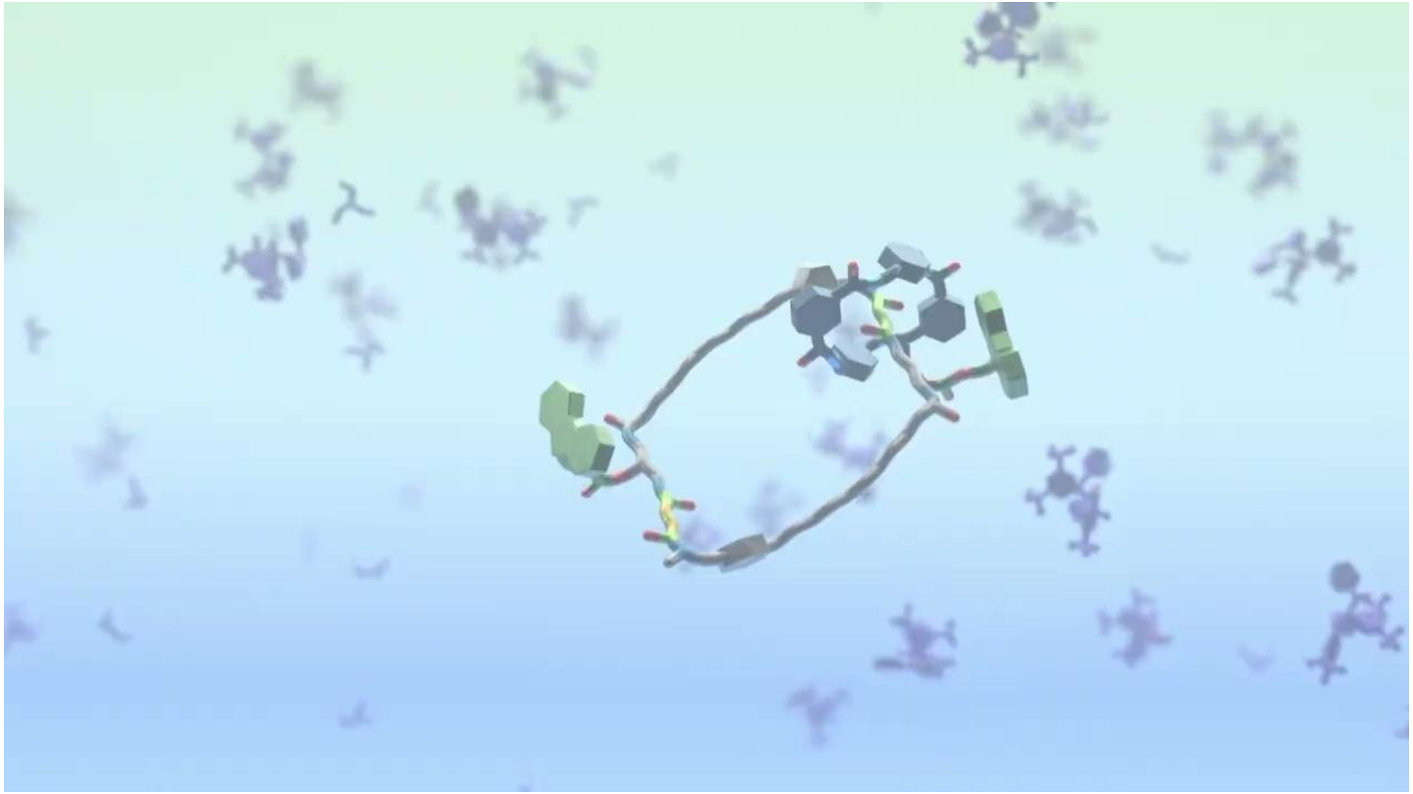
A challenging task, due to error propagation across steps.

UCLA Samueli School of Engineering

Example: Multiscale Dynamics in a Molecular Machine

Video Source:

[Wilson et al., 2016]



Motivations and Challenges

Challenge 1: **Computational Cost**

- MD simulations require millions of small timesteps (femtoseconds)
- Biologically relevant phenomena occur at nano to microsecond scales

Challenge 2: **Multi-scale Nature**

- Fast dynamics: Atomic vibrations (femtoseconds)
- Slow dynamics: Conformational changes (nano to microseconds)

Challenge 3: **Resolution Tradeoff**

- Small timesteps: High accuracy but computationally expensive
- Large timesteps: Efficient but loses fine-grained details

Molecular Dynamics as Dynamic Graphs

Graph representation:

– $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

- $\mathcal{V} = v_i$ are atoms
- $\mathcal{E} = \langle i, j \rangle$ are interactions (bonds, electrostatics)

Node attributes at time t:

– Features $\mathbf{x}_i^t$ (e.g. velocity, charge)

– Position $\mathbf{r}_i(t) \in \mathbb{R}^3$

Prediction Task Formulation

- Observed input: full sequence of positions & velocities

$$\{(\mathbf{r}_i(t), \mathbf{x}_i(t))\}_{i=1}^{N} \quad \text{for } t = 1, \ldots, 2000$$

- Prediction target: positions & velocities

$$\left\{\left(\hat{\mathbf{r}}_i(t), \hat{\mathbf{x}}_i(t)\right)\right\}_{i=1}^{N} \quad \text{for } t = 2001, \ldots, 10000$$

- Evaluation: error measured on predicted positions $\hat{\mathbf{r}}_i(t)$ only

Current Approaches & Limitations

Traditional Autoregressive Model:

- Roll-out step-by-step, lead to high cumulative error

Neural ODE-based models [Chen et al., 2018; Huang et al., 2020]

- Pro: Continuous formulation avoids autoregressive rollout issues
- Con: Lack explicit mechanisms for capturing fast/slow dynamics

Wavelet-based methods [Kriechbaumer et al., 2014]

- Pro: Decompose time series into frequency components
- Con: Fixed frequency decompositions miss complex nonlinearities

Generative MD approaches [Schreiner et al., 2023]

- Pro: Enable efficient long-range sampling
- Con: Can suffer from long-term instability

The General Graph ODE Framework: LG-ODE (Huang et al., 2020)

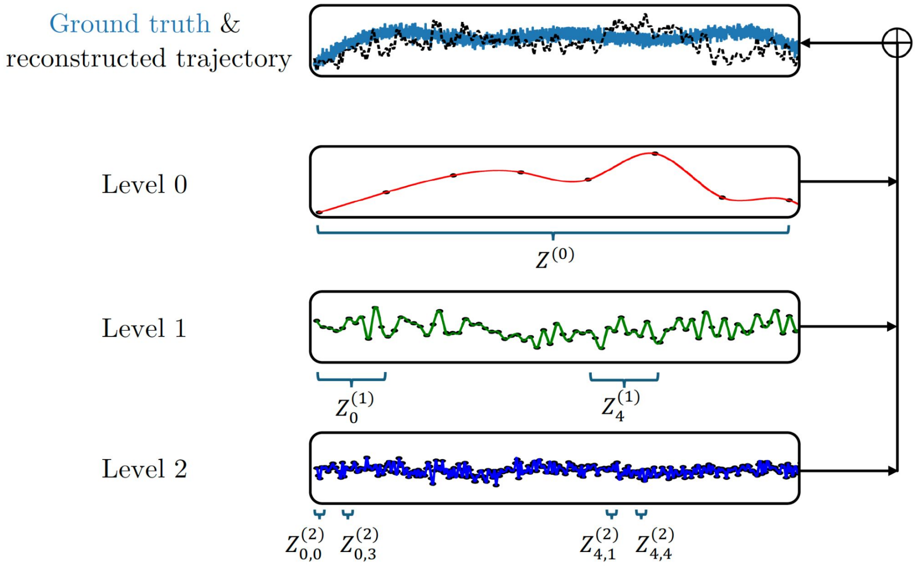
A Generative Model

- **Step 1**: Initial State Encoder to get a distribution for $\boldsymbol{Z}^0$
- **Step 2**: Sample $\boldsymbol{Z}^0$. and use ODESolver to infer $\boldsymbol{Z}^t$ at any observed t

$$z_i^T = z_i^0 + \int_{t=0}^{T} g_i\left(z_1^t, z_2^t \cdots z_N^t\right) dt$$

- **Step 3**: Generate observations from latent representation
- Loss = reconstruction loss + regularization term:
  $\mathrm{ELBO}(\theta, \phi)$
  $= \mathbb{E}_{\boldsymbol{Z}^0 \sim q_\phi(\boldsymbol{Z}^0 \mid o_1, \cdots o_N)}\left[\log p_\theta\left(o_1, \ldots, o_N\right)\right] - \mathrm{KL}\left[q_\phi\left(\boldsymbol{Z}^0 \mid o_1, \cdots, o_N\right) \| p\left(\boldsymbol{Z}^0\right)\right]$

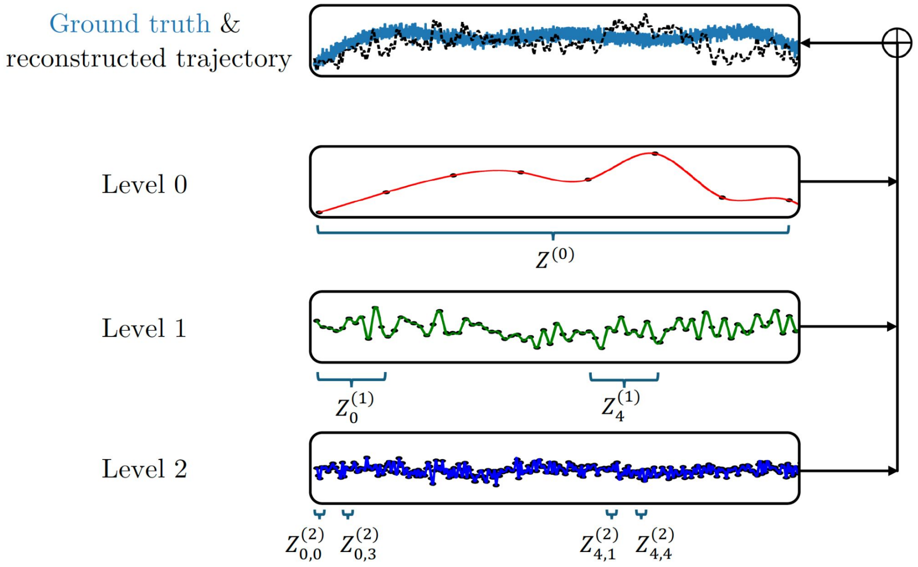
Key Insight: Multi-scale Temporal Resolution



- MD trajectories contain coexisting fast and slow modes
- Different dynamics require different temporal resolutions
- Hierarchical decomposition can capture both scales efficiently
- Progressive refinement through "down-scaling"

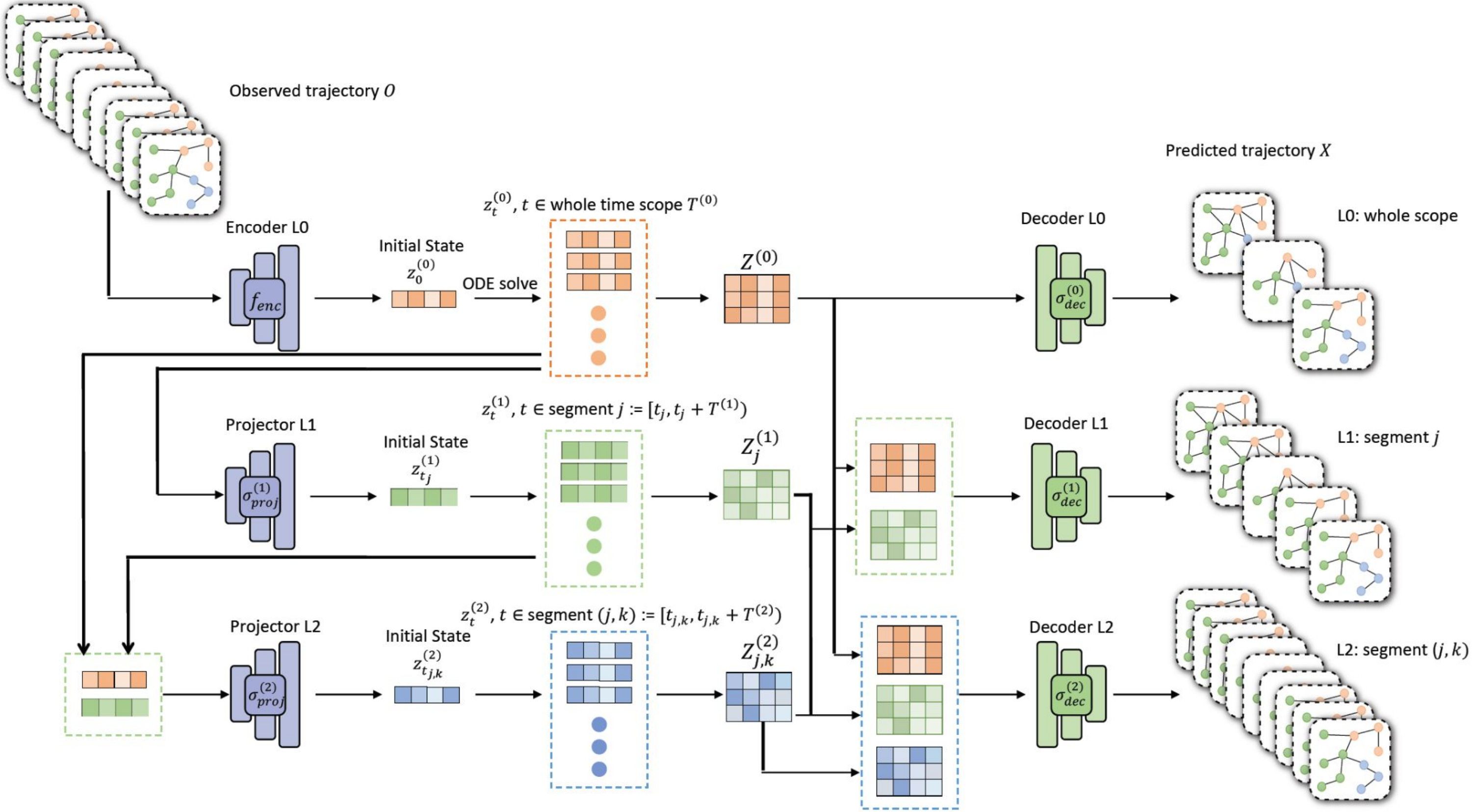
UCLA Samueli School of Engineering

Down-scaling Approach



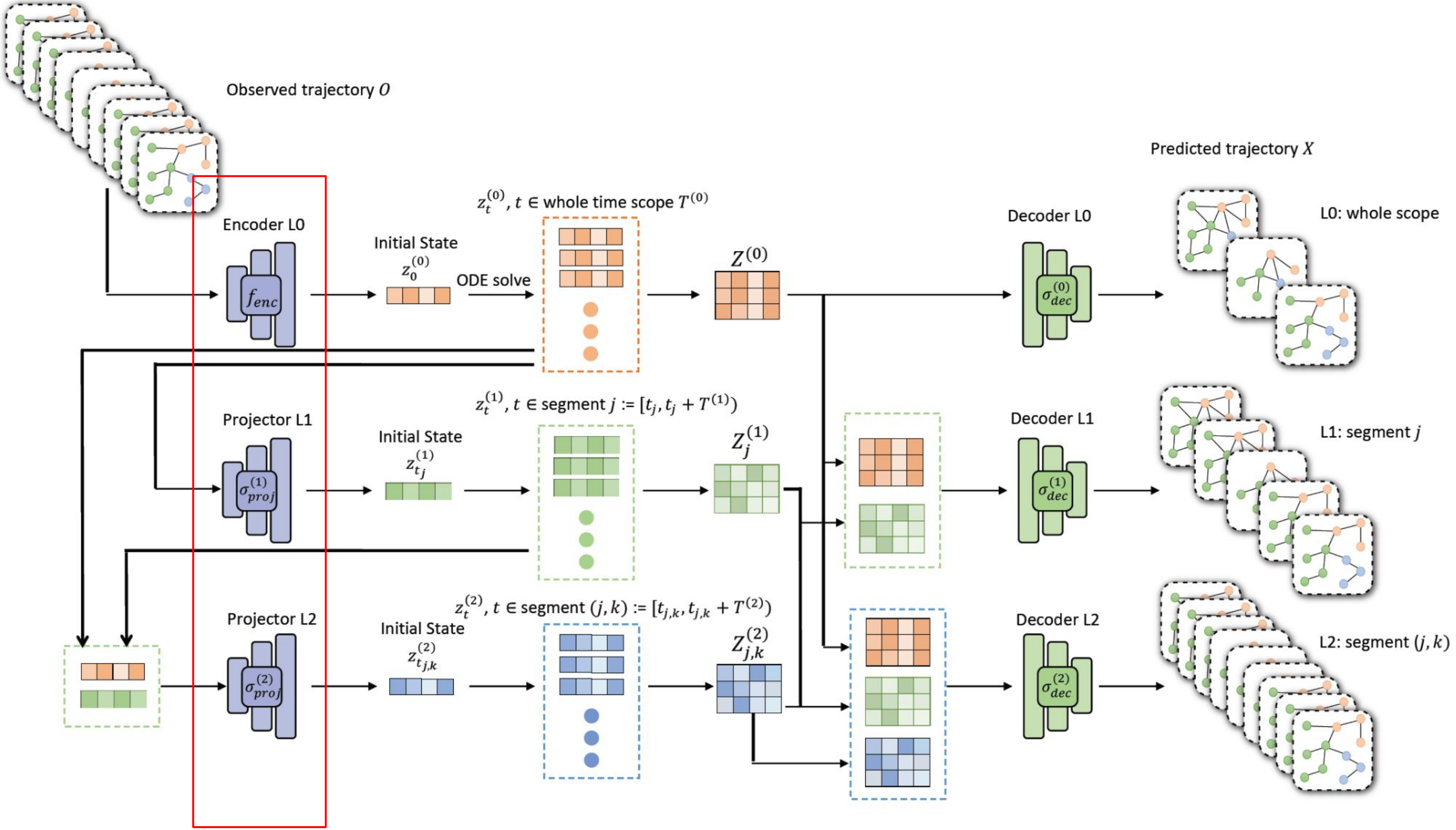
- Level 0: Captures slowest, large-scale motions
- Level 1, 2, etc.: Add progressively finer temporal details
- Only selective sampling needed at each level
- Efficient computation while maintaining accuracy

DoMiNO Framework Overview



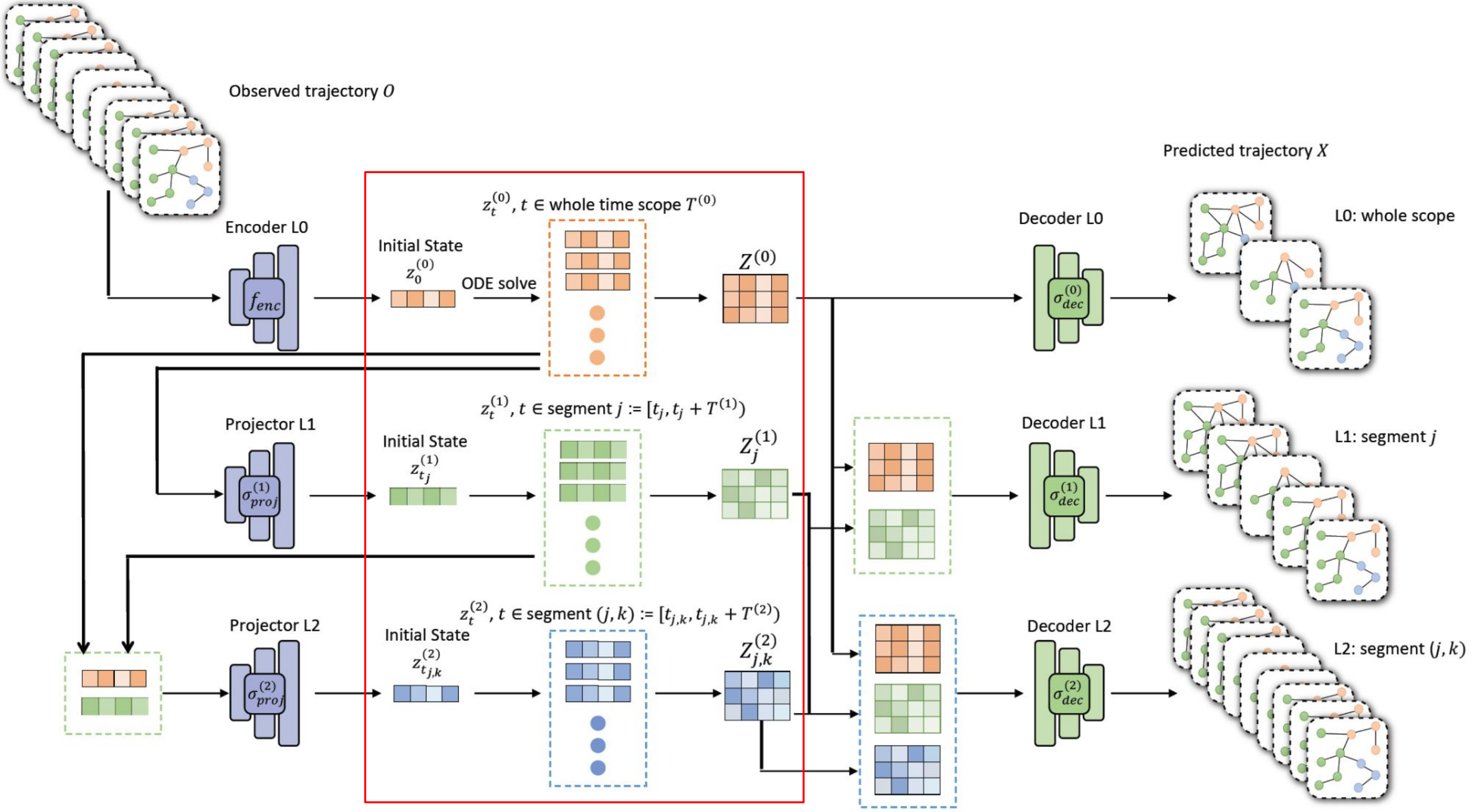
UCLA

Encoder: PaiNN-based equivariant GNN extracts initial latent states



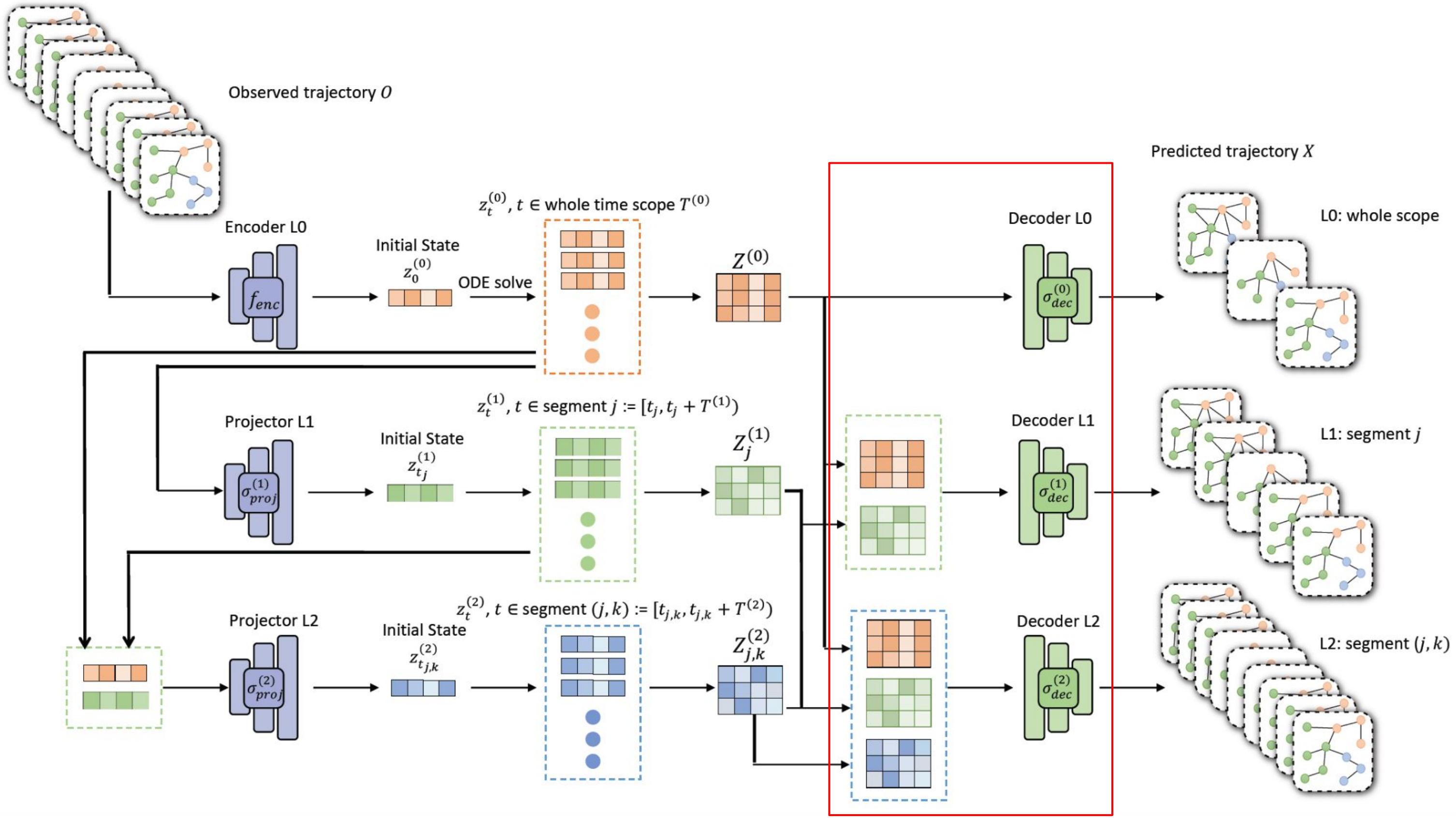
UCLA

Hierarchical Neural Graph ODEs: Multi-level continuous-time dynamics modeling



UCLA

Decoder: Fuses multi-level latent features for trajectory reconstruction



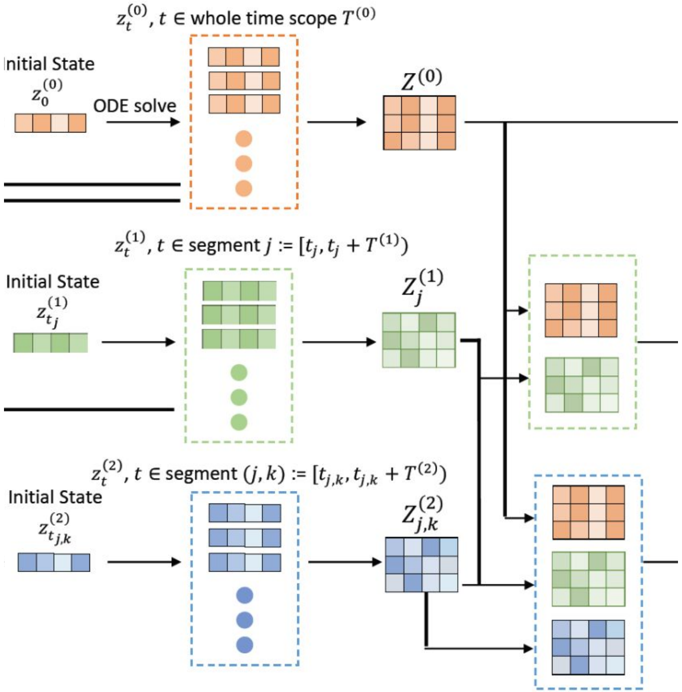
Mathematical Formulation for Downscaling ODEs

Level 0 ODE: $z_t^{(0)} = \mathrm{ODESolve}(z_0^{(0)}, f_\theta^{(0)}, t)$

Higher levels: $z_0^{(i)} = \sigma_{\text{proj}}^{(i)}(z_{t_j}^{(i-1)})$

Evolution: $z_t^{(i)} = \mathrm{ODESolve}(z_0^{(i)}, f_\theta^{(i)}, t)$

Multi-level fusion: $z_{\text{concat}}^t = [z_t^{(0)}, z_t^{(1)}, \ldots, z_t^{(N)}]$



Evaluation & Datasets

Small Molecules:

- Lennard-Jones (LJ) system, 250 argon atoms
- TIP3P water, 258 water molecules
- TIP4P water, 258 water molecules

Proteins:

- Alanine dipeptide (ALA2)

Evaluation Strategies:

- Full domain evaluation: we randomly sample time points across the 8000 time steps
- Selected time domain evaluation: we focus on particular time regions

Train-test Split

Small-molecule systems (LJ, TIP3P, TIP4P)
– 8 independent trajectories (length 10 000, random inits)
– Train on 5 trajectories, test on 3

Protein system (ALA2)
– 3 open-source trajectories
– Train on 2, test on 1

Note: no temporal dependency between train/test splits; each trajectory used wholly for one split.

Baseline Comparisons

Wavelet ARIMA [Kriechbaumer et al., 2014]: Multiscale statistical model with wavelet decomposition

DESCINet [Silva et al., 2023]: Hierarchical deep Convolutional Network for long time series forecasting

ITO [Schreiner et al., 2023]: Generative model using denoising diffusion with SE(3) equivariance

LG-ODE [Huang et al., 2020]: Neural-ODE based framework for continuous multi-agent dynamics

Overall performance based on sampled future timestamps

DoMiNO consistently outperforms baselines across all datasets

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**

Long-term stability

DoMiNO maintains consistently lower MSE at extended time horizons

Demonstrates robustness in handling long-sequence data

Captures both short and long-range molecular dynamics effectively

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**

Analysis of Temporal MSE Variation

Two Key Hardness Factors:

1. **Dynamic Regime:**
   - **Early:** system still "settling" from initial conditions → rapid, nonlinear changes → high error
   - **Middle:** dynamics approach quasi-equilibrium → smoother evolution → lower error
   - **Late:** error accumulation over long rollouts dominates → rising MSE
2. **Roll-out Error Accumulation:**
   - Even in converged regimes, small per-step ODE integration errors compound over millions of femtoseconds

Ablation Study: Impact of Hierarchical Levels

Compared three configurations:

- Only Level 0 (L0)
- Levels 0 + 1 (L0+L1)
- Full model with all levels (L0+L1+L2)

Full hierarchical model consistently achieves lowest MSE

Validates effectiveness of multi-scale approach

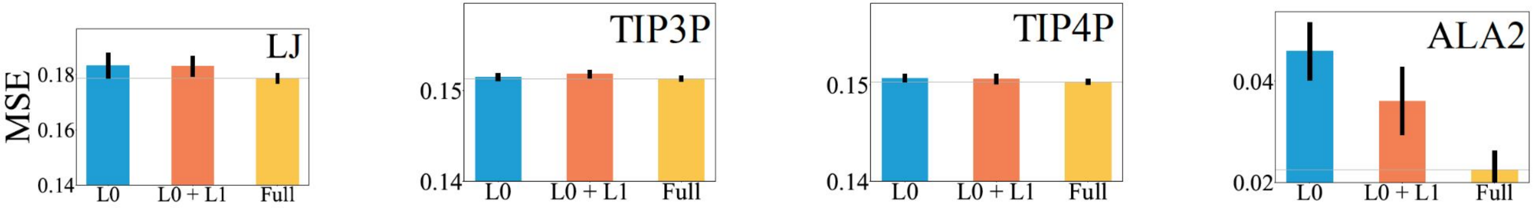


Figure 3: **Ablation study on hierarchical latent encoding.** The results show that incorporating multiple encoding levels leads to a consistent reduction in error, with the full hierarchical model (L0 + L1 + L2) achieving the best performance.

UCLA

Key Contributions

New perspective: Fast-slow dichotomy in MD simulations with hierarchical downscaling

Novel architecture: Multi-level Neural Graph ODEs with learned transitions

Improved performance: Enhanced long-term stability and accuracy

Flexibility: Can predict both short and long-range dynamics with high fidelity

References

[Wilson et al., 2016] "An autonomous chemically fuelled small-molecule motor." *Nature* 534.7606 (2016): 235-240.

[Agarwal, 2006] "Enzymes: An integrated view of structure, dynamics and function." *Microbial cell factories* 5 (2006): 1-12.

[Chen et al., 2018] "Neural ordinary differential equations." *Advances in neural information processing systems* 31 (2018).

[Huang et al., 2020] "Learning continuous system dynamics from irregularly-sampled partial observations." *Advances in Neural Information Processing Systems* 33 (2020): 16177-16187.

[Kriechbaumer et al., 2014] "An improved wavelet–ARIMA approach for forecasting metal prices." Resources Policy 39 (2014): 32-41.

[Schreiner et al., 2023] "Implicit transfer operator learning: Multiple time-resolution models for molecular dynamics." Advances in Neural Information Processing Systems 36 (2023): 36449-36462.

[Silva et al.. 2023] "DESCINet: A hierarchical deep convolutional neural network with skip connection for long time series forecasting." *Expert Systems with Applications* 228 (2023): 120246.

Q&A