Consistent projection of Langevin dynamics: Preserving thermodynamics and kinetics in coarse-grained models

Downloaded from: https://research.chalmers.se, 2026-11-15 02:48 UTC

Citation for the original published paper (version of record):
Nateghi, V., Neureither, L., Moqvist, S. et al (2026). Consistent projection of Langevin dynamics: Preserving thermodynamics and kinetics in
coarse-grained models. Physical Review E, 113(5). http://dx.doi.org/10.1103/wckl-dz9d

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Consistent projection of Langevin dynamics: Preserving thermodynamics and kinetics in coarse-grained models

Vahid Nateghi [1], Lara Neureither [2], Selma Moqvist [3], Carsten Hartmann [2], Simon Olsson [3] and Feliks Nüske [1,*]

[1] *Max-Planck-Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany*

[2] *Institute of Mathematics, Brandenburg University of Technology Cottbus-Senftenberg, Cottbus, Germany*

[3] *Department of Computer Science and Engineering, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden*

(Received 2 December 2025; accepted 6 May 2026; published 28 May 2026)

Coarse graining (CG) is an important task for efficient modeling and simulation of complex multiscale systems, such as the conformational dynamics of biomolecules. This work presents a projection-based coarse-graining formalism for general underdamped Langevin dynamics. Following the Zwanzig projection approach, we derive a closed-form expression for the coarse-grained dynamics. In addition, we show how the generator extended dynamic mode decomposition method, which was developed in the context of Koopman operator methods, can be used to model the CG dynamics and evaluate its kinetic properties, such as transition timescales. Finally, we combine our approach with thermodynamic interpolation, a generative approach to transform samples between thermodynamic conditions, to extend the scope of the approach across thermodynamic states without repeated numerical simulations. Using a two-dimensional model system, we demonstrate that the proposed method allows to accurately capture the thermodynamic and kinetic properties of the full-space model.

DOI: 10.1103/wckl-dz9d

I. INTRODUCTION

The task of finding a simpler representation for a complex dynamical system, while preserving essential features of the full system, is referred to as coarse graining (CG) or model reduction [1–3]. CG models are central across many scientific disciplines because they enable interpretability and simulation of large-scale systems at greatly reduced computational cost. In molecular dynamics, CG modeling has long been used to study biomolecular systems whose size and timescales are inaccessible to all-atom simulations [4]. Since simulations of biomolecules often exhibit rare events and metastable states, constructing high-quality CG models that can reduce the computational burden is of particular importance.

The quality of a CG model depends both on the function that is used to map a full-state configuration to its reduced representation (CG map), as well as the dynamical model defined on the reduced space (CG dynamics). While numerous methods exist for defining effective CG maps [5–7], the accurate construction of CG dynamics remains more challenging. Different metrics have been suggested to assess the quality of a CG model. For equilibrium systems, reproducing the marginal steady-state distribution is a common criterion of quality as formalized by the multiscale CG method [8] and its machine-learning-based variants [9,10]. However, ensuring that the CG dynamics also reproduces the kinetic properties of the full model—such as transition rates and relaxation timescales—requires a more rigorous dynamical framework. The Mori-Zwanzig projection formalism [11,12] provides an analytical foundation, offering a systematic way to derive reduced equations that retain essential dynamical information. Several variants of this approach have been developed, including the approach by Gyöngy [13], the Zwanzig projection as formalized by Legoll and Lelièvre [14,15], as well as hybrid formats [16]. For the Zwanzig projection which we consider in this paper, called effective dynamics in Ref. [14], there exist various theoretical results for the error between the original and the CG dynamics [17–20], see also Ref. [21] for results specific to the Langevin equation and affine CG maps. Importantly, it has been shown that the ability of a CG model to capture rare-event dynamics depends critically on how well the CG map resolves the leading eigenfunctions of the Fokker-Planck or Koopman operator [15,22,23].

Since the mid-2010s, machine-learning approaches have emerged as powerful tools for constructing dynamically accurate CG models. These include spectral matching [24,25]; generator extended dynamic mode decomposition (gEDMD) [26], which is based on the Koopman operator framework [27,28]; molecular latent space simulators [29]; implicit transfer operators [30–32]; and reaction coordinate flows [33]. Recently, the gEDMD method was used to directly approximate the coarse-grained dynamics implied by the Zwanzig projection for reversible systems. In the same study, a learning scheme for the parameters of the CG dynamics enabled direct validation of the learned model through rare-event

*Contact author: nueske@mpi-magdeburg.mpg.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by Max Planck Society.

transition timescales [34]. Here we extend this approach to nonreversible underdamped Langevin dynamics—commonly used to model molecular systems in thermal equilibrium—and integrate it with recent advances in generative modeling, yielding a data-efficient framework for learning coarse-grained Langevin dynamics.

The contributions of this paper are as follows:

(1) We derive, within the Zwanzig projection formalism, explicit analytical expressions for the CG dynamics corresponding to the underdamped Langevin equation. The resulting dynamics retain the structure of the full-space system but feature generalized forces and state-dependent diffusion. We further analyze the properties of the associated CG generator.

(2) We show that thermodynamic interpolation (TI) [35], a recently proposed generative modeling approach based on Boltzmann generators [36–38] and the stochastic interpolant framework [39], can efficiently generate accurate position-space samples for numerically approximating the CG Langevin equation.

(3) We employ the gEDMD method to learn models for the CG Langevin dynamics and to analyze its dynamical properties, such as rare event transition timescales. Learning algorithms for the CG parameters are provided, and both the gEDMD models and parameters are trained on TI-generated data, removing the need for prohibitively long full-space simulations.

The structure of the paper is as follows: We introduce the underdamped Langevin dynamics and the key properties of its generator in Sec. II, along with the definition of the phase-space CG map which is used throughout this work. The analysis of the CG Langevin dynamics and its parameters is given in Sec. III. In Sec. IV, we introduce TI as a sampling method, along with the gEDMD algorithm and the learning methods for the CG parameters. Using a two-dimensional model system, we demonstrate the capabilities of the proposed methods numerically in Sec. V. Detailed proofs for our analytical results, as well as additional simulation and learning details are provided in the appendix.

II. LANGEVIN EQUATION AND COARSE GRAINING THEORY

A. Langevin dynamics

Underdamped Langevin dynamics are a stochastic version of the Hamiltonian dynamics on full phase-space coordinates $(\mathbf{q}, \mathbf{p}) \in \mathbb{R}^{2d}$. The dynamical equations are

$$\begin{aligned} \mathrm{d}\mathbf{q}_t &= \mathbf{M}^{-1}\mathbf{p}_t\,\mathrm{d}t \\ \mathrm{d}\mathbf{p}_t &= -[\nabla_{\mathbf{q}} V(\mathbf{q}_t) + \gamma \mathbf{p}_t]\,\mathrm{d}t + \sqrt{2\gamma \mathbf{M}\beta^{-1}}\,\mathrm{d}\mathbf{B}_t. \end{aligned} \tag{1}$$

Here $\gamma, \beta > 0$ are constants (friction and inverse temperature), $\mathbf{M} \in \mathbb{R}^{d\times d}$ is a diagonal mass matrix with entries $m_i > 0$, while $\mathbf{B}_t$ is d-dimensional Brownian motion. From now on, we will drop the subscript $\mathbf{q}$ in $\nabla_{\mathbf{q}} V$ as the potential V is only a function of the position. We apply the same convention to any function that depends only on $\mathbf{q}$, such as the spatial CG map ξ introduced later.

We will frequently write (1) in the more compact form

$$\mathrm{d}\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}\mathbf{A}}\,\mathrm{d}\mathbf{W}_t, \tag{2}$$

where $\mathbf{x} = (\mathbf{q}, \mathbf{p}) \in \mathbb{R}^{2d}$, $\mathbf{W}_t = (\tilde{\mathbf{B}}_t, \mathbf{B}_t) \in \mathbb{R}^{2d}$ is $2d$-dimensional Brownian motion and the drift and diffusion coefficients are given by

$$\mathbf{b}(\mathbf{q}, \mathbf{p}) = \begin{pmatrix} \mathbf{M}^{-1}\mathbf{p} \\ -\nabla V(\mathbf{q}) - \gamma \mathbf{p} \end{pmatrix} \in \mathbb{R}^{2d},$$

$$\mathbf{A} := \sigma\sigma^{\top} = \begin{pmatrix} 0_{d\times d} & 0_{d\times d} \\ 0_{d\times d} & \gamma \mathbf{M} \end{pmatrix} \in \mathbb{R}^{2d\times 2d}.$$

Note that we can also write the drift field $\mathbf{b}$ in the Hamiltonian form

$$\mathbf{b}(\mathbf{x}) = (\mathbf{J} - \mathbf{A})\nabla H(\mathbf{x}),$$

where the canonical Hamiltonian and skew-symmetric matrix $\mathbf{J}$ are given by

$$H(\mathbf{x}) = V(\mathbf{q}) + \frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} \quad \text{and}$$

$$\mathbf{J} = \begin{pmatrix} 0_{d\times d} & I_d \\ -I_d & 0_{d\times d} \end{pmatrix} \in \mathbb{R}^{2d\times 2d}.$$

The invariant distribution of the process (1) is the canonical distribution with density

$$\begin{aligned} \mathrm{d}\mu(\mathbf{q}, \mathbf{p}) &\propto \exp\left(-\beta H(\mathbf{q}, \mathbf{p})\right)\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p} \\ &= \exp\left(-\beta\left(\frac{1}{2}\mathbf{p}^T\mathbf{M}^{-1}\mathbf{p} + V(\mathbf{q})\right)\right)\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p} \\ &= \mu_{\mathbf{p}}(\mathbf{p})\mu_{\mathbf{q}}(\mathbf{q})\,\mathrm{d}\mathbf{q}\,\mathrm{d}\mathbf{p}. \end{aligned} \tag{3}$$

B. The generator

An important tool for large parts of this paper is the evolution of observables under the Langevin dynamics. Observables are functions ϕ of the phase space, which we require to be contained in the weighted space $L^2_\mu(\mathbb{R}^{2d})$:

$$L^2_\mu(\mathbb{R}^{2d}) = \left\{\phi : \mathbb{R}^{2d} \mapsto \mathbb{R} : \left| \int \phi^2(\mathbf{x}) e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x} < \infty \right.\right\}. \tag{4}$$

The *generator* is a differential operator acting on observables $\phi \in L^2_\mu(\mathbb{R}^{2d})$:

$$\begin{aligned} \mathcal{L}\phi &= \mathbf{b}\cdot\nabla\phi + \beta^{-1}\mathbf{A} : \nabla^2\phi \\ &= \mathbf{J}\nabla H\cdot\nabla\phi - \mathbf{A}\nabla H\cdot\nabla\phi + \beta^{-1}\mathbf{A} : \nabla^2\phi, \end{aligned} \tag{5}$$

where the notation $\mathbf{A} : \nabla^2\phi$ denotes the Frobenius inner product between $\mathbf{A}$ and the Hessian matrix of ϕ:

$$\mathbf{A} : \nabla^2\phi = \sum_{i,j=1}^{2d} \mathbf{A}_{ij}\frac{\partial^2}{\partial \mathbf{x}_i \partial \mathbf{x}_j}\phi.$$

The first term in (5) is the Hamiltonian part of the generator

$$\mathcal{L}_H = \mathbf{J}\nabla H\cdot\nabla = \mathbf{M}^{-1}\mathbf{p}\cdot\nabla_{\mathbf{q}} - \nabla V\cdot\nabla_{\mathbf{p}}, \tag{6}$$

while the next two terms form the Ornstein-Uhlenbeck part of the generator

$$\mathcal{L}_{\mathrm{OU}} = -\mathbf{A}\nabla H \cdot \nabla + \beta^{-1}\mathbf{A} : \nabla^2 = -\gamma \mathbf{p} \cdot \nabla_{\mathbf{p}} + \frac{\gamma}{\beta}\mathbf{M} : \nabla_{\mathbf{p}}^2. \tag{7}$$

An equivalent way of decomposing the generator is into its antisymmetric and symmetric parts:

$$\mathcal{L} = \mathbf{J}^{\mathrm{eq}} \cdot \nabla + \frac{1}{\beta} e^{\beta H} \nabla \cdot (e^{-\beta H}\mathbf{A}\nabla),$$

where $\mathbf{J}^{\mathrm{eq}}$ is the stationary probability flow

$$\mathbf{J}^{\mathrm{eq}} = \mathbf{b} - \frac{1}{\beta} e^{\beta H}\nabla \cdot (\mathbf{A}e^{-\beta H}) = \begin{pmatrix} \mathbf{M}^{-1}\mathbf{p} \\ -\nabla V(\mathbf{q}) \end{pmatrix}. \tag{8}$$

It can be verified directly that $\mathbf{J}^{\mathrm{eq}} = \mathbf{J}\nabla H$, and therefore $\mathcal{L}_H = \mathbf{J}^{\mathrm{eq}} \cdot \nabla$ coincides with the antisymmetric part. Similarly, the Ornstein-Uhlenbeck generator $\mathcal{L}_{\mathrm{OU}}$ coincides with the symmetric part. As a consequence,

$$\langle \mathcal{L}\phi,\ \psi\rangle_\mu = \langle \mathbf{J}^{\mathrm{eq}} \cdot \nabla\phi,\ \psi\rangle_\mu - \frac{1}{\beta}\langle \mathbf{A}\nabla\phi,\ \nabla\psi\rangle_\mu, \tag{9}$$

where the rightmost term follows from integration by parts:

$$\begin{aligned}\langle \mathcal{L}_{\mathrm{OU}}\phi,\ \psi\rangle_\mu &= -\frac{\gamma}{\beta}\int_{\mathbb{R}^{2d}} \nabla_{\mathbf{p}}\phi \cdot \mathbf{M} \cdot \nabla_{\mathbf{p}}\psi \,\mathrm{d}\mu(\mathbf{q},\mathbf{p}) \\ &= -\frac{1}{\beta}\langle \mathbf{A}\nabla\phi,\ \nabla\psi\rangle_\mu = \langle \phi,\ \mathcal{L}_{\mathrm{OU}}\psi\rangle_\mu.\end{aligned}$$

Therefore, in order to approximate the generator, we can use numerical approximations that only require first-order derivatives.

For systems exhibiting metastable states and undergoing rare transitions, we are generally interested in eigenvalues of the generator close to zero. Thus, we expect to find a number of real eigenvalues $0 = \kappa_0 < \kappa_1 \leqslant \ldots \leqslant \kappa_M$ and eigenfunctions ψ_i, $i = 0, \ldots, M$, such that

$$-\mathcal{L}\psi_i = \kappa_i \psi_i, \tag{10}$$

and $\kappa_i \ll 1$. These eigenvalues can be interpreted as rates of transition, their associated implied timescales are defined as [40,41]

$$t_i = \frac{1}{\kappa_i}.$$

C. CG map and potential of mean force

Next, we consider coarse graining of the Langevin dynamics in Eq. (1). We assume the CG map is defined based on a smooth mapping

$$\xi : \mathbb{R}^d \to \mathbb{R}^k,\ \mathbf{q} \mapsto \xi(\mathbf{q}),\ k \leqslant d,$$

called the *spatial CG map*, which transforms the spatial degrees of freedom by mapping them to a lower-dimensional space. In order to arrive at a coarse-grained dynamics which is again of Langevin type, we need to define CG velocity. Since formally $\frac{d}{dt}\xi(\mathbf{q}) = \nabla\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p}$, where $\nabla\xi \in \mathbb{R}^{k\times d}$ is the Jacobian of the spatial CG map, it has been suggested [21,42] to define the coarse-grained velocity analogously:

$$\mathbf{v}_i = \dot{\xi}_i(\mathbf{q},\mathbf{p}) = \nabla\xi_i(\mathbf{q})\mathbf{M}^{-1}\mathbf{p}, \tag{11}$$

leading to an augmented $2k$-dimensional coarse-graining map called the *phase-space CG map*,

$$\Xi(\mathbf{q},\mathbf{p}) = \begin{bmatrix}\mathbf{z}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\xi(\mathbf{q})\\ \nabla\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p}\end{bmatrix} \in \mathbb{R}^{2k}. \tag{12}$$

We use the following notation for its Jacobian determinant, assuming it has maximum rank almost everywhere (i.e., nonzero determinant for almost all $\mathbf{x}$):

$$\mathrm{Jac}(\mathbf{x}) = \det[\nabla\Xi(\mathbf{x})\nabla\Xi^\top(\mathbf{x})].$$

To proceed further, we need to introduce some abstract concepts. It is often helpful to break up integrals over full phase space into an integral over all points that are assigned to the same point in CG space and then integrate out over the CG variables. Formally, we write:

$$\int_{\mathbb{R}^{2d}} g(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^{2k}}\left[\int_{\mathbb{R}^{2d}} g(\mathbf{x})\delta(\mathbf{y} - \Xi(\mathbf{x}))\,\mathrm{d}\mathbf{x}\right]\mathrm{d}\mathbf{y}.$$

To make this notation more transparent, we consider the nonlinear preimage of a point $\mathbf{y} = (\mathbf{z},\mathbf{v}) \in \mathbb{R}^{2k}$ under the CG map as

$$\Xi^{-1}(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^{2d} : \Xi(\mathbf{x}) = \mathbf{y}\}.$$

That is, $\Xi^{-1}(\mathbf{y})$ is the set of phase-space elements which collapse to the same point $\mathbf{y}$ in CG space. Writing $\sigma_{\mathbf{y}}$ for the Hausdorff measure on the nonlinear manifold $\Xi^{-1}(\mathbf{y})$, we introduce the weighted surface measure:

$$\mathrm{d}\rho_{\mathbf{y}}(\mathbf{x}) = \mathrm{Jac}^{-1/2}(\mathbf{x})\,\mathrm{d}\sigma_{\mathbf{y}}(\mathbf{x}).$$

By means of the measure $\rho_{\mathbf{y}}$, we now rewrite the previous integration formula by splitting up integrals along preimages of the CG map, known as the *coarea formula*:

$$\int_{\mathbb{R}^{2d}} g(\mathbf{x})\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^{2k}}\int_{\Xi^{-1}(\mathbf{y})} g(\mathbf{x})\mathrm{Jac}^{-1/2}(\mathbf{x})\,\mathrm{d}\sigma_{\mathbf{y}}(\mathbf{x})\,\mathrm{d}\mathbf{y}. \tag{13}$$

We can now proceed to introduce the *free energy* or potential of mean force on the CG space:

$$F(\mathbf{y}) = -\frac{1}{\beta}\log\left[\int_{\Xi^{-1}(\mathbf{y})} e^{-\beta H(\mathbf{x})}\,\mathrm{d}\rho_{\mathbf{y}}(\mathbf{x})\right], \tag{14}$$

with associated probability measure $\nu(\mathbf{y}) \sim e^{-\beta F(\mathbf{y})}\mathrm{d}\mathbf{y}$. The coarea formula implies that for two functions $\phi,\ \psi \in L^2_\nu(\mathbb{R}^{2k})$, depending only on the CG variables, we have the following identity between inner products on full space and on CG space:

$$\begin{aligned}\langle \phi,\ \psi\rangle_\nu &= \frac{1}{Z}\int_{\mathbb{R}^{2k}} \phi(\mathbf{y})\psi(\mathbf{y})e^{-\beta F(\mathbf{y})}\,\mathrm{d}\mathbf{y} \\ &= \frac{1}{Z}\int_{\mathbb{R}^{2d}} \phi(\Xi(\mathbf{x}))\psi(\Xi(\mathbf{x}))e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x} \\ &= \langle \phi\circ\Xi,\ \psi\circ\Xi\rangle_\mu.\end{aligned}$$

D. Projection operators and CG dynamics

To define dynamics on the CG space, given a phase-space CG map Ξ and the SDE in Eq. (2), the starting point is usually to apply Ito's lemma to the observed process $\Xi(\mathbf{X}_t)$:

$$\mathrm{d}\Xi(\mathbf{X}_t) = \mathcal{L}\Xi(\mathbf{X}_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\nabla\Xi(\mathbf{X}_t)\cdot\sqrt{\mathbf{A}}\,\mathrm{d}\mathbf{W}_t.$$

This equation is not closed as it depends on the full state $\mathbf{X}_t$. A general procedure to close the equation is to apply a projection operator to the infinitesimal generator $\mathcal{L}$. A linear orthogonal projection $\mathcal{P}$ is a linear map from functions in $L^2_\mu(\mathbb{R}^{2d})$ to functions in the CG domain $L^2_\nu(\mathbb{R}^{2k})$, which is self-adjoint on $L^2_\mu(\mathbb{R}^{2d})$, and satisfies $\mathcal{P}^2 = \mathcal{P}$. A widely used choice for $\mathcal{P}$ is *Zwanzig's projection* operator [43]:

$$\begin{aligned}\mathcal{P}\phi(\mathbf{y}) &= \mathbb{E}[\phi \mid \Xi(\mathbf{x}) = \mathbf{y}] \\ &= \frac{1}{e^{-\beta F(\mathbf{y})}} \int_{\Xi^{-1}(\mathbf{y})} \phi(\mathbf{x}) e^{-\beta H(\mathbf{x})}\, \mathrm{d}\rho_{\mathbf{y}}(\mathbf{x}).\end{aligned} \tag{15}$$

The Zwanzig projector represents the best approximation of ϕ by a function that only depends on the CG variables $\mathbf{y}$. Following Ref. [15], we choose $\mathcal{P}$ according to Eq. (15), and define the generator in CG space by

$$\mathcal{L}^\Xi = \mathcal{P}\mathcal{L}\mathcal{P}. \tag{16}$$

It can be shown that $\mathcal{L}^\Xi$ is indeed the generator of an effective SDE

$$\mathrm{d}\mathbf{Y}_t = \bar{\mathbf{b}}(\mathbf{Y}_t)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\bar{\sigma}(\mathbf{Y}_t)\,\mathrm{d}\mathbf{B}_t \tag{17}$$

on the CG space $\mathbb{R}^{2k}$ [15].

III. ANALYTICAL STRUCTURE OF THE COARSE-GRAINED LANGEVIN DYNAMICS

Here we report the main analytical results of the paper on the construction of a CG dynamics which preserve structural and dynamical properties of the full-space model. Following Ref. [15], the effective drift and diffusion of the coarse-grained Langevin equation (17) are given by the abstract analytical expressions below:

$$\begin{aligned}\bar{\mathbf{b}}(\mathbf{y}) &= \mathcal{P}[\mathcal{L}\Xi](\mathbf{y}), \\ \bar{\sigma}(\bar{\sigma})^\top(\mathbf{y}) = \bar{\mathbf{A}}(\mathbf{y}) &= \mathcal{P}[\nabla\Xi\mathbf{A}\nabla\Xi^\top](\mathbf{y}).\end{aligned} \tag{18}$$

These parameters are spelled out in detail in the following result:

Proposition 1. The effective drift and effective diffusion of the coarse-grained Langevin equation can be calculated explicitly as

$$\bar{\mathbf{b}}(\mathbf{z},\mathbf{v}) = \begin{pmatrix} \mathbf{v} \\ \mathbf{f}(\mathbf{z},\mathbf{v}) - \gamma\mathbf{v} \end{pmatrix}, \quad \bar{\mathbf{A}}(\mathbf{z},\mathbf{v}) = \begin{pmatrix} 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & \gamma\mathbf{D}(\mathbf{z},\mathbf{v}) \end{pmatrix}, \tag{19}$$

where the effective force and diffusion fields are given by

$$\begin{aligned}\mathbf{f}(\mathbf{z},\mathbf{v}) &= \mathcal{P}(-\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla V(\mathbf{q}) \\ &\qquad + \mathbf{p}^\top\mathbf{M}^{-1}\partial^2\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p})(\mathbf{z},\mathbf{v}), \\ \mathbf{D}(\mathbf{z},\mathbf{v}) &= \mathcal{P}(\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla\xi(\mathbf{q})^\top)(\mathbf{z},\mathbf{v}).\end{aligned} \tag{20}$$

In the definition of $\mathbf{f}$ we use a compact notation for the Hessian matrix of ξ, meaning that the effective force is given component-wise as

$$\mathbf{f}_i = \mathcal{P}(-\nabla\xi_i\mathbf{M}^{-1}\nabla V + \mathbf{p}^\top\mathbf{M}^{-1}\nabla^2\xi_i\mathbf{M}^{-1}\mathbf{p}).$$

Proposition 1 states that the effective SDE (17) is again a Langevin dynamics, only with a modified force $\mathbf{f}$ and a state-dependent diffusion field $\mathbf{D}$ acting on the velocity variables. The CG force and diffusion fields in (20) essentially capture two effects: first, application of a nonlinear change of variables, which manifests in the presence of derivatives of ξ, and, second, averaging over all phase-space points, which are collapsed under the action of the CG map Ξ, manifest in the action of the projection operator $\mathcal{P}$.

The structure of the dynamical parameters in Eqs. (19) and (20) carries over to the coarse-grained generator $\mathcal{L}^\Xi$ and the invariant distribution of the CG Langevin dynamics:

Proposition 2. The effective dynamics in Eq. (17) admits the invariant measure $e^{-\beta F(\mathbf{y})}$. The associated stationary probability flow is

$$\bar{\mathbf{J}}^{\mathrm{eq}}(\mathbf{z},\mathbf{v}) = \mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^\top)(\mathbf{z},\mathbf{v}) = \begin{pmatrix} \mathbf{v} \\ \mathbf{f}(\mathbf{z},\mathbf{v}) \end{pmatrix}.$$

Moreover, the projected generator $\mathcal{L}^\Xi$ admits a decomposition:

$$\mathcal{L}^\Xi = \mathcal{L}^\Xi_a + \mathcal{L}^\Xi_s,$$

with antisymmetric and symmetric parts:

$$\begin{aligned}\mathcal{L}^\Xi_a\phi &= \bar{\mathbf{J}}^{\mathrm{eq}} \cdot \nabla_{\mathbf{y}}\phi, \\ \mathcal{L}^\Xi_s\phi &= \frac{\gamma}{\beta} e^{\beta F(\mathbf{y})}\nabla_{\mathbf{v}} \cdot [\mathbf{D}(\mathbf{y}) e^{-\beta F(\mathbf{y})}\nabla_{\mathbf{v}}\phi].\end{aligned} \tag{21}$$

Proposition 2 shows that the invariant distribution of the CG dynamics is induced by the potential of mean force, ensuring thermodynamic consistency between the full-space and CG dynamics. As in the full-space setting, the stationary probability flow $\bar{\mathbf{J}}^{\mathrm{eq}}$ associated to the invariant distribution determines the antisymmetric part of the CG generator, as it accounts for the "Hamiltonian" part of the CG drift, consisting of the velocity and generalized force terms. The symmetric part is in turn determined by the potential of mean force and the CG diffusion field, and it accounts for the friction term and the generalized diffusion term acting on the CG velocity component. Since $\mathcal{L}^\Xi_s$ is also given in a divergence form, it admits an integration by parts formula analogous to (9), which can be exploited for numerical approximation of the CG generator.

In summary, Propositions 1 and 2 show that the coarse-grained Langevin dynamics retains much of the structure of the full state dynamics (1). Detailed proofs of these results can be found in Appendix A.

IV. METHODS

We now turn to the numerical approximation and analysis of the coarse-grained Langevin equation (17). This entails three essential steps: first, generation of sufficient training data; second, analysis of the coarse-grained generator $\mathcal{L}^\Xi$; and, third, learning of the parameters $\bar{\mathbf{b}}$ and $\bar{\mathbf{A}}$ of the CG dynamics.

A. Data generation

In order to infer properties of the coarse-grained Langevin equation (17), we require samples from the coarse-grained equilibrium distribution $\nu(\mathbf{y})$. As this distribution is the marginal of the full-state canonical distribution μ, it is sufficient to generate data sampling from μ and subsequently map

them by means of Ξ. Finally, as the canonical distribution in full space factors into positional component $\mu_{\mathbf{q}}$ and momentum component $\mu_{\mathbf{p}}$, it is enough to generate samples from $\mu_{\mathbf{q}}$ and complement them using samples from the multivariate normal distribution $\mu_{\mathbf{p}}$. To obtain position space samples, we consider three methods:

1. Traditional methods

(a) *Rejection sampling.* Rejection sampling or acceptance-rejection sampling is a standard method to generate independent and identically distributed samples from a distribution with known density function (up to normalization constant). On a low-dimensional state space, it is usually an efficient approach, but it is not scalable to higher-dimensional spaces. We use rejection sampling only as a baseline to compare to the performance of other, more scalable methods.

(b) *Ergodic integration of the Langevin dynamics.* Another standard approach is to integrate the SDE (1) over a sufficiently long time horizon T, using an Euler scheme or higher-order integration method. By the ergodic theorem, all snapshots from the trajectory will then form an approximate (correlated) sample from the canonical distribution. As is well known, at high values of β (low temperature), a large time horizon T will be needed to explore the entire position space and obtain a near-equilibrated dataset. To circumvent these problems, we employ a generative modeling approach as outlined in the next section.

2. Thermodynamic interpolation

Thermodynamic interpolation was proposed in Ref. [35] for equilibrium sampling of the Boltzmann distribution across different thermodynamic states. The method is based on flow models where we learn the velocity field $\mathbf{b}^{(\theta)} : [0, 1] \times \mathbb{R}^d \to \mathbb{R}^d$ of an ordinary differential equation

$$\frac{\mathrm{d}\mathbf{x}(\tau)}{\mathrm{d}\tau} = \mathbf{b}^{(\theta)}(\tau, \mathbf{x}(\tau)), \mathbf{x}(0) = \mathbf{x}_0, \tag{22}$$

such that its solution is a flow map $\Phi_\tau^{(\theta)}$ with a push-forward $\Phi_{1\#}^{(\theta)}\rho_0 = \rho_1$. In other words, if we draw a sample $\mathbf{x}_0 \sim \rho_0$, and integrate Eq. (22) for τ from 0 to 1, then we obtain a sample from $\mathbf{x}_1 \sim \rho_1$. In Ref. [35], two versions of latent and ambient TI were proposed. Here we apply the ambient version, meaning that the initial distribution is the canonical position space distribution $\rho_0 = \mu_{\mathbf{q}}(\beta_0)$ at a high temperature β_0, while the target is the canonical distribution at a low temperature, i.e., $\rho_1 = \mu_{\mathbf{q}}(\beta_1)$. We choose the ambient version over the latent one because it extrapolates better to β values outside the training range and exhibits superior empirical performance.

Our implementation of TI is based on the two-sided interpolant [39],

$$\mathbf{x}(\tau) = I(\tau, \mathbf{x}_0, \mathbf{x}_1) + \gamma(\tau)\mathbf{z}, \tag{23}$$

where $\mathbf{z} \sim \mathcal{N}(0, \mathrm{Id})$, $\mathbf{x}_0 \sim \rho_0$, $\mathbf{x}_1 \sim \rho_1$. The function I is a deterministic interpolant that satisfies $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{x}(1) = \mathbf{x}_1$, and $\gamma(\tau)$ is a smooth function controlling the noise level at intermediate times (see Ref. [39] for details.) We train the model according to the stochastic interpolant loss function

$$\mathbf{L}[\mathbf{b}^{(\theta)}] = \mathbb{E}_{\tau,\mathbf{z},\mathbf{x}_0,\mathbf{x}_1}\big[\tfrac{1}{2}|\mathbf{b}^{(\theta)}(\tau, \mathbf{x}(\tau); \beta_0, \beta_1)|^2 - \mathbf{b}^{(\theta)}(\tau, \mathbf{x}(\tau); \beta_0, \beta_1) \cdot (\partial_\tau I(\tau, \mathbf{x}_0, \mathbf{x}_1) + \dot{\gamma}(\tau)\mathbf{z})\big], \tag{24}$$

employing antithetic sampling for smoother convergence, as previously described [39]. The loss function essentially minimizes the error between the learned velocity field and true one based on the interpolant. The expectation in the loss is taken independently over time τ, the random variable z, x_0 drawn from initial distribution, and x_1 drawn from the target distribution.

Once the TI model has been trained, we can generate samples from the target distribution (lower temperature) given initial states from the initial distribution (higher temperature). We also compute the *importance weight* of each sample as proposed in Ref. [35], in order to account for any bias or inaccuracy in the model output distribution. These weights are later used to correct towards the equilibrium expectation of mass and stiffness matrices $\mathbf{G}$ and $\mathbf{A}$ within the gEDMD algorithm (see the next section), in the same fashion as in Ref. [36]. It is worth noting that the weights are computed with respect to the unnormalized distribution, as the normalization constant cancels out in the gEDMD algorithm.

B. gEDMD

In order to verify if metastable states, eigenvalues and implied timescales are preserved by the CG dynamics, we need to numerically approximate the coarse-grained infinitesimal generator $\mathcal{L}^\Xi$ in Eq. (16). As shown in Ref. [26], $\mathcal{L}^\Xi$ can be approximated by a data-driven method called gEDMD. Remarkably, the gEDMD algorithm can be applied directly on the full-space data without the need to compute the coefficients $\bar{\mathbf{b}}$ and $\bar{\mathbf{A}}$ of the CG dynamics; see also Ref. [34].

The algorithm requires n basis functions acting on the CG phase-space variables $\mathbf{z}$, $\mathbf{v}$,

$$\phi(\mathbf{z}, \mathbf{v}) = \{\phi_1(\mathbf{z}, \mathbf{v}), \ldots, \phi_n(\mathbf{z}, \mathbf{v})\},$$

along with position space samples $\{\mathbf{q}_l\}_{l=1}^m$ and the Jacobians of the CG map $\nabla\xi(\mathbf{q}_l)$ evaluated at each sample. We then augment these data by drawing momenta from the canonical momentum distribution, yielding phase-space samples $\mathbf{x}_l = (\mathbf{q}_l, \mathbf{p}_l)$. We project these samples into CG phase space as

$$\mathbf{y}_l = \begin{pmatrix} \mathbf{z}_l \\ \mathbf{v}_l \end{pmatrix} = \begin{pmatrix} \xi(\mathbf{q}_l) \\ \nabla\xi(\mathbf{q}_l)\mathbf{M}^{-1}\mathbf{p}_l \end{pmatrix}.$$

Next, we build $n \times m$-dimensional data matrices

$$\mathbf{\Phi} = [\phi(\mathbf{z}_1, \mathbf{v}_1) \quad | \quad \cdots \quad | \quad \phi(\mathbf{z}_m, \mathbf{v}_m)]$$
$$\mathcal{L}\mathbf{\Phi} = [[\mathcal{L}\phi](\mathbf{q}_1, \mathbf{p}_1) \quad | \quad \cdots \quad | \quad [\mathcal{L}\phi](\mathbf{q}_m, \mathbf{p}_m)].$$

To form the matrix $\mathcal{L}\mathbf{\Phi}$, we apply the full-state generator $\mathcal{L}$, which requires derivatives of the spatial CG map. Note that when we apply the full-state generator to functions acting on the CG space, we treat them as the composition $\phi \circ \Xi$. The final matrix approximation $\mathbf{L}$ to the CG generator $\mathcal{L}^\Xi$ is then given by

$$\widehat{\mathbf{L}} = \widehat{\mathbf{G}}^{-1}\widehat{\mathbf{A}}, \tag{25}$$

ALGORITHM 1. Numerical approximation of the coarse-grained Langevin equation.

Input: position space samples $\{\mathbf{q}_l\}_{l=1}^m$, basis functions $\phi(\mathbf{z}, \mathbf{v}) \in \mathbb{R}^n$.
1: **Data Augmentation:**
2: Complement position space samples by random momenta: $\mathbf{x}_l = (\mathbf{q}_l, \mathbf{p}_l)$.
3: Convert augmented samples into CG phase-space samples: $\mathbf{y}_l = (\mathbf{z}_l, \mathbf{v}_l) = \Xi(\mathbf{q}_l, \mathbf{p}_l)$.
4: **Generator Learning (gEDMD):**
5: Compute mass and stiffness matrices for the generator:

$$\hat{\mathbf{G}} = \frac{1}{m}\sum_{l=1}^{m} \phi(\mathbf{y}_l) \otimes \phi(\mathbf{y}_l), \qquad \hat{\mathbf{A}} = \frac{1}{m}\sum_{l=1}^{m} \phi(\mathbf{y}_l) \otimes [\mathcal{L}\phi](\mathbf{x}_l).$$

6: Compute generator matrix $\hat{\mathbf{L}} = \hat{\mathbf{G}}^{-1}\hat{\mathbf{A}}$.
7: **Parameter Estimation:**
8: Learn a parametric model for the effective force ($\bar{\mathbf{f}}_m$) and diffusion ($\bar{\mathbf{A}}_m$), as in Eq. (29).

where

$$\widehat{\mathbf{G}} = \frac{1}{m}\sum_{l=1}^{m} \phi(\mathbf{y}_l) \otimes \phi(\mathbf{y}_l), \quad \widehat{\mathbf{A}} = \frac{1}{m}\sum_{l=1}^{m} \phi(\mathbf{y}_l) \otimes [\mathcal{L}\phi](\mathbf{x}_l) \tag{26}$$

are called mass and stiffness matrices, respectively. To solve for $\widehat{\mathbf{L}}$ numerically, we do not explicitly form $\widehat{\mathbf{G}}$, instead, we perform a whitening transformation by removing the small singular values of $\mathbf{\Phi}$:

$$\mathbf{\Phi} = \mathbf{U}\mathbf{S}\mathbf{V}^\top, \quad \mathbf{R} = \mathbf{U}\mathbf{S}^{-1} \in \mathbb{R}^{n\times r}, \quad \widehat{\mathbf{L}}_r = \mathbf{R}^\top \widehat{\mathbf{A}}\, \mathbf{R}, \tag{27}$$

where $r \leqslant n$. In practice, we do not fix r in advance and instead set a threshold to cut off small singular values. Slowest timescales t_i of the system, or dominant eigenvalues κ_i of the generator, can be computed by diagonalization of the reduced matrix $\widehat{\mathbf{L}}_r$.

In this study, we use random Fourier features (RFFs) [44] as basis functions:

$$\phi(\mathbf{y}) = \{\cos(\omega_1^\top \mathbf{y}), \sin(\omega_1^\top \mathbf{y}), \ldots, \cos(\omega_n^\top \mathbf{y}), \sin(\omega_n^\top \mathbf{y})\},$$

where $\{\omega_1, \ldots, \omega_n\}$ are random frequency vectors drawn from a spectral distribution ρ. RFFs represent a low-rank approximation to a reproducing kernel function, e.g., associated to a Gaussian kernel [44]. The number of basis functions n, or number of random frequencies in the case of RFFs, and any parameter associated to the underlying kernel are hyperparameters that need to be optimized.

C. Parameter estimation for CG dynamics

To produce simulations of the coarse-grained Langevin SDE, we need to learn models for their diffusion and drift coefficients defined in Eq. (19). Let us define the *local force* $\mathbf{f}_{\text{loc}}$ and *local diffusion* $\mathbf{D}_{\text{loc}}$ as follows:

$$\mathbf{f}_{\text{loc}}(\mathbf{q}, \mathbf{p}) = -\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla V(\mathbf{q}) + \mathbf{p}^\top\mathbf{M}^{-1}\partial^2\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p},$$
$$\mathbf{D}_{\text{loc}}(\mathbf{q}, \mathbf{p}) = \nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla\xi(\mathbf{q})^\top. \tag{28}$$

These are the terms which are projected onto CG space in Eq. (20) to form the learnable parts of the effective drift (velocity component) and diffusion (lower right block). Using available training data, we can learn the effective drift and diffusion as best approximations of the local force and diffusion by solving data-driven regression problems, similarly to the one proposed in Ref. [34]:

$$\mathbf{D}_m = \underset{\tilde{\mathbf{D}}=\tilde{\mathbf{D}}(\mathbf{z},\mathbf{v})}{\arg\min} \frac{1}{m}\sum_{l=1}^{m} \|\tilde{\mathbf{D}}(\mathbf{z}_l, \mathbf{v}_l) - \mathbf{D}_{\text{loc}}(\mathbf{q}_l, \mathbf{p}_l)\|_F^2$$
$$\mathbf{f}_m = \underset{\tilde{\mathbf{f}}=\tilde{\mathbf{f}}(\mathbf{z},\mathbf{v})}{\arg\min} \frac{1}{m}\sum_{l=1}^{m} \|\tilde{\mathbf{f}}(\mathbf{z}_l, \mathbf{v}_l) - \mathbf{f}_{\text{loc}}(\mathbf{q}_l, \mathbf{p}_l)\|_F^2, \tag{29}$$

where the regression is carried out against m samples. Note that the learned effective force and diffusion are only functions of the CG variables. The complete effective drift and diffusion fields can then be constructed as

$$\bar{\mathbf{b}}_m = \begin{pmatrix} \mathbf{v} \\ \mathbf{f}_m(\mathbf{z}, \mathbf{v}) - \gamma\mathbf{v} \end{pmatrix}, \quad \bar{\mathbf{A}}_m = \begin{pmatrix} 0_{k\times k} & 0_{k\times k} \\ 0_{k\times k} & \gamma\mathbf{D}_m(\mathbf{z}, \mathbf{v}) \end{pmatrix}.$$

In the numerical examples, we parametrize the effective force and diffusion either as a linear combination of random Fourier features or as the output of a shallow neural network (NN). We found that the solution of the RFF regression tends to be accurate within data-rich regions, while the NN provides better extrapolation accuracy at the cost of increased computational cost. During simulations of the CG dynamics, we apply a hybrid method to leverage the strengths of both models by a switching rule: The RFF-based solution is used within data-rich regions, while the more accurate NN is used for extrapolation outside this region.

The summary of the numerical approximation of the CG dynamics is given in Algorithm 1.

V. NUMERICAL EXAMPLE

We illustrate the coarse graining scheme discussed in Secs. II and III and the proposed numerical methods in Sec. IV using the two-dimensional lemon slice potential.

A. Lemon slice potential

The lemon slice potential field is a function of two-dimensional polar coordinates r and θ by

$$V(r, \theta) = 1.2 \cdot [\cos(4\theta) + 10(r-1)^2]. \tag{30}$$

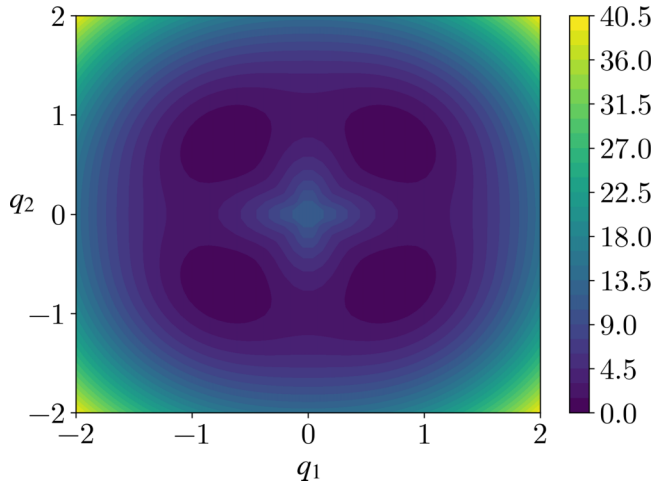

FIG. 1. Potential field of the lemon slice system.

The energy landscape of the system is shown in Fig. 1, consisting of four minima, which correspond to metastable states at low temperature.

The potential field gives rise to underdamped Langevin dynamics (1) on four-dimensional phase space $(\mathbf{q}, \mathbf{p}) \in \mathbb{R}^4$.

Following previous work [23,34], the polar angle θ is a suitable spatial CG coordinate for this system as it separates the four minima. According to Eq. (11), we can define the phase-space CG map as

$$\Xi(\mathbf{q}, \mathbf{p}) = \begin{pmatrix} \xi(\mathbf{q}) \\ \nabla\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p} \end{pmatrix}, \quad \xi(\mathbf{q}) = \theta = \operatorname{atan}\left(\frac{\mathbf{q}_2}{\mathbf{q}_1}\right).$$

Figure 2 shows an illustration of the phase-space CG map for $\mathbf{M} = \mathrm{Id}$. For a fixed point in CG space at $(\mathbf{z}, \mathbf{v}) = (\theta_0, 0)$, the level set $\Xi^{-1}(\theta_0, 0)$ corresponds to the Cartesian product of the blue lines shown in position and momentum space, respectively. We prepare training datasets based on all methods reported in Sec. IV A: rejection sampling, ergodic simulations, and thermodynamic interpolation (see next section for more details). The simulation details, as well as the selected parameters are given in Appendix B 2.

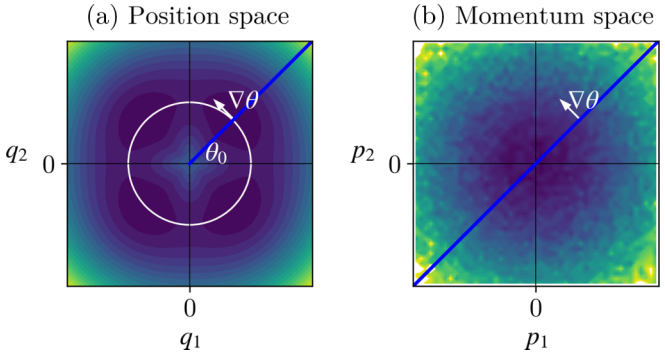

FIG. 2. CG definition for the lemon slice system, in position and momentum space. Points belonging to the blue line in positional space have spatial CG component of $\xi(r, \theta) = \theta_0$, and points belonging to the blue line in the momentum space are perpendicular to $\nabla\theta$ and therefore have zero velocity CG component. Note that for $\mathbf{v} \neq 0$, the preimage in momentum space would be more complex as the magnitude of the gradient $\nabla\theta$ depends on r.

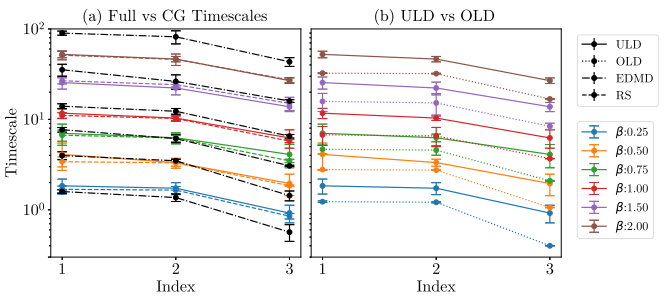

FIG. 3. (a) Slowest timescales t_1 to t_3 of the coarse-grained Langevin equation (17) with the lemon slice potential, corresponding to different values of the inverse temperature β, computed on rejection sampling data (RS, dashed lines) or on simulations of the underdamped Langevin SDE (ULD, solid lines), compared against the baseline timescales from the EDMD method (EDMD, black dash-dotted lines). (b) Timescales of the underdamped CG Langevin dynamics (ULD, solid lines) versus overdamped CG Langevin dynamics (OLD, dotted lines).

B. Timescale analysis

We begin by comparing the implied timescales of the full four-dimensional Langevin dynamics (1) and the coarse-grained dynamics (17). We also study their dependence on the inverse temperature β.

1. Identical masses

We first study the case where $\mathbf{M}$ is the identity matrix. To obtain a baseline, we apply the EDMD method [45] to the long ergodic simulation data and compute the eigenvalues and timescales of the full system's Koopman operator. The basis set for EDMD consists of RFF on four-dimensional phase space, which have been shown to provide accurate approximations of eigenvalues and timescales [46]. The dominant timescales for inverse temperature values in a range $\beta \in [0.25, 2.0]$ are shown in Fig. 3. The results show that for the selected inverse temperature range, the timescales vary by about two orders of magnitude, which proves this setting to be an interesting test case for the effectiveness of sampling algorithms.

Next, we compute an approximation of the projected generator $\mathcal{L}^{\Xi}$ following Algorithm 1, using both datasets obtained by rejection sampling and by ergodic simulation of the full Langevin dynamics. Again we use RFF basis functions, their main hyperparameters (the number of random features and the kernel bandwidth) are optimized using cross validation based on the VAMP score. Details of this analysis are given in Appendix B 1. We then solve the associated eigenvalue problem for the projected generator, which provides the eigenvalues and implied transition timescales of the CG dynamics (17). The results depicted in Fig. 3 show agreement of the CG generator's timescales with the baseline, regardless of the data generation scheme. We conclude that if the spatial CG map is chosen well, then the underdamped CG dynamics (17) can be expected to retain slow transition timescales of the full system.

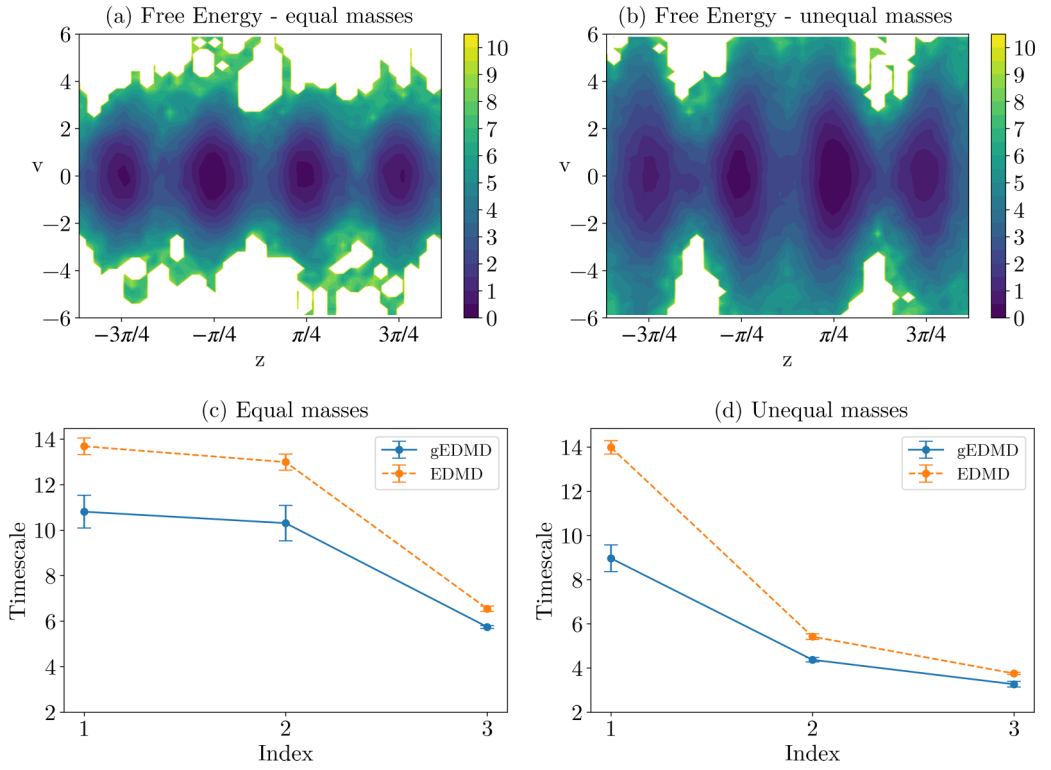

FIG. 4. Top row: Empirical free-energy surface for lemon slice with identity mass matrix in (a) and with unequal masses in (b), both at $\beta = 1$. Bottom row: Slowest timescales t_1 to t_3 of the coarse-grained Lemon slice system with equal masses in (c) and unequal masses in (d), both based on the integration of the underdamped Langevin SDE at $\beta = 1$.

Moreover, we can use the gEDMD algorithm to analyze the properties of the CG Langevin dynamics.

We also compute the timescales of the system if it was governed by overdamped Langevin dynamics in position space only. The results show a similar temperature dependence and are almost identical but slightly smaller compared to the underdamped case. In previous work, it was observed that overdamped dynamics lead to an acceleration of the timescales [34]. The effect is not drastic for this model system but can be expected to be much more noticeable for larger-scale systems.

2. *Unequal masses*

Next, we study the case where different masses act along different coordinate directions by setting $\mathbf{M} = \mathrm{diag}(1.00, 0.25)$, meaning that the mass in the horizontal direction is larger than along the vertical direction. The free-energy surface in CG space is depicted in Figs. 4(a) and 4(b). The full space momentum distribution is more spread out in the $\mathbf{p}_2$ direction for decreased m_2, and hence we expect to see a larger spread of the $\mathbf{v}$ distribution in CG space, which is indeed observed in the free-energy surface.

Next, we again approximate the projected generator $\mathcal{L}^{\Xi}$ to obtain the timescales of this system. Figure 4(d) show the dominant timescales computed via the gEDMD algorithm compared to EDMD timescales serving as a reference. Both sets of timescales are once again in agreement. With the introduction of unequal masses, the first and second nontrivial timescales are no longer identical, as the unequal masses break the symmetry of the system. As a result, the transition from the left- to the right-hand side of the potential field is slower than the transition from the bottom to top part of the potential field, leading to a down-shift of t_2.

C. Application of thermodynamic interpolation

Next, we show that thermodynamic interpolation can produce accurate position space samples outside its training regime, allowing us to estimate transition timescales using gEDMD. Returning to the case where $\mathbf{M} = \mathrm{Id}$, the TI model is trained against a dataset consisting of samples from a simulation of the lemon slice system at high temperatures $\beta \in [0.25, 0.50, 0.75, 1.00]$. In each training step, we choose two specific values of β from this set, along with one sample corresponding to each of these values, and ask the model to interpolate between them. This process is repeated while shuffling through different combinations of β. The model is then validated against simulation data corresponding to low temperature $\beta = 2.00$ interpolated from $\beta = 1.00$. Further details on the TI architecture and training are provided in Appendix B 4.

Once we have trained the generative model, we use it to produce samples at various temperatures $\beta \in [0.25, 0.50, 0.75, 1.00, 1.50, 2.00]$ interpolated from $\beta = 1.00$. Note that the sampler provides the positional information of the samples, so as before, we complement these samples by drawing their momentum components according to a Gaussian distribution with zero mean and covariance $\frac{1}{\beta}\delta_{ij}$. Figure 5 shows the histogram of the positional data obtained

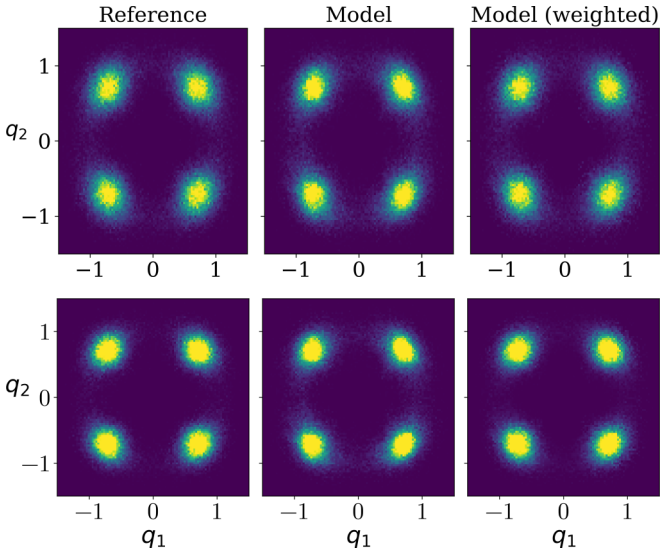

FIG. 5. Histograms comparing the dataset obtained via rejection sampling (left) with data generated by the TI model without (middle) and with (right) importance sampling weights. First row is for the case of $\beta = 1.50$ and the second row is for the case of $\beta = 2.00$.

via TI model for $\beta \in [1.50, 2.00]$. One can see that the TI model produces samples whose distribution is very similar to the one of the rejection sampling scheme. The last column of the figure shows that results improve further after taking the importance weights of the samples into account, making the distributions near-indistinguishable.

Using the augmented TI dataset, we perform gEDMD as in Algorithm 1 and compute the timescales of the system via estimation of the generator. Figure 6 shows the timescales corresponding to the temperature values listed above. Here we incorporated the sample weights computed via importance sampling only for $\beta \in [1.50, 2.00]$ as it does not affect other

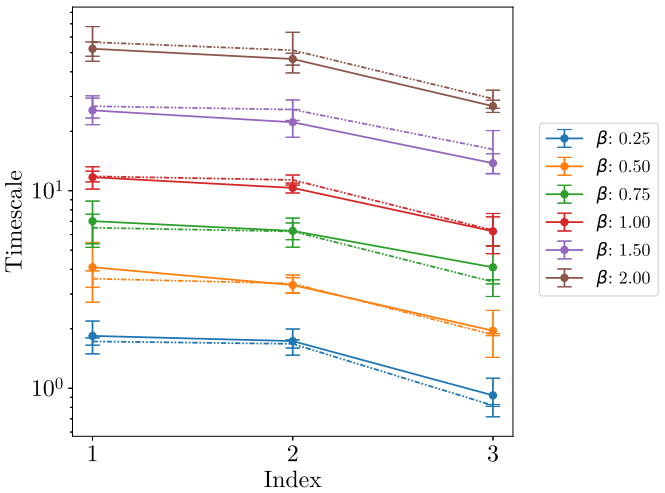

FIG. 6. Slowest timescales t_1 to t_3 of the Lemon slice potential corresponding to different values of the inverse temperature β, computed using data generated via simulations of the underdamped Langevin SDE (solid lines) and via the TI generative model (dash-dot-dotted lines).

cases which were included in the training phase. The figure shows that dataset generated via TI allows to recover the timescales computed via other sources of data up to statistical error. Thus, the TI data allows to predict the kinetic properties of a system that takes an order of magnitude longer to equilibrate relative to the training data. It is worth highlighting that the information about the case with $\beta = 1.50$ has neither been involved in the training nor validation of the TI model, which emphasizes the extrapolation capabilities of TI. This is a significant result since the sole usage of the sampler to generate equilibrated samples takes considerably less time than integrating the SDE governing the underlying underdamped Langevin dynamics at unseen β.

D. CG dynamics learning

We now turn our attention to identifying the CG dynamics by solving the two minimization problems (29) in Sec. IV C. Figure 7 shows the learned effective force and diffusion for $\beta = 1$. Results for learned diffusion show that it is almost constant as a function of $\mathbf{z}$ and within the strongly populated region $\mathbf{v} \in [-2, 2]$, while it keeps increasing in magnitude towards the tail of CG velocity distribution. A similar behavior can be observed for the learned drift, with the difference that it exhibits a sinusoidal behavior as a function of $\mathbf{z}$. Note that the CG drift is two-dimensional, but we only need to learn the velocity component. We repeat the same process for the case of $\beta = 2$, and it results in similar patterns as shown in Fig. 8.

E. Analysis of learned CG dynamics

After learning the coefficients $\bar{\mathbf{b}}_m$ and $\bar{\mathbf{A}}_m$ of the CG Langevin dynamics, we integrate the learned dynamics to verify its statistical and dynamical properties agree with reference values. Figure 9(a) shows the position space time series of two independent simulations of the CG dynamics at $\beta = 1$ for 10^6 time steps. The trajectories visit the expected metastable regions which are highlighted by horizontal dashed lines. We also compute an empirical free-energy surface in the CG space based on these simulation data, as shown in Fig. 9(b), which is in agreement with the potential of mean force obtained from the full dynamics.

In Fig. 9(d), we compare multiple ways of assessing the slowest timescales of the CG Langevin process. First, we show the same timescales as in Fig. 3 (blue lines), which were obtained by applying Algorithm 1 to the original data in full space. These timescales represent a direct approximation of the CG generator $\mathcal{L}^{\Xi}$ using the full space training data. Next, we can project the full space data into CG space, and then use the gEDMD method with these data and the learned parameters $\bar{\mathbf{A}}_m$ and $\bar{\mathbf{b}}_m$, to estimate a model for the generator of the learned CG dynamics. Note that this model for the learned CG generator, and its implied transition timescales, can be estimated without explicit integration of the CG Langevin dynamics. The corresponding implied timescales are shown by the orange line in Fig. 9(d).

Using the CG simulation data and the gEDMD algorithm, we can compute yet another model for the learned generator

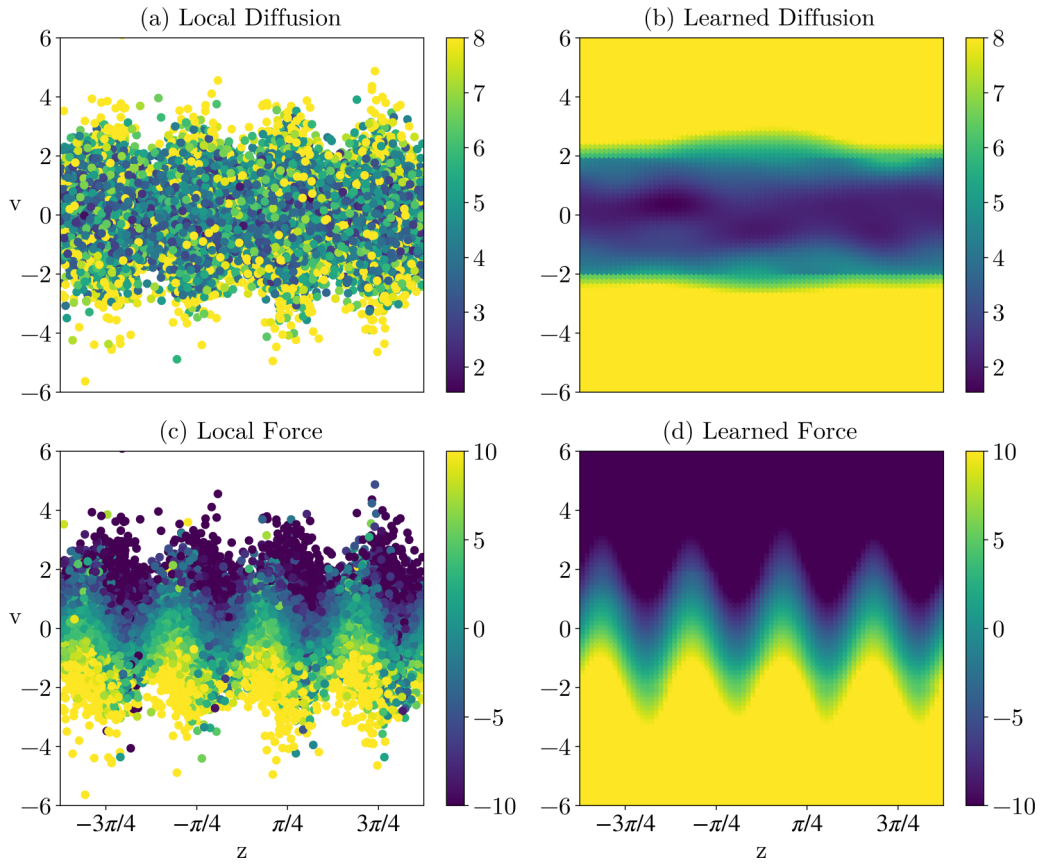

FIG. 7. Learning of the effective diffusion ($\bar{\mathbf{A}}_m$) and effective drift ($\bar{\mathbf{b}}_m$) for $\beta = 1$. (a) shows a scatter plot of the local diffusion $\mathbf{D}_{\text{loc}}$, while (b) shows the learnable component $\mathbf{D}_m$ of the effective diffusion. Similarly, (c) shows a scatter of the local force $\mathbf{f}_{\text{loc}}$, while (d) shows the learnable component $\mathbf{f}_m$ of the effective drift.

in CG space. The corresponding timescales are depicted by the green line. Finally, we also apply the EDMD method at finite lag time to the CG simulation data, these timescales are represented by the red line. It is clear that all of these timescale estimates are in agreement within statistical error. These results show that we can evaluate transition timescales of the CG dynamics *a priori* by applying gEDMD to the full space training data. We can also assess the quality of the learned CG dynamics before running CG simulations. All of these results are consistent with the timescales we obtain after actually integrating the CG dynamics in time.

We also apply robust Perron cluster cluster analysis (PCCA+) algorithm [47] to the eigenvectors of the CG generator built upon the integrated dataset to assign to each sample point a degree of membership to each metastable state. As depicted in Fig. 9(c), we can confirm that metastable states are also recovered. We conclude that all dynamical properties related to slow transitions are successfully retained by the learned dynamics.

Last, we repeat the same experiment for the case of $\beta = 2$. Figure 10 shows the results, which confirm all of our findings for the previous setting. We note that the CG dynamics are significantly more metastable than for $\beta = 1$, as the oscillations around the minima are smaller and transitions occur rarely. In fact, extremely long CG simulations were needed to obtain statistically significant reference values. This highlights the value of our proposed approach: We can faithfully predict CG dynamics without long-time simulations and their transition timescales without explicit integration of the CG dynamics.

VI. SUMMARY

We investigated coarse graining of the underdamped Langevin dynamics based on a phase-space CG map which uses both a coarse-grained position and velocity component. We showed that application of the Zwanzig projection leads to a coarse-grained dynamics with a similar structure as the full Langevin equation. We have also shown that the CG generator can be decomposed into symmetric and antisymmetric parts resembling the Hamiltonian and Ornstein-Uhlenbeck parts of the full generator. Moreover, we showed that thermodynamic interpolation, a generative modeling technique that interpolates in temperature space, can be used to produce accurate samples of the position space distribution after training on high-temperature simulations. These samples can be combined with data-driven learning techniques, in particular the gEDMD method, to analyze kinetic properties of the CG dynamics, and to estimate its parameters, without explicit integration of the coarse-grained model. By means of a two-dimensional model potential, we demonstrated that the CG dynamics reproduce thermodynamic and kinetic properties of the full-space dynamics.

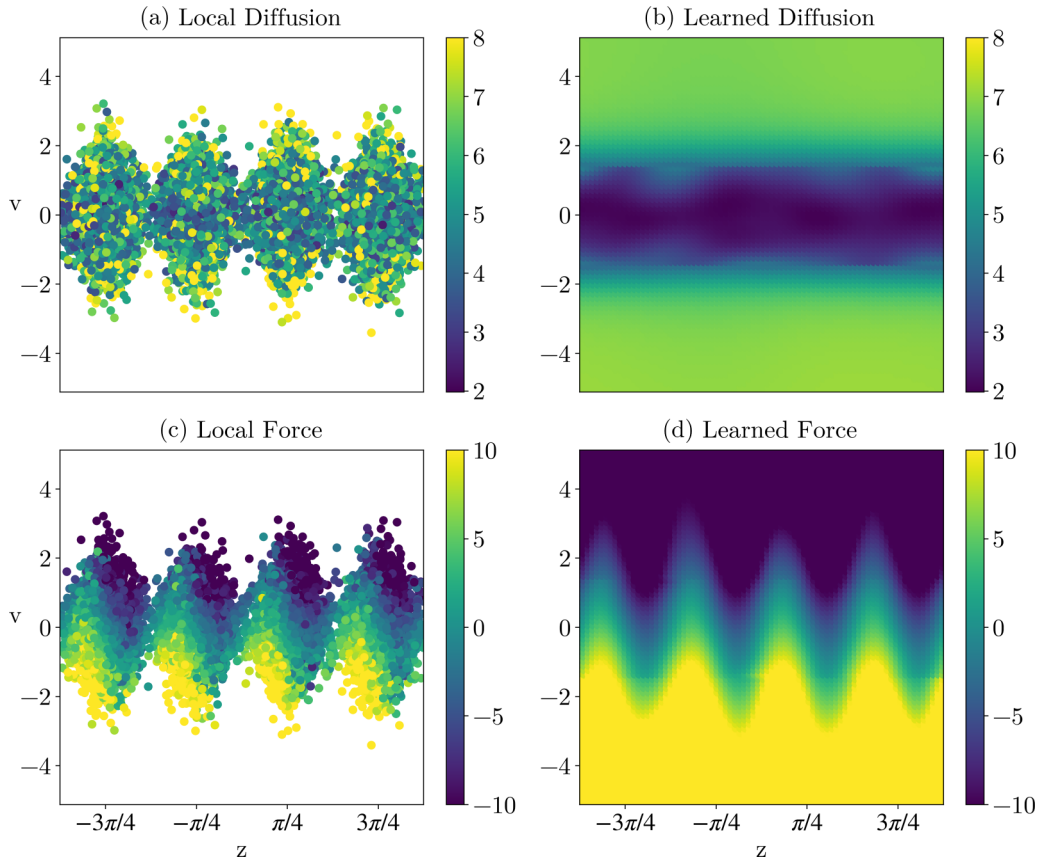

FIG. 8. Learning of the effective diffusion ($\bar{\mathbf{A}}_m$) and effective drift ($\bar{\mathbf{b}}_m$) for $\beta = 2$. (a) shows a scatter plot of the local diffusion $\mathbf{D}_{\text{loc}}$, while (b) shows the learnable component $\mathbf{D}_m$ of the effective diffusion. Similarly, (c) shows a scatter of the local force $\mathbf{f}_{\text{loc}}$, while (d) shows the learnable component $\mathbf{f}_m$ of the effective drift.

ACKNOWLEDGMENTS

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation (to S.O.) and by a project grant from the Knut and Alice Wallenberg Foundation "From Atom to Organism: Bridging the Scales in the Design of Ion Channel Drugs" (to S.M.). The work of L.N. and C.H. has been partially supported by the DFG Collaborative Research Center 1114 "Scaling Cascades in Complex Systems," Project No. 235221301; Project No. A05, "Probing Scales in Equilibrated Systems by Optimal Nonequilibrium Forcing"; and by the German Federal Government, the Federal Ministry of Research, Technology and Space and the State of Brandenburg within the framework of the joint project EIZ: Energy Innovation Center (Projects No. 85056897 and No. 03SF0693A) with funds from the Structural Development Act (Strukturstärkungsgesetz) for coal-mining regions.

DATA AVAILABILITY

The data that support the findings of this article are openly available [48].

APPENDIX A: PROOFS FOR ANALYTICAL RESULTS

In this section, we provide technical proofs for Propositions 1 and 2. We start by verifying the structure of the CG parameters (20), which follows quite directly after calculating the derivatives of the phase-space CG map. The proof of Proposition 2 is along the lines of Proposition 4 in Ref. [22], which deals with CG for SDEs with nondegenerate noise. We spell out these results for the underdamped Langevin equation. In a few places, we use the summation convention that repeated indices are being summed over, as well as the shorthands $\partial_i = \frac{\partial}{\partial \mathbf{q}_i}$, $\partial_{ij} = \frac{\partial^2}{\partial \mathbf{q}_i \partial \mathbf{q}_j}$.

1. Proof of Proposition 1

Let us compute the derivatives of the CG map Ξ. We have for any $l \in \{1, \dots, k\}$

$$\nabla \mathbf{z}_l = \begin{pmatrix} \nabla \xi_l^\top \\ 0_d \end{pmatrix} \in \mathbb{R}^{2d}, \quad \nabla^2 \mathbf{z}_l = \begin{pmatrix} \nabla^2 \xi_l & 0_{d\times d} \\ 0_{d\times d} & 0_{d\times d} \end{pmatrix} \in \mathbb{R}^{2d\times 2d}$$

$$\nabla \mathbf{v}_l = \begin{pmatrix} \nabla^2 \xi_l \mathbf{M}^{-1} \mathbf{p} \\ \nabla \xi_l^\top \mathbf{M}^{-1} \end{pmatrix} \in \mathbb{R}^{2d},$$

$$\nabla^2 \mathbf{v}_l = \begin{pmatrix} \partial_{ijk} \xi_l m_i \mathbf{p}_i & \nabla^2 \xi_l \mathbf{M}^{-1} \\ \nabla^2 \xi_l \mathbf{M}^{-1} & 0_{d\times d} \end{pmatrix} \in \mathbb{R}^{2d\times 2d}.$$

The key observation is that the lower right blocks of the Hessians of all CG variables Ξ_j vanish, which means that $\mathbf{A} : \nabla^2 \Xi_j \equiv 0$ for any $j \in \{1, \dots, 2k\}$. Therefore, all second-order terms vanish when computing the action of the generator $\mathcal{L}\Xi$ on the phase-space CG map. With this in mind, the

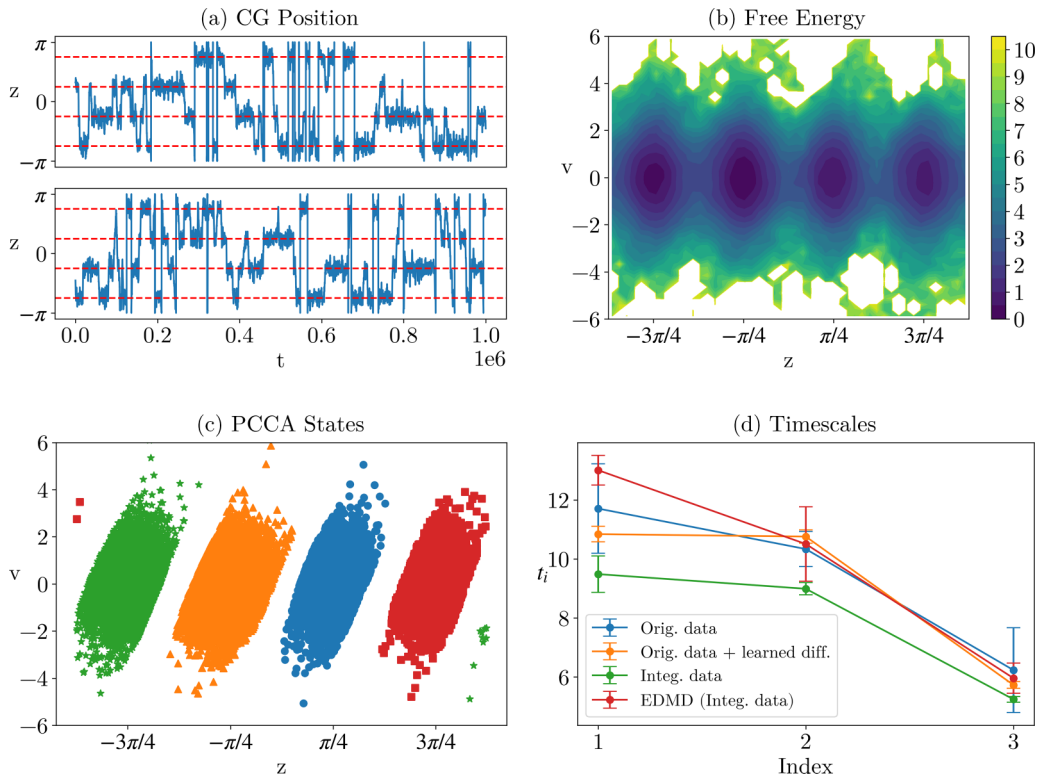

FIG. 9. Integration of coarse-grained SDE for $\beta = 1$. Time evolution of the CG position for exemplary simulations in (a), empirical free-energy surface in (b), PCCA+ analysis of the CG dynamics in (c), and a comparison for the timescales of the coarse-grained system in (d). For a detailed description of panel (d), see the main text.

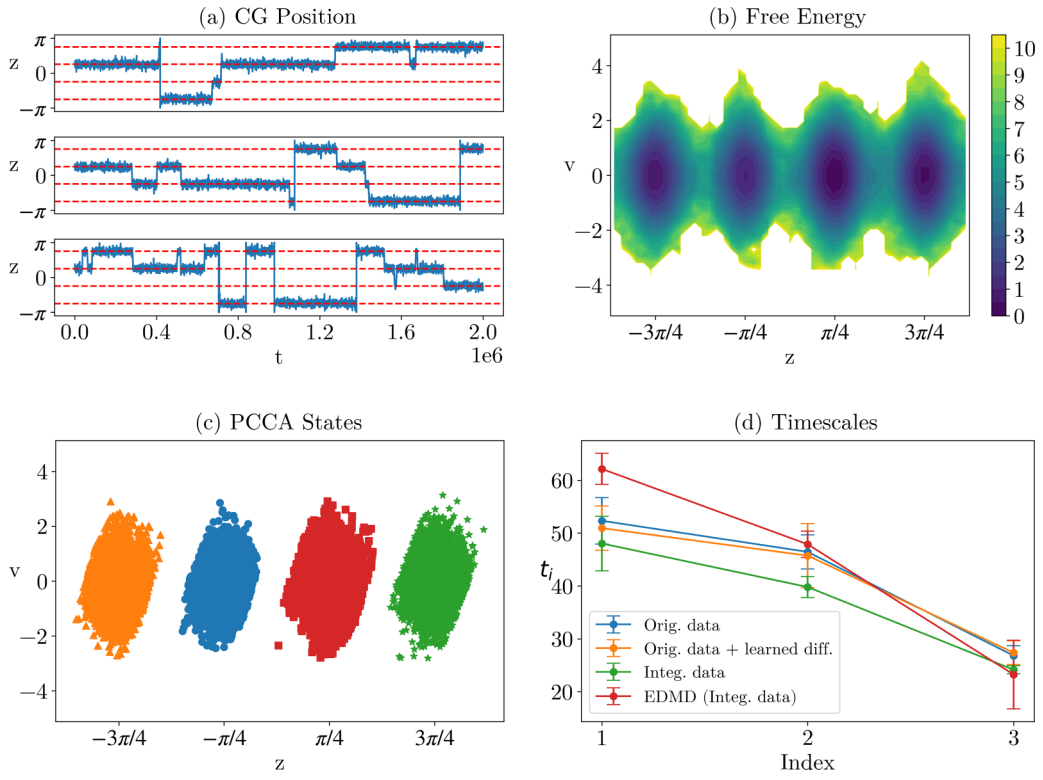

FIG. 10. Integration of coarse-grained SDE for $\beta = 2$. Time evolution of the CG position for exemplary simulations in (a), empirical free-energy surface in (b), PCCA+ analysis of the CG dynamics in (c), and a comparison for the timescales of the coarse-grained system in (d). For a detailed description of panel (d), see the main text.

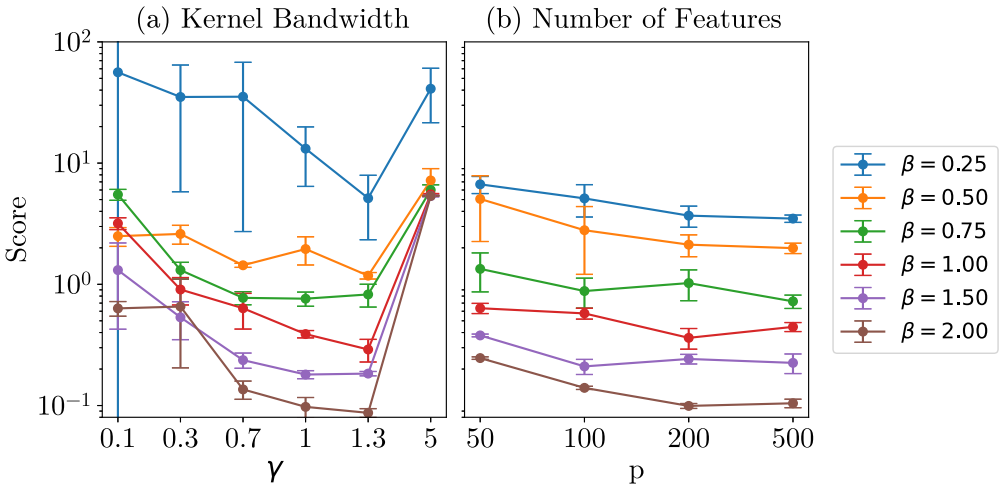

FIG. 11. VAMP-score analysis. Score corresponding to various values of the kernel bandwidth and number of Fourier features in (a) and (b), respectively.

computation of the effective drift $\bar{\mathbf{b}}$ according to Eq. (18) simplifies to

$$\begin{aligned}\bar{\mathbf{b}}(\mathbf{z},\mathbf{v}) &= \mathcal{P}(\mathcal{L}\Xi) = \mathcal{P}(\nabla\Xi\mathbf{b}+0) \qquad \text{(A1)}\\ &= \mathcal{P}\left(\begin{pmatrix}\nabla\xi(\mathbf{q}) & 0_{k\times d}\\ \mathbf{p}^\top\mathbf{M}^{-1}\partial^2\xi(\mathbf{q}) & \nabla\xi(\mathbf{q})\mathbf{M}^{-1}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{M}^{-1}\mathbf{p}\\ -\nabla V(\mathbf{q})-\gamma\mathbf{p}\end{pmatrix}\right)\\ &= \begin{pmatrix}\mathcal{P}(\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p})\\ \mathcal{P}(-\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla V(\mathbf{q})+\mathbf{p}^\top\mathbf{M}^{-1}\partial^2\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p}-\gamma\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p})\end{pmatrix}\\ &= \begin{pmatrix}\mathbf{v}\\ \mathbf{f}(\mathbf{z},\mathbf{v})-\gamma\mathbf{v}\end{pmatrix}. \qquad \text{(A2)}\end{aligned}$$

A similar calculation shows that the effective diffusion coefficient $\bar{\mathbf{A}}$ is given by

$$\bar{\mathbf{A}} := \mathcal{P}(\nabla\Xi\mathbf{A}\nabla\Xi^\top) = \begin{pmatrix}0_{k\times k} & 0_{k\times k}\\ 0_{k\times k} & \gamma\mathcal{P}(\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla\xi(\mathbf{q})^\top)\end{pmatrix} = \begin{pmatrix}0_{k\times k} & 0_{k\times k}\\ 0_{k\times k} & \gamma\mathbf{D}(\mathbf{z},\mathbf{v})\end{pmatrix},$$

with $\mathbf{D}(\mathbf{z},\mathbf{v})$ defined as $\mathbf{D}(\mathbf{z},\mathbf{v}) = \mathcal{P}(\nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla\xi(\mathbf{q})^\top)(\mathbf{z},\mathbf{v})$.

2. Proof of Proposition 2

The decomposition (21) of the coarse-grained generator $\mathcal{L}^\Xi$ is obtained in multiple steps.

a. A different expression for the effective drift

First, we derive an alternative expression for the effective drift $\bar{\mathbf{b}}$, which is given by the projection of the action of $\mathcal{L}$ on the phase-space CG map. We derive this relation in a weak form, that is, we multiply $\bar{\mathbf{b}}$ by an arbitrary function h on the CG space, and integrate against the Boltzmann factor $e^{-\beta F}$. An application of the coarea formula (13) together with the relation (9) then lead us to the following expression in terms of integrals on full phase space:

$$\begin{aligned}&\int_{\mathbb{R}^{2k}}\bar{\mathbf{b}}(\mathbf{y})h(\mathbf{y})e^{-\beta F(\mathbf{y})}\mathrm{d}\mathbf{y}\\ &\quad= \int_{\mathbb{R}^{2k}}\mathcal{P}(\mathcal{L}\Xi)(\mathbf{y})h(\mathbf{y})e^{-\beta F(\mathbf{y})}\mathrm{d}\mathbf{y}\\ &\quad= \int_{\mathbb{R}^{2d}}\mathcal{L}\Xi(\mathbf{x})h(\Xi(\mathbf{x}))e^{-\beta H(\mathbf{x})}\mathrm{d}\mathbf{x}\\ &\quad= \left\langle\mathbf{J}^{\mathrm{eq}}\nabla\Xi^\top,\, h\circ\Xi\right\rangle_\mu - \frac{1}{\beta}\left\langle\mathbf{A}\nabla\Xi^\top,\,\nabla(h\circ\Xi)\right\rangle_\mu.\end{aligned}$$

The gradient $\nabla(h\circ\Xi)$ is taken with respect to the full phase-space variables $(\mathbf{q},\,\mathbf{p})$ and by the chain rule we have $\nabla_{\mathbf{x}}(h\circ\Xi) = \nabla_{\mathbf{y}}h\,\nabla\Xi$. Inserting this relation, and applying the coarea formula in the reverse direction, we reexpress the above in terms of CG space integrals and projections:

$$\begin{aligned}&\langle\mathbf{J}^{\mathrm{eq}}\nabla\Xi^\top,\, h\circ\Xi\rangle_\mu - \frac{1}{\beta}\langle\mathbf{A}\nabla\Xi^\top,\,\nabla(h\circ\Xi)\rangle_\mu\\ &\quad= \int_{\mathbb{R}^{2k}}\mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^\top)(\mathbf{y})h(\mathbf{y})e^{-\beta F(\mathbf{y})}\mathrm{d}\mathbf{y}\\ &\qquad- \frac{1}{\beta}\int_{\mathbb{R}^{2k}}\mathcal{P}(\nabla\Xi\mathbf{A}\nabla\Xi^\top)(\mathbf{y})\nabla_{\mathbf{y}}h(\mathbf{y})e^{-\beta F(\mathbf{y})}\mathrm{d}\mathbf{y}.\end{aligned}$$

It remains to apply a partial integration to the second term, which provides:

$$\begin{aligned}&\int_{\mathbb{R}^{2k}}\bar{\mathbf{b}}(\mathbf{y})h(\mathbf{y})e^{-\beta F(\mathbf{y})}\mathrm{d}\mathbf{y}\\ &\quad= \int_{\mathbb{R}^{2k}}\Big[\mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^\top)(\mathbf{y}) + \frac{1}{\beta}e^{\beta F(\mathbf{y})}\,\nabla_{\mathbf{y}}\\ &\qquad\times\,[\mathcal{P}(\nabla\Xi\mathbf{A}\nabla\Xi^\top)(\mathbf{y})e^{-\beta F(\mathbf{y})}]\Big]h(\mathbf{y})e^{-\beta F(\mathbf{y})}\mathrm{d}\mathbf{y}.\end{aligned}$$

As the function h is arbitrary, and recognizing the definition (18) of the effective diffusion field, we have shown that

$$\bar{\mathbf{b}}(\mathbf{y}) = \mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})(\mathbf{y}) + \frac{1}{\beta}e^{\beta F(\mathbf{y})}\nabla_{\mathbf{y}}\cdot[\bar{\mathbf{A}}(\mathbf{y})e^{-\beta F(\mathbf{y})}]. \quad \text{(A3)}$$

Using that only the lower right block of $\bar{\mathbf{A}}$ is nonzero, the second term simplifies to

$$\frac{1}{\beta}e^{\beta F(\mathbf{y})}\nabla_{\mathbf{y}}\cdot[\bar{\mathbf{A}}(\mathbf{y})e^{-\beta F(\mathbf{y})}] = \frac{\gamma}{\beta}e^{\beta F(\mathbf{y})}\begin{pmatrix}0\\ \nabla_{\mathbf{v}}\cdot[\mathbf{D}(\mathbf{y})e^{-\beta F(\mathbf{y})}]\end{pmatrix}.$$

This furnishes the final expression for the effective drift,

$$\bar{\mathbf{b}}(\mathbf{y}) = \mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})(\mathbf{y}) + \frac{\gamma}{\beta}e^{\beta F(\mathbf{y})}\begin{pmatrix}0\\ \nabla_{\mathbf{v}}\cdot[\mathbf{D}(\mathbf{y})e^{-\beta F(\mathbf{y})}]\end{pmatrix}. \quad \text{(A4)}$$

b. A first decomposition of the effective generator

Using Eq. (A4), we can start from the general definition of the CG generator $\mathcal{L}^{\Xi}$, and rewrite it as follows:

$$\begin{aligned}\mathcal{L}^{\Xi} &= \bar{\mathbf{b}}\cdot\nabla_{\mathbf{y}} + \frac{1}{\beta}\bar{\mathbf{A}} : \nabla_{\mathbf{y}}^2\\ &= \mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})\cdot\nabla_{\mathbf{y}} + \frac{\gamma}{\beta}e^{\beta F(\mathbf{y})}\nabla_{\mathbf{v}}\cdot[\mathbf{D}(\mathbf{y})e^{-\beta F(\mathbf{y})}]\\ &\quad\times\nabla_{\mathbf{v}} + \frac{\gamma}{\beta}\mathbf{D}(\mathbf{y}) : \nabla_{\mathbf{v}}^2.\end{aligned}$$

To rewrite the last term, we have used again that only the lower right block of $\bar{\mathbf{A}}(\mathbf{y})$ is nonzero. The second and third term can be summarized as follows:

$$\mathcal{L}^{\Xi} = \mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})\cdot\nabla_{\mathbf{y}} + \frac{\gamma}{\beta}e^{\beta F(\mathbf{y})}\nabla_{\mathbf{v}}\cdot[\mathbf{D}(\mathbf{y})e^{-\beta F(\mathbf{y})}\nabla_{\mathbf{v}}]. \quad \text{(A5)}$$

The decomposition (A5) is intriguing as the second term is in divergence form, which implies that it is automatically symmetric with respect to the inner product defined by $e^{-\beta F}$. The first term remains as a candidate for the antisymmetric part, which is going to be verified next.

c. Continuity equation

Rearranging Eq. (A3), we see that

$$\mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})(\mathbf{y}) = \bar{\mathbf{b}}(\mathbf{y}) - \frac{1}{\beta}e^{\beta F(\mathbf{y})}\nabla_{\mathbf{y}}\cdot[\bar{\mathbf{A}}(\mathbf{y})e^{-\beta F(\mathbf{y})}],$$

which is exactly the definition of the *probability flow* associated to the Boltzmann factor $e^{-\beta F}$ under the CG dynamics [15]. If we can show that $\mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})$ satisfies the continuity equation,

$$-\nabla_{\mathbf{y}}\cdot[\mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})(\mathbf{y})e^{-\beta F(\mathbf{y})}] = 0,$$

then it follows that $e^{-\beta F}$ is the invariant distribution for the CG dynamics. The continuity equation follows from a similar sequence of arguments that led us to the decomposition (A4).

TABLE I. Simulation setup.

Parameter	Value
Kernel bandwidth (γ)	1
Number of features (p)	400
Tolerance in SVD truncation	10^{-4}

Again, we test against an arbitrary function h and integrate over CG space:

$$\begin{aligned}&-\int_{\mathbb{R}^{2k}}\nabla_{\mathbf{y}}\cdot[\mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})(\mathbf{y})e^{-\beta F(\mathbf{y})}]\,h(\mathbf{y})\,\mathrm{d}\mathbf{y}\\ &\quad=\int_{\mathbb{R}^{2k}}\mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})(\mathbf{y})e^{-\beta F(\mathbf{y})}\nabla_{\mathbf{y}}\,h(\mathbf{y})\,\mathrm{d}\mathbf{y}\\ &\quad=\int_{\mathbb{R}^{2d}}\mathbf{J}^{\mathrm{eq}}(\mathbf{x})\nabla\Xi^{\top}(\mathbf{x})\nabla_{\mathbf{y}}h(\Xi(\mathbf{x}))e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x},\end{aligned}$$

where we applied partial integration and the coarea formula. Using the chain rule and another partial integration, we arrive at the continuity equation in full space, which is zero as $\mathbf{J}^{\mathrm{eq}}$ is already the stationary probability for the full system:

$$\begin{aligned}&\int_{\mathbb{R}^{2d}}\mathbf{J}^{\mathrm{eq}}(\mathbf{x})\nabla\Xi^{\top}(\mathbf{x})\nabla_{\mathbf{y}}h(\Xi(\mathbf{x}))e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x}\\ &\quad=\int_{\mathbb{R}^{2d}}\mathbf{J}^{\mathrm{eq}}(\mathbf{x})\cdot\nabla_{\mathbf{x}}h(\Xi(\mathbf{x}))e^{-\beta H(\mathbf{x})}\,\mathrm{d}\mathbf{x}\\ &\quad=-\int_{\mathbb{R}^{2d}}\nabla_{\mathbf{x}}\cdot[\mathbf{J}^{\mathrm{eq}}(\mathbf{x})e^{-\beta H(\mathbf{x})}]h(\Xi(\mathbf{x}))\,\mathrm{d}\mathbf{x}=0.\end{aligned}$$

Indeed, $e^{-\beta F}$ is the invariant distribution of the CG dynamics, and we can introduce the notation

$$\bar{\mathbf{J}}^{\mathrm{eq}}(\mathbf{y}) := \mathcal{P}(\mathbf{J}^{\mathrm{eq}}\nabla\Xi^{\top})(\mathbf{y})$$

for its stationary probability flow. The continuity equation also implies directly that the first term in Eq. (A5) is antisymmetric:

$$\begin{aligned}&\int_{\mathbb{R}^{2k}}\bar{\mathbf{J}}^{\mathrm{eq}}(\mathbf{y})\cdot\nabla_{\mathbf{y}}\phi(\mathbf{y})\,\psi(\mathbf{y})e^{-\beta F(\mathbf{y})}\,\mathrm{d}\mathbf{y}\\ &\quad=-\int_{\mathbb{R}^{2k}}\nabla_{\mathbf{y}}\cdot[\bar{\mathbf{J}}^{\mathrm{eq}}(\mathbf{y})e^{-\beta F(\mathbf{y})}]\phi(\mathbf{y})\,\psi(\mathbf{y})\,\mathrm{d}\mathbf{y}\\ &\qquad-\int_{\mathbb{R}^{2k}}\bar{\mathbf{J}}^{\mathrm{eq}}(\mathbf{y})\cdot\nabla_{\mathbf{y}}\psi(\mathbf{y})\,\phi(\mathbf{y})e^{-\beta F(\mathbf{y})}\,\mathrm{d}\mathbf{y}\\ &\quad=-\int_{\mathbb{R}^{2k}}\bar{\mathbf{J}}^{\mathrm{eq}}(\mathbf{y})\cdot\nabla_{\mathbf{y}}\psi(\mathbf{y})\,\phi(\mathbf{y})e^{-\beta F(\mathbf{y})}\,\mathrm{d}\mathbf{y}.\end{aligned}$$

d. Summary of the results

We have shown that Eq. (A5) furnishes the decomposition of the CG generator into its symmetric and antisymmetric parts, with invariant distribution $e^{-\beta F}$ and stationary probability flow $\bar{\mathbf{J}}^{\mathrm{eq}}$:

$$\begin{aligned}\mathcal{L}^{\Xi} &= \mathcal{L}_a^{\Xi} + \mathcal{L}_s^{\Xi},\\ \mathcal{L}_a^{\Xi}\phi &= \bar{\mathbf{J}}^{\mathrm{eq}}\cdot\nabla_{\mathbf{y}}\phi,\\ \mathcal{L}_s^{\Xi}\phi &= \frac{\gamma}{\beta}e^{\beta F(\mathbf{y})}\nabla_{\mathbf{v}}\cdot[\mathbf{D}(\mathbf{y})e^{-\beta F(\mathbf{y})}\nabla_{\mathbf{v}}\phi].\end{aligned}$$

TABLE II. Simulation setup.

Parameter	Value
Inverse temperature (β)	[0.25, 0.5, 0.75, 1.0, 1.5, 2.0]
Time constant ($1/\gamma$)	0.2
Time step	10^{-3}
Simulation time steps	10^{6}–10^{7}

TABLE III. Hyperparameters for the CG dynamics model.

Parameter	Value
Batch size	200
Number of hidden layer	1
Hidden layer size	8
Learning rate	$(10^{-4}, 10^{-1})$
Optimizer	Adam
Number of epochs	200

Last, we note that $\bar{\mathbf{J}}^{\mathrm{eq}}$ indeed accounts for the Hamiltonian part of the CG drift:

$$\begin{aligned}\bar{\mathbf{J}}^{\mathrm{eq}} &= \mathcal{P}(\nabla\Xi\mathbf{J}^{\mathrm{eq}})\\ &= \mathcal{P}\left(\begin{pmatrix}\nabla\xi(\mathbf{q}) & 0_{k\times d}\\ \mathbf{p}^{\top}\mathbf{M}^{-1}\partial^2\xi(\mathbf{q}) & \nabla\xi(\mathbf{q})\mathbf{M}^{-1}\end{pmatrix}\cdot\begin{pmatrix}\mathbf{M}^{-1}\mathbf{p}\\ -\nabla V(\mathbf{q})\end{pmatrix}\right)\\ &= \begin{pmatrix}\mathbf{v}\\ \mathcal{P}(\mathbf{p}^{\top}\mathbf{M}^{-1}\partial^2\xi(\mathbf{q})\mathbf{M}^{-1}\mathbf{p} - \nabla\xi(\mathbf{q})\mathbf{M}^{-1}\nabla V(\mathbf{q}))\end{pmatrix}\\ &= \begin{pmatrix}\mathbf{v}\\ \mathbf{f}\end{pmatrix}.\end{aligned}$$

APPENDIX B: SIMULATION AND LEARNING DETAILS

1. Vamp-score analysis

We tune the parameters of the random Fourier feature basis for the approximation of the generator based on the VAMP variational principle proposed in Ref. [49]. The two main parameters are the number of random Fourier features and the kernel bandwidth. Figure 11 shows the result for these two parameters at given temperature. This leads to the following optimal parameters used to produce the timescale plots shown in the main text.The VAMP-score analysis for the kernel bandwidth shows that there is a range of values that results in almost the same score, leading to flexibility in the choice of the parameter. For the number of features, we take the lowest value in the range where the score becomes constant. The summary of parameters relevant to the analysis is provided in Table I.

2. Parameters for Langevin simulations

We produced simulation data for the lemon slice system by integrating the Langevin SDE (1) using Euler-Maruyama scheme, with the input setting summarized in Table II.

TABLE IV. Hyperparameters and TI model settings.

Parameter	Value
Batch size	1024
Number of layers in MLP	5
Positional embedding size, N	256
Hidden layer size	256
Learning rate	10^{-4}
Optimizer	Adam
Number of epochs	75
Positional encoder scale, l_β	350
Positional encoder scale, l_t	10
Positional encoder scale, $l_{\mathbf{x}}$	50
Numerical solver	Euler

3. Neural network architecture for parameters of the effective dynamics

To learn the parameters $\bar{\mathbf{b}}_m$ and $\bar{\mathbf{A}}_m$ of the coarse-grained dynamics, we adopt a shallow fully connected feed-forward neural network with the setup summarized in Table III.

4. TI

For the TI model we employ the two-sided linear interpolant with a sine-squared noise schedule [39], implying that the interpolant is defined as

$$\mathbf{x}(\tau) = (1-\tau)\mathbf{x}_0 + \tau\mathbf{x}_1 + \sin^2(\pi\tau)\mathbf{z}, \tag{B1}$$

for $\tau \in [0,\ 1]$, $\mathbf{x}_0 \sim \rho_0(\mathbf{x}_0)$, $\mathbf{x}_1 \sim \rho_1(\mathbf{x}_1)$ and $\mathbf{z} \sim \mathcal{N}(0,\ \mathrm{Id})$.

To encourage extrapolation of the TI model to temperatures unseen during training, we embed the inverse temperatures β using a positional embedding $\lambda_{\mathrm{pos}}(\beta; l_\beta)$ [35]. Similarly, we also use positional embeddings $\lambda_{\mathrm{pos}}(\tau; l_\tau)$ and $\lambda_{\mathrm{pos}}(\mathbf{x}(\tau); l_{\mathbf{x}})$ to embed the positions $\mathbf{x}$ and times τ. We define the embeddings for some model feature $\mathbf{x}$ as $\lambda_{\mathrm{pos}}(\mathbf{x}(\tau); l_{\mathbf{x}}) = \{\lambda^n_{\mathrm{pos}}(\mathbf{x}; l_{\mathbf{x}})\}_{n=0}^{N}$, where

$$\lambda^n_{\mathrm{pos}}(\mathbf{x}; l) = \begin{cases}\cos\left(\left(1+\frac{n}{2}\right)+\frac{\pi\beta}{l}\right) & n \text{ even}\\ \sin\left(\left(1+\frac{n-1}{2}\right)+\frac{\pi\beta}{l}\right) & n \text{ odd},\end{cases} \tag{B2}$$

for some maximal embedding dimension N. Then, the embedded τ, β_0, β_1 and $\mathbf{x}(t)$ are projected through individual linear layers, concatenated and fed into a shared MLP. The network architecture and the corresponding hyperparameters are summarized in Table IV.

[1] P. Das, M. Moll, H. Stamati, L. E Kavraki, and C. Clementi, Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction, Proc. Natl. Acad. Sci. USA **103**, 9885 (2006).

[2] C. Clementi, Coarse-grained models of protein folding: Toy models or predictive tools? Curr. Opin. Struct. Biol. **18**, 10 (2008).

[3] P. Benner, S. Gugercin, and K. Willcox, A survey of projection-based model reduction methods for parametric dynamical systems, SIAM Rev. **57**, 483 (2015).

[4] J. Jin, A. J. Pak, A. E. P. Durumeric, T. D. Loose, G. A. Voth, Bottom-up coarse-graining: Principles and perspectives, J. Chem. Theory Comput. **18**, 5759 (2022).

[5] M. A. Rohrdanz, W. Zheng, and C. Clementi, Discovering mountain passes via Torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions, Annu. Rev. Phys. Chem. **64**, 295 (2013).

[6] H. Sidky, W. Chen, and A. L. Ferguson, Machine learning for collective variable discovery and enhanced sampling

in biomolecular simulation, Mol. Phys. **118**, e1737742 (2020).

[7] P. Gkeka, G. Stoltz, A. B. Farimani, Z. Belkacemi, M. Ceriotti, J. D. Chodera, A. R. Dinner, A. L. Ferguson, J.-B. Maillet, H. Minoux, C. Peter, F. Pietrucci, A. Silveira, A. Tkatchenko, Z. Trstanova, R. Wiewiora, and T. Lelièvre, Machine learning force fields and coarse-grained variables in molecular dynamics: Application to materials and biological systems, J. Chem. Theory Comput. **16**, 4757 (2020).

[8] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models, J. Chem. Phys. **128**, 244114 (2008).

[9] J. Wang, S. Olsson, C. Wehmeyer, A. Pérez, N. E. Charron, G. de Fabritiis, F. Noé, and C. Clementi, Machine learning of coarse-grained molecular dynamics force fields, ACS Cent. Sci. **5**, 755 (2019).

[10] N. E. Charron, K. Bonneau, A. S. Pasos-Trejo, A. Guljas, Y. Chen, F. Musil, J. Venturin, D. Gusew, I. Zaporozhets, A. Krämer, *et al.*, Navigating protein landscapes with a machine-learned transferable coarse-grained model, Nat. Chem. **17**, 1284 (2025).

[11] H. Mori, Transport, Collective motion, and Brownian motion, Prog. Theor. Phys. **33**, 423 (1965).

[12] R. Zwanzig, Nonlinear generalized Langevin equations, J. Stat. Phys. **9**, 215 (1973).

[13] I. Gyöngy, Mimicking the one-dimensional marginal distributions of processes having an ito differential, Probab. Th. Rel. Fields **71**, 501 (1986).

[14] F. Legoll and T. Lelièvre, Effective dynamics using conditional expectations, Nonlinearity **23**, 2131 (2010).

[15] W. Zhang, C. Hartmann, and C. Schütte, Effective dynamics along given reaction coordinates, and reaction rate theory, Farad. Discuss. **195**, 365 (2016).

[16] C. Ayaz, L. Scalfi, B. A. Dalton, and R. R. Netz, Generalized Langevin equation with a nonlinear potential of mean force and nonlinear memory friction from a hybrid projection scheme, Phys. Rev. E **105**, 054138 (2022).

[17] F. Legoll, T. Lelièvre, and U. Sharma, Effective dynamics for non-reversible stochastic differential equations: A quantitative study, Nonlinearity **32**, 4779 (2019).

[18] T. Lelievre and W. Zhang, Pathwise estimates for effective dynamics: The case of nonlinear vectorial reaction coordinates, Multiscale Model. Simul. **17**, 1019 (2019).

[19] F. Legoll, T. Lelievre, and S. Olla, Pathwise estimates for an effective dynamics, Stochast. Process. Appl. **127**, 2841 (2017).

[20] C. Hartmann, L. Neureither, and U. Sharma, Coarse graining of nonreversible stochastic differential equations: Quantitative results and connections to averaging, SIAM J. Math. Anal. **52**, 2689 (2020).

[21] M. H. Duong, A. Lamacz, M. A. Peletier, A. Schlichting, and U. Sharma, Quantification of coarse-graining error in Langevin and overdamped Langevin dynamics, Nonlinearity **31**, 4517 (2018).

[22] W. Zhang and C. Schütte, Reliable approximation of long relaxation timescales in molecular dynamics, Entropy **19**, 367 (2017).

[23] F. Nüske, P. Koltai, L. Boninsegna, and C. Clementi, Spectral properties of effective dynamics from conditional expectations, Entropy **23**, 134 (2021).

[24] D. Crommelin and E. Vanden-Eijnden, Diffusion estimation from multiscale data by operator eigenpairs, Multiscale Model. Simul. **9**, 1588 (2011).

[25] F. Nüske, L. Boninsegna, and C. Clementi, Coarse-graining molecular systems by spectral matching, J. Chem. Phys. **151**, 044116 (2019).

[26] S. Klus, F. Nüske, S. Peitz, J.-H. Niemann, C. Clementi, and C. Schütte, Data-driven approximation of the Koopman generator: Model reduction, system identification, and control, Physica D **406**, 132416 (2020).

[27] B. O. Koopman, Hamiltonian systems and transformation in Hilbert space, Proc. Natl. Acad. Sci. USA **17**, 315 (1931).

[28] I. Mezić, Spectral properties of dynamical systems, model reduction and decompositions, Nonlin. Dyn. **41**, 309 (2005).

[29] H. Sidky, W. Chen, and A. L. Ferguson, Molecular latent space simulators, Chem. Sci. **11**, 9459 (2020).

[30] J. V. Diez, M. Schreiner, and S. Olsson, Transferable generative models bridge femtosecond to nanosecond time-step molecular dynamics, Sci. Adv. **12**, eaed2333 (2026).

[31] J. V. Diez, M. Schreiner, O. Engkvist, and S. Olsson, Boltzmann priors for implicit transfer operators, in *The Thirteenth International Conference on Learning Representations*, edited by Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (ICLR, 2025), pp. 38454–38482.

[32] M. Schreiner, O. Winther, and S. Olsson, Implicit transfer operator learning: Multiple time-resolution models for molecular dynamics, in *Advances in Neural Information Processing Systems*, edited by A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Curran Associates, Inc., 2023), Vol. 36, pp. 36449–36462.

[33] H. Wu and F. Noé, Reaction coordinate flows for model reduction of molecular kinetics, J. Chem. Phys. **160**, 044109 (2024).

[34] V. Nateghi and F. Nüske, Kinetically consistent coarse graining using kernel-based extended dynamic mode decomposition, J. Chem. Theory Comput. **21**, 7236 (2025).

[35] S. Moqvist, W. Chen, M. Schreiner, F. Nüske, and S. Olsson, Thermodynamic interpolation: A generative approach to molecular thermodynamics and kinetics, J. Chem. Theory Comput. **21**, 2535 (2025).

[36] F. Noé, S. Olsson, J. Köhler, and H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, Science **365**, eaaw1147 (2019).

[37] M. Dibak, L. Klein, A. Krämer, and F. Noé, Temperature steerable flows and Boltzmann generators, Phys. Rev. Res. **4**, L042005, (2022).

[38] J. F. Gloy and S. Olsson, Hollowflow: Efficient sample likelihood evaluation using hollow message passing, in *Advances in Neural Information Processing Systems* (NeurIPS, 2025), Vol. 38, p. 35416.

[39] M. Albergo, N. M. Boffi, and E. Vanden-Eijnden, Stochastic interpolants: A unifying framework for flows and diffusions, J. Machine Learning Res. **26**, 1 (2025).

[40] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard, A direct approach to conformational dynamics based on hybrid Monte Carlo, J. Comput. Phys. **151**, 146 (1999).

[41] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, Markov models of molecular kinetics: Generation and validation, J. Chem. Phys. **134**, 174105 (2011).

[42] T. Lelievre, M. Rousset, and G. Stoltz, Langevin dynamics with constraints and computation of free energy differences, Math. Comput. **81**, 2071 (2012).

[43] R. Zwanzig, Memory effects in irreversible thermodynamics, Phys. Rev. **124**, 983 (1961).

[44] A. Rahimi and B. Recht, Random features for large-scale kernel machines, in *Proceedings of the 21st International Conference on Neural Information Processing Systems* (Curran Associates Inc., Red Hook, NY, 2007), pp. 1177–1184.

[45] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, A data-driven approximation of the Koopman operator: Extending dynamic mode decomposition, J. Nonlin. Sci. **25**, 1307 (2015).

[46] F. Nüske and S. Klus, Efficient approximation of molecular kinetics using random Fourier features, J. Chem. Phys. **159**, 074105 (2023).

[47] P. Deuflhard and M. Weber, Robust Perron cluster analysis in conformation dynamics, Linear Algebr. Appl. **398**, 161 (2005).

[48] V. Nateghi, L. Neureither, S. Moqvist, C. Hartmann, S. Olsson, and F. Nüske, Consistent projection of Langevin dynamics: Preserving thermodynamics and kinetics in coarse-grained models, [Data set], Zenodo (2026), doi:10.5281/zenodo.19184692.

[49] H. Wu and F. Noé, Variational approach for learning Markov processes from time series data, J. Nonlin. Sci. **30**, 33 (2020).