Generative Model-based Collective Variable Learning with Metastable State Identification in Molecular Dynamics

Liyao Lyu
Department of Computational Mathematics, Science and Engineering
Michigan State University
East Lansing, MI 48824
lyuliyao@msu.edu

Zhiyuan She
Department of Computational Mathematics, Science and Engineering
Michigan State University
East Lansing, MI 48824
shezhiyu@msu.edu

Huan Lei
Department of Computational Mathematics, Science and Engineering
Department of Statistics and Probability
Michigan State University
East Lansing, MI 48824
leihuan@msu.edu

Abstract

We propose a generative model-based framework for learning collective variables (CVs) that faithfully capture the individual metastable states of the full-dimensional molecular dynamics (MD) systems. Unlike most existing approaches based on various feature extraction strategies, the new framework transfers the exhausting efforts of feature selection into a generative task of reconstructing the full-dimensional probability density function (PDF) from a set of CVs under a prior distribution with pre-assigned local maxima. By pairing the CVs with a set of auxiliary Gaussian random variables, we seek an invertible mapping that recovers the full-dimensional PDF and meanwhile, preserves the correspondence between the metastable states of the MD space and individual local maxima of the prior distribution. Through identifying the metastable states within MD space that are generally unknown and imposing the correspondence between the two spaces, the constructed CVs retain clear physical interpretations and provide kinetic insight for the molecular systems on the collective scale. We demonstrate the effectiveness of the proposed method with the alanine dipeptide in the aqueous environment. The constructed CVs faithfully capture the essential metastable states of the full MD systems, which show good agreement with kinetic properties such as the transition from the ballistic to the plateau regime for the mean square displacement. The code is available in https://anonymous.4open.science/r/generative-metastable-F834/README.md.

Preprint. Under review.

1 Introduction

Molecular dynamics (MD) provides a unique computational approach Frenkel and Smit (2023) to probe the micro-scale insights of complex processes in various scientific disciplines related to physics, chemistry, and biology. In many real applications, such as phase transitions, chemical reactions, protein folding, and ligand binding, one essential challenge arises from the prevalence of local minima within the high-dimensional energy landscape. In particular, the energy barriers separating individual metastable states could be much greater than the thermal energy k_BT; direct simulations of the transition processes often become computationally intractable, imposing a severe limitation in predicting many important phenomena on the scale of interest. Existing approaches often resort to various enhanced sampling methods to facilitate the exploration of the energy landscape. Regardless of the detailed form, these approaches generally rely on choosing a small set of collective variables (CVs) defined as a vector-valued function of the high-dimensional MD coordinates, i.e., a coarse-grained representation of the full MD states. By introducing certain biased potentials in the CV space, the MD simulations may efficiently cross the energy barriers and achieve converged statistical properties within a short time. On the other hand, the performance of these methods crucially depends on the specific choice of the CVs, which, however, remains largely empirical.

Ideally, the selected CVs should provide a faithful low-dimensional representation of the full MD system (Bussi and Laio, 2020), i.e. transforming the high-dimensional metastable landscape into separable basins in low dimensions. Traditionally, the choice of CVs mainly relies on physical intuition, which becomes increasingly infeasible for many complex systems. Recent data-driven approaches propose using the covariance function to identify the important directions based on the principal component analysis (PCA) and the time-lagged independent component analysis (TICA). However, the variance-based metric is merely indirectly related to the local minima, and may not explicitly preserve the metastable structures. Similarly, the CVs obtained from auto-encoder approaches may not discriminate the large-variance and metastable nature and generally lack physical interpretation. Alternatively, linear discriminant analysis (LDA) can distinguish metastable states when those states are known in advance, but its applicability could be limited in real applications where the metastable basins are often unknown *a priori*.

To address the above challenges, the central problem is to construct CVs that can resolve all metastable states as distinct basins in the reduced space. However, this further relies on the efficient identification of the local minima within the full-dimensional energy landscape, which, to the best of our knowledge, remains an open problem. In this study, we fill this gap and present a generative-model-based approach for learning the CVs that faithfully inherit the metastable states of the full MD space. Unlike most existing approaches that focus on various feature extractions, our approach is based on a different perspective by pursuing the following question: given a set of CVs with pre-assigned local minima, can we learn an invertible mapping (with a set of auxiliary variables) to reconstruct the full MD probability density function (PDF), and meanwhile pair these local minima in two spaces? This new perspective motivates us to transfer the exhausting efforts of the feature selection task into the joint learning of both the mapping between the CVs and the MD coordinates for the PDF estimation, and the various metastable states in the full MD space. In particular, we choose the normalizing flows to illustrate the essential ideas; other continuous invertible maps can also be used. The constructed CVs retain clear physical interpretation and encode the local minima with explicit PDFs in both the CV and MD space. While this study focuses on CV construction from existing MD data, the method can be naturally extended for enhanced sampling, where the CV learning and biased sampling can be alternated. This work achieves the following main contribution:

- We unify the CV construction and metastable structure preservation into a generative modeling framework, which provides a new perspective distinct from the existing feature extraction based approaches.

- We introduce a novel approach for identifying local energy minima directly from high-dimensional MD data without prior knowledge or additional simulations, which provides kinetic insight into the system's collective behavior.

2 Background and Related work

Molecular dynamics and the metastability. MD is a widely used computational approach to study various physical and chemical phenomena based on micro-scale descriptions. At a high level, MD integrates Newton's equations $M_i\ddot{\mathbf{x}}_i = -\nabla_{\mathbf{x}_i} U(\mathbf{x}_i)$ for each particle i in the configuration space $(\mathbf{x}_1, \cdots, \mathbf{x}_N) \in \mathbb{R}^{3N}$, where M_i is the mass of each atom and $U : \mathbb{R}^{3N} \to \mathbb{R}$ is the potential function. In practice, these equations are modified to include the friction and thermal fluctuations in form of the Langevin equation to model the coupling with the environment at a given temperature, i.e.,

$$\mathrm{d}\mathbf{x}_i = \mathbf{v}_i \mathrm{d}t \quad M_i \mathrm{d}\mathbf{v}_i = -\nabla_{\mathbf{x}_i} U \mathrm{d}t - \gamma \mathbf{v}_i \mathrm{d}t + \sqrt{2\gamma k_B T}\mathrm{d}\mathbf{W}_t, \tag{1}$$

where $\mathbf{v}_i$ is the velocity, γ is the friction constant and $\mathbf{W}_t$ is the Wiener process. Under equilibrium, this process converges to the Boltzmann distribution $p(\mathbf{x}) = {\exp(-U(\mathbf{x})/k_B T)}/{Z}$, where $Z = \int \exp(-U(\mathbf{x})/k_B T)\mathrm{d}\mathbf{x}$ is the normalization constant. Despite its broad application, one essential challenge of the MD simulation is the prevalence of the local minima separated by large energy barriers within the high-dimensional potential landscape $U(\mathbf{x})$. By Kramers' theory (Kramers, 1940), the escape rate over a barrier is proportional to $\exp({-V_b(\mathbf{x})}/{k_B T})$, where V_b is the energy barrier. This implies the escape rate decreases exponentially with respect to the height of the energy barrier, posing a severe challenge to simulate the transition processes among the individual metastable states as a rare event. In recent decades, enhanced sampling methods have become one of the main approaches to accelerate the simulation efficiency. These developed methods are mainly based on the introduction of a set of CVs defined as functions of the atomic coordinates $\mathbf{z} = \eta(\mathbf{x})$, where $\eta : \mathbb{R}^{3N} \to \mathbb{R}^d$ with $d \ll 3N$. These CVs essentially serve as a coarse-grained description of the full MD configuration. Accordingly, a bias potential function in terms of these CVs is introduced into the MD potential to alleviate the energy barriers and facilitate the exploration of the configuration space (Laio and Parrinello, 2002; Dama et al., 2014; Barducci et al., 2008; Valsson and Parrinello, 2014; Huber et al., 1994; Darve and Pohorille, 2001; Wang and Landau, 2001; Hansmann and Wille, 2002; Maragakis et al., 2009; Piana and Laio, 2007; Valsson et al., 2016; Lyu and Lei, 2023b). Therefore, the effectiveness of these methods crucially relies on the selection of proper CVs, which needs to faithfully retain the metastable structure, i.e., encapsulate the difficult-to-sample modes of the full MD space (Bussi and Laio, 2020). However, this CV-selection task remains a challenging problem and largely empirical, particularly when the system is complex and the transition pathways are unidentified.

Related work on data-driven CVs. The task of learning data-driven CVs beyond intuition-based selection aligns with the long-standing problem on the dimension reduction of high-dimensional data (Roweis and Saul, 2000; Tenenbaum et al., 2000; Weinberger and Saul, 2006; Donoho and Grimes, 2003; Maaten and Hinton, 2008). For MD modeling, one common approach is to use the covariance-informed metrics based on PCA (Ichiye and Karplus, 1991; García, 1992; Amadei et al., 1993; Lange et al., 2008; Clarage et al., 1995) and TICA (M. Sultan and Pande, 2017; Schultze and Grubmüller, 2021). However, the obtained CVs correspond to the principal directions of the largest variation and may not inherit the local minimum information. Similarly, methods based on auto-encoder (Wehmeyer and Noé, 2018; Chen and Ferguson, 2018; Ribeiro et al., 2018) and manifold learning (Ceriotti et al., 2011; Das et al., 2006; Ferguson et al., 2010) may not retain the metastable structure and the invertibility; the CVs defined in the latent space often lack physical interpretations. Conversely, the LDA-based methods show the promise of identifying CVs that separate different labeled classes of data da Hora et al. (2024); Sasmal et al. (2023); Bonati et al. (2020); Mendels et al. (2018). However, this approach relies on the prior knowledge or the clustering of metastable states, which often becomes infeasible for high-dimensional complex systems.

In this work, we propose a generative modeling-based approach for both learning the CVs and identifying the metastable states from the full MD samples. Unlike existing methods based on various indirect metrics, the present method directly preserves the metastable structure in the CV space. In particular, we use the framework of normalizing flows (Rezende and Mohamed, 2015) to learn the invertible mapping between the MD and CV space. Rather than the standard Gaussian distribution, we modify the basic space following non-Gaussian prior with pre-assigned local maxima and further impose their correspondence to the metastable regions in the MD space; see Sec. 3 for details.

Preliminary on Normalizing Flows. Normalizing flows (NFs) are learnable invertible functions f, usually represented by a neural network, pushing forward basic space $\mathbf{z}$ with a prior measure $p_z(\mathbf{z})$ towards the target space $\mathbf{x}$ with unknown complex measure (Rezende and Mohamed, 2015; Tabak and Turner, 2013; Dinh et al., 2014, 2016). Using the change of variable rule, these models provide

exact densities of the generated samples $\mathbf{x} = f(\mathbf{z})$, which is given by

$$p(\mathbf{x}) = p_z(f^{-1}(\mathbf{x}))|\det(J_{\mathbf{x}})|, \tag{2}$$

where f^{-1} is the invert map of f and $J_{\mathbf{x}}$ is the Jacobian evaluated at $\mathbf{x}$. Many designs have been proposed to parameterize a family of invertible functions f_θ. The parameters θ are then optimized to maximize the likelihood of the training dataset. NFs enjoy a great interest in physical science for generative modeling and importance sampling (Li and Wang, 2018; Noé et al., 2019; Albergo et al., 2019; Nicoli et al., 2020; Garcia Satorras et al., 2021; Kanwar et al., 2020; Köhler et al., 2020; Klein et al., 2023). However, these works only focus on the PDF estimation and sampling with given variables. In contrast, we focus on learning CVs from the data.

3 Method

3.1 CV construction: Mapping with Non-Gaussian prior

In principle, a set of well-chosen CVs should effectively capture the key features of the high-dimensional MD energy landscape such that the individual basins can be separated in the reduced representation. This motivates us to seek a mapping between the MD and CV space that can pair the local energy minima of the pre-assigned distributions in the two spaces. This *distribution-then-mapping* perspective differs from common methods based on *mapping-then-distribution*, i.e., determining the CV mapping and then estimating their marginal distribution (free energy).

To pair local minima, we note that a collection of neighboring points near a metastable state in full MD space should remain the neighboring points near a metastable state in the CV space. More precisely, if we define a connected set E near a local minimum in MD, $f^{-1}(E)$ should also form a connected set near a local minimum in CV space. Therefore, it is natural to learn a continuous, invertible mapping between the two spaces. From a physical perspective, the mapping essentially forms a convection equation, preserving the proximity of nearby points in the two spaces. In contrast, the Fokker-Planck equation adopted in the diffusion model (Sohl-Dickstein et al., 2015) would cause the mixture of points due to the non-negative entropy introduced by the noise term. Mathematically, a more formal explanation can be understood by the following theorem.

Theorem 3.1. *If f is a continuous invertible mapping of a metric space $\mathcal{X}$ into a metric space $\mathcal{Y}$, and E is a connected subset of X, then $f(E)$ is connected.*

Proof. The proof can be seen in Theorem 4.22 in Rudin (1976), here we present it for completeness. Let us assume, on the contrary, that $f(E)$ can be separated by two sets $A \cup B$, where A and B are nonempty separated subsets of Y. Let

$$G = E \cap f^{-1}(A), \quad H = E \cap f^{-1}(B),$$

then $E = G \cup H$, and neither G nor H is empty. Since $A \subset \overline{A}$ (the closure of A), we have $G \subset f^{-1}(\overline{A})$, since f is continuous and the latter set is closed. It follows that $f(\overline{G}) \subset \overline{A}$. Since $f(H) = B$ and $\overline{A} \cap B$ is empty, we conclude that $\overline{G} \cap H$ is empty. The same argument shows that $G \cap \overline{H}$ is empty. Thus G and H are separated. This contradicts the fact that E is connected. □

By Thm. 3.1, a natural choice is to learn a normalizing flow that preserves both the connectivity and metastable structure, and hence, establishes a correspondence between the high-probability regions in both spaces. To proceed, one crucial challenge is that the local energy minima (i.e., probability maxima) in the MD space are generally unknown *a prior*.

To circumvent this difficulty, a key observation is that we can impose the metastable structure in an inverse way, i.e., through the mapping from the prior base space to the MD space. In contrast to the common methods that first define the CV mapping and then estimate the reduced free energy, we first specify the free energy of the CV space as the prior base distribution and then determine the mapping that preserves the correspondence of the local minima. Without loss of generality, we choose the multimodal prior distribution defined by

$$p_z(\mathbf{z}) \propto \prod_{i=1}^{d} \sum_{k=1}^{K} \frac{\pi_{i,k}}{\sqrt{2\pi\sigma_{i,k}^2}} \exp\left(-\frac{(\mathbf{z}_i - \tilde{\mathbf{z}}_{i,k})^2}{2\sigma_{i,k}^2}\right) \cdot \prod_{i=d+1}^{D} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\mathbf{z}_i^2}{2}\right), \tag{3}$$

where d and D denote the dimension associated with metastable structure and the full space, respectively. $\pi_{i,k}$, $\mathbf{z}_{i,k}$, $\sigma_{i,k}^2$ represent the mixture weights, centers, and variances of the stable modes, respectively. Assuming that $\tilde{\mathbf{z}}_{i,k} \neq \tilde{\mathbf{z}}_{i,l}$ for all $i = 1 \cdots, d$ and $k, l = 1, \cdots, K$, the distribution p_z exhibits $M = K^d$ distinct high-probability regions. We denote the locations of these local maxima by $\hat{\mathbf{z}}^m \in \mathbb{R}^d$ for $m = 1, \cdots, M$. We note that p_z is expressed as a direct product of the full dimensions, which allows the local maxima to be computed independently in each dimension.

With the pre-assigned prior $p_z(\mathbf{z})$, we seek an invertible mapping $\mathbf{x} = f(\mathbf{z})$ to best match the target MD distribution $p(\mathbf{x})$ and the individual local maxima, enabling both the identification and the preservation of metastable states as detailed in following Sec. 3.2. Finally, with the established mapping, we choose the first $\tilde{d}$ components of the inverse map $f^{-1}(\mathbf{x})$ as the CVs.

3.2 Loss function and Training

We construct the invertible mapping $\mathbf{x} = f(\mathbf{z})$ based on the KR-net (Tang et al., 2020) as a variant of the real NVP (Dinh et al., 2016) framework. To learn $f(\mathbf{z})$, the main difficulty arises from the fact that we need to estimate the PDF from the MD samples and simultaneously identify the local maxima, i.e., high-probability regions of this unknown measure in the MD space. Essentially, this requires $f(\mathbf{z})$ can accurately capture both the global PDF $p(\mathbf{x})$ and the local metastable structure. In particular, due to the high dimensionality and the non-smooth nature of the distribution in the MD space, directly computing the Hessian matrix of $p(\mathbf{x})$ is often numerically unstable. To reconcile these challenges, we propose the loss function consisting of three components:

$$\mathcal{L} = \lambda_{\text{global}}\mathcal{L}_{\text{global}} + \lambda_{\text{local}}\mathcal{L}_{\text{local}} + \lambda_{\text{repulsion}}\mathcal{L}_{\text{repulsion}}, \tag{4}$$

where $\lambda_{\text{global}}, \lambda_{\text{local}}, \lambda_{\text{repulsion}}$ are three hyper-parameters.

Global loss To capture the overall shape of the MD distribution, we minimize the cross-entropy between the model density $p(\mathbf{x})$ and the empirical MD samples. Given the observation $\{\mathbf{x}^s\}_{s=1}^S$, the global loss is simply the cross entropy between the empirical samples and $p(\mathbf{x})$ in Eq. (2), i.e.,

$$\begin{aligned}\mathcal{L}_{\text{global}} &= -\frac{1}{S}\sum_{s=1}^{S}\log\left(p\left(\mathbf{x}^s\right)\right) \\ &= -\frac{1}{S}\sum_{s=1}^{S}\left(\log\left(p_z\left(f^{-1}\left(\mathbf{x}^s\right)\right)\right) + \log\left(\left|\det\left(J_{\mathbf{x}^s}\right)\right|\right)\right),\end{aligned} \tag{5}$$

which is equivalent to maximizing the likelihood in standard PDF estimation.

Local loss To identify metastable states in the MD space, we impose the correspondence such that an individual local maximum point $\hat{\mathbf{z}}$ of $p_z(\mathbf{z})$ defined in Eq. (3) remains a local maximum of $p(\mathbf{x})$. This is achieved by a loss term enforcing the local maximal structure, i.e.,

$$\mathcal{L}_{local} = \frac{1}{M}\sum_{m=1}^{M}\|\nabla_{\mathbf{x}}\log p(\hat{\mathbf{x}}^m)\| + \frac{1}{M}\sum_{m=1}^{M}\frac{1}{J}\sum_{j=1}^{J}\max\left(0, \log(p(\hat{\mathbf{x}}^{m,j}) - \log(p(\hat{\mathbf{x}}^m)\right), \tag{6}$$

where $\hat{\mathbf{x}}^m = f(\hat{\mathbf{z}}^m)$ denotes the candidate of a metastable configuration mapped from metastable state $\hat{\mathbf{z}}^m$ in CV space, and $\hat{\mathbf{x}}^{m,j} = f(\hat{\mathbf{z}}^m + \boldsymbol{\epsilon}^j)$ are perturbed neighbors. Here $\boldsymbol{\epsilon}^j$ is sampled from the normal distribution $\mathcal{N}(0, \sigma_{\mathbf{x}})$, where $\sigma_{\mathbf{x}}$ specifies the perturbation range. Thm. 3.1 ensures that the neighboring points $\{\hat{\mathbf{x}}^{m,j}\}_{j=1}^J$ remains a connected set under the mapping f and therefore enables us to impose the metastable structure by the loss (6). Specifically, the first term penalizes non-zero gradients (i.e., $\nabla p(\hat{\mathbf{x}}^m) \approx 0$) to promote $\hat{\mathbf{x}}^m$ as a stationarity point, and the second term penalizes any neighboring points having lower probability than $p(\hat{\mathbf{x}}^m)$, and thereby promote the identification as a candidate of a true local maximal point.

Repulsion loss To prevent mode collapse and facilitate the exploration of the local maxima in the MD space, we introduce a repulsion term

$$\mathcal{L}_{\text{repulsion}} = \sum_{m_1=1}^{M}\sum_{m_2=m_1+1}^{M}\max\left(0, r_c - \|\hat{\mathbf{x}}^{m_1} - \hat{\mathbf{x}}^{m_2}\|\right)^2, \tag{7}$$

where r_c is a threshold value ensuring that mode candidates remain sufficiently separated in configuration space. By combining these three loss functions, we optimize the network representation of $\mathbf{x} = f(\mathbf{z})$ using the Adam optimizer (Kingma and Ba, 2014), as detailed in Appendix A.

3.3 Identification of the local maximal points

In practice, one potential challenge is the number of true local maximal points M_{MD} in the full MD space is unknown *a priori*. Fortunately, this caveat can be efficiently addressed by a post-process. Without loss of generality, we begin by choosing the prior distribution $p_z(\mathbf{z})$ with a redundant set of local maxima points in the CV space, i.e., $M_{\mathrm{MD}} \leq M$. For each obtained candidate $\hat{\mathbf{x}}^m = f(\hat{\mathbf{z}}^m)$, we analyze the one-dimensional marginal distribution of the MD samples along a set of projection directions $\mathbf{v}_i$ for $i = 1, \cdots D$. In particular, $\mathbf{v}_i$ represents the eigenvector of the Hessian at $\hat{\mathbf{x}}^m$, which now can be efficiently computed using the obtained invertible mapping $\mathbf{x} = f(\mathbf{z})$. Specifically, we consider the projected distance

$$(\mathbf{x} - \hat{\mathbf{x}}^m)^T \mathbf{v}_i, \tag{8}$$

and examine the shape of the projected distribution. In particular, if a peak appears near zero within the perturbation range σ_x, we keep the point as the candidate for a true local maximum. Otherwise, the candidate is discarded. We emphasize that the above post-process is computationally efficient since we only examine the 1D distributions. Furthermore, we may re-train $\mathbf{x} = f(\mathbf{z})$ with the updated local loss function (6) using the true local maximal points and repeat the process if necessary.

3.4 Kinetic insights of the MD systems

Besides the CV construction, the present framework essentially provides a unique capability to identify the kinetically important metastable states directly from high-dimensional MD data without additional simulations. This enables us to probe the kinetic insights of the MD systems on the collective scale which may not be captured by other CV construction methods. In this work, we examine the mean square displacement of molecular conformation in Sec. 4.2. Further studies on transition dynamics and rare-event sampling will be pursued in future work.

4 Experiments

4.1 Müller potential

We first illustrate our method using a two-dimensional Müller potential (Müller and Brown, 1979) $U(\mathbf{x})$ with three metastable states separated by high energy barriers. The sample points are collected from the Langevin dynamics in Eq. (1) with $k_B T = 10$. The prior of base space is defined as

$$p(\mathbf{z}_1, \mathbf{z}_2) \propto \sum_{k=1}^{3} \frac{\pi_{1,k}}{\sqrt{2\pi\sigma_{1,k}^2}} \exp\left(-\frac{(\mathbf{z}_1 - \tilde{\mathbf{z}}_{1,k})^2}{2\sigma_{1,k}^2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\mathbf{z}_2^2}{2}\right), \tag{9}$$

where $\pi_1 = (0.5, 0.2, 0.3)$, $\tilde{\mathbf{z}}_1 = (-3, 0, 3)$ and $\sigma_1 = (1.0, 0.5, 0.8)$. In the loss function, the parameters λ_{global}, λ_{local} and $\lambda_{\text{repulsion}}$ are set to 1, 0.1 and 1 respectively. For each training step, we randomly select $J = 1000$ samples as the neighboring points near each local maxima to compute the local loss. As shown in Figure 1, three independent samples from the pre-assigned triple Gaussian prior can be mapped to three distinct metastable regions in the configuration x-space. Notably, the first component $\mathbf{z}_1$ serves as an effective CV separating the different metastable states. Furthermore, the high-probability regions in configuration space corresponds to the maximum points in the prior triple Gaussian distribution. This result reveals a unique capability of the present method. The established mapping $\mathbf{x} = f(\mathbf{z})$ not only identifies the CVs separating the different metastable states but simultaneously pinpoints the most important configurations in the full space, which is not achievable by previous approaches.

4.2 Alanine Dipeptide

We study an alanine dipeptide molecule solvated in an aqueous environment. The training samples of the molecule configuration are collected by MD simulation in a $2.304 \times 2.304 \times 2.304$ nm box containing 386 water molecules. The atomistic interactions of the alanine dipeptide are modeled using the Amber99 (Hornak et al., 2006) force field, and the TIP3P model (Jorgensen et al., 1983) is employed for the water molecules. While the two backbone dihedral angles (ψ, ϕ) are often used

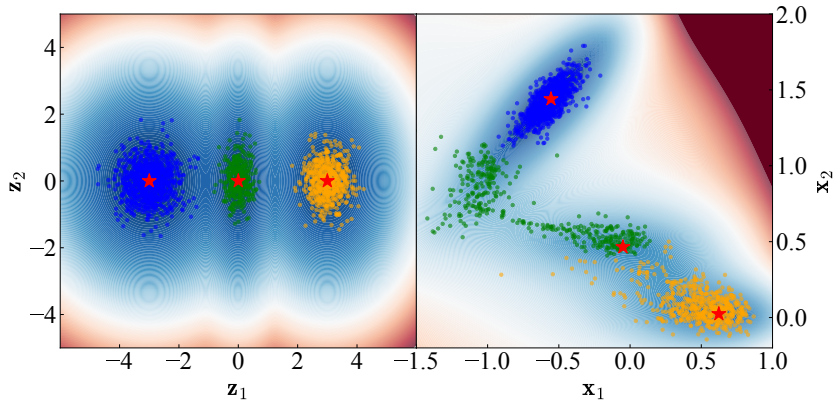

Figure 1: Mapping from a triple Gaussian prior distribution to the Boltzmann distribution of the Müller potential. Left: the negative logrithm of the prior distribution $-\log(p_{\mathbf{z}})$ along with three independent sample sets drawn from the three Gaussian components. Right: the Müller potential and the corresponding points $\mathbf{x} = f(\mathbf{z})$ mapped in the configuration space.

as the CVs for this system, we deliberately exclude this knowledge to test our method's ability to discover CVs from atomic positions alone. The base distribution is defined as a double Gaussian for the first three components (i.e., the CVs of the present method) and the standard Gaussian for the remaining 12 components, corresponding to auxiliary noise-like degrees of freedom.

First, we examine the time auto-correlation $C_{zz}(t) = \langle \mathbf{z}(t)\mathbf{z}(0) \rangle$ to verify if the first three variables $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)$ serve as a set of effective CVs; see Appendix E.3 for details. As shown in Figure 2, the autocorrelation of the first three variables decays much slower than the other variables. This separation of scales provides a hallmark of good CVs indicating that they effectively capture the slow collective motions associated with transitions between metastable states. Also, we evaluate the variance in the configuration space with respect to different values of the latent variables $\mathbf{z}$,

$$\mathcal{V}(\bar{\mathbf{z}}_i, \bar{\mathbf{z}}_j) = \int f(\mathbf{z})^T f(\mathbf{z}) \delta(\mathbf{z}_i - \bar{\mathbf{z}}_i) \delta(\mathbf{z}_j - \bar{\mathbf{z}}_j) p_z(\mathbf{z}) \mathrm{d}\mathbf{z}. \tag{10}$$

As shown in Figure 3, the configuration variance with fixed $(\mathbf{z}_1, \mathbf{z}_2)$ is significantly smaller than the ones with fixed $(\mathbf{z}_4, \mathbf{z}_5)$ and $(\mathbf{z}_5, \mathbf{z}_6)$. This result indicates that $(\mathbf{z}_1, \mathbf{z}_2)$ correspond to stable, slowly varying modes—consistent with their role as effective CVs while other variables primarily capture fast, less relevant fluctuations.

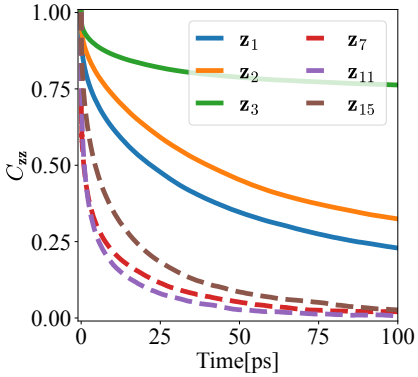

Figure 2: Normalized autocorrelation functions $C_{zz}(t)$ for the learned CVs $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)$ (solid lines) and auxiliary variables $(\mathbf{z}_7, \mathbf{z}_{11}, \mathbf{z}_{15})$ (dashed lines). The CV correlations decay significantly slower than the ones of the auxiliary variables. The scale separation indicates that the CVs capture the dynamics associated with metastable states.

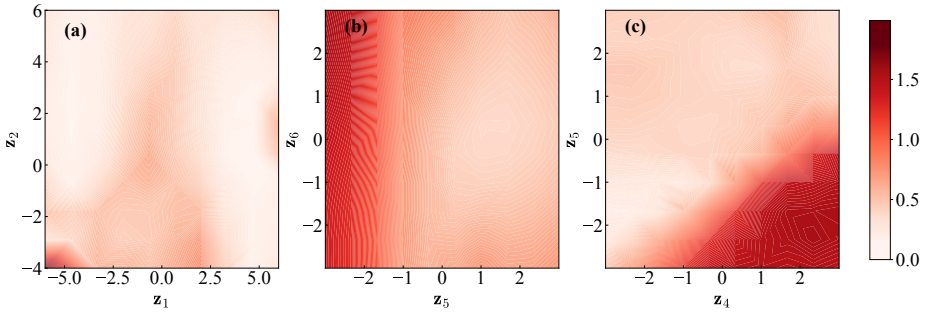

Figure 3: Configuration variance $\mathcal{V}(\bar{\mathbf{z}}_i, \bar{\mathbf{z}}_j)$ defined by Eq. (10) with fixed latent variables $(\mathbf{z}_i, \mathbf{z}_j)$: (a) $\mathbf{z}_1$ and $\mathbf{z}_2$, (b) $\mathbf{z}_5$ and $\mathbf{z}_6$, (c) $\mathbf{z}_4$ and $\mathbf{z}_5$.

Furthermore, we examine the preservation of the metastable structure — a unique capability of the present method. To verify this capability, we collect a set of neighboring points near each of the 8 local probability maxima in the z-space. Then we use the constructed $\mathbf{x} = f(\mathbf{z})$ to map these points back to the configuration space and further compute the corresponding values of two torsional angles $\psi(\mathbf{x})$ and $\phi(\mathbf{x})$. As a comparative study, we use PCA to reduce the dimension and select the first three components as CVs. Since PCA can not identify the local probability maximum points, we use the same eight points obtained by our method. Similarly, we collect a set of neighbor points for each local maximum using the three PCA-CVs and map the neighboring points back to the configuration space, by which we compute the two torsion angles. Figure 4 shows the scatter plot of the points in the ψ-ϕ plane. In particular, each set of neighboring points generated by the present method is located near a free energy local minima for the two torsion angles. In contrast, the individual set of neighboring points obtained by PCA may locate in multiple different local minima. This difference verify the unique capablity of our method in strictly preserving the metastable structure of the full-dimensional energy landscape, where the common methods show limitations.

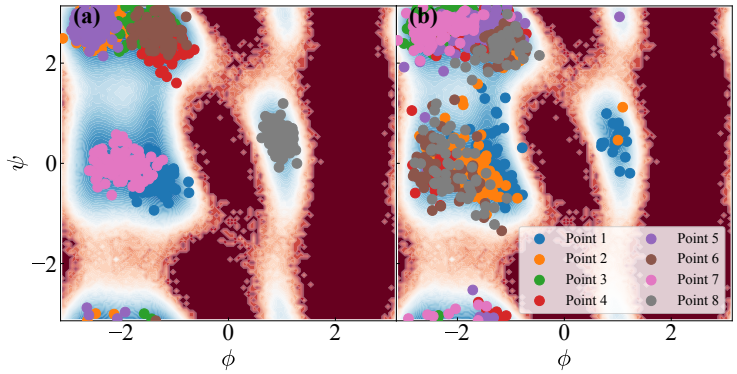

Figure 4: The distribution of the eight sets of neighboring points in the dihedral angle $\phi - \psi$ plane. Each set corresponds to the neighboring points near a local probability maximum in the prior $\mathbf{z}$-space. The points are mapped to the MD space by (a) $\mathbf{x} = f(\mathbf{z})$ of the present method where each set locates in a free energy local minima; (b) PCA using the first three components as CVs to determine the neighboring points, where the individual set may locate in multiple energy local minima.

Finally, we examine if the present method can identify the metastable states of the full MD space. This task is essentially beyond the standard CV construction and paves the way towards quantifying the kinetic insight of the molecular systems on the collective scale. Since the number of true metastable states is generally unknown, we deliberately choose redunt local probability maximal points $\{\hat{\mathbf{z}}^m\}_{m=1}^M$ in the CV space the latent modes and filter invalid candidates as a post-process. Following Sec. 3.3, for each obtained candidate $\hat{\mathbf{x}}^m = f(\hat{\mathbf{z}}^m)$, we project the MD samples to the principal directions $\mathbf{v}_i$ and examining the scalar projection $(\mathbf{x} - \hat{\mathbf{x}}^m)^T \mathbf{v}_i$. If the candidate is a valid metastable point, there exists a peak of the 1D distribution centered at zero within the range of $\hat{\mathbf{x}}^m$. Figure 5 shows 1D distribution of the MD samples projected to the first two principle directions.

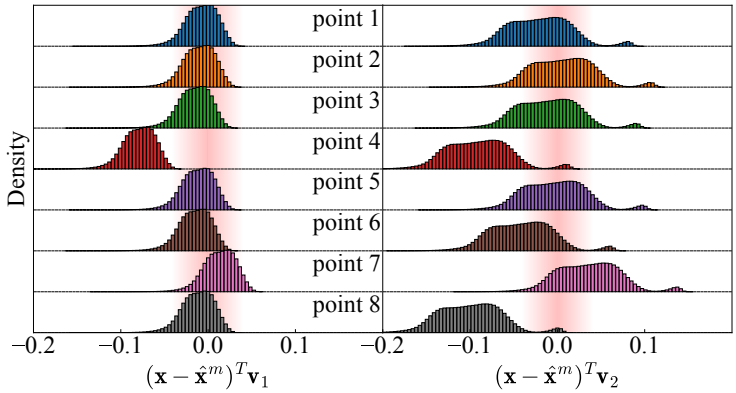

Figure 5: Distribution of the projected distance $(\mathbf{x}-\hat{\mathbf{x}}^m)^T\mathbf{v}$ for each metastable candidate $\hat{\mathbf{x}}^m$, where $\mathbf{v}$ is the principle directions determined by the Hessian of $p(\hat{\mathbf{x}}^m)$: (a) $\mathbf{v}_1$ and (b) $\mathbf{v}_2$ are eigenvectors associated with the first two largest eigenvalue. The peak off the zero region (the red shadow area) for candidate 4 and 7 suggests that these two points are not metastable state in the MD space.

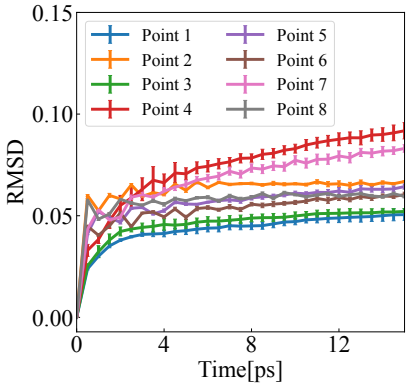

Figure 6: RMSD for molecule configurations initialized from the eight metastable candidates. The curves represent the mean RMSD, with error bars representing the standard deviation over 10 independent trajectories. The RMSDs of point 4 and 7 exhibit a continuous increase, which verify that they are outside of stable basins and are consistent with the distribution analysis in Fig. 5.

In particular, there exits no peak near zero for candidate 4 and 7, suggesting they are outside of a local stable region and should be discarded. In contrast, candidate 8 shows a local maximum in the projected distribution but resides in the second, which implies that it's a valid metastable but could be rarely visited due to the high energy barriers.

We further validate the kinetics of each candidate by running short-time MD simulations and computing the root mean square displacement (RMSD) of trajectories initiated from these points. As shown in Figure 6, most candidates exhibit stable plateau region over time, confirming their metastability nature. In contrast, the RMSD of point 4 and 7 show a continuous increase, indicating that they lie outside stable basins. These results show good agreement with the analysis of projection-based distribution in Fig. 5 and further verify the capability of the present method in identifying the true metastable states of the full MD space.

5 Discussion

In this work, we present a data-driven method for learning CVs from full-dimensional samples based on the generative model. Unlike common methods based on empirical feature extraction strategies, the present method directly preserves the metastable structure of the full MD energy landscape within the low-dimensional CV space. This unique capability is achieved by imposing the non-Gaussian

prior with pre-assigned probability maxima and seeking an invertible mapping that pairs the local maxima between the two spaces, as opposed to the common approaches that start with CV mapping followed by the marginal distribution estimation. Numerical results show that the CVs constructed by the present method can distinguish the different metastable states where the correspondence between the high-probability regions of the two spaces can be faithfully retained. Moreover, the present method provides a novel approach to identifying the metastable stables directly from full-dimensional MD data, which, to the best of our knowledge, is generally inaccessible by current methods.

While this work focuses on the CV construction from a given sample set, the present method can be further coupled with enhanced sampling methods so that the CV learning and phase space exploration can be optimized alternatively. Furthermore, we aim to leverage the constructed CVs to investigate the dynamical behaviors related to conformation relaxation and transition dynamics (Lyu and Lei, 2023a; Ge et al., 2024), enabling a comprehensive understanding of the complex molecule kinetics on the collective scale.

References

Michael S Albergo, Gurtej Kanwar, and Phiala E Shanahan. Flow-based generative models for markov chain monte carlo in lattice field theory. *Physical Review D*, 100(3):034515, 2019.

Andrea Amadei, Antonius BM Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4):412–425, 1993.

Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical review letters*, 100(2):020603, 2008.

Luigi Bonati, Valerio Rizzi, and Michele Parrinello. Data-driven collective variables for enhanced sampling. *The journal of physical chemistry letters*, 11(8):2998–3004, 2020.

Giovanni Bussi and Alessandro Laio. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics*, 2(4):200–212, 2020.

Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1), 2007.

Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences*, 108(32):13023–13028, 2011.

Wei Chen and Andrew L. Ferguson. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *Journal of Computational Chemistry*, 39(25):2079–2102, 2018.

James B Clarage, Tod Romo, B Kim Andrews, B Montgomery Pettitt, and George N Phillips Jr. A sampling problem in molecular dynamics simulations of macromolecules. *Proceedings of the National Academy of Sciences*, 92(8):3288–3292, 1995.

Gabriel CA da Hora, Myongin Oh, John DM Nguyen, and Jessica MJ Swanson. One descriptor to fold them all: Harnessing intuition and machine learning to identify transferable lasso peptide reaction coordinates. *The Journal of Physical Chemistry B*, 128(17):4063–4075, 2024.

James F Dama, Michele Parrinello, and Gregory A Voth. Well-tempered metadynamics converges asymptotically. *Physical review letters*, 112(24):240602, 2014.

Eric Darve and Andrew Pohorille. Calculating free energies using average force. *The Journal of chemical physics*, 115(20):9169–9183, 2001.

Payel Das, Mark Moll, Hernan Stamati, Lydia E Kavraki, and Cecilia Clementi. Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction. *Proceedings of the National Academy of Sciences*, 103(26):9885–9890, 2006.

Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

David L Donoho and Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591–5596, 2003.

Ulrich Essmann, Lalith Perera, Max L Berkowitz, Tom Darden, Hsing Lee, and Lee G Pedersen. A smooth particle mesh ewald method. *The Journal of chemical physics*, 103(19):8577–8593, 1995.

Andrew L. Ferguson, Athanassios Z. Panagiotopoulos, Pablo G. Debenedetti, and Ioannis G. Kevrekidis. Systematic determination of order parameters for chain dynamics using diffusion maps. *Proceedings of the National Academy of Sciences*, 107(31):13597–13602, 2010.

Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

Angel E García. Large-amplitude nonlinear motions in proteins. *Physical review letters*, 68(17):2696, 1992.

Victor Garcia Satorras, Emiel Hoogeboom, Fabian Fuchs, Ingmar Posner, and Max Welling. E(n) equivariant normalizing flows. *Advances in Neural Information Processing Systems*, 34:4181–4192, 2021.

Pei Ge, Zhongqiang Zhang, and Huan Lei. Data-driven learning of the generalized langevin equation with state-dependent memory. *Phys. Rev. Lett.*, 133:077301, 2024.

Ulrich HE Hansmann and Luc T Wille. Global optimization by energy landscape paving. *Physical review letters*, 88(6):068105, 2002.

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.

Thomas Huber, Andrew E Torda, and Wilfred F Van Gunsteren. Local elevation: a method for improving the searching properties of molecular dynamics simulation. *Journal of computer-aided molecular design*, 8:695–708, 1994.

Toshiko Ichiye and Martin Karplus. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Structure, Function, and Bioinformatics*, 11(3):205–217, 1991.

William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

Gurtej Kanwar, Michael S Albergo, Denis Boyda, Kyle Cranmer, Daniel C Hackett, Sébastien Racaniere, Danilo Jimenez Rezende, and Phiala E Shanahan. Equivariant flow-based sampling for lattice gauge theory. *Physical Review Letters*, 125(12):121601, 2020.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Leon Klein, Andrew Foong, Tor Fjelde, Bruno Mlodozeniec, Marc Brockschmidt, Sebastian Nowozin, Frank Noé, and Ryota Tomioka. Timewarp: Transferable acceleration of molecular dynamics by learning time-coarsened dynamics. *Advances in Neural Information Processing Systems*, 36: 52863–52883, 2023.

Jonas Köhler, Leon Klein, and Frank Noé. Equivariant flows: exact likelihood generative learning for symmetric densities. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org, 2020.

Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *physica*, 7(4):284–304, 1940.

Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the national academy of sciences*, 99(20):12562–12566, 2002.

Oliver F Lange, Nils-Alexander Lakomek, Christophe Fares, Gunnar F Schroder, Korvin FA Walter, Stefan Becker, Jens Meiler, Helmut Grubmuller, Christian Griesinger, and Bert L De Groot. Recognition dynamics up to microseconds revealed from an rdc-derived ubiquitin ensemble in solution. *science*, 320(5882):1471–1475, 2008.

Shuo-Hui Li and Lei Wang. Neural network renormalization group. *Physical review letters*, 121(26): 260601, 2018.

Lindahl, Abraham, Hess, and van der Spoel. Gromacs 2019.2 source code, April 2019. URL `https://doi.org/10.5281/zenodo.2636382`.

Liyao Lyu and Huan Lei. Construction of coarse-grained molecular dynamics with many-body non-markovian memory. *Phys. Rev. Lett.*, 131:177301, 2023a.

Liyao Lyu and Huan Lei. Consensus-based construction of high-dimensional free energy surface. *arXiv preprint arXiv:2311.05009*, 2023b.

Mohammad M. Sultan and Vijay S. Pande. tICA-metadynamics: Accelerating metadynamics by using kinetically selected collective variables. *Journal of Chemical Theory and Computation*, 13 (6):2440–2447, 2017.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

Paul Maragakis, Arjan van der Vaart, and Martin Karplus. Gaussian-mixture umbrella sampling. *The Journal of Physical Chemistry B*, 113(14):4664–4673, 2009.

Dan Mendels, GiovanniMaria Piccini, and Michele Parrinello. Collective variables from local fluctuations. *The journal of physical chemistry letters*, 9(11):2776–2781, 2018.

Klaus Müller and Leo D Brown. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica chimica acta*, 53:75–93, 1979.

Kim A Nicoli, Shinichi Nakajima, Nils Strodthoff, Wojciech Samek, Klaus-Robert Müller, and Pan Kessel. Asymptotically unbiased estimation of physical observables with neural samplers. *Physical Review E*, 101(2):023304, 2020.

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

Stéphane Nonnenmacher. Lecture notes for the course introduction to spectral theory. 2021.

Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics*, 52(12):7182–7190, 1981.

Stefano Piana and Alessandro Laio. A bias-exchange approach to protein folding. *The journal of physical chemistry B*, 111(17):4553–4559, 2007.

Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

João Marcelo Lamim Ribeiro, Pablo Bravo, Yihang Wang, and Pratyush Ribeiro. Reweighted autoencoded variational bayes for enhanced sampling (rave). *The Journal of Chemical Physics*, 149(7):072301, 2018.

Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill, 1976.

Subarna Sasmal, Martin McCullagh, and Glen M Hocky. Reaction coordinates for conformational transitions using linear discriminant analysis on positions. *Journal of chemical theory and computation*, 19(14):4427–4435, 2023.

Steffen Schultze and Helmut Grubmüller. Time-lagged independent component analysis of random walks and protein dynamics. *Journal of Chemical Theory and Computation*, 17(9):5766–5776, 2021.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France, 07–09 Jul 2015. PMLR.

Esteban G Tabak and Cristina V Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013.

Keju Tang, Xiaoliang Wan, and Qifeng Liao. Deep density estimation via invertible block-triangular mapping. *Theoretical and Applied Mechanics Letters*, 10(3):143–148, 2020.

Kejun Tang, Xiaoliang Wan, and Qifeng Liao. Adaptive deep density approximation for fokker-planck equations. *Journal of Computational Physics*, 457:111080, 2022.

Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

Omar Valsson and Michele Parrinello. Variational approach to enhanced sampling and free energy calculations. *Physical review letters*, 113(9):090601, 2014.

Omar Valsson, Pratyush Tiwary, and Michele Parrinello. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. *Annual review of physical chemistry*, 67 (1):159–184, 2016.

Xiaoliang Wan and Kejun Tang. Augmented krnet for density estimation and approximation. *arXiv preprint arXiv:2105.12866*, 2021.

Fugao Wang and David P Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical review letters*, 86(10):2050, 2001.

Christoph Wehmeyer and Frank Noé. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *The Journal of Chemical Physics*, 148(24):241703, 2018.

Kilian Q Weinberger and Lawrence K Saul. Unsupervised learning of image manifolds by semidefinite programming. *International journal of computer vision*, 70:77–90, 2006.

A Network Architecture

Following the framework of RealNVP Dinh et al. (2016) and KRnet Tang et al. (2020), we seek an invertible map $\mathbf{x} = f(\mathbf{z})$ with a tractable Jacobian. To learn the map, we construct its inverse $\mathbf{z} = g(\mathbf{x})$ as a composition of a sequence of simpler invertible transformations:

$$g = g^{(L)} \circ g^{(L-1)} \circ \cdots \circ g^{(1)}, \tag{11}$$

where each map $g^{(l)}$ updates a block of coordinates by a composition of transformations:

$$g^{(l)} = \mathcal{Z}^{(l)} \circ \mathcal{A}^{(l)} \circ \mathcal{S}^{(l)} \circ \mathcal{R}^{(l)}, \tag{12}$$

In what follows, we describe each transformation in detail. The variables $\mathbf{x}$ and $\mathbf{y}$ in the equations below refer to intermediate latent representations, passed sequentially through the transformation pipeline. They do not correspond to the original input $\mathbf{z}$ or final output $\mathbf{x}$.

- $\mathcal{R}^{(l)}$: a learnable linear transformation that mixes all latent dimensions. It is parameterized via LU decomposition as

  $$\mathcal{R}^{(l)}(\mathbf{x}) = W^{(l)}\mathbf{x}, \quad \text{with} \quad W^{(l)} = \text{diag}(L^{(l)}, I) \cdot \text{diag}(U^{(l)}, I)$$

  where $L^{(l)}$ is a strictly lower triangular matrix with trainable off-diagonal entries and unit diagonal and $U^{(l)}$ is an upper triangular matrix with learnable entries. This structure ensures efficient computation of the Jacobian determinant:

  $$\log \left| \det \left(\frac{\partial \mathcal{R}^{(l)}}{\partial \mathbf{x}} \right) \right| = \sum_{i=1}^{d} \log |U_{ii}^{(l)}|.$$

- $\mathcal{S}^{(l)}$: The Scale-and-Bias layer is a data-dependent normalization layer used to improve the conditioning and training stability of normalizing flows, particularly deep architectures

  $$\mathbf{y} = \mathcal{S}^{(l)}(\mathbf{x}) = \mathbf{a}^{(l)} \odot \mathbf{x} + \mathbf{b}^{(l)}, \tag{13}$$

where $\mathbf{a}^{(l)}, \mathbf{b}^{(l)} \in \mathbb{R}^d$ are learned scale and bias parameters, and $\odot$ denotes elementwise multiplication. These parameters are initialized using the first minibatch of training data:

$$\mathbf{a}^{(l)} = \frac{s}{\boldsymbol{\sigma} + \varepsilon}, \qquad \mathbf{b}^{(l)} = -\boldsymbol{\mu}, \tag{14}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are the per-dimension mean and standard deviation of the input batch, s is a fixed scaling constant (e.g., 1.0), and ε is a small constant to prevent numerical instability.

Once initialized, $\mathbf{a}^{(l)}$ and $\mathbf{b}^{(l)}$ remain fixed as learnable parameters and are updated via gradient descent. The inverse transformation is given by:

$$\mathbf{x} = (\mathcal{S}^{(l)})^{-1}(\mathbf{y}) = \frac{\mathbf{y} - \mathbf{b}^{(l)}}{\mathbf{a}^{(l)}}. \tag{15}$$

The log-determinant of the Jacobian of $\mathcal{S}^{(l)}$ is easily computed as:

$$\log \left| \det \left(\frac{\partial \mathcal{S}^{(l)}}{\partial \mathbf{x}} \right) \right| = \sum_{i=1}^{d} \log |\mathbf{a}_i^{(l)}|. \tag{16}$$

- $\mathcal{A}^{(l)}$: The affine coupling layer transforms part of the input vector conditioned on the other part while preserving invertibility and allowing expressive nonlinear mappings. Let the input be partitioned as $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2] \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, where $d_1 + d_2 = d$. Define two neural networks $s(\cdot), t(\cdot) : \mathbb{R}^{d_1} \to \mathbb{R}^{d_2}$, realized by multi-layer perceptrons (MLPs). The transformation is given by:

$$[\mathbf{y}_1, \mathbf{y}_2] = \mathcal{A}^{(l)}(\mathbf{x}_1, \mathbf{x}_2) = [\mathbf{x}_1,\ \mathbf{x}_2 \odot (1 + \alpha \tanh(s(\mathbf{x}_1))) + \boldsymbol{\gamma} \tanh(t(\mathbf{x}_1))], \tag{17}$$

where $\alpha \in \mathbb{R}$ is a fixed scalar (e.g., $\alpha = 0.6$), and $\boldsymbol{\gamma} \in \mathbb{R}^{d_2}$ is a learnable scaling vector shared across training samples.

The transformation is invertible as long as $1 + \alpha \tanh(s_j(\mathbf{x}_1)) \neq 0$ for all j, and the inverse is computed as:

$$\mathbf{x}_2 = \frac{\mathbf{y}_2 - \boldsymbol{\gamma} \tanh(t(\mathbf{x}_1))}{1 + \alpha \tanh(s(\mathbf{x}_1))}. \tag{18}$$

Since the Jacobian of $\mathcal{A}^{(l)}$ is triangular, the log-determinant simplifies to:

$$\log \left| \det \left(\frac{\partial \mathcal{A}^{(l)}}{\partial \mathbf{x}} \right) \right| = \sum_{j=1}^{d_2} \log |1 + \alpha \tanh(s_j(\mathbf{x}_1))|. \tag{19}$$

- $\mathcal{Z}^{(l)}$: The squeezing layer reduces the active dimensionality of the input vector in a reversible manner. Let the input at stage l be $\mathbf{x} \in \mathbb{R}^d$, and let the squeezing layer remove $c^{(l)}$ dimensions. Partition the input as:

$$\mathbf{x} = [\mathbf{x}_{\text{active}},\ \mathbf{x}_{\text{inactive}}], \quad \mathbf{x}_{\text{inactive}} \in \mathbb{R}^{c^{(l)}},\ \mathbf{x}_{\text{active}} \in \mathbb{R}^{d - c^{(l)}}. \tag{20}$$

The squeezing operation retains only the active part:

$$\mathcal{Z}^{(l)}(\mathbf{x}) = \mathbf{x}_{\text{active}}, \quad \text{and stores } \mathbf{x}_{\text{inactive}}. \tag{21}$$

At the inverse stage, the stored variables are reattached:

$$(\mathcal{Z}^{(l)})^{-1}(\mathbf{x}_{\text{active}}) = [\mathbf{x}_{\text{active}},\ \mathbf{x}_{\text{inactive}}]. \tag{22}$$

Since this transformation is a coordinate projection and reordering, the Jacobian is identity (or a permutation matrix), and:

$$\log \left| \det \left(\frac{\partial \mathcal{Z}^{(l)}}{\partial \mathbf{x}} \right) \right| = 0. \tag{23}$$

Further details on the construction and implementation of each layer can be found in (Tang et al., 2020; Wan and Tang, 2021; Tang et al., 2022).

B Algorithms

We train the invertible map $\mathbf{z} = g(\mathbf{x})$ discussed in Appendix A by minimizing a loss function that balances the maximum-likelihood estimation and the preservation of the local metastable structure. As discussed in Sec. 3, the loss function consists of three terms. The global loss term represents the PDF estimation of the MD samples. The local loss term preserves the local probability maximal structure under the mapping. Finally, a repuslion loss term is introduced to prevent the collapse of the individual local maximum points; see Eqs. (4) (5) (6) (7) for the detailed formulation. For each training step, the computation of the loss term is outlined in Algorithm 1. The training details are presented in Appendix C.

Algorithm 1 Computation of Loss function $\mathcal{L}$

Input: invertible flow g {Defined in Section A},
training batch $\mathcal{D} = \{\mathbf{x}_i\}_{i=1}^N$, $\mathbf{x}_i \in \mathbb{R}^d$, noise scale σ,
maxima on prior $\mathbf{z}_{\text{max}} = \{\mathbf{z}_{\text{max}}^m\}_{m=1}^M$, $\mathbf{z}_{\text{max}}^m \in \mathbb{R}^d$,
prior density $p_{\mathbf{z}}$, weights λ_{local},$\mathcal{L}_{\text{repel}}$, number of neighbours J,
minimum separation threshold δ
Output: Total loss $\mathcal{L}_{\text{total}}$

$\{\mathbf{z}_i\}_{i=1}^N \leftarrow g(\mathcal{D})$
$\mathcal{L}_{\text{NLL}} \leftarrow -\frac{1}{N}\sum_{i=1}^{N}\left(\log p_{\mathbf{z}}(\mathbf{z}_i) + \log\left|\det\left(\frac{\partial \mathbf{z}_i}{\partial \mathbf{x}_i}\right)\right|\right)$
$\mathcal{L}_{\text{grad}} \leftarrow 0$
$\mathcal{L}_{\text{contrast}} \leftarrow 0$
for $m = 1$ to M **do**
  $\mathbf{x}_{\text{max}}^m \leftarrow g^{-1}(\mathbf{z}_{\text{max}}^m)$
  $g_m \leftarrow \left\|\nabla_{\mathbf{x}} \log p_z(g(\mathbf{x}_{\text{max}}^m)) + \log\left|\det\left(\frac{\partial \mathbf{z}_{\text{max}}^m}{\partial \mathbf{x}_{\text{max}}^m}\right)\right|\right\|$
  Generate J perturbed neighbors: $\hat{\mathbf{x}}^{(m,j)} \leftarrow \mathbf{x}_{\text{max}}^m + \varepsilon^{(m,j)}$, where $\varepsilon^{(m,j)} \sim \mathcal{N}(0, \sigma^2 I)$
  for $j = 1$ to J **do**
    $\hat{\mathbf{z}}^{(m,j)} \leftarrow g(\hat{\mathbf{x}}^{(m,j)})$
    $\ell_{mj} \leftarrow \log p_z(\hat{\mathbf{z}}^{(m,j)}) + \log\left|\det\left(\frac{\partial \hat{\mathbf{z}}^{(m,j)}}{\partial \hat{\mathbf{x}}^{(m,j)}}\right)\right|$
    $\ell_{\text{max}} \leftarrow \log p_z(\mathbf{z}_{\text{max}}^m) + \log\left|\det\left(\frac{\partial \mathbf{z}_{\text{max}}^m}{\partial \mathbf{x}_{\text{max}}^m}\right)\right|$
    $\delta_{mj} \leftarrow \max\left(0,\ \ell_{mj} - \ell_{\text{max}}\right)$
  end for
  $\mathcal{L}_{\text{grad}} \mathrel{+}= \frac{1}{M} g_m$
  $\mathcal{L}_{\text{contrast}} \mathrel{+}= \frac{1}{MJ}\sum_{j=1}^{J} \delta_{mj}$
end for
$\mathcal{L}_{\text{repel}} \leftarrow 0$
for $m = 1$ to M **do**
  for $m' = m + 1$ to M **do**
    $\mathbf{x}_m \leftarrow g^{-1}(\mathbf{z}_{\text{max}}^m)$
    $\mathbf{x}_{m'} \leftarrow g^{-1}(\mathbf{z}_{\text{max}}^{m'})$
    $d \leftarrow \|\mathbf{x}_m - \mathbf{x}_{m'}\|$
    $\mathcal{L}_{\text{repel}} \mathrel{+}= \max(0,\ \delta - d)^2$
  end for
end for
$\mathcal{L}_{\text{total}} \leftarrow \lambda_{\text{global}}\mathcal{L}_{\text{NLL}} + \lambda_{\text{local}}(\mathcal{L}_{\text{contrast}} + \mathcal{L}_{\text{grad}}) + \lambda_{\text{repel}}\mathcal{L}_{\text{repel}}$

return $\mathcal{L}_{\text{total}}$

C Training details

For the case of the Müller potential (Müller and Brown, 1979), the network of the invertible map consists of $S = 2$ stages and each stage has depth 32 and width 128. In the loss function, the parameters λ_{global}, λ_{local}, and $\lambda_{\text{repulsion}}$ are set to $1, 0.1$ and 1, respectively. To find the local maxima, $J = 1000$ samples are randomly selected near each local maximum in every training iteration. The training dataset size consists of 5×10^6 samples. The ADAM optimizer (Kingma and Ba, 2014) is used with a batch size of 50000 samples for all training processes. Each training consists of 200 epochs, and the KL divergence is computed using a validation dataset containing 4×10^6 samples.

For the case of the Alanine Dipeptide molecule, the network of the invertible map consists of $S = 8$ stages and each stage has depth 12 and width 64. In the loss function, the parameters λ_{global}, λ_{local}, and $\lambda_{\text{repulsion}}$ are set to 1, 5 and 5, respectively. The training dataset consists of 5×10^6 samples. The ADAM optimizer (Kingma and Ba, 2014) is used with a batch size of 1000 samples throughout the training processes. Each training consists of 400 epochs.

D Computational resource

The MD simulations were performed on a machine equipped with an AMD EPYC 7H12 128-core process with the base clock speed 2.6 GHz, up to 3.3 GHz. For parallelization, each simulation utilized 4 MPI tasks with no OpenMP used for threading. The training is All computations were performed on the computational resources and services provided by the Institute for Cyber-Enabled Research at Michigan State University. The training was performed on a system equipped with an Intel(R) Xeon(R) Platinum 8260 CPU (2.40 GHz) and a single NVIDIA V100 GPU with 32,768 MB of memory.

E Experimental Details

E.1 MD setup

The MD simulations were conducted using the GROMACS 2019.2 software package (Lindahl et al., 2019). The Ace-Ala-Nme (ala2) molecule was modeled with the Amber99SB force field (Hornak et al., 2006), and solvated in a periodic simulation box containing 386 water molecules.

For the van der Waals interaction, a cutoff radius of 0.9 nm was employed. Electrostatic interactions were treated using the smooth particle mesh Ewald (PME) method (Essmann et al., 1995), with a real-space cutoff of 0.9 nm and a reciprocal space grid spacing of 0.12 nm. The PME interpolation order was set to 4, and the Fourier grid spacing was set to 0.16 nm. A force-switch modifier was applied to the van der Waals interactions between 1.0 and 1.2 nm to smoothly transition from short-range to long-range interactions. Neighbor searching was performed using a grid-based scheme with the Verlet cutoff. The cutoff radius for both van der Waals and Coulomb interactions was set to 1.2 nm.

The MD simulation consists of both the equilibrium stage and the production stage. For the equilibrium stage, energy minimization was performed using the steepest descent algorithm for 50,000 steps, with a maximum force convergence criterion of 1000.0 kJ mol^{-1} nm^{-1} and a minimization step size of 0.01 nm. Consequently, an NPT equilibration was conducted to stabilize the system under constant pressure and temperature conditions. The equilibration ran for 0.1 ns, corresponding to 100,000 steps with a timestep of 1 fs, using the leap-frog integrator. The temperature was controlled at 350 K using the velocity-rescale thermostat Bussi et al. (2007), with separate coupling groups for the solute and solvent. Each group was coupled to a reference temperature with a time constant of 0.1 ps. Pressure coupling was applied using the Parrinello-Rahman barostat Parrinello and Rahman (1981), with isotropic scaling of the box vectors, set to a reference pressure of 1 bar. Finally, an additional NVT equilibration was performed to ensure thermal stability. This simulation ran for $50,000$ steps with a timestep of 2 fs. The temperature was controlled by the v-rescale thermostat, with separate coupling groups for water and non-water molecules, each set to a reference temperature of 350 K with a time constant of 0.2 ps.

During the production stage, the equations of motion were integrated using the leap-frog algorithm with a timestep of 2 fs. The initial configuration was derived from the result of the equilibration stage. To ensure a diverse exploration of the configuration space, the velocities of the system were

randomly regenerated following a Maxwellian distribution at a temperature of 350 K, using 100 different random seeds. To enhance the ergodicity of the configuration space, data was collected in parallel from these 100 different initializations. Each initialization run $10,000,000$ steps and data collected in every 100 steps to generate the samplings.

E.2 Frame indifference distance

To compare the configurations based on their internal structural differences, we compute the displacement between their molecular positions while eliminating the contributions from translational and rotational degrees of freedom. First, both sets of coordinates, P and $Q \in \mathbb{R}^{N\times 3}$, are translated so that their centroids coincide with the origin of the coordinate system. This is done by subtracting the centroid coordinates from each point. Next, we calculate the singular value decomposition (SVD) of the covariance matrix $H = P^TQ = U\Sigma V^T$. The optimal rotation matrix R is then defined as:

$$R = V \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & d \end{pmatrix} U^T,$$

where $d = \det(U)\det(V)$ records if the orthogonal matrices contain a reflection. This rotation matrix R aligns the two configurations and meanwhile preserves their relative internal structure, allowing for a meaningful comparison based solely on the internal differences. Finally, the distance between the two configurations can be defined as:

$$Id(Q, P) = \|RQ - P\|.$$

This formulation ensures that the comparison focuses purely on the structural differences, excluding translational and rotational artifacts.

E.3 Time correlation

Although our method relies solely on time discrete samples, the constructed CVs have a clear physical interpretation and exhibit slow-decay mode in the dynamical properties such as autocorrelation function (see Sec. 4). In this section, we further discuss the rationale of the slow decay as a metric of well-defined CVs, and elaborate the implication related to the collective dynamics of the system.

Regardless of the detailed form of the dynamics, there exists a family of linear propagators $\mathcal{P}(\tau)$, that evolve the probability density of states ρ_τ from ρ_0 through

$$\rho_\tau(\mathbf{x}) = \int P(\mathbf{x}, \mathbf{y}; \tau)\rho_0(\mathbf{y})\mathrm{d}\mathbf{y} := \mathcal{P}(\tau)\rho_0,$$

where $P(\mathbf{x}, \mathbf{y}; \tau)$ is the probability density of finding the process at points $\mathbf{x}$ and $\mathbf{y}$ at a time spacing of τ under the dynamics and $\mathcal{P}(\tau)$ satisfies the a Chapman-Kolmogorov equation: for all τ_1, τ_2, $\mathcal{P}(\tau_1 + \tau_2) = \mathcal{P}(\tau_1)\mathcal{P}(\tau_2)$. For simplicity, we assume that the atomic positions are sufficiently ergodic such that a unique stationary density, $p(\mathbf{x})$, exists. This assumption leads to the following result.

Proposition E.1. *When the propagator operator $\mathcal{P}(\tau)$ is compact and self-adjoint, it can be decomposed into the spectral components of the propagator:*

$$\rho_\tau = p(\mathbf{x}) + \sum_{i=2}^{m} a_i[\rho_0]\lambda_i(\tau)l_i + \mathcal{P}_{\textit{fast}}(\tau)\rho_0,$$

where $l_2, \cdots, l_m$ are the eigenfunctions of the propagator, $\lambda_i(\tau) = \exp(-\kappa_i\tau)$ (sorted in non-ascending order) are the real-valued eigenvalues of the propagator that decay exponentially with time, and $a_i(\rho_0)$ are factors that depend on the initial density ρ_0.

Proof. The proof follows directly from the spectral properties of compact self-adjoint operators, as outlined in Theorem 5.3.4 in Nonnenmacher (2021). Due to the compactness and self-adjointness of $\mathcal{P}(\tau)$, it possesses a sequence of real eigenvalues $\lambda_i(\tau)$, with $|\lambda_i(\tau)| \leq 1$ and $|\lambda_i(\tau)| \to 0$ as $i \to \infty$. Additionally, from the existence of the equilibrium distribution $p(\mathbf{x}) = \mathcal{P}(\tau)p(\mathbf{x})$, we conclude that $p(\mathbf{x})$ is one eigenfunction with eigenvalue to be 1. Because of the Chapman-Kolmogorov equation, each eigenvalue $\lambda_i(\tau)$ decays exponentially in time, i.e. we have $\lambda_i(\tau) = \exp(\kappa_i\tau)$, for some $\kappa_i \geq 0$.

As a result, we can express the action of the operator $\mathcal{P}(\tau)$ on a given distribution ρ_τ as a summation over its eigenfunctions:

$$\begin{aligned}\rho_\tau &= \sum_{i=1}^{\infty} \mathcal{P}(\tau) a_i[\rho_0] l_i \\ &= \sum_{i=1}^{\infty} \lambda_i(\tau) a_i[\rho_0] l_i \\ &= \sum_{i=1}^{\infty} \exp(-\kappa_i \tau) a_i[\rho_0] l_i \\ &\simeq p(\mathbf{x}) + \sum_{i=2}^{m} a_i[\rho_0] \lambda_i(\tau) l_i + \mathcal{P}_{\text{fast}}(\tau)\rho_0,\end{aligned} \tag{24}$$

where $a_i[\rho_0]$ is the projection of ρ_0 on the basis and $\mathcal{P}_{\text{fast}}(\tau)$ represents the contribution of the fast-decaying modes, which is very small for lag times $\frac{1}{\kappa_{m+1}}$. □

Based on this result, we can conclude that the slow modes play a crucial role in determining the collective dynamics of the system. Let f be a real-valued function of state, its autocorrelation (denoted by $\text{acf}(f;\tau)$) is given by

$$\text{acf}(f;\tau) = \mathbb{E}[f(\mathbf{x}_0) f(\mathbf{x}_\tau)] = \int f(\mathbf{x}) p(\mathbf{x}) \mathcal{P}_\tau p(\mathbf{y}) f(\mathbf{y}) \mathrm{d}\mathbf{x} \mathrm{d}\mathbf{y}. \tag{25}$$

The autocorrelation function is related to the slow modes through the following proposition:

Proposition E.2. *The autocorrelation function of a weighted eigenfunction $r_k(\mathbf{x}) = p(\mathbf{x})^{-1} l_k(\mathbf{x})$ is its eigenvalue $\lambda_k(\tau)$, i.e.*

$$\text{acf}(r_k;\tau) = \lambda_k(\tau)$$

Proof. The proof follows directly from the definition of the autocorrelation function and can be obtained through straightforward calculation. □

Therefore, the characteristic relaxation time of the autocorrelation function provides an essential metric for the chosen CVs. In practice, if statistically sufficient realizations of $\mathbf{x}_t$ are available, the autocorrelation function can be estimated via:

$$\text{acf}(f;\tau) = \mathbb{E}(f(\mathbf{x}_0) f(\mathbf{x}_\tau)) \approx \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_0) f(\mathbf{x}_\tau), \tag{26}$$

where N is the number of simulated time windows of length τ.

F Additional results

F.1 Addistion result for Müller

To verify the accuracy of the trained model on the Müller potential, we evaluate the PDF obtained via the learned inverse map $f^{-1}(\mathbf{x})$, where $\mathbf{x}$ is sampled from the Müller distribution. As illustrated in Figure 7, the distribution of the first component of $\mathbf{z} = f^{-1}(\mathbf{x})$ (i.e., the CV) shows good agreement with the prior base distribution.

Furthermore, we compare the performance of models trained with a standard Gaussian prior and the non-Gaussian prior used in our method. As shown in Figure 8, the KL divergence between $p(\mathbf{x})$ by Eq. (2) and the empirical distribution of the present method is consistently lower than the one using the standard Gaussian. This shows that an informed prior may further improve the learning of complex underlying PDFs.

Besides the KL divergence associated with the global loss (5), we further examine the preservation of the metastable structure between the two spaces. Figure 9 shows the distribution of neighboring points in the prior $\mathbf{z}$-space mapped from the points near two energy local minima in configuration

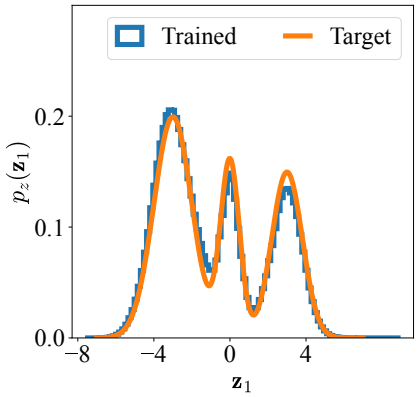

Figure 7: Comparison of the PDF of the first component of $\mathbf{z} = f^{-1}(\mathbf{x})$ (trained), where $\mathbf{x}$ is drawn from the Boltzmann distribution of the Müller potential with $k_B T = 10$ and the triple-Gaussian prior base distribution defined by Eq. (9) (target).

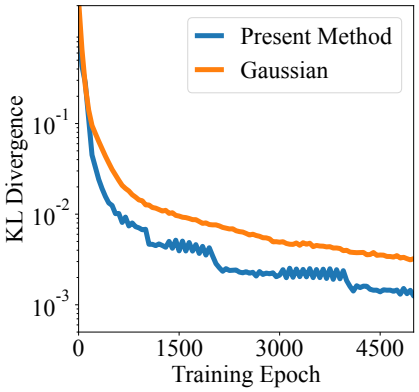

Figure 8: Comparison of the KL divergence between $p(\mathbf{x})$ (by Eq. (2)) and the empirical distribution during training using a standard Gaussian prior and the present non-Gaussian prior distribution.

$\mathbf{x}$-space using the standard Gaussian prior. The set of neighboring points associated with two different energy basins are mixed in the prior space. In contrast, as shown in Fig. 1, the map constructed in the present method ensures the separation of different basins in both spaces.

Finally, we compare the time autocorrelation functions of the CV $\mathbf{z}_1$ obtained by the present method. For comparison, we also compute the correlation of auxiliary variables $\mathbf{z}_2$ which follows a Gaussian prior. As shown in Figure 10, $\mathbf{z}_1$ exhibits a significantly slower decay in its autocorrelation function, capturing long-timescale dynamics. This behavior indicates that $\mathbf{z}_1$ encodes the dominant slow modes of the system and thus serves as an effective CV.

F.2 Addition result for the alanine dipeptide molecule system

We present supplementary results supporting the analysis in Figure 3. In Figure 11, we report the configuration variances $\mathcal{V}(\bar{\mathbf{z}}_i, \bar{\mathbf{z}}_j)$ defined by Eq. (10) with fixed variables. Consistent with the result in Figure 3, the variables $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ yield substantially lower variance, indicating reduced configurational freedom. This result further supports that these three variables form an effective set of CVs for capturing the essential conformation dynamics of the system.

Further, we analyze the preservation of the metastable structure for the present method. In Fig. 5, we present the distribution of individual sets of neighboring points projected onto the plane of the two dihedral angles in the configuration space. As a supplementary result, we further examine their distributions projected onto the plane L_e-R_g, where L_e represents the end-to-end distance of the molecule and R_g represents the radius of gyration. As shown in Figure 12, the set of points obtained

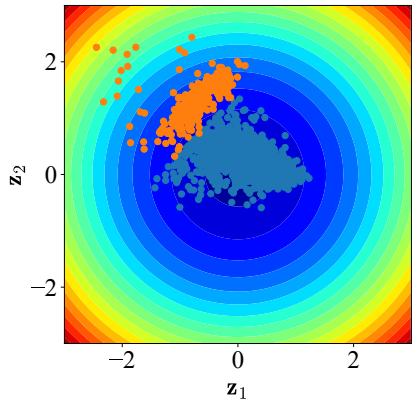

Figure 9: Locations of neighboring points in the prior $\mathbf{z}$-space mapped from the points near two individual energy local minima in configuration $\mathbf{x}$-space using the standard Gaussian prior.

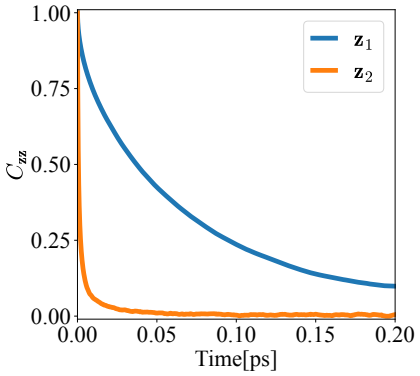

Figure 10: The time autocorrelation function $C_{\mathbf{zz}}(t)$ of the two variables for $\mathbf{z}_1$ and $\mathbf{z}_2$ of the Müller potential system. The time correlation of the CV $\mathbf{z}_1$ decays significantly slower than that of the auxiliary variables $\mathbf{z}_2$.

by the invertible map of the present method is more concentrated than the points obtained by the PCA. This result is consistent with Figure 5 and further verifies the unique capablity of the present method in preserving the metastable structure through the CV construction process.

Finally, we present supplementary results on the time autocorrelation functions of the learned variables. As shown in Figure 13, the first three variables exhibit significantly slower decay compared to the auxiliary variables. This behavior indicates that they capture the dominant slow dynamics of the system, further supporting their role as effective CVs.

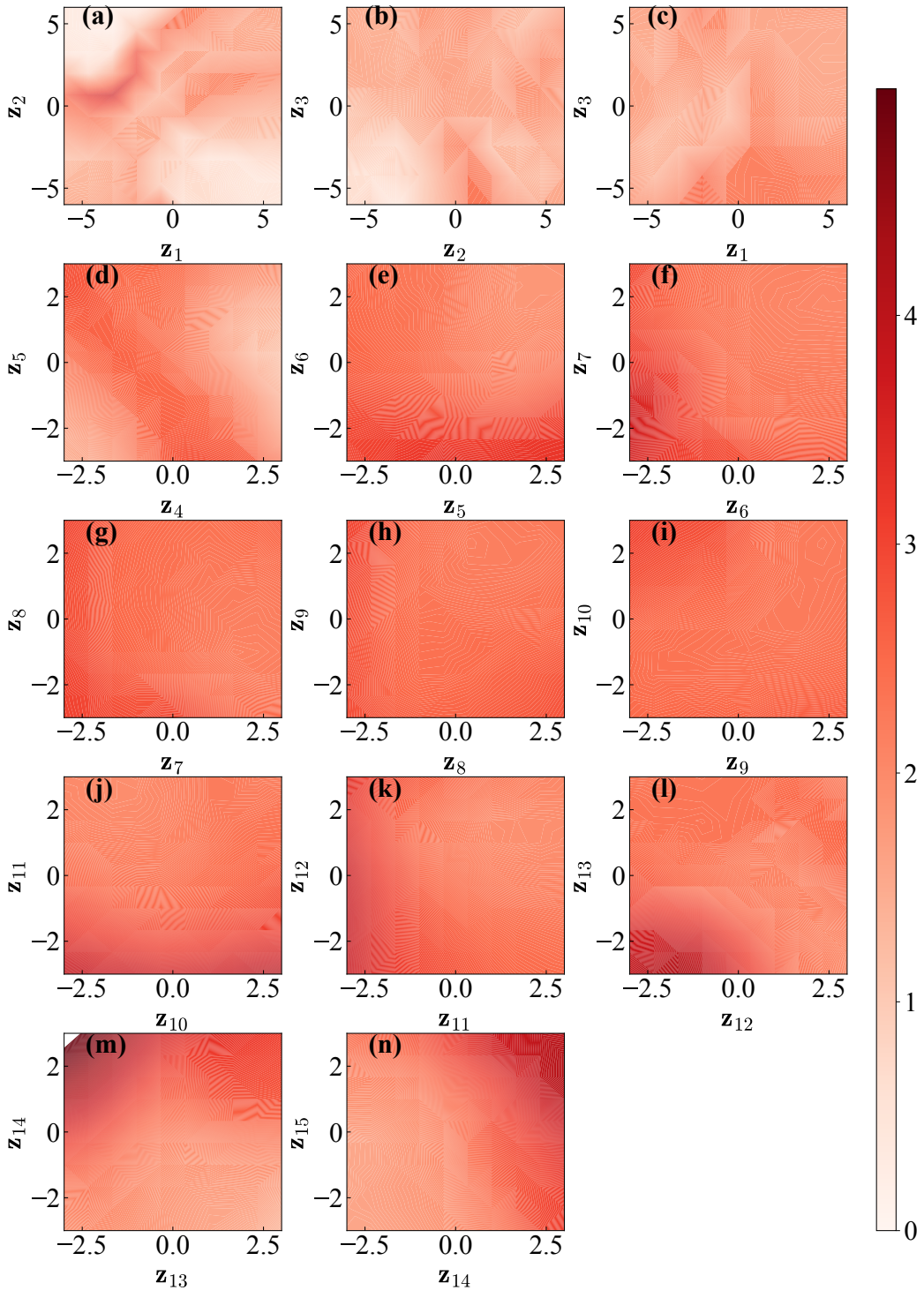

Figure 11: Supplementary analysis of configuration variances $\mathcal{V}(\bar{\mathbf{z}}_i, \bar{\mathbf{z}}_j)$ defined by Eq. (10) with various fixed latent variables $(\mathbf{z}_i, \mathbf{z}_j)$. The lower variances associated with $\mathbf{z}_1$, $\mathbf{z}_2$, and $\mathbf{z}_3$ suggest their effectiveness as collective variables.

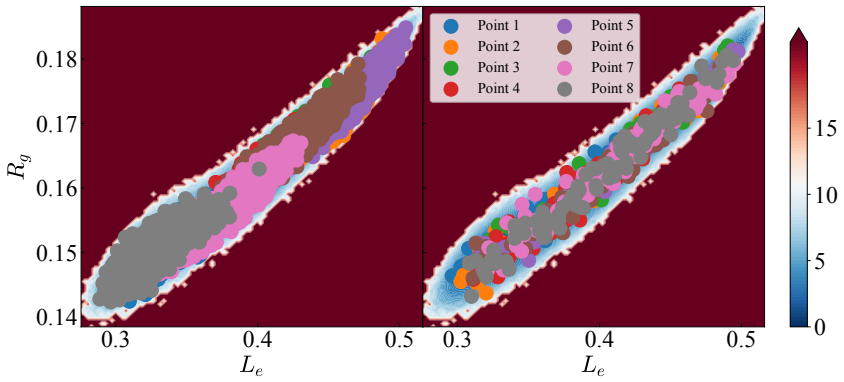

Figure 12: The distribution of the eight sets of neighboring points projected on the L_e-R_g plane, where L_e and R_g represent the end-to-end distance and the radius of gyration, respectively. Each set corresponds to the neighboring points near a local probability maximum in the prior $\mathbf{z}$-space. The points are mapped to the MD space by $\mathbf{x} = f(\mathbf{z})$ of the present method (left) and PCA (right) using the first three components as CVs to determine the neighboring points.

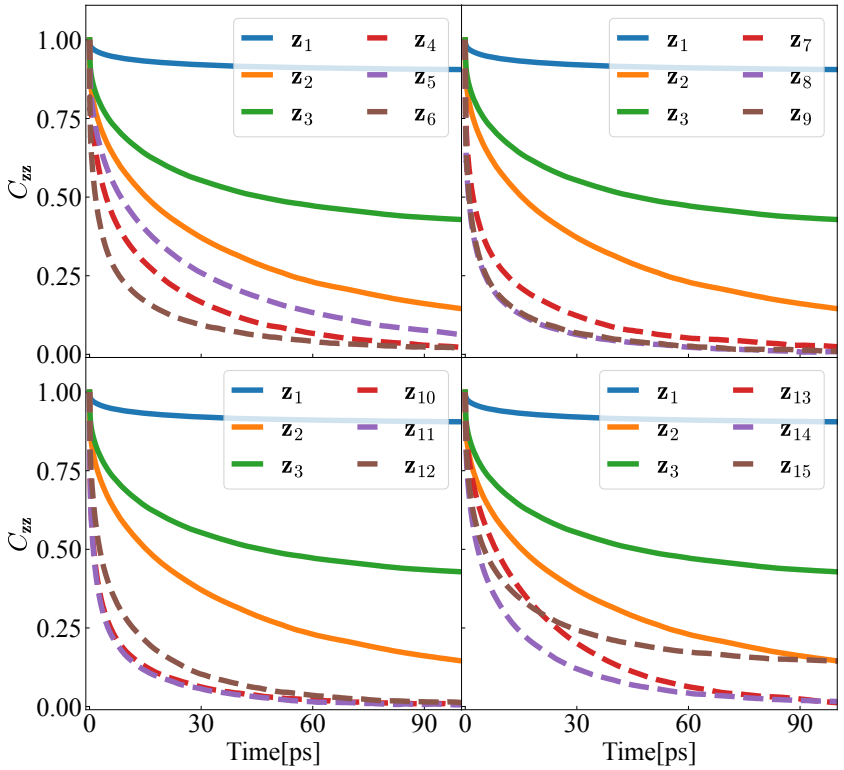

Figure 13: Normalized time autocorrelation functions for the learned CVs $(\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)$, compared with the ones for other auxiliary variables.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The contributions and scope of the paper are included in the abstract and Introduction. Please refer to the first and last paragraph of Section 1 for scope and contributions respectively.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: The limitation of the paper has been discussed in the Section 5.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The assumption of the Theorem 3.1 has been provided.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We introduce the details of the experiment, such as the simulation setup and training parameters, in Section 4.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The code and data are available on the Github we provided.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide the details of the experimental settings, such as optimizers and parameters in Training.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: In Figure 6 the standard deviation is used to provide statistical significance by error bar.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification:The CPUs or GPUs of the compute workers have been allocated in Section D.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research complies with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: This work focuses on advancing core methodological research in scientific machine learning. While the proposed framework may have potential applications across scientific and engineering domains, it is primarily intended as a foundational contribution, and we do not identify any specific broader societal impacts at this stage.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: This work does not involve the release of data or models that pose a significant risk of misuse.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer:[Yes]

Justification: Our work use the package Gromacs to conduct the MD simulations and have cited the package in the reference.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: This work shared the code via a public repository with instructions for replicating training and evaluation.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This study does not involve human subjects, personal data, or interactions requiring IRB approval.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [No]

    Justification: LLM was not used as part of the core methods, experiments, or scientific contributions of this work.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.