

A unified framework for data-driven construction of stochastic reduced models with state-dependent memory

Zhiyuan She^{a,1}, Liyao Lyu^{a,b,1}, Huan Lei^{a,c}

^a*Department of Computational Mathematics, Science & Engineering, Michigan State University, 428 S Shaw Ln, East Lansing, MI, USA*

^b*Department of Mathematics, University of California, Los Angeles, 520 Portola Plaza, Los Angeles, CA, USA*

^c*Department of Statistics & Probability, Michigan State University, 619 Red Cedar Road, East Lansing, MI, USA*

Abstract

We present a unified framework for the data-driven construction of stochastic reduced models with state-dependent memory for high-dimensional Hamiltonian systems. The method addresses two key challenges: (i) accurately modeling heterogeneous non-Markovian effects where the memory function depends on the coarse-grained (CG) variables beyond the standard homogeneous kernel, and (ii) efficiently exploring the phase space to sample both equilibrium and dynamical observables for reduced model construction. Specifically, we employ a consensus-based sampling method to establish a shared sampling strategy that enables simultaneous construction of the free energy function and collection of conditional two-point correlation functions used to learn the state-dependent memory. The reduced dynamics is formulated as an extended Markovian system, where a set of auxiliary variables, interpreted as non-Markovian features, is jointly learned to systematically approximate the memory function using only two-point statistics. The constructed model yields a generalized Langevin-type formulation with an invariant distribution consistent with the full dynamics. We demonstrate the effectiveness of the proposed framework on a two-dimensional CG model of an alanine dipeptide molecule. Numerical results on the transition dynamics between metastable states show that accurately capturing state-dependent memory is essential for predicting non-equilibrium kinetic properties, whereas the standard generalized Langevin model with a homogeneous kernel exhibits significant discrepancies.

Keywords:

1. Introduction

The construction of reduced models for high-dimensional dynamical systems poses a fundamental problem in computational mathematics and multiscale modeling. In many real applications related to fluid physics, materials science, and molecular modeling, the direct simulation of the full-dimensional micro-models often becomes prohibitively expensive. Existing approaches often resort to modeling the evolution of a set of coarse-grained (CG) variables to capture the collective dynamics on the resolved scale of interest. Mathematically, this can be viewed as

Email addresses: lyuliyao@math.ucla.edu (Liyao Lyu), leihuan@msu.edu (Huan Lei)

¹These authors contributed equally to this work.

projecting the full dynamics onto a low-dimensional space, yielding an effective reduced model that encodes the effects of the unresolved dynamics. While the heterogeneous multiscale method [9, 10] and equation-free [21, 22] provide rigorous frameworks for bridging micro- and macro-scales, these approaches generally rely on a clear scale separation. For Hamiltonian systems without scale separation, the reduced dynamics generally exhibits memory effects and stochastic behavior due to the unresolved degrees of freedom. For such cases, the generalized Langevin equation (GLE) provides a mathematically principled structure for the reduced model. This model can be derived from the Mori–Zwanzig (MZ) projection formalism [39, 54] and describes the effective dynamics of the CG variables in terms of a conservative force, a memory kernel, and a stochastic noise term that are coupled through a fluctuation–dissipation relation to ensure the thermodynamic consistency.

Despite its formal appeal, the practical construction of GLE-type reduced models presents two major challenges. The first lies in the complexity of the reduced model formulation. In particular, the rigorous MZ-based studies [8, 17, 29, 48, 2, 33] show that the memory term may depend on the CG variables in a complex way where the analytical form is generally unknown. To construct the closed-form reduced models for stochastic simulations, most existing approaches further simplify the memory term into a homogeneous kernel independent of the CG variables, namely, the standard GLE. Several approaches based on the hierarchical construction [6, 34, 52, 35, 19, 53] and data-driven parameterization [27, 5, 3, 30, 41, 20, 28, 15, 24, 47, 42, 51, 49, 50] have been proposed to construct the homogeneous memory kernel. While such models capture non-Markovian effects beyond mean-field approximations, they are generally insufficient to capture the heterogeneous nature of the energy dissipation processes over the CG space. To address this issue, our recent work [14] proposed a data-driven approach for constructing the GLE with a state-dependent memory term and showed that this broadly over-simplified effect can play a crucial role in accurately predicting the non-equilibrium transition dynamics, where the standard GLE fails to reproduce. On the other hand, that study focused on a one-dimensional CG variable, and the approach relies on evaluating three-point correlation functions conditioned on various initial states, which can become computationally expensive for systems with multi-dimensional CG variables.

The second major challenge lies in the efficient exploration and sampling of the phase space with complex energy landscapes. In particular, the energy landscape may consist of multiple metastable states separated by large energy barriers; direct sampling could be trapped in a local minimum. This difficulty has been well recognized for the construction of the free energy function. Several enhanced sampling approaches, such as the umbrella sampling [45, 25], metadynamics [26], adaptive biasing force [7], and temperature accelerated molecular dynamics [1, 36] have been developed. However, the construction of the state-dependent memory term further requires sampling dynamical observables, such as time correlation functions over the CG space. This makes many biased-sampling approaches less suitable. In particular, existing strategies based on the direct estimation of the probability density function (PDF) of the CG variables or thermodynamic perturbations cannot be readily applied. This calls for a sampling strategy that is capable of concurrently collecting both equilibrium samples for free energy construction and unbiased dynamical data for memory estimation in a unified and consistent manner.

In this work, we aim to address the above two challenges through a unified framework for data-driven construction of stochastic reduced models with complex free energy landscapes and state-dependent memory. Specifically, we employ a consensus-based enhanced sampling method [32] to facilitate exploration of the phase space. Unlike most existing approaches for the free energy construction, the present method is gradient-free where the sampling dynamics is driven by

the local residual and therefore independent of the underlying potential of the full model. This feature enables the efficient sampling across stiff energy landscapes and, importantly, supports the concurrent acquisition of equilibrium data (for free energy estimation) and dynamical observables (for memory construction) from a shared set of simulation trajectories. To construct the reduced model, we adopt a data-driven approach through the joint learning of both a set of non-Markovian features and the extended Markovian dynamics of both the CG variables and these features. In contrast to the formulation in Ref. [42], the proposed model enables systematic approximation of state-dependent memory, while introducing a coherent multiplicative noise term that preserves the consistent invariant distribution. Furthermore, the parameterization relies only on two-point correlation functions, in contrast to the three-point statistics required in Ref. [14]. We demonstrate the effectiveness of the proposed approach by constructing a two-dimensional reduced model of an alanine dipeptide molecule in aqueous solution. Numerical results show that the constructed model accurately reproduces correlation functions conditioned on various initial states. More importantly, comparison of the predicted transition times with full MD simulations reveals that accurately modeling the broadly over-simplified state-dependent memory is essential for modeling non-equilibrium collective behaviors, whereas the standard GLE with a homogeneous kernel shows limitations.

2. Methods

2.1. Preliminaries

The full model under consideration is a Hamiltonian system with $2N$ -dimensional phase space vector $\mathbf{Z} = [\mathbf{Q}; \mathbf{P}]$, where $\mathbf{Q}, \mathbf{P} \in \mathbb{R}^N$ represent the position and momentum vector, respectively. With Hamiltonian $H(\mathbf{Z}) = \frac{1}{2}\mathbf{P}^T\mathbf{M}^{-1}\mathbf{P} + V(\mathbf{Q})$, the governing equation follows

$$\dot{\mathbf{Z}} = \mathbf{S}\nabla H(\mathbf{Z}), \quad \mathbf{Z}(0) = \mathbf{Z}_0, \quad (1)$$

where $\mathbf{M} = \text{diag}(\mathbf{M}_1, \dots, \mathbf{M}_N)$ is the constant mass matrix, $V : \mathbb{R}^N \rightarrow \mathbb{R}$ is the potential function and \mathbf{S} is the symplectic matrix. Our goal is to construct a reduced stochastic model that captures the effective dynamics of a set of CG variables $\mathbf{z} = [\mathbf{q}; \mathbf{p}]$ by defining a mapping $\phi : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2m}$, where $\mathbf{q} = \phi^q(\mathbf{Q})$ and $\mathbf{p} = \phi^p(\mathbf{Q}, \mathbf{P})$ represent the reaction coordinates and the corresponding momenta, respectively. By the Koopman operator, the time evolution of \mathbf{z} can be written as $\dot{\mathbf{z}} = \mathcal{L}\mathbf{z}$, where $\mathcal{L}\phi(\mathbf{Z}) = -((\nabla H(\mathbf{Z}_0))^T \mathbf{S} \nabla_{\mathbf{Z}_0})\phi(\mathbf{Z})$ represents the Liouville operator that depends on the full-dimensional phase vector \mathbf{Z} .

To construct the closed form of the reduced model, we follow Zwanzig's projection formalism by defining the projection operator that acts on a phase-space function as a condition expectation with respect to the CG variables $\mathbf{z}(0) = \mathbf{z}$, i.e.,

$$\mathcal{P}_z f := \mathbb{E}[f(\mathbf{Z}) | \phi(\mathbf{Z}) = \mathbf{z}] = \frac{\int \delta(\phi(\mathbf{Z}) - \mathbf{z}) f(\mathbf{Z}) \rho_0(\mathbf{Z}) d\mathbf{Z}}{\int \delta(\phi(\mathbf{Z}) - \mathbf{z}) \rho_0(\mathbf{Z}) d\mathbf{Z}},$$

where $\rho_0(\mathbf{Z}) \propto \exp(-\beta H(\mathbf{Z}))$ represents the equilibrium distribution. By the Duhamel-Dyson formula, the reduced dynamics follows:

$$\dot{\mathbf{z}}(t) = e^{\mathcal{L}t} \mathcal{P}_z \mathcal{L} \mathbf{z}(0) + \int_0^t ds e^{\mathcal{L}(t-s)} \mathcal{P}_z \mathcal{L} e^{\mathcal{Q}_z \mathcal{L} s} \mathcal{Q}_z \mathcal{L} \mathbf{z}(0) + e^{\mathcal{Q}_z \mathcal{L} t} \mathcal{Q}_z \mathcal{L} \mathbf{z}(0),$$

where $\mathcal{Q}_z = \mathbf{I} - \mathcal{P}_z$ and the right-hand-side represents the mean-field, the memory and the fluctuation term that arises from the initial condition, respectively. By assuming that the memory term only depends on the CG coordinates \mathbf{q} and interpreting the fluctuation terms as a stochastic process, the reduced dynamics can be simplified as

$$\begin{aligned}\dot{\mathbf{q}} &= \nabla_{\mathbf{p}} F(\mathbf{p}, \mathbf{q}), \\ \dot{\mathbf{p}} &= -\nabla_{\mathbf{q}} F(\mathbf{p}, \mathbf{q}) + \int_0^t \mathbf{K}(\mathbf{q}(s), t-s) \dot{\mathbf{q}}(s) ds + \mathcal{R}_t,\end{aligned}\quad (2)$$

where $F(\mathbf{q}, \mathbf{p})$ is the total free energy, $\mathbf{K}(\mathbf{q}, t)$ is the memory kernel, and \mathcal{R}_t is the multiplicative noise term related to the memory kernel through the fluctuation-dissipation relation. Specifically, the total free energy $F(\mathbf{q}, \mathbf{p})$ takes the form

$$F(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + \frac{1}{2} \mathbf{p}^\top \mathbf{m}(\mathbf{q})^{-1} \mathbf{p} + \frac{1}{2} \beta^{-1} \ln \det(\mathbf{m}(\mathbf{q})),$$

where $U(\mathbf{q})$ is the conservation free energy with respect to \mathbf{q} and is determined by the marginal density function $\rho_q(\mathbf{q})$, i.e.,

$$U(\mathbf{q}) = -\beta^{-1} \log \rho_q(\mathbf{q}), \quad \rho_q(\mathbf{q}) = \int \delta(\phi_q(\mathbf{Q}) - \mathbf{q}) \rho_0(\mathbf{Q}) d\mathbf{Q}.\quad (3)$$

$\mathbf{m}(\mathbf{q})$ is the position-dependent mass matrix given by

$$\mathbf{m}(\mathbf{q})^{-1} = \beta \int \dot{\phi}_q(\mathbf{Q}) \dot{\phi}_q(\mathbf{Q})^\top \delta(\phi_q(\mathbf{Q}) - \mathbf{q}) \rho_0(\mathbf{Q}) d\mathbf{Q}.\quad (4)$$

The memory kernel $\mathbf{K}(\mathbf{q}, t)$ is defined by the Zwanzig projection operator \mathcal{P}_z , i.e.,

$$\mathbf{K}(\mathbf{q}, t) = \mathcal{P}_z \left[(e^{\mathcal{Q}_z \mathcal{L} t} \mathcal{Q}_z \mathcal{L} \mathbf{p}) (\mathcal{Q}_z \mathcal{L} \mathbf{p})^\top \right].\quad (5)$$

In particular, by simplifying $\mathbf{K}(\mathbf{q}, t)$ into a homogeneous kernel $\mathbf{K}(\mathbf{q}, t) \approx \boldsymbol{\theta}(t)$, Eq. (2) reduces to the standard GLE. In this work, we do not take this assumption and aim to accurately model the state-dependent form $\mathbf{K}(\mathbf{q}, t)$ to capture the heterogeneous energy dissipation in the CG space. Accordingly, the construction of the reduced model (2) lies in two essential challenges: (I) dynamical exploration of the complex energy landscape for efficient sampling of CG observables following $\rho_q(\mathbf{q})$ in Eqs (3)(4)(5); (II) accurate approximation of the memory kernel (5) preserving the state-dependent non-Markovian effects with a coherent noise term.

2.2. Consensus-based enhanced sampling

To efficiently explore the CG space and sample both equilibrium and dynamical observables, we adopt the consensus-based adaptive sampling (CAS) method developed in Ref. [32]. This approach formulates sampling as a residual-driven minimax optimization problem and evolves an interacting particle system according to a McKean-type stochastic differential equation (SDE). Unlike conventional enhanced sampling methods [26, 36], the dynamical exploration of the CAS method is gradient-free and independent of the underlying full potential function. Furthermore, unlike the adaptive sampling strategies [43, 44, 13, 12] for solving high-dimensional partial differential equations, the method does not rely on the free query of arbitrary sampling points and

is well-suited for systems with constrained phase space where thermodynamically accessible regions are unknown *a priori*.

Without loss of generality, let $X(\mathbf{q})$ denote an quantity of interest and $X_{\mathcal{N}}(\mathbf{q})$ its surrogate. The CAS method iteratively solves the following minimax problem

$$\min_{X_{\mathcal{N}}} \max_{\rho} (\mathcal{L}_{\mathcal{N}}, \rho), \quad (6)$$

where the residual is defined as $\mathcal{L}_{\mathcal{N}}(\mathbf{q}) = \|X(\mathbf{q}) - X_{\mathcal{N}}(\mathbf{q})\|^2$, and the maximization objective represents the weighted value under the sampling distribution $\rho(\mathbf{q})$, i.e.,

$$(\mathcal{L}_{\mathcal{N}}, \rho) = \int (\mathcal{L}_{\mathcal{N}}(\mathbf{q}) - \kappa_h^{-1} \ln \rho(\mathbf{q})) \rho(\mathbf{q}) d\mathbf{q}. \quad (7)$$

Here, the second term serves as an entropy regularization. The rationale of this term can be understood by examining the max-problem of Eq. (6), which is convex with the analytical solution

$$\rho^*(\mathbf{q}) := \arg \max_{\rho} (\mathcal{L}_{\mathcal{N}}, \rho) \propto \exp(-\kappa_h \mathcal{L}_{\mathcal{N}}^-(\mathbf{q})), \quad (8)$$

where $\mathcal{L}_{\mathcal{N}}^-(\mathbf{q}) = -\mathcal{L}_{\mathcal{N}}(\mathbf{q})$ and κ_h can be interpreted as the inverse of the temperature. For the low temperature limit ($\kappa_h \rightarrow \infty$), ρ^* concentrates as a Dirac measure at $\mathbf{q}^* = \arg \max \mathcal{L}_{\mathcal{N}}(\mathbf{q})$, suppressing the dynamical exploration. Conversely, the high-temperature limit ($\kappa_h \rightarrow 0$) yields a more diffuse distribution and enables the efficient exploration of the uncharted sample space.

In this work, we choose $X(\mathbf{q}) = -\nabla U(\mathbf{q})$ i.e., the mean force for the free energy, which naturally targets kinetically important regions. Moreover, this choice also enables concurrent collection of data for dynamical observables such as time correlation functions conditioned with various initial states can be conveniently collected; see Remark 1 for further discussion. The mean force $\mathbf{F}(\mathbf{q}^*) = -\nabla_{\mathbf{q}^*} U(\mathbf{q})$ at a target point \mathbf{q}^* can be estimated from the restrained dynamics by introducing a biased quadratic potential to the full MD system, i.e.,

$$\tilde{V}(\mathbf{Q}; k, \mathbf{q}^*) = V(\mathbf{Q}) + V_b(\mathbf{Q}; k, \mathbf{q}^*) \quad V_b(\mathbf{Q}; k, \mathbf{q}^*) = \frac{k}{2} |\phi^q(\mathbf{Q}) - \mathbf{q}^*|^2, \quad (9)$$

where k is the magnitude of the biased potential. $\mathbf{F}(\mathbf{q}^*)$ can be sampled from the equilibrium distribution under biased potential $\rho_b(\mathbf{Q}; k, \mathbf{q}^*) \propto \exp(-\tilde{V}(\mathbf{Q}; k, \mathbf{q}^*))$, i.e.,

$$\begin{aligned} \mathbf{F}(\mathbf{q}^*) &= \lim_{k \rightarrow \infty} \int -\nabla_{\mathbf{Q}} V_b(\mathbf{Q}; k, \mathbf{q}^*) \rho_b(\mathbf{Q}; k, \mathbf{q}^*) d\mathbf{Q} \\ \lim_{k \rightarrow \infty} \int \rho_b(\mathbf{Q}; k, \mathbf{q}^*) d\mathbf{Q} &= \int \delta(\phi_q(\mathbf{Q}) - \mathbf{q}^*) \rho_0(\mathbf{Q}) d\mathbf{Q} := \rho_q(\mathbf{q}^*), \end{aligned} \quad (10)$$

and we refer to Ref. [37] for the proof. Accordingly, the sampling of $\rho_b(\mathbf{Q}; k, \mathbf{q}^*)$ by Eq. (9) provides training data not only for constructing the free energy, but also for computing dynamical quantities. Specifically, for each \mathbf{q}^* , it also provides a natural approach to collecting the training samples conditional on the initial state $\phi_q(\mathbf{Q}) = \mathbf{q}^*$

$$S(\mathbf{q}^*) = \left\{ \mathbf{q}^{(j,l)} = \phi_q(\mathbf{Q}^{(l)}(t_j)) \mid \rho(\mathbf{Q}^{(l)}(0)) = \rho_b(\mathbf{Q}; \mathbf{q}^*) \right\}_{j,l=1}^{N_j, N_l}, \quad (11)$$

which are used to estimate the mass matrix $\mathbf{m}(\mathbf{q})$ via Eq. (4) and the state-dependent correlations for learning the memory term $\mathbf{K}(\mathbf{q}, t)$ discussed in Section 2.3. Therefore, the CAS framework

establishes a shared sampling strategy which provides the training data for all the reduced modeling terms $U(\mathbf{q})$, $\mathbf{m}(\mathbf{q})$, and $\mathbf{K}(\mathbf{q}, t)$. This avoids the need to explicitly evaluate the full orthogonal dynamics $e^{Q\mathcal{L}_t}$ and ensures consistent data collection across modeling terms.

To collect $\mathbf{F}(\mathbf{q}^*)$ and $S(\mathbf{q}^*)$ over the CG space, we need to establish efficient sampling of $\rho^*(\mathbf{q})$ by Eq. (8). However, the analytical solution is only formal since the residual $\mathcal{L}_N(\mathbf{q})$ is unknown *a priori* and the numerical query at an arbitrary point can be computationally expensive or even thermodynamically inaccessible. As a result, common approaches such as Markovian Chain Monte Carlo and Langevin-type algorithms can not be directly used. Instead, the CAS method introduces an interacting particle system $\{\mathbf{q}_i\}_{i=1}^{N_w}$ and adaptively seeks a local quadratic approximation of $\mathcal{L}_N(\mathbf{q})$ near the max-residual region, i.e.,

$$G(\mathbf{q}; \mathcal{M}, \mathcal{V}) = \frac{1}{2}(\mathbf{q} - \mathcal{M})^\top \mathcal{V}^{-1}(\mathbf{q} - \mathcal{M}).$$

Specifically, \mathcal{M} and \mathcal{V} represent the mean and covariance matrix of a weighted empirical distribution of the particle system, i.e.,

$$\begin{aligned} \rho_l(\mathbf{q}) &= \sum_{i=1}^{N_w} \delta(\mathbf{q} - \mathbf{q}_i) \exp(-\kappa_l \mathcal{L}_N^-(\mathbf{q}_i)) / \sum_{i=1}^{N_w} \exp(-\kappa_l \mathcal{L}_N^-(\mathbf{q}_i)) \\ \mathcal{M}[\rho_l] &= \int \mathbf{q} \rho_l(\mathbf{q}) d\mathbf{q} \\ \mathcal{V}[\rho_l] &= \kappa_l \int (\mathbf{q} - \mathcal{M}[\rho_l]) \otimes (\mathbf{q} - \mathcal{M}[\rho_l]) \rho_l(\mathbf{q}) d\mathbf{q}, \end{aligned} \quad (12)$$

where κ_l and κ_t are parameters interpreted as the inverse of the temperature. Following the Laplace's principle, by taking the low temperature limit ($\kappa_l \rightarrow \infty$), $G(\mathbf{q})$ provides a quadratic approximation of $\mathcal{L}_N^-(\mathbf{q})$ near the max-residual point.

To sample $\rho^*(\mathbf{q})$, we treat each particle as a random walker governed by the following McKean-type SDE

$$d\mathbf{q}_i(t) = \frac{1}{\gamma} \nabla_{\mathbf{q}} G(\mathbf{q}_i(t); \mathcal{M}_t, \mathcal{V}_t) dt + \sqrt{\frac{2}{\kappa_h \gamma}} d\mathbf{W}(t), \quad i = 1, \dots, N_w, \quad (13)$$

where $\mathcal{M}_t = \mathcal{M}[\rho_t^i]$, $\mathcal{V}_t = \mathcal{V}[\rho_t^i]$ and ρ_t^i represents the instantaneous weighted empirical distribution of $\{\mathbf{q}_i(t)\}_{i=1}^{N_w}$. Consequently, the mean-field potential $G(\mathbf{q})$ exploits the local quadratic approximation of the residual \mathcal{L}_N under a low-temperature limit and drives the random walkers towards \mathcal{M}_t , which represents the maximal points conditioned on the information exploited so far. Meanwhile, the second stochastic term promotes the exploration of uncharted space, where γ denotes the friction coefficient and $\mathbf{W}(t)$ denotes the standard Brownian motion.

Model (13) modulates the coupling between the exploitation and exploration (i.e., the conservative and stochastic term) via a high temperature κ_h^{-1} consistent with the target distribution $\rho^*(\mathbf{q})$ defined by Eq. (8). Specifically, we show that by choosing $\kappa_t = \kappa_h + \kappa_l$, the distribution will converge to the target one under appropriate conditions.

Proposition 1. Suppose $\mathcal{L}_N^-(\mathbf{q})$ takes a local quadratic approximation in form of $\frac{1}{2}(\mathbf{q} - \mu)^\top \Sigma^{-1}(\mathbf{q} - \mu)$. If the dynamics converge to an invariant distribution, then the stationary density is given by

$$\rho_\infty = \frac{\exp(-\kappa_h \mathcal{L}_N^-(\mathbf{z}))}{\int \exp(-\kappa_h \mathcal{L}_N^-(\mathbf{z})) d\mathbf{z}}, \quad (14)$$

by choosing $\kappa_t = \kappa_l + \kappa_h$.

Proof. Let $\rho_\infty(\mathbf{z})$ denote the invariant distribution of Eq. (13). Then $\rho_\infty(\mathbf{z})$ must be the invariant distribution of the following SDE

$$d\mathbf{q} = -\frac{1}{\gamma} \mathcal{V}_{\kappa_l, \infty}^{-1} (\mathbf{q} - \mathcal{M}_{\kappa_l, \infty}) dt + \sqrt{\frac{2}{\gamma \kappa_h}} d\mathbf{W}_t, \quad (15)$$

where $\mathcal{M}_{\kappa_l, \infty}$ and $\kappa_t^{-1} \mathcal{V}_{\kappa_l, \infty}$ are the mean and the covariance matrix of the re-weighted density $\propto \rho_\infty(\mathbf{z}) e^{-\kappa_l \mathcal{L}_N^-(\mathbf{z})}$. With the fluctuation-dissipation relation for Eq. (15), we can show $\rho_\infty(\mathbf{z})$ follows the Gaussian distribution with mean $\mathcal{M}_{\kappa_l, \infty}$ and covariance matrix $\kappa_h^{-1} \mathcal{V}_{\kappa_l, \infty}$.

Since $\mathcal{L}_N(\mathbf{q}) = \frac{1}{2} (\mathbf{q} - \mu)^\top \Sigma^{-1} (\mathbf{q} - \mu)$ is quadratic, the re-weighted density of a Gaussian distribution $\rho(\mathbf{z}) \sim \mathcal{N}(\mathcal{M}, \mathcal{V})$ remains Gaussian, i.e., $\rho(\mathbf{z}) e^{-\kappa_l \mathcal{L}_N^-(\mathbf{z})} \propto \mathcal{N}(\mathcal{M}_{\kappa_l}, \mathcal{V}_{\kappa_l})$, where \mathcal{M}_{κ_l} and \mathcal{V}_{κ_l} are defined by

$$\begin{aligned} \mathcal{M}_{\kappa_l}(\mathcal{M}, \mathcal{V}) &= (\mathcal{V}^{-1} + \kappa_l \Sigma^{-1})^{-1} (\kappa_l \Sigma^{-1} \mu + \mathcal{V}^{-1} \mathcal{M}), \\ \mathcal{V}_{\kappa_l}(\mathcal{M}, \mathcal{V}) &= (\mathcal{V}^{-1} + \kappa_l \Sigma^{-1})^{-1}. \end{aligned} \quad (16)$$

Therefore, the mean and covariance of the steady-state Gaussian distribution satisfy

$$\begin{aligned} \mathcal{M}_{\kappa_l, \infty} &= (\kappa_h \mathcal{V}_{\kappa_l, \infty}^{-1} + \kappa_l \Sigma^{-1}) (\kappa_l \Sigma^{-1} \mu + \kappa_h \mathcal{V}_{\kappa_l, \infty}^{-1} \mathcal{M}_{\kappa_l, \infty}), \\ \mathcal{V}_{\kappa_l, \infty} &= \kappa_t (\kappa_h \mathcal{V}_{\kappa_l, \infty}^{-1} + \kappa_l \Sigma^{-1})^{-1}. \end{aligned}$$

It is easy to show that by choosing $\kappa_t = \kappa_l + \kappa_h$, $\mathcal{M}_{\kappa_l, \infty}$ and $\mathcal{V}_{\kappa_l, \infty}$ recovers μ and Σ , respectively, and the invariant distribution takes the form

$$\rho_\infty(\mathbf{z}) \sim \mathcal{N}(\mu, \kappa_h^{-1} \Sigma).$$

□

While the proof is only formal, it illustrates the essential idea of the present CAS method. A more rigorous version can be found in Proposition 2.5 of Ref. [32]. Specifically, we show that by choosing the parameter $\kappa_t = \kappa_l + \kappa_h$ for \mathcal{V} in Eq. (12), the empirical distribution of the random walker (13) converges to the target distribution $\rho^*(\mathbf{q})$ with the first and second moments converging at an exponential rate under appropriate conditions. In particular, unlike common enhanced sampling methods, Eq. (13) only depends on the local residual and does not explicitly rely on the underlying atomistic potential function. This unique feature enables us to choose a larger time step and achieve efficient sampling for systems with complex energy landscapes.

Algorithm. 1 summarizes the detailed sampling process.

Remark 1. In principle, the memory term $\mathbf{K}(\mathbf{q}, t)$ can be directly evaluated by approximating the orthogonal dynamics e^{QLt} in Eq. (5) (e.g., see Refs. [17, 33]) where the residual between $\mathbf{K}(\mathbf{q}, t)$ and its surrogate can be incorporated into the adaptive metric $X(\mathbf{q})$. In this work, we only impose the sampling adaptivity based on the free energy by choosing $X(\mathbf{q}) = -\nabla U(\mathbf{q})$. This choice ensures that samples are concentrated in kinetically important regions while simultaneously enabling the collection of dynamical observables for $\mathbf{m}(\mathbf{q})$ and $\mathbf{K}(\mathbf{q}, t)$. This shared sampling strategy enables us to avoid dealing with the full-dimensional orthogonal dynamics e^{QLt} and establish an efficient data-driven construction of $\mathbf{K}(\mathbf{q}, t)$ presented in Sec. 2.3.

Algorithm 1 Consensus-based enhanced sampling.

Require: Initial sampling point $\mathbf{q}_{i,0}$, for $i = 1, \dots, N_w$

Require: Initial NN parameter θ_0

Require: The number of training iterations N_{train}

Require: The number of data collected N_{data} in each training iteration

$j \leftarrow 0, t \leftarrow 0$

$T \leftarrow \lceil \frac{N_{data}}{N_w} \rceil$

while $j < N_{train}$ **do**

while $t \leq T$ **do**

 calculate the mean force $\mathbf{F}(\mathbf{q}_{i,t})$ and the training samples $S(\mathbf{q}_{i,t})$

 calculate the predicted force $\mathcal{F}_\theta(\mathbf{q}_{i,t}) = \nabla_{\mathbf{q}} X_{\mathcal{N}}(\mathbf{q}_{i,t}; \theta_j)$

$L^i \leftarrow \mathcal{L}_{\mathcal{N}}(\mathbf{q}_{i,t})$

$w^i \leftarrow \frac{\exp(\kappa_l L^i)}{\sum_i \exp(\kappa_l L^i)}$

$\mathcal{M}_{t+1} \leftarrow \beta_1 \mathcal{M}_t + (1 - \beta_1) \sum_i \mathbf{q}_{i,t} w^i$

$\mathcal{V}_{t+1} \leftarrow \beta_2 \mathcal{V}_t + (\kappa_l + \kappa_h)(1 - \beta_2) \sum_i (\mathbf{q}_{i,t} - \mathbf{m}_t)^2 w^i$

$\mathcal{M} \leftarrow \frac{\mathcal{M}_{t+1}}{1 - \beta_1}$

$\mathcal{V} \leftarrow \frac{\mathcal{V}_{t+1}}{1 - \beta_2}$

$\mathbf{q}_{i,t+1} \leftarrow \mathbf{q}_{i,t} - \frac{\delta t}{\gamma} (\mathbf{q}_{i,t} - \mathcal{M}) \oplus \mathcal{V} + \sqrt{\frac{2\delta t}{\gamma \kappa_h}} \eta_{i,t}, \eta_{i,t} \sim \mathcal{N}(0, 1)$

$t \leftarrow t + 1$

end while

 Save the training dataset $\mathcal{D}_j = \{\mathbf{q}_{i,t}, \mathbf{F}(\mathbf{q}_{i,t}), S(\mathbf{q}_{i,t})\}_{t=0}^T$

 Optimize θ_{j+1} using the generated training set \mathcal{D}_l for $l = 0, \dots, j$.

$j \leftarrow j + 1$

end while

2.3. Data-driven learning of the state-dependent memory

Let $\mathcal{S} = \{\mathbf{q}^{(k)}, \mathbf{F}(\mathbf{q}^{(k)}), S(\mathbf{q}^{(k)})\}_{k=1}^{N_s}$ and $U(\mathbf{q})$ denote the training set and the learned free energy function obtained from the CAS-based sampling process. The surrogate for the mass matrix $\mathbf{m}(\mathbf{q})$ is constructed by minimizing the empirical discrepancy between the exact definition (4) and the velocity correlation function

$$\mathcal{L}_m = \sum_{k=1}^{N_s} \left\| \mathbf{m}(\mathbf{q}^{(k)}) - \beta^{-1} \mathbf{C}_{vv}(0, \mathbf{q}^{(k)}) \right\|^2, \quad (17)$$

where $\mathbf{C}_{vv}(t, \mathbf{q}^{(k)})$ is estimated from the sample set $S(\mathbf{q}^{(k)})$ introduced in Eq. (11). In this subsection, we focus on the approximation of the memory term $\mathbf{K}(\mathbf{q}, t)$.

To capture the state-dependent non-Markovian nature, we embed $\mathbf{K}(\mathbf{q}, t)$ into an extended Markovian system in terms of $\hat{\mathbf{z}} = [\mathbf{q}; \mathbf{p}; \boldsymbol{\chi}]$, where $\boldsymbol{\chi} = [\chi_1; \dots; \chi_n]$ represents a set of non-Markovian features and the individual feature $\chi_i \in \mathbb{R}^m$ will be learned from the full model as specified later. The extended dynamics takes the general form

$$d \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \\ \boldsymbol{\chi} \end{pmatrix} = \begin{pmatrix} 0 & \mathbf{I} & 0 \\ -\mathbf{I} & \mathbf{J}(\mathbf{q}) & \\ 0 & & \mathbf{\Sigma}(\mathbf{q}) \end{pmatrix} \nabla \hat{F}(\mathbf{q}, \mathbf{p}, \boldsymbol{\chi}) dt + \begin{pmatrix} 0 & 0 & 0 \\ 0 & & \mathbf{\Sigma}(\mathbf{q}) \\ 0 & & \end{pmatrix} d\mathbf{W}_t, \quad (18)$$

where $\hat{F}(\mathbf{q}, \mathbf{p}, \boldsymbol{\chi})$ is the free energy function of the extended system defined by

$$\hat{F}(\mathbf{q}, \mathbf{p}, \boldsymbol{\chi}) = U(\mathbf{q}) + \frac{1}{2} \mathbf{p}^\top \mathbf{m}(\mathbf{q})^{-1} \mathbf{p} + \frac{1}{2} \beta^{-1} \ln \det [\mathbf{m}(\mathbf{q})] + \frac{1}{2} \boldsymbol{\chi}^\top \boldsymbol{\chi}. \quad (19)$$

The matrices $\mathbf{J}(\mathbf{q})$ and $\mathbf{\Sigma}(\mathbf{q})$ govern the energy dissipation and the noise, and are constructed as

$$\mathbf{J}(\mathbf{q}) = \begin{pmatrix} 0 & -\mathbf{H}(\mathbf{q}) \\ \mathbf{H}(\mathbf{q})^\top & \mathbf{\Gamma}(\mathbf{q}) \end{pmatrix}, \quad \mathbf{\Sigma}(\mathbf{q}) = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{D}(\mathbf{q}) \end{pmatrix}, \quad (20)$$

where $\mathbf{H} : \mathbb{R}^m \rightarrow \mathbb{R}^{n \times nm}$ encodes the coupling between the CG momentum and the auxiliary variables and $\mathbf{\Lambda}, \mathbf{D} : \mathbb{R}^m \rightarrow \mathbb{R}^{nm \times nm}$ characterize the dissipation and noise, respectively. In particular, with the proper choice of the multiplicative noise, we can show that the reduced model (18) retains a consistent invariant distribution function of the full MD system.

Proposition 2. By choosing $\mathbf{D}(\mathbf{q})\mathbf{D}(\mathbf{q})^\top = -\beta^{-1} (\mathbf{\Gamma}(\mathbf{q}) + \mathbf{\Gamma}(\mathbf{q})^\top)$, Eq. (18) has an invariant distribution

$$\rho_0(\mathbf{q}, \mathbf{p}, \boldsymbol{\chi}) \propto \frac{1}{(2\pi)^{m/2} \det[\mathbf{m}(\mathbf{q})]^{1/2}} \exp \left[-\beta \left(U(\mathbf{q}) + \frac{1}{2} \mathbf{p}^\top \mathbf{m}(\mathbf{q})^{-1} \mathbf{p} + \frac{1}{2} \boldsymbol{\chi}^\top \boldsymbol{\chi} \right) \right] \quad (21)$$

Proof. For simple notation, we write Eq. (18) as

$$d\hat{\mathbf{z}} = \hat{\mathbf{J}}(\mathbf{q}) \nabla \hat{F}(\hat{\mathbf{z}}) dt + \hat{\mathbf{\Sigma}}(\mathbf{q}) d\mathbf{W}_t.$$

The Fokker-Planck equation is given by

$$\partial_t \rho = \nabla \cdot \left(-\hat{\mathbf{J}}(\mathbf{q}) \nabla \hat{F}(\hat{\mathbf{z}}) \rho + \frac{1}{2} \nabla \cdot \left(\hat{\mathbf{\Sigma}}(\mathbf{q}) \hat{\mathbf{\Sigma}}(\mathbf{q})^\top \rho \right) \right).$$

With $\mathbf{J}(\mathbf{q})$ and $\mathbf{\Sigma}(\mathbf{q})$ taking the form (20), we can show that $\nabla \cdot ((\mathbf{J}(\mathbf{q}) - \mathbf{J}(\mathbf{q})^\top) \nabla g(\hat{\mathbf{z}})) \equiv 0$ for any scalar function g and $\nabla \cdot (\hat{\mathbf{\Sigma}}(\mathbf{q}) \hat{\mathbf{\Sigma}}(\mathbf{q})^\top \rho) = \hat{\mathbf{\Sigma}}(\mathbf{q}) \hat{\mathbf{\Sigma}}(\mathbf{q})^\top \nabla \rho$. By plugging ρ_0 into the right-hand side, we have

$$\begin{aligned} \nabla \cdot \left(-\hat{\mathbf{J}}(\mathbf{q}) \nabla \hat{F}(\hat{\mathbf{z}}) \rho_0 + \frac{1}{2} \nabla \cdot (\hat{\mathbf{\Sigma}}(\mathbf{q}) \hat{\mathbf{\Sigma}}(\mathbf{q})^\top \rho_0) \right) &= \nabla \cdot \left(\beta^{-1} \hat{\mathbf{J}}(\mathbf{q}) \nabla \rho_0 + \frac{1}{2} \hat{\mathbf{\Sigma}}(\mathbf{q}) \hat{\mathbf{\Sigma}}(\mathbf{q})^\top \nabla \rho_0 \right) \\ &= \nabla \cdot \left(\frac{\hat{\mathbf{J}}(\mathbf{q}) + \hat{\mathbf{J}}(\mathbf{q})^\top}{2\beta} \nabla \rho_0 + \frac{1}{2} \hat{\mathbf{\Sigma}}(\mathbf{q}) \hat{\mathbf{\Sigma}}(\mathbf{q})^\top \nabla \rho_0 \right) \\ &\equiv 0 \end{aligned}$$

□

By proposition 2, it is natural to construct $\mathbf{\Gamma}(\mathbf{q})$ by

$$\mathbf{\Gamma}(\mathbf{q}) = -\mathbf{L}(\mathbf{q})\mathbf{L}(\mathbf{q})^\top + \mathbf{J}^a(\mathbf{q}), \quad \mathbf{D}(\mathbf{q}) = \beta^{-1/2}\mathbf{L}(\mathbf{q})$$

where $\mathbf{L}(\mathbf{q})$ is a lower-triangular matrix and $\mathbf{J}^a(\mathbf{a})$ is a skew-symmetric matrix represented by neural networks. As a special case, by taking $\mathbf{H}(\mathbf{q})$ and $\mathbf{\Lambda}(\mathbf{q})$ as constant matrices, the reduced model (18) recovers the standard GLE with a homogeneous kernel [42]. In this study, we do not make such a simplification; the embedded memory enables us to capture the heterogeneous energy dissipation process overlooked in the standard GLE.

To train the reduced model, we establish a joint learning of both the auxiliary variables $\{\chi_i\}_{i=1}^n$ and the functions $\mathbf{H}(\mathbf{q})$ and $\mathbf{\Lambda}(\mathbf{q})$. Specifically, $\chi_i(t)$ essentially serves as a set of non-Markovian features that encode the unresolved dynamics orthogonal to $\mathbf{z}(t) = \mathcal{P}\mathbf{Z}(t)$. However, the direct construction in terms of $\mathbf{Q}\mathbf{Z}(t)$ can be computationally expensive due to the high-dimensionality of $\mathbf{Z}(t)$. Alternatively, one important observation is that the time history of $\mathbf{p}(t)$ naturally encodes the unresolved dynamics and can be readily obtained from the training set \mathcal{S} discussed in Sec. 2.2. Hence, we construct $\chi_i(t)$ by

$$\begin{aligned} \chi_i(t) &= \int_0^\infty \omega_i(t) \mathbf{p}(t - \tau) d\tau \\ &\approx \sum_{j=1}^{N_w} \mathbf{w}_{ij} \mathbf{p}(t - t_j), \quad i = 1, \dots, n, \end{aligned} \tag{22}$$

where $\omega_i \in \mathbb{R}^+ \rightarrow \mathbb{R}^{m \times m}$ is the non-Markovian encoders and $\{\mathbf{w}_{ij}\}_{j=1}^{N_w}$ represents the discrete weights whose values will be determined.

To proceed, we multiply Eq. (18) by $\mathbf{v}(0) := \dot{\mathbf{q}}(0)$ and take the conditional expectation with $\mathbf{q}(0) = \mathbf{q}^*$. The correlation functions follow

$$\begin{aligned} \frac{d}{dt} \langle \hat{\mathbf{z}}(t) \mathbf{v}(0)^\top | \mathbf{q}^* \rangle &= \langle \mathbf{J}(\mathbf{q}(t)) \nabla \hat{F}(\hat{\mathbf{z}}(t)) \mathbf{v}(0)^\top | \mathbf{q}^* \rangle \\ &\approx \mathbf{J}(\mathbf{q}^*) \langle \nabla \hat{F}(\hat{\mathbf{z}}(t)) \mathbf{v}(0)^\top | \mathbf{q}^* \rangle \end{aligned} \tag{23}$$

for $\hat{\mathbf{z}}$ taking \mathbf{p} and χ . In particular, the approximation of the second equation is based on the assumption that the position correlation function $\mathbf{C}_{qq}(t)$ decays much slower than the velocity correlation function $\mathbf{C}_{vv}(t)$ for the CG variables. The separation enables us to efficiently evaluate

Eq. (23) with pre-computed correlation functions rather than the on-the-fly computation from the time-series samples of \mathbf{q} and \mathbf{p} ; see Fig. 4 for the numerical verification of this assumption. By using Eq. (20), we have

$$\frac{d}{dt} \underbrace{\begin{pmatrix} \langle \mathbf{p}(t)\mathbf{v}(0)^\top | \mathbf{q}^* \rangle \\ \langle \chi_1(t)\mathbf{v}(0)^\top | \mathbf{q}^* \rangle \\ \vdots \\ \langle \chi_n(t)\mathbf{v}(0)^\top | \mathbf{q}^* \rangle \end{pmatrix}}_{\mathbf{C}_1(t, \mathbf{q}^*)} + \underbrace{\begin{pmatrix} \langle \nabla_{\mathbf{q}(t)} F(\mathbf{q}, \mathbf{p})\mathbf{v}(0)^\top | \mathbf{q}^* \rangle \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\mathbf{C}_0(t, \mathbf{q}^*)} = \mathbf{J}(\mathbf{q}^*) \underbrace{\begin{pmatrix} \langle \mathbf{v}(t)\mathbf{v}(0)^\top | \mathbf{q}^* \rangle \\ \langle \chi_1(t)\mathbf{v}(0)^\top | \mathbf{q}^* \rangle \\ \vdots \\ \langle \chi_n(t)\mathbf{v}(0)^\top | \mathbf{q}^* \rangle \end{pmatrix}}_{\mathbf{C}_2(t, \mathbf{q}^*)}, \quad (24)$$

where the correlation $\langle \chi_i(t)\mathbf{v}(0)^\top | \mathbf{q}^* \rangle$ can be obtained from the non-Markovian weights \mathbf{w}_{ij} and $\mathbf{C}_{pv}(t; \mathbf{q}^*)$ by Eq. (22). Accordingly, the correlation matrices $\mathbf{C}_0(t, \mathbf{q}^*)$, $\mathbf{C}_1(t, \mathbf{q}^*)$ and $\mathbf{C}_2(t, \mathbf{q}^*)$ can be directly evaluated for each \mathbf{q}^* in the training set. This enables us to establish a joint learning of both the non-Markovian weights $\{\mathbf{w}_{ij}\}_{i,j=1}^{n, N_w}$ and the function $\mathbf{J}(\mathbf{q})$ (i.e., $\mathbf{H}(\mathbf{q})$, $\mathbf{L}(\mathbf{q})$ and $\mathbf{J}^a(\mathbf{q})$) by minimizing the empirical loss function

$$\begin{aligned} L_k &= \lambda_c L_c + \lambda_\Lambda L_\Lambda \\ L_c &= \sum_{k=1}^{N_s} \sum_{j=1}^{N_t} \left\| \frac{d}{dt} \mathbf{C}_1(t_j, \mathbf{q}^{(k)}) + \mathbf{C}_0(t_j, \mathbf{q}^{(k)}) - \mathbf{J}(\mathbf{q}^{(k)}) \mathbf{C}_2(t_j, \mathbf{q}^{(k)}) \right\|^2 \\ L_\Lambda &= \sum_{k=1}^{N_s} \left\| \mathbf{\Lambda}(\mathbf{q}^{(k)}) - \mathbf{I} \right\|^2 \quad \mathbf{\Lambda}(\mathbf{q}^*) = \langle \chi(t)\chi(t)^\top | \mathbf{q}(t) = \mathbf{q}^* \rangle, \end{aligned}$$

where λ_c and λ_Λ are hyperparameters. The second loss term imposes a constraint such that the covariance matrix of the auxiliary variables χ constructed from the full model is close to \mathbf{I} , which is consistent with the pre-assigned free energy function $\hat{F}(\mathbf{q}, \mathbf{p}, \chi)$ and the invariant distribution $\rho_0(\mathbf{q}, \mathbf{p}, \chi)$ of the reduced model specified in Eqs. (19) and (21). We use the ADAM optimizer [23] to train the model and refer to Appendix ?? for the details.

3. Numerical Result

In this section, we examine the effectiveness of the constructed reduced models in comparison with the full MD model. We use SD-GLE to denote the present state-dependent GLE model and SI-GLE to denote the standard state-independent GLE. For fair comparison, both reduced models use the same free energy function $U(\mathbf{Q})$.

3.1. Polymer Chain

Let us start with a polymer chain consisting of $N = 16$ atoms. The full atomistic potential function is governed by standard intramolecular interactions, including non-bonded Lennard-Jones, bond stretching, angular bending, and dihedral torsion potential functions; see Appendix A.1 for details. We choose the CG coordinate as the end-to-end distance defined by the Euclidean distance between the first and last atoms of the polymer chain $q = \|\mathbf{Q}_1 - \mathbf{Q}_N\|$. Accordingly, the mass \mathbf{m} is independent of q since $\langle \dot{q}\dot{q} \rangle = \beta^{-1}(M_1^{-1} + M_N^{-1})$. The training set consists of 25 distinct configurations sampled from the atomistic simulations. The number of auxiliary variables $\{\chi_i\}_{i=1}^n$ is set to be $n = 4$ for both the SI-GLE and SD-GLE models.

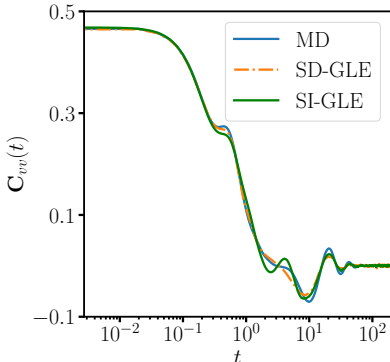


Figure 1: Velocity autocorrelation function $C_{vv}(t)$ obtained from the full MD, the present model (SD-GLE) and the standard GLE (SI-GLE), where the CG coordinate q is defined as the end-to-end distance of a polymer molecule.

First, we examine the overall velocity autocorrelation function defined as $C_{vv}(t) = \langle \dot{q}(t)\dot{q}(0) \rangle$. As shown in Fig. 1, the predictions from both reduced models show good agreement with the full MD results, with the present SD-GLE model yielding improved accuracy in the intermediate regime for $t \sim O(1)$. This result is not surprising since $C_{vv}(t)$ represents a statistical property taking the ensemble average over the full CG space and can be well captured by the standard GLE with a homogeneous memory function.

On the other hand, the differences between the two reduced models become more pronounced by examining the conditional velocity correlation $C_{vv}(t, q^*) = \langle \dot{q}(t)\dot{q}(0) | q(0) = q^* \rangle$. Fig. 2 shows the obtained $C_{vv}(t, q^*)$ for representative values q^* ranging from 3 to 16. The apparent dispersive behavior indicates pronounced heterogeneous energy dissipation manifested by the state-dependent memory. The standard SI-GLE is insufficient to capture such behavior. In contrast, the present SD-GLE shows good agreement with the full MD results.

Finally, we study the rare event behavior for the large extension of the polymer configuration. In particular, we use $\rho(T|q > 15)$ to denote the distribution of the time duration of the polymer with $q > 15$. As shown in Fig. 3, the prediction of the present SD-GLE model shows good agreement with the full MD results. In contrast, the prediction of the standard SI-GLE shows apparent deviations, which indicates the crucial role of the state-dependent memory function for capturing the conformation dynamics of the molecular system on the collective scale.

3.2. Alanine Dipeptide

Next, we consider an alanine dipeptide molecule solvated in 384 water molecules at 300 K. The AMBER99SB force field [18] is used to model the intramolecular and intermolecular interactions, and the TIP3P water model is used to model the solvent; we refer to Appendix A.2 for the details of the MD setup.

We choose the CG coordinates $\mathbf{q} = [\phi, \psi]$, where ϕ and ψ represent the two backbone torsion angles. We use 10 random walkers to sample the configuration space. The initial positions of the walkers at the k -th iteration are designated as the terminal states of the preceding iteration. The inverse low and high temperatures are set to $\kappa_l = 10$ and $\kappa_h = 1$, respectively. For each sample point, we conduct 2 ps of restrained dynamics to calculate the average force and mass, followed by 4 ps of unrestrained dynamics to determine the conditional correlation functions.

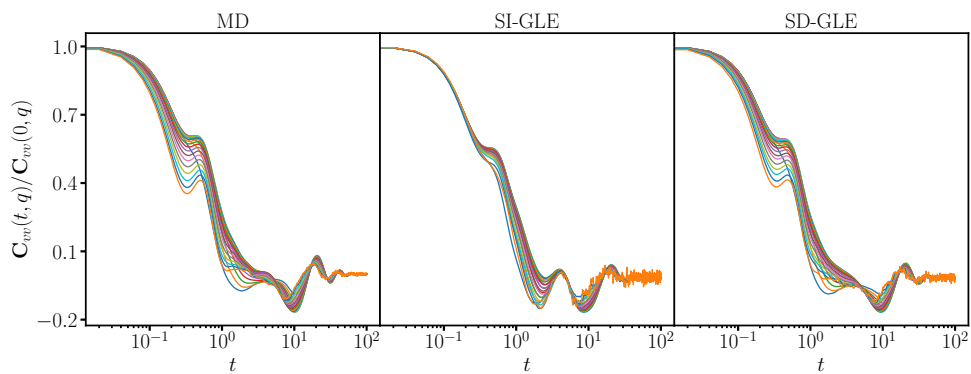


Figure 2: Conditional velocity autocorrelation function $C_{vv}(t, q^*)$ of a polymer molecule with the initial state of the end-to-end distance $q(0) = q^*$. Each curve corresponds to the correlation function taking a specific value ranging between 3 and 16. The dispersive behavior indicates the pronounced state-dependent memory effect over the CG space.

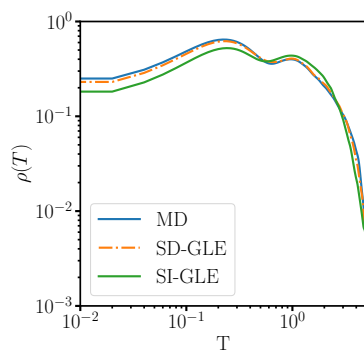


Figure 3: The rare-event distribution of the time duration for the polymer molecule under a large extension with the end-to-end distance $q > 15$.

Fig. 4 shows the constructed free energy function $U(\mathbf{q})$ and the overall correlation functions $C_{qq}(t)$ and $C_{vv}(t)$. The four red points represent the individual metastable states separated by energy barriers where the transition among these states will be systematically investigated. In addition, the position correlation function $C_{qq}(t)$ decays much slower than the velocity correlation $C_{vv}(t)$, which verifies the separation of $\mathbf{J}(\mathbf{q})$ from the conditional correlation function in Eq. (23). Fig. 5 shows the components of the mass matrix $\mathbf{m}(\mathbf{q})$. Unlike the above polymer system, $\mathbf{m}(\mathbf{q})$ exhibits a non-negligible state-dependent nature where the off-diagonal component $\mathbf{m}_{\phi,\psi}$ is about 10% of the diagonal components.

To further examine the localized memory effects, Figs. 6-9 shows the obtained conditional correlation functions $C_{vv}(t, \mathbf{q}^*)$ with \mathbf{q}^* taking the four metastable points shown in Fig. 4. For all the cases, the predictions from the present SD-GLE model show good agreement with the full MD results. In contrast, the standard SI-GLE model yields apparent deviations.

Finally, we examine the transition dynamics among the individual metastable states. To faithfully characterize the non-equilibrium processes, we consider two different transition times between metastable states defined by

$$\begin{aligned} T_{1,A \rightarrow B} &= \inf \{t > 0 \mid \mathbf{q}(t) \in B\}, \quad \mathbf{q}(0) \in A \\ T_{2,A \rightarrow B} &= \inf \{t > 0 \mid \mathbf{q}(t) \in B, \mathbf{q}(s) \notin A, \forall s \in (0, t)\}, \quad \mathbf{q}(0) \in A, \end{aligned}$$

where $T_{1,A \rightarrow B}$ represents the standard first passage time, i.e., the first time the process reaches set B starting from set A. On the other hand, $T_{2,A \rightarrow B}$ is the committor-based transition time [11] that measures the first arrival at B after the system has last exited A, without re-entering A along the way. It characterizes the duration of the actual transition event without the recrossings. Fig. 10 shows the distribution of first passage times (FPTs), defined as the time required for the system to travel from one configuration to another without visiting other intermediate configurations. In addition, we examine the residence time of the system within a metastable state prior to transition, as illustrated in Fig. 11. The SI-GLE model with a spatially homogeneous memory is insufficient to account for the local dynamical environments across the CG space. Consequently, it cannot capture heterogeneous dissipation and memory effects that are crucial for accurately resolving metastable transitions. In contrast, the present SD-GLE model shows consistent improvement in capturing these dynamical behaviors, which demonstrates the crucial role of the state-dependent memory term in predicting the non-equilibrium properties of molecular systems at the CG level.

4. Summary

We have proposed a data-driven framework for constructing stochastic reduced models with state-dependent memory that accurately capture both equilibrium and non-equilibrium dynamics of molecular systems. The resulting SD-GLE model extends the standard GLE formulation by incorporating a memory kernel that varies across the CG space, thereby accounting for heterogeneous energy dissipation and complex metastable transitions. To address the dual challenges of efficient sampling and non-Markovian modeling, we adopt a consensus-based sampling method that enables the efficient exploration of the phase space and the concurrent collection of free energy gradients, mass matrices, and conditional correlation functions through restrained dynamics. This shared-sampling approach makes the construction of the SD-GLE model a natural extension of existing GLE-based workflows. To approximate the state-dependent memory, we formulate an extended Markovian representation by embedding a set of auxiliary variables into

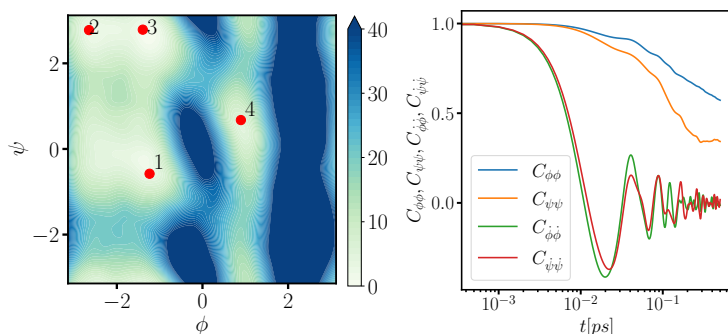


Figure 4: Left: The 2D free energy function $U(\mathbf{q})$ for the reduced model of an alanine dipeptide molecule in solvent, where the CG coordinates $\mathbf{q} = [\phi, \psi]$ denote the two backbone torsion angles of the molecule. The four red points denote the individual metastable states separated by energy barriers. Right: The overall position correlation function $C_{qq}(t)$ and the velocity correlation function $C_{vv}(t)$.

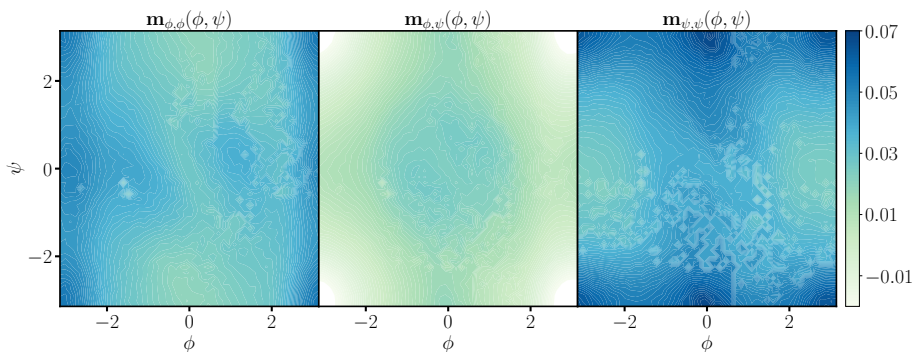


Figure 5: The 2D state-dependent mass matrix $\mathbf{m}(\mathbf{q})$ for the reduced model of an alanine dipeptide molecule in solvent.

the reduced dynamics. The memory term is learned jointly with these non-Markovian features through a conditional correlation matching scheme that requires only two-point statistics.

We demonstrate the accuracy and effectiveness of the proposed SD-GLE model on both a polymer chain and an alanine dipeptide molecule in solvent. The numerical results show clear improvements over standard GLE models in predicting the non-equilibrium properties, especially for the rare events and the transition dynamics. The present framework naturally extends existing workflows based on free energy modeling and provides a systematic framework towards data-driven modeling of state-dependent memory effects in complex molecular systems. Future directions include generalizing the framework to higher-dimensional CG representations and extending the methodology to nonequilibrium systems under external force fields. Also, it could be interesting to incorporate the approximation residual of the memory term into the adaptive sampling process. We leave these for future studies.

References

- [1] Jerry B Abrams and Mark E Tuckerman. Efficient and direct generation of multidimensional free energy surfaces via adiabatic dynamics without coordinate transformations. *The Journal of Physical Chemistry B*, 112(49):15742–

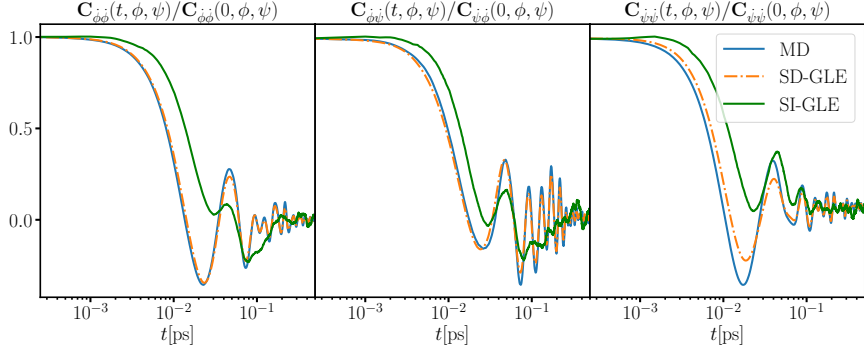


Figure 6: The conditional velocity correlation function $C_{vv}(t, \phi, \psi)$ at a metastable point $\phi = 0.887, \psi = 0.675$.

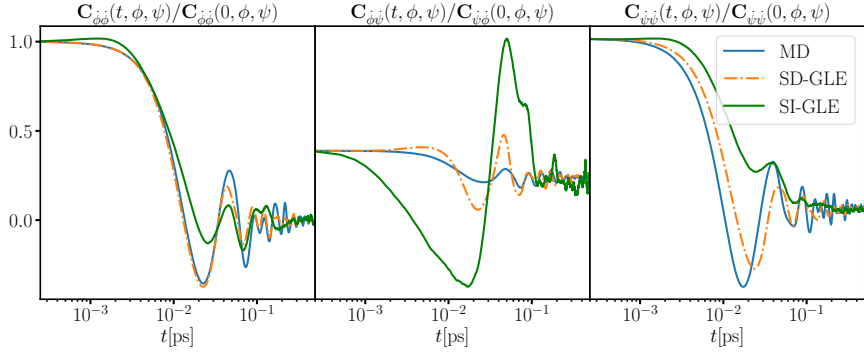


Figure 7: The conditional velocity correlation function $C_{vv}(t, \phi, \psi)$ at a metastable point at $\phi = -1.594, \psi = 2.785$.

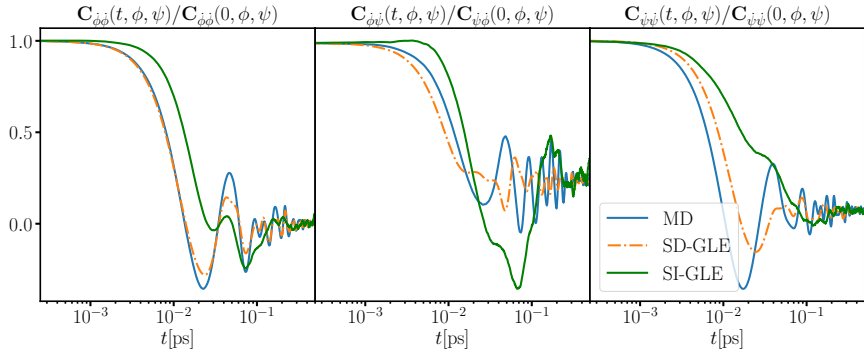


Figure 8: The conditional velocity correlation function $C_{vv}(t, \phi, \psi)$ at a metastable point $\phi = -2.751, \psi = 2.776$.

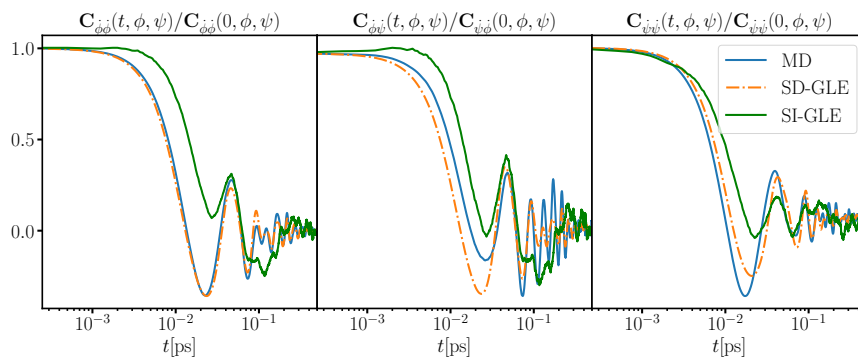


Figure 9: The conditional velocity correlation function $C_{vv}(t, \phi, \psi)$ at a metastable point $\phi = -1.236, \psi = -0.578$.

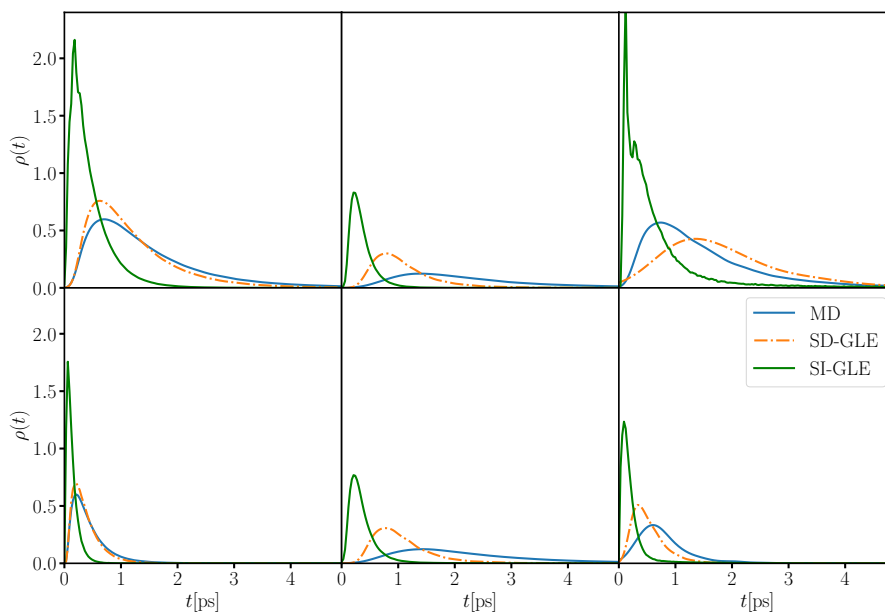


Figure 10: The distribution of first passage times, defined as the time it takes for the molecule to transfer from one configuration to another, without reaching other configurations along the way.

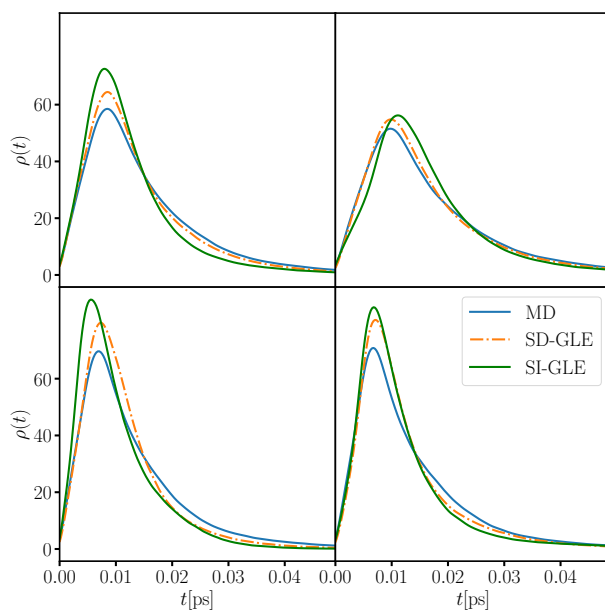


Figure 11: Histogram of the time the system remains in one metastable state before transition.

- 15757, 2008.
- [2] Cihan Ayaz, Laura Scalfi, Benjamin A. Dalton, and Roland R. Netz. Generalized langevin equation with a non-linear potential of mean force and non-linear memory friction from a hybrid projection scheme. *Physical Review E*, 105:054138, 2022.
 - [3] Andrew D Baczewski and Stephen D Bond. Numerical integration of the extended variable generalized Langevin equation with a positive Prony representable memory kernel. *The Journal of chemical physics*, 139(4):044107, 2013.
 - [4] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), 2007.
 - [5] Michele Ceriotti, Giovanni Bussi, and Michele Parrinello. Langevin equation with colored noise for constant-temperature molecular dynamics simulations. *Physical review letters*, 102(2):020601, 2009.
 - [6] Minxin Chen, Xiantao Li, and Chun Liu. Computation of the memory functions in the generalized Langevin models for collective dynamics of macromolecules. *J. Chem. Phys.*, 141:064112, 2014.
 - [7] Eric Darve and Andrew Pohorille. Calculating free energies using average force. *The Journal of Chemical Physics*, 115(20):9169–9183, 2001.
 - [8] Eric Darve, Jose Solomon, and Amirali Kia. Computing generalized Langevin equations and generalized fokker-planck equations. *Proc. Natl. Acad. Sci.*, 106(27):10884–10889, 2009.
 - [9] Weinan E and Bjorn Engquist. The heterogenous multiscale methods. *Commun. Math. Sci.*, 1(1):87–132, 03 2003.
 - [10] Weinan E, Bjorn Engquist, Xiantao Li, Weiqing Ren, and Eric Vanden-Eijnden. Heterogeneous multiscale methods: A review. *Communications in Computational Physics*, 2(3):367–450, 2007.
 - [11] Weinan E and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, 61(1):391–420, 2010.
 - [12] Wenhan Gao and Chunmei Wang. Active learning based sampling for high-dimensional nonlinear partial differential equations. *Journal of Computational Physics*, 475:111848, 2023.
 - [13] Zhiwei Gao, Liang Yan, and Tao Zhou. Failure-informed adaptive sampling for pinns. *SIAM Journal on Scientific Computing*, 45(4):A1971–A1994, 2023.
 - [14] Pei Ge, Zhongqiang Zhang, and Huan Lei. Data-driven learning of the generalized langevin equation with state-dependent memory. *Physical Review Letters*, 133(7):077301, 2024.
 - [15] Francesca Grogan, Huan Lei, Xiantao Li, and Nathan A. Baker. Data-driven molecular modeling with the generalized langevin equation. *J. Comput. Phys.*, 418:109633–109641, 2020.

- [16] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.
- [17] Carmen Hijón, Pep Español, Eric Vanden-Eijnden, and Rafael Delgado-Buscalioni. Mori–Zwanzig formalism as a practical computational tool. *Faraday discuss.*, 144:301–322, 2010.
- [18] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.
- [19] Thomas Hudson and Xingjie H. Li. Coarse-Graining of Overdamped Langevin Dynamics via the Mori–Zwanzig Formalism. *Multiscale Modeling & Simulation*, 18(2):1113–1135, 2020.
- [20] Gerhard Jung, Martin Hanke, and Friederike Schmid. Iterative reconstruction of memory kernels. *Journal of Chemical Theory and Computation*, 13(6):2481–2488, 2017.
- [21] Ioannis G. Kevrekidis, C. William Gear, James M. Hyman, Panagiotis G Kevrekidis, Olof Runborg, and Constantinos Theodoropoulos. Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.*, 1(4):715–762, 12 2003.
- [22] Ioannis G. Kevrekidis and Giovanni Samaey. Equation-free multiscale computation: Algorithms and applications. *Annual Review of Physical Chemistry*, 60(1):321–344, 2009.
- [23] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 12 2015.
- [24] Viktor Klippenstein and Nico F. A. van der Vegt. Cross-correlation corrected friction in (generalized) langevin models. *The Journal of Chemical Physics*, 154(19):191102, 2021.
- [25] Shankar Kumar, John M. Rosenberg, Djamal Bouzida, Robert H. Swendsen, and Peter A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.
- [26] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.
- [27] Oliver F Lange and Helmut Grubmüller. Collective Langevin dynamics of conformational motions in proteins. *J. Chem. Phys.*, 124:214903, 2006.
- [28] Hee Sun Lee, Surl-Hee Ahn, and Eric F. Darve. The multi-dimensional generalized Langevin equation for conformational motion of proteins. *The Journal of Chemical Physics*, 150(17):174113, 2019.
- [29] H. Lei, B. Caswell, and G. E. Karniadakis. Direct construction of mesoscopic models from microscopic simulations. *Phys. Rev. E*, 81:026704, 2010.
- [30] Huan Lei, N. A. Baker, and X. Li. Data-driven parameterization of the generalized Langevin equation. *Proc. Natl. Acad. Sci.*, 113(50):14183–14188, 2016.
- [31] Lindahl, Abraham, Hess, and van der Spoel. Gromacs 2019.2 source code, April 2019.
- [32] Liyao Lyu and Huan Lei. Consensus-based construction of high-dimensional free energy surface. *arXiv preprint arXiv:2311.05009*, 2023.
- [33] Liyao Lyu and Huan Lei. Construction of coarse-grained molecular dynamics with many-body non-markovian memory. *Physical Review Letters*, 131(17):177301, 2023.
- [34] Lina Ma, Xiantao Li, and Chun Liu. The derivation and approximation of coarse-grained dynamics from Langevin dynamics. *The Journal of Chemical Physics*, 145(20):204117, 2016.
- [35] Lina Ma, Xiantao Li, and Chun Liu. Coarse-graining langevin dynamics using reduced-order techniques. *Journal of Computational Physics*, 380:170–190, 2019.
- [36] Luca Maragliano and Eric Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical Physics Letters*, 426(1):168–175, 2006.
- [37] Luca Maragliano and Eric Vanden-Eijnden. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical Physics Letters*, 426(1):168 – 175, 2006.
- [38] Shuichi Miyamoto and Peter A Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.
- [39] Hazime Mori. Transport, collective motion, and Brownian motion. *Progress of Theoretical Physics*, 33(3):423–455, 1965.
- [40] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.
- [41] Antonio Russo, Miguel A Durán-Olivencia, Ioannis G Kevrekidis, and Serafim Kalliadasis. Deep learning as closure for irreversible processes: A data-driven generalized Langevin equation. *arXiv preprint arXiv:1903.09562*, 2019.
- [42] Zhiyuan She, Pei Ge, and Huan Lei. Data-driven construction of stochastic reduced dynamics encoded with non-markovian features. *The Journal of Chemical Physics*, 158(3), 2023.
- [43] Kejun Tang, Xiaoliang Wan, and Qifeng Liao. Adaptive deep density approximation for fokker-planck equations. *Journal of Computational Physics*, 457:111080, 2022.

- [44] Kejun Tang, Jiayu Zhai, Xiaoliang Wan, and Chao Yang. Adversarial adaptive sampling: Unify pinn and optimal transport for the approximation of pdes. *arXiv preprint arXiv:2305.18702*, 2023.
- [45] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.
- [46] Gareth A Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer physics communications*, 185(2):604–613, 2014.
- [47] Hadrien Vroylandt, Ludovic Goudenège, Pierre Monmarché, Fabio Pietrucci, and Benjamin Rotenberg. Likelihood-based non-markovian models from molecular dynamics. *Proceedings of the National Academy of Sciences*, 119(13):e2117586119, 2022.
- [48] Hadrien Vroylandt and Pierre Monmarché. Position-dependent memory kernel in generalized Langevin equations: Theory and numerical estimation. *The Journal of Chemical Physics*, 156(24):244105, 06 2022.
- [49] Pinchen Xie, Roberto Car, and Weinan E. Ab initio generalized langevin equation. *Proceedings of the National Academy of Sciences*, 121(14):e2308668121, 2024.
- [50] Pinchen Xie and Weinan E. Coarse-graining conformational dynamics with multidimensional generalized langevin equation: How, when, and why. *Journal of Chemical Theory and Computation*, 20, 2024.
- [51] Yuanran Zhu, Yu-Hang Tang, and Changho Kim. Learning stochastic dynamics with statistics-informed neural network. *Journal of Computational Physics*, 474:111819, 2023.
- [52] Yuanran Zhu and Daniele Venturi. Faber approximation of the Mori-Zwanzig equation. *Journal of Computational Physics*, 372:694–718, 2018.
- [53] Yuanran Zhu and Daniele Venturi. Generalized langevin equations for systems with local interactions. *Journal of Statistical Physics*, 178:1217, 2020.
- [54] R. Zwanzig. Statistical mechanics of irreversibility. *Lectures in Theoretical Physics*, 3:106–141, 1961.

Acknowledgments

The work is supported in part by the National Science Foundation under Grant DMS-2110981 and the ACCESS program through allocation MTH210005. The authors also acknowledge the support from the Institute for Cyber-Enabled Research at Michigan State University.

Appendix A. Simulation Setup

Appendix A.1. polymer molecule

The polymer molecule is represented as a bead–spring chain consisting of four sub-units, each containing four atoms. The total potential energy is expressed as

$$V_{\text{mol}}(\mathcal{Q}) = \sum_{i \neq j} V_p(\mathcal{Q}_{ij}) + \sum_{i=1}^{N_b} V_b(l_i) + \sum_{i=1}^{N_a} V_a(\theta_i) + \sum_{i=1}^{N_d} V_d(\phi_i), \quad (\text{A.1})$$

where V_p , V_b , V_a , and V_d denote the pairwise, bond, angle, and dihedral interactions, respectively. The detailed functional forms and parameters are given below.

The pairwise interaction V_p is modeled by the truncated Lennard–Jones potential:

$$V_p(\mathcal{Q}) = \begin{cases} 4\varepsilon \left[\left(\frac{\sigma}{\mathcal{Q}} \right)^{12} - \left(\frac{\sigma}{\mathcal{Q}} \right)^6 \right] - 4\varepsilon \left[\left(\frac{\sigma}{\mathcal{Q}_c} \right)^{12} - \left(\frac{\sigma}{\mathcal{Q}_c} \right)^6 \right], & \mathcal{Q} < \mathcal{Q}_c, \\ 0, & \mathcal{Q} \geq \mathcal{Q}_c, \end{cases} \quad (\text{A.2})$$

with parameters $\varepsilon = 0.005$, $\sigma = 1.8$, and cutoff distance $\mathcal{Q}_c = 10.0$.

Bonded interactions are modeled by the finite extensible nonlinear elastic (FENE) potential:

$$V_b(l) = -\frac{k_s l_0^2}{2} \log \left(1 - \frac{l^2}{l_0^2} \right), \quad (\text{A.3})$$

where three different bond types are defined. Within each sub-unit, the atoms 1-2, 3-4 are connected by type 1. The atoms 2-3 are connected by type-2 bond. Finally, the sub-unit groups are connected by type-3. The parameter values are summarized in Table A.1.

Table A.1: Parameters of the FENE bond interactions.

Type	k_s	l_0
1	0.40	1.8
2	0.64	1.6
3	0.32	1.8

Angular interactions are described by a harmonic potential:

$$V_a(\theta) = \frac{k_a}{2}(\theta - \theta_0)^2, \quad (\text{A.4})$$

where two different types are used. Within each sub-unit group, the bond angles formed by 1-2-3 and 2-3-4 are imposed by type-1 potential. The bond angles formed by atoms of different sub-unit groups (e.g. 3-4-5,4-5-6) are imposed by type-2 potential. Parameters are listed in Table A.2.

Table A.2: Parameters of the harmonic angle interactions.

Type	k_a	θ_0 (deg)
1	1.2	114.0
2	1.5	119.7

The dihedral potential is modeled by a multiharmonic expansion:

$$V_d(\phi) = \sum_{n=1}^6 A_n \cos^{n-1}(\phi), \quad (\text{A.5})$$

where two different types are used. Type 1 dihedral potential is imposed to dihedral angles formed by 2-3-4-5, 4-5-6-7, \dots , and Type 2 dihedral angle are imposed to dihedral angles formed by 3-4-5-6, 7-8-9-10, \dots . Parameters are given in Table A.3.

Table A.3: Parameters of the multiharmonic dihedral interactions.

Type	A_1	A_2	A_3	A_4	A_5	A_6
1	0.0673	1.8479	0.0079	-2.2410	-0.0058	0.0051
2	0.1602	-3.9993	0.2483	6.2837	0.0165	-0.0146

Appendix A.2. alanine dipeptide

The MD simulations were performed with GROMACS 2019.2[31] in combination with the open-source PLUMED library [46]. The system was modeled using the Amber99SB force field [18], with the alanine dipeptide solvated in an aqueous environment containing 383 TIP3P water molecules. Periodic boundary conditions were applied in all three spatial directions. A

cutoff of 0.9 nm was used for van der Waals interactions. Long-range electrostatics were treated with the smooth particle-mesh Ewald (PME) method, using a real-space cutoff of 0.9 nm and a reciprocal-space grid spacing of 0.12 nm. The equations of motion were integrated with the leap-frog scheme and a time step of 2 fs. The system temperature was maintained at 300 K using a velocity-rescale thermostat [4] with a relaxation time of 0.2 ps. Pressure was controlled at 1 bar with the Parrinello–Rahman barostat [40] employing a relaxation time of 1.5 ps and an isothermal compressibility of $4.5 \times 10^{-5} \text{ bar}^{-1}$. Bond lengths involving hydrogen atoms were constrained with the LINCS algorithm [16], while the H–O bond lengths and H–O–H angles in water molecules were constrained with the SETTLE algorithm [38]. All simulations were carried out on an Intel(R) Xeon(R) Platinum 8260 CPU.