Scalable learning of macroscopic stochastic dynamics

Mengyi Chen [1] **Pengru Huang** [2] **Kostya S. Novoselov** [2 3] **Qianxiao Li** [1 2]

[1]*Department of Mathematics, National University of Singapore, Singapore, Singapore* [2]*Institute for Functional Intelligent Materials, National University of Singapore, Singapore, Singapore* [3]*Materials Science and Engineering, National University of Singapore, Singapore, Singapore*. Correspondence to: Qianxiao Li qianxiao@nus.edu.sg.

1. Introduction

Macroscopic dynamical descriptions of complex physical systems are crucial for understanding and controlling material behavior. With the growing availability of data and compute, machine learning has become a promising alternative to first-principles methods to build accurate macroscopic models from microscopic trajectory simulations. However, for spatially extended systems, direct simulations of sufficiently large microscopic systems that inform macroscopic behavior are prohibitive. In this work, we propose a framework that learns the macroscopic dynamics of large stochastic microscopic systems using only small-system simulations. Our framework employs a partial evolution scheme to generate training data pairs by evolving large-system snapshots within local patches. We subsequently derive the closure variables associated with the macroscopic observables and learn the macroscopic dynamics using a custom loss. Furthermore, we introduce a hierarchical upsampling scheme that enables efficient generation of large-system snapshots from small-system snapshots. We empirically demonstrate the accuracy and robustness of our framework through a variety of stochastic spatially extended systems, including those described by stochastic partial differential equations, idealised lattice spin systems, and a more realistic NbMoTa alloy system.

2. METHODOLOGY

We focus on microscopic systems that are spatially extended, including SPDE systems, the Ising model, and alloy systems. We assume the microscopic time evolution can be modeled as a Markov process of a random variable supported on a finite but large lattice structure. Let the microscopic state be $\mathbf{x} = (\mathbf{x}_1, \cdots, \mathbf{x}_n) \in \mathbb{R}^n$, where n represents the number of lattice sites of the system. We assume the lattice sites are arranged on a regular lattice structure, and $\mathbf{x}_i$ represents some physical quantity associated with the i-th lattice site.

In many real applications, we are interested in the dynamics of some macroscopic observables, denoted by $\mathbf{z}^* = \boldsymbol{\varphi}^*(\mathbf{x})$. The form of $\boldsymbol{\varphi}^*$ is given beforehand, and $\boldsymbol{\varphi}^*$ can be applied to different system sizes. Since the underlying microscopic system is spatially extended, we are interested in the intensive quantities that do not scale with system size.

Assume we are given a microscopic simulator $\mathcal{S}_{n_s}$, which can accurately simulate the microscopic dynamics of a small system up to size $n_s \ll n$ due to computational constraints. From this simulator, we obtain the dataset D_s of the small system composed of snapshots sampled from trajectories of the small system. The goal of this work is to derive the macroscopic dynamics of a large system of size n using only such small-scale simulations.

Partial evolution scheme. We propose a scheme for locally evolving the microscopic dynamics of a large system within a small spatial patch, which we refer to as the *partial evolution scheme*. The purpose of this scheme is to generate locally evolved training data pairs $\{\mathbf{x}_t, \mathbf{x}_{t+\delta t, \mathcal{I}}\}$ by evolving the microscopic dynamics on a small patch for a short time interval.

Autoencoder for discovering closure variables. We employ an autoencoder architecture to discover closure variables $\hat{\mathbf{z}}$ associated with the macroscopic observables $\mathbf{z}^*$. The closure variables will capture the unresolved information by $\mathbf{z}^*$, and ensure the dynamics of $\mathbf{z} = (\mathbf{z}^*, \hat{\mathbf{z}})$ depend only on itself. F

Since the training data pairs take the form of $\{\mathbf{x}_t, \mathbf{x}_{t+\delta t, \mathcal{I}}\}$, we want the closure function to be well-defined for both the microscopic state $\mathbf{x}$ and the microscopic state $\mathbf{x}_{\mathcal{I}}$, which are of different dimensions. Denote the closure function by $\hat{\boldsymbol{\varphi}}$. To achieve this, $\hat{\boldsymbol{\varphi}}$ is directly applied to $\mathbf{x}_{\mathcal{I}}$, yielding $\hat{\boldsymbol{\varphi}}(\mathbf{x}_{\mathcal{I}})$. The closure representation for the full state $\mathbf{x}$ is defined as the average of $\hat{\boldsymbol{\varphi}}(\mathbf{x}_{\mathcal{I}})$ over all the patches:

$$\hat{\boldsymbol{\varphi}}(\mathbf{x}) = \frac{1}{K} \sum_{\mathcal{I} \in \{\mathcal{I}^1, \cdots, \mathcal{I}^K\}} \hat{\boldsymbol{\varphi}}(\mathbf{x}_{\mathcal{I}}). \tag{1}$$

We denote the closure variables by $\hat{\mathbf{z}} = \hat{\boldsymbol{\varphi}}(\mathbf{x})$, and concatenate it with the macroscopic observable $\mathbf{z}^* = \boldsymbol{\varphi}^*(\mathbf{x})$ to form the full latent state $\mathbf{z} = (\mathbf{z}^*, \hat{\mathbf{z}}) = (\boldsymbol{\varphi}^*(\mathbf{z}), \hat{\boldsymbol{\varphi}}(\mathbf{x}))$. Denote the encoder by $\boldsymbol{\varphi} = (\boldsymbol{\varphi}^*, \hat{\boldsymbol{\varphi}})$ and the decoder by $\boldsymbol{\psi}$, where $\boldsymbol{\varphi}^*$ is the predefined macroscopic observable function with no trainable parameters. The functions $\hat{\boldsymbol{\varphi}}$ and $\boldsymbol{\psi}$ are parameterized by neural networks and are trained jointly. We omit the explicit dependence on the parameters for notational simplicity. The autoencoder is trained by minimizing the reconstruction loss:

$$\mathcal{L}_{\text{recon}} = \mathbb{E}_{\mathbf{x}_t} ||\boldsymbol{\psi} \circ \boldsymbol{\varphi}(\mathbf{x}_t) - \mathbf{x}_t||_2^2. \tag{2}$$

Once the autoencoder is trained, we will generate the latent training data pair $\{\mathbf{z}_t, \mathbf{z}_{t+\delta t, \mathcal{I}}\}$ for the macroscopic dynamics derivation:

$$\begin{aligned} &\mathbf{z}_t = \boldsymbol{\varphi}(\mathbf{x}_t), \\ &\mathbf{z}_{t+\delta t, \mathcal{I}} \coloneqq \mathbf{z}_t + \left(\boldsymbol{\varphi}(\mathbf{x}_{t+\delta t, \mathcal{I}}) - \boldsymbol{\varphi}(\mathbf{x}_{t, \mathcal{I}})\right), \end{aligned} \tag{3}$$

where $\mathbf{x}_{t,\mathcal{I}}$ denotes the restriction of $\mathbf{x}_t$ to the local patch $\mathcal{I}$. Next, we introduce the process of deriving macroscopic dynamics.

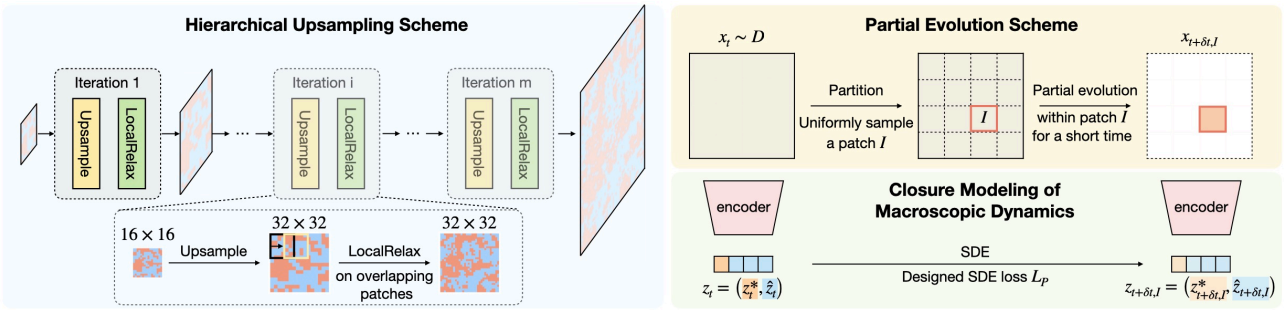

Fig. 1: Schematic illustration of our framework. The hierarchical upsampling scheme generates the large-system dataset D from the small-system dataset D_s through multiple iterations, each consisting of an UPSAMPLE and a LOCALRELAX step. An example of one iteration for the Ising model is shown. For the partial evolution scheme, for every $\mathbf{x}_t \in D$, a patch $\mathcal{I}$ is first uniformly sampled, then the microscopic dynamics is evolved locally within the patch $\mathcal{I}$ for a short time to yield $\mathbf{x}_{t+\delta t,\mathcal{I}}$. For the closure modeling, an autoencoder is trained to discover the closure variables to the macroscopic observables, and the macroscopic dynamics are derived with the designed loss $\mathcal{L}_p$.

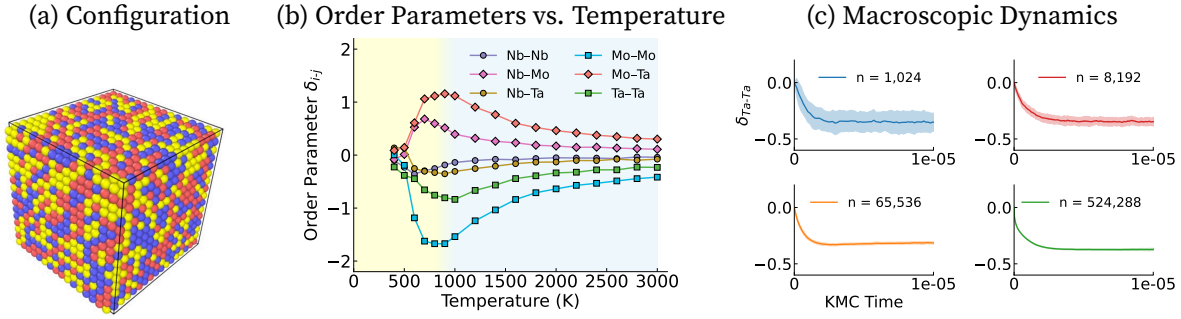

Fig. 2: Results of the NbMoTa equimolar alloy system. (a) Microscopic configuration of NbMoTa alloy with 8,192 atoms. (b) Equilibrium order parameters as a function of temperature, obtained from macroscopic dynamics simulations by the trained SDE model. The microscopic system contains $n = 8192$ atoms. (c) Macroscopic dynamics of $\delta_{\text{Ta-Ta}}$ for microscopic systems of varying sizes when $T = 2000K$. The mean and standard deviations are calculated over 100 trajectories.

Macroscopic dynamics derivation We model the Macroscopic dynamics with SDE:

$$\mathrm{d}\mathbf{z}_t = \boldsymbol{\mu}(\mathbf{z}_t)\mathrm{d}t + \Sigma^{1/2}(\mathbf{z}_t)\mathrm{d}\mathbf{B}_t, \tag{4}$$

where $\boldsymbol{\mu}$ is the drift term and Σ is the diffusion term. In most experiments, we adopt fully connected networks for both $\boldsymbol{\mu}$ and Σ. Existing works train the SDE by minimizing the negative log-likelihood [1, 2, 3, 4]:

$$\begin{aligned}\mathcal{L}[\boldsymbol{\mu}, \Sigma] = \mathbb{E}_{\mathbf{z}_t, \mathbf{z}_{t+\delta t}}[-2\log \\ p\left(\mathbf{z}_{t+\delta t} \,\middle|\, \mathbf{z}_t + \boldsymbol{\mu}(\mathbf{z}_t)\delta t,\, \Sigma(\mathbf{z}_t)\delta t\right)],\end{aligned} \tag{5}$$

where $\mathbf{z}_{t+\delta t}$ denotes the latent state obtained by evolving the full system from $\mathbf{z}_t$ over a time step δt. The conditional distribution p is given by the Gaussian distribution $\mathcal{N}(\mathbf{z}_{t+\delta t}; \mathbf{z}_t + \boldsymbol{\mu}(\mathbf{z}_t)\delta t, \Sigma(\mathbf{z}_t)\delta t)$, obtained by discretizing the SDE with the Euler-Maruyama scheme.

In our setting, however, $\mathbf{z}_{t+\delta t,\mathcal{I}}$ is not obtained by evolving the full system by δt. Instead, it results from a partial evolution over a localized spatial patch. To account for this, we adapt the SDE loss as follows:

$$\begin{aligned}\mathcal{L}_p[\boldsymbol{\mu}, \Sigma] = \mathbb{E}_{\mathbf{z}_t, \mathbf{z}_{t+\delta t,\mathcal{I}}}[-2\log \\ p(\mathbf{z}_{t+\delta t,\mathcal{I}} \,|\, \mathbf{z}_t + \boldsymbol{\mu}(\mathbf{z}_t)\delta t,\, K\Sigma(\mathbf{z}_t)\delta t)].\end{aligned} \tag{6}$$

The only difference between $\mathcal{L}$ and $\mathcal{L}_p$ is that the covariance term in $\mathcal{L}_p$ is multiplied by a factor K, where K denotes the number of patches introduced earlier. We can interpret the influence of K qualitatively. Since $K > 1$ scales the diffusion term, K will reduce the magnitude of the diffusion term. During the data generation of $\mathbf{x}_{t+\delta t,\mathcal{I}}$ from $\mathbf{x}_t$, we introduce additional randomness by performing partial evolution. Therefore, in the derivation of macroscopic dynamics, we multiply the diffusion term by K to correct for the extra stochasticity. We provide a theoretical justification for the loss $\mathcal{L}_p$ under the suitable conditions.

3. Experiments

We empirically validate the accuracy and robustness of our method across various microscopic systems. We first demonstrate our method on a SPDE system and spin systems, and then validate it on a more realistic NbMoTa alloy system. The result of the NbMoTa alloy system is shown in Fig. 2.

Acknowledgments

This research is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG3-RP-2022-028), the Ministry of Education, Singapore, under its funding for the Research Centre of Excellence Institute for Functional Intelligent Materials (Project No. EDUNC-33-18-279-V12), and Academic Research Fund Tier 3 Grant (Project No. MOET32024-0002).

References

[1] Xiaoli Chen, Beatrice W. Soh, Zi-En Ooi, Eleonore Vissol-Gaudin, Haijun Yu, Kostya S. Novoselov, Kedar Hippalgaonkar, and Qianxiao Li. Constructing custom thermodynamics using deep learning. *Nature Computational Science*, 4(1):66–85, 2023.

[2] Felix Dietrich, Alexei Makeev, George Kevrekidis, Nikolaos Evangelou, and Tom Bertalan. Learning effective stochastic differential equations from microscopic simulations: Linking stochastic numerics to deep learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(2):023121, 2023.

[3] Zhen Zhang, Yeonjong Shin, and George Em Karniadakis. Gfinns: Generic formalism informed neural networks for deterministic and stochastic dynamical systems. *Philosophical Transactions of the Royal Society A*, 380(2229):20210207, 2022.

[4] Ting-Ting Gao, Baruch Barzel, and Gang Yan. Learning interpretable dynamics of stochastic complex systems from experimental data. *Nature Communications*, 15(1):6029, 2024.