UvA-DARE (Digital Academic Repository)

Reweighting non-equilibrium steady-state dynamics along collective variables

Bause, M.; Bereau, T.

Link to publication

Reweighting non-equilibrium steady-state dynamics along collective variables

Marius Bause[1,a) ID and Tristan Bereau[1,2,b) ID

AFFILIATIONS

[1] Max Planck Institute for Polymer Research, 55128 Mainz, Germany

[2] Van 't Hoff Institute for Molecular Sciences and Informatics Institute, University of Amsterdam, Amsterdam 1098 XH, The Netherlands

[a)] Electronic mail: mariusbause@gmx.de
[b)] Author to whom correspondence should be addressed: t.bereau@uva.nl

ABSTRACT

Computer simulations generate microscopic trajectories of complex systems at a single thermodynamic state point. We recently introduced a Maximum Caliber (MaxCal) approach for dynamical reweighting. Our approach mapped these trajectories to a Markovian description on the configurational coordinates and reweighted path probabilities as a function of external forces. Trajectory probabilities can be dynamically reweighted both from and to equilibrium or non-equilibrium steady states. As the system's dimensionality increases, an exhaustive description of the microtrajectories becomes prohibitive—even with a Markovian assumption. Instead, we reduce the dimensionality of the configurational space to collective variables (CVs). Going from configurational to CV space, we define local entropy productions derived from configurationally averaged mean forces. The entropy production is shown to be a suitable constraint on MaxCal for non-equilibrium steady states expressed as a function of CVs. We test the reweighting procedure on two systems: a particle subject to a two-dimensional potential and a coarse-grained peptide. Our CV-based MaxCal approach expands dynamical reweighting to larger systems, for both static and dynamical properties, and across a large range of driving forces.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0042972

I. INTRODUCTION

Dynamical processes are used to describe complex behavior in a number of fields; examples are transition state dynamics of chemical reactions[1] or photosynthesis.[2] Many processes are influenced by external driving and operate away from equilibrium. Long-time driving often leads to systems eventually settling in a non-equilibrium steady state (NESS). Application of NESS includes the description of lasers,[3] photosynthesis,[4] gene regulatory circuits,[5] or constant pulling experiments.[6,7] Despite our current lack of a universal theory for statistical mechanics of equilibrium (or NESS),[8] computer simulations can provide microscopic insight into these complex processes. Unfortunately, limited computational power often prevents molecular simulations from reaching the experimentally relevant timescales, or alternatively, requires them to operate at artificially large driving forces.[9] A formalism to reweight non-equilibrated dynamics across these driving forces is needed.

Several reweighting schemes for dynamic and static information in equilibrium are known. The Ferrenberg–Swendsen reweighting[10] is frequently used on stationary probability distributions drawn from simulation in equilibrium. Potential and force-based reweighting schemes for equilibrium dynamics have been of recent interest and are based on Kramer's rule,[11,12] maximum likelihood methods,[13–15] the Girsanov–Radon derivative,[16] or Maximum Caliber (MaxCal) methods.[17] A method similar to the Rosenbluth algorithm performs reweighting in NESS for minimal processes such as birth–death processes.[18] Another method based on iterative trajectory weighting is expected to scale to NESS systems;[19] however, these methods have not yet been shown to reweight complex systems across non-equilibrium conditions. We recently introduced a method based on MaxCal ansatz, which is capable of reweighting the dynamics of minimal systems in NESSs.[20] This paper extends this method to reweight dynamical information of complex systems described by *collective* coordinates.

Designed as an extension of the Ferrenberg–Swendsen method, our reweighting scheme is based on the Gibbs maximum entropy approach. While the maximum entropy claims that a physical system is in a state where it can be realized by the highest number of microstates (i.e., highest entropy), MaxCal aims at extending this idea to microtrajectories. The extension to microtrajectories is motivated by systems out of equilibrium being characterized by probability currents. The currents cannot be modeled by microstates alone and need microtrajectories for a complete description.

Jaynes introduced MaxCal as a theoretical framework for all dynamical processes.[21] The method was shown to recover physical relations of equilibrium;[22] model dynamical complex systems from limited information;[23,24] correct dynamic information by inferring physical information;[25,26] and more applications on statistical systems in physics, chemistry, and biology.[27] We use MaxCal as the basis for our NESS reweighting method.

The MaxCal formalism requires us to choose a set of implied constraints based on the physical manipulations made on a system in NESS. A driving force exerted on the system will affect the heat exchange of each pathway. The microscopic characterization of heat exchange is described by the local entropy production.[20] The NESS system is also constrained by global balance to preserve probability fluxes. The dynamics can be separated into two parts: dissipative and non-dissipative contributions.[28] The dissipative contribution is determined by the target NESS, accessed via the local entropy production. The non-dissipative contribution, on the other hand, is drawn from the reference data itself. We highlighted an invariant, which contains the *time-symmetric* contributions—they do not change under driving. The invariant acts similar to the density of states in equilibrium. MaxCal, combined with the appropriate constraints, opens the possibility to reweight dynamical information across external forces as a function of the system's configurational space.

Because the reweighting is performed at a microscopic level, it requires the consideration of large numbers of microstates. The sheer number of microtrajectories becomes computationally intractable for all but the smallest of systems, and they are here instead coarse-grained by Markov state models (MSMs). MSMs describe the system's dynamics by coarse-grained time and space. They discretize the configurational space in microstates and model the Markovian probability of transitions between these states. We performed space discretization based on configurational coordinates.[20] Computational aspects typically limit the size of the transition probability matrix to $\sim 10^3$ microstates.[29] The representation of molecular systems with a large number of particles rapidly becomes problematic. Instead, the configurational space is often projected down to a set of low-dimensional collective variables (CVs).[30] The application of CVs to MaxCal-based dynamical reweighting is the topic of this study.

This paper is structured as follows: First, we will introduce the reweighting method and show that it is applicable to CVs without loss of generality in Sec. II. Second, the models investigated and first-passage-time distributions used to analyze the dynamics are introduced in Sec. III. In Sec. IV, we will apply the reweighting to a toy model in full coordinates and along collective variables to show how the choice of variables impacts the accuracy of the methodology. Reweighting is then applied to a molecular system: a

tetra-alanine peptide. We apply the reweighting along two collective variables, testing both conservative and non-conservative forces.

II. THEORY

Steady states are a special case of non-equilibrium, where heat is supplied to and withdrawn from the system from an unlimited reservoir at the same rate. The amount of heat flowing from and to the system is controlled by the entropy. The system will eventually settle in a state with constant, positive total entropy production $dS_{tot} > 0$, without the system undergoing changes—in a steady state $dS_{sys} = 0$. The system is characterized by steady currents from a macroscopic point of view. These dynamical currents are described by ensembles of microtrajectories, each with a time-independent weight. Maintaining the currents results in positive entropy production in the reservoir. The system remains off equilibrium but loses time-dependence because the macroscopic system does not change in time.

The resulting time-independent set of microtrajectories is mapped onto a discrete Markov process. The configurational space is discretized into the so-called microstates (i.e., collection of microscopic states), and time is discretized in steps of constant duration τ (i.e., the lag-time).[31] All the observed transitions from microtrajectories are collected to infer a transition probability matrix $p_{ij}(\tau)$, where i and j label microstates. This coarsening of microtrajectories leaves us with the easier task of sampling transition probabilities and subsequently constructing microtrajectories out of the combination of individual micro-transitions. This mapping has been proven to reach timescales that are out of range of brute-force computer simulations.[32]

A. Maximum Caliber (MaxCal)

The maximum entropy formalism by Gibbs states that an equilibrium system is in a state where it can realize the highest number of microscopic configurations, subject to external constraints such as the mean energy.[33] Analogously, MaxCal proposes a framework to study dynamical systems by replacing microstates with microtrajectories.[34] In doing so, MaxCal moves away from Gibbs' physical argument to an information theoretic point of view: Based on partial information, what is the most likely state the system is in? Jaynes answers this question by assuming the most uncertain (or highest entropy) probability distribution as noncommittal as possible regarding unknown information. This point of view boils down to a general inference method only subject to adequate physical constraints. We take advantage of this formalism by generalizing equilibrium reweighting, which focuses on the static distributions of microstates, to *dynamical* reweighting of NESS.

An adequate choice of physical constraints forms an essential element of MaxCal.[35] For dynamical reweighting to another NESS, we recently proposed the combined use of the local entropy production and global balance.[20] These constraints focus on the interactions of the system with its environment:

1. Heat exchange is described at the microscopic level by entropy production, itself constrained by microscopic reversibility.[36,37] The system's spatial heterogeneity, as well as the need to describe dissipative dynamics, requires a *local* constraint.[35]

The local entropy production, ΔS_{ij}, between microstates i and j is constrained by the following relation:[20]

$$\langle \Delta S_{ij} \rangle = \ln \frac{p_{ij}}{p_{ji}}, \qquad (1)$$

where p_{ij} denotes the probability to jump between microstates i and j.[20] By making use of a microscopic expression for ΔS, we integrate the conservative *and* non-conservative force contributions along a trajectory [see Eq. (B1)].[38]

2. To connect all local changes, we add global balance, $\pi_i = \sum_k \pi_k p_{ki}$, for each microstate i. Global balance ensures the conservation of probability flux.[39] It connects a single state on the left-hand side of the equation to all other states and couples both stationary and dynamical properties.

Including adequate normalization constraints, the caliber functional becomes

$$
\begin{aligned}
\mathcal{C} = &- \sum_{i,j} \pi_i p_{ij} \ln \frac{p_{ij}}{q_{ij}} + \sum_i \mu_i \pi_i \left(\sum_j p_{ij} - 1 \right) \\
&+ \zeta \left(\sum_i \pi_i - 1 \right) + \sum_j \nu_j \left(\sum_i \pi_i p_{ij} - \pi_j \right) \\
&+ \sum_{ij} \pi_i \alpha_{ij} \left(\ln \left(\frac{p_{ij}}{p_{ji}} \right) - \Delta S_{ij} \right).
\end{aligned} \qquad (2)
$$

Here, the first term represents the relative-entropy term on pathways, specifically between the target (MSM-based) transition probability p_{ij} with its reference counterpart, q_{ij}. The other terms consist of constraints, expressed as Lagrange multipliers: first, normalization constraints on the transition probability, p_{ij}, and the steady-state distribution, π_i, with associated parameters μ_i and ζ, respectively. The last two terms constrain the global-balance condition and local entropy production with Lagrangian multipliers ν_i and α_{ij}, respectively. The parameters α_{ij} and μ_i were both rescaled by π_i. Maximization is described in Appendix A and yields

$$
\begin{aligned}
p_{ij} &= q_{ij} \exp \left(\zeta + \frac{1}{2} \left(c_i + c_j + \Delta S_{ij} - \Delta S_{ij}^q \right) \right) \\
&= \sqrt{q_{ij} q_{ji}} \exp \left(\zeta + \frac{1}{2} \left(c_i + c_j + \Delta S_{ij} \right) \right),
\end{aligned} \qquad (3)
$$

where S_{ij}^q is the local entropy production of the reference system and c_i are constants to be determined. This shows that we have two options for the input parameters: The reweighting depends on either the total entropy production ΔS_{ij} of the target system or the difference in local entropy production $\Delta S_{ij} - \Delta S_{ij}^q$ between target and reference systems. The unknowns c are calculated by enforcing the relation $\sum_j p_{ij} = 1$,

$$ 1 = \sum_j \sqrt{q_{ij} q_{ji}} \exp \left(\zeta + \frac{1}{2} \left(c_i + c_j + \Delta S_{ij} \right) \right). \qquad (4) $$

This is a convex set of equations that can be solved by numerical iteration, for instance, by least-squares.[40]

B. Collective variables

To reduce the number of microstates, describing complex systems by collective variables (CVs) is essential to make the system computationally accessible. Examples of CVs include the description of a magnet by its magnetization while ignoring the influence of local dipole fluctuations[41] or the crystallization of particles described by the closest radial environment of each crystallizing particle.[42] Many fast and local processes are averaged out when settling on a set of collective variables. The mesoscopic descriptors or collective variables are inherently system-specific and limit the view on the system: The crystallization described by the local environment of the particles holds a detailed view on the crystalline phase but only holds limited information on the liquid phase.[42] Furthermore, a poor choice of CVs can hide important processes and free-energy barriers or cause an inaccurate estimation of implied timescales.[43–45] The adequate choice of collective variables is a widely discussed research field on its own and is applied to describe complex systems in chemistry, biology, and physics.[46]

To extend our reweighting procedure from configurational coordinates, x, to CVs, z, we need to adapt the expression for the change in local-entropy production [Eq. (B1)]. CVs and configurational coordinates are related by a mapping operator, $z = M(x)$. The potential energy is replaced by the potential of mean force,[47]

$$ G(z) = -k_B T \ln \int dx \, \delta(M(x) - z) \pi(x). \qquad (5) $$

The change in entropy production due to a trajectory $z(t)$ with starting and end points z_0 and z_T, respectively, yields (see Appendix B)

$$
\begin{aligned}
&\Delta S(z_0, z_T) - \Delta S^q(z_0, z_T) \\
&= \frac{1}{k_B T} \Big[G(z_T) - G^q(z_T) - \big(G(z_0) - G^q(z_0) \big) \\
&\quad + (z_T - z_0) \cdot (f - f^q) \Big],
\end{aligned} \qquad (6)
$$

where $z(t)$ is the D-dimensional CV vector, f is the non-conservative force, and superscript q indicates the reference system. Conceptually, adapting ΔS from configurational to CV space amounts to replacing the potential energy by the potential of mean force. The expression holds for an arbitrary system with or without boundary conditions but only for driving forces along the CVs. While Eq. (6) assumes constant forces, it can readily be generalized, i.e., $f(z)$, analogous to the full-configurational case.[20]

III. METHODS

The reweighting procedure for CVs is tested on two systems. The first model is a non-interacting particle subject to a two-dimensional potential. The potential consists of three Gaussian potential wells of varying depths. All boundaries are periodic, and the external force is applied along the x-direction. Results for this system are presented in reduced units, where the box size is set to $3\mathcal{L} \times 1\mathcal{L}$, the mass of the particle is set to $\mathcal{M}$, and energy is measured in ϵ. The temperature is $T = 1 \, \epsilon/k_B$, and the unit of time is $\mathcal{T} = \mathcal{L}\sqrt{\mathcal{M}/\epsilon}$. The integration time step is set to $\delta t = 10^{-5} \, \mathcal{T}$, the

non-conservative force is varied between 0 and $9\epsilon/\mathcal{L}$, the microstates consists of 30×10 squares of equal size, and the lag-time is chosen at $0.02\,\mathcal{T}$. The potential minima are Gaussian functions with depths 3ϵ, 5ϵ, and 7ϵ and are located at $x = \{0.5, 1.5, 2.5\}$ and $y = 0.5$. The standard deviation of the Gaussian is $0.2\,\mathcal{L}$ in both directions. By integrating out the y-dimension orthogonal to the driving force, $\boldsymbol{F}_{2D}(x, y)$, we imitate a reduction in variables, providing a testing ground for the reweighting along CVs. The mean force is calculated via the stationary distribution,

$$\langle \boldsymbol{F}(x) \rangle = \int \mathrm{d}y\, \boldsymbol{F}_{2D}(x, y)\pi(x, y). \tag{7}$$

Both full and reduced descriptions will be analyzed along x. All dynamics are extracted from the same reference simulations. An MSM is constructed with the same lag-time $\tau = 0.02\,\mathcal{T}$ and the same 30 equisized microstates in the x-direction. The lag-time for MSM is validated by the Chapman–Kolmogorov test,[48] for both the full 2D and reduced 1D systems.[49]

The second system represents a tetra-alanine peptide consisting of four amino acids and 52 atoms. Each amino acid is coarse-grained to one bead centered at the backbone of the peptide. The coarse-grained force field for the molecule solvated in water consists of three pair potentials along the backbone, two bending-angle interactions, a dihedral angle φ, and an effective pairwise interaction between the first and last beads, R_{14}.[49] Simulations were run with ESPResSo++.[50]

The MSM is constructed using two CVs: the end-to-end distance, R_{14}, and the dihedral angle, φ (see Fig. 1).[51,52] The unperturbed equilibrium system is called the reference system. Driven systems consist of constant forces along either CV in either direction. We define 15 microstates over the range $[-\pi, +\pi]$ in the φ-direction and 15 microstates in the range [0.45, 1.15] nm in the R_{14}-direction. Two additional sets of microstates were added to collect end-to-end distances outside this range. Energies are given in $\epsilon = \frac{\mathrm{kJ}}{\mathrm{mol}}$, and the system is simulated at temperature $T = 2.479\,\frac{\epsilon}{k_B}$. A lag-time for the MSM is chosen using lag-time analysis and the Chapman–Kolmogorov test.[48,49] Metastable states for tetra-alanine are defined by PCCA+.[53] The metastable state analysis relies on equilibrium dynamics, satisfying detailed balance. Thus, the analysis is performed for the reference system, and the same metastable states are chosen for the driven systems.

The dynamics are analyzed by using first-passage-time distributions (FPTDs) between metastable states. It is defined by the
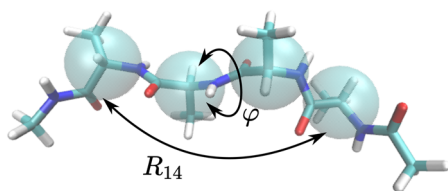

FIG. 1. Atomistic and coarse-grained representation of tetra-alanine. Atoms are shown in licorice, where turquoise, white, blue, and red represent C, H, N, and O, respectively. The transparent beads show the coarse-grained representation of the system. The end-to-end distance R_{14} and the dihedral angle φ are defined based on the coarse-grained representation.

distribution of time that a process starting from metastable state A needs to reach metastable state B. FPTDs are widely used to characterize processes in biology, chemistry, and physics and are often associated with a free-energy barrier a system has to overcome. The FPTD contains detailed transition information by collecting numerous realizations of a process. Often, few observed realizations limit the analysis to the mean of the distribution.[54] Given an MSM with identified metastable states, the FPTD between all metastable states can be calculated directly.[55] The collection of initial states is denoted by I, the collection of final states is denoted by F, and the FPTD is denoted by $p_{FPT}(I \rightarrow F, t)$. Knowing the FPTD, all moments of the distribution can be calculated by

$$M_{I \rightarrow F}^{(n)} = \sum_t p_{FPT}(I \rightarrow F, t)t^n. \tag{8}$$

In particular, we will make use of the quantities

$$\mu_{I \rightarrow F} = M_{I \rightarrow F}^{(1)},$$

$$\sigma_{I \rightarrow F} = \sqrt{M_{I \rightarrow F}^{(2)} - \mu_{I \rightarrow F}^2},$$

$$\kappa_{I \rightarrow F} = \frac{M_{I \rightarrow F}^{(3)} - 3\mu_{I \rightarrow F}\sigma_{I \rightarrow F}^2 - \mu_{I \rightarrow F}^3}{\sigma_{I \rightarrow F}^3}, \tag{9}$$

where $\mu_{I \rightarrow F}$ is the mean, $\sigma_{I \rightarrow F}$ is the standard deviation, and $\kappa_{I \rightarrow F}$ is the standardized skewness, defined by the expectation value of $\left(\frac{t - \mu}{\sigma}\right)^3$. These moments are used to compare FPTDs throughout this paper to capture the main features and draw physical information from the distribution.

IV. RESULTS

A. Particle in a two-dimensional potential

We first consider a toy model: a particle in a multi-well. The system is originally in two dimensions, but we also consider a reduced one-dimensional description. We perform dynamical reweighting for both descriptions from and to equilibrium and a driven NESS. Figures 2(a) and 2(b) show the potential of the full and reduced single-particle system. Dynamical reweighting leads to an accurate reproduction of the stationary distribution, as seen for two different driving forces [Fig. 2(c)]. Reweighting also leads to an accurate reproduction of the FPTD [Fig. 2(d)], as shown for the process C → B at both equilibrium and under driving and for both the full and reduced descriptions. While longer timescales are reproduced accurately, the reweighting for short processes of $1–5\tau$ shows small deviations. These are caused by the spatial discretization, especially in highly populated areas. Overall, the dynamical-reweighting scheme performs in both full-configurational and CV spaces as well. We note that the present methodology requires external forces to be aligned with the CVs.

In the following, we reduce all FPTDs to the first three moments and the stationary distribution of metastable states for a comparison between simulation and reweighting (Fig. 3). The largest deviation can be seen for the process A → B, where the reweighting error in the
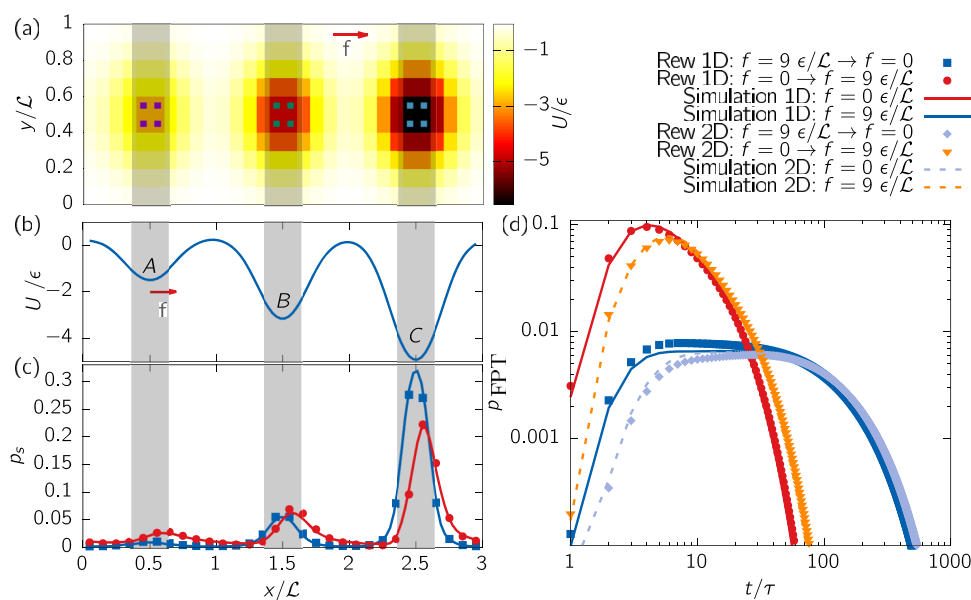
FIG. 2. (a) The 2D potential with the three metastable states indicated by squares. Integrating along the y-dimension gives (b) the mean potential of the equilibrium system. The gray area represents the new metastable states A, B, and C. The area of the metastable state is effectively increased. (c) The stationary distribution of the reduced system. (d) FPTD of the process C → B. The lines in (c) and (d) represent the results for a single particle in reduced space without (blue) and with (red) external force. The dots represent the results from reweighting the systems into each other. The orange and light-blue dashed lines show the same process for the underlying 2D process with dots representing the reweighted FPTD.

1D and 2D systems is comparable, as well as the occupation probability of state C. The discrepancy in the stationary distribution at heavy driving is also shown in more detail in Fig. 2(c). A metastable state in the reduced system covers two microstates of the MSM and is thus susceptible to discretization errors. Despite minor deviations, dynamic and static data are reweighted virtually perfectly into each other. We conclude that use of collective variables of the system did not affect the accuracy of the reweighting process. Hence, it can be applied to the same extent as the reweighting in configurational space.

We now more closely compare the dynamics for the two system descriptions (Fig. 3). While the processes remain qualitatively similar irrespective of representation, the 2D processes are consistently *slower* than those in the reduced representation. These accelerated dynamics are common in coarse-grained modeling.[56–59] The reduced roughness in the free-energy surface results in a decrease in the effective friction. For our simplified model, this effect reduces the effective potential barriers, which leads to the acceleration of the coarse-grained process. Similar effects can be found for the standard deviation (STD) and skewness, though to smaller extents. More details on the skewness can be found in the supplementary material.

The occupation probability of the metastable states is significantly larger for the reduced system. The metastable states are effectively smaller for the 2D system because they do not span the whole y-direction. The reduction to the x axis enlarges the metastable states effectively, and the occupation probability increases. The trend of decreasing occupation in C and increasing occupation in A and B is the same for both systems.

B. Tetra-alanine peptide

To further challenge the reweighting procedure, we apply it to a coarse-grained tetra-alanine peptide. This system is of higher complexity than the previous model by showing rougher free-energy landscapes and many-body interactions. External global forces are applied along the CVs to alter the dynamics. Physically, these forces may represent an optical tweezer controlling atom distances. The external forces are chosen to test the effectiveness of the reweighting procedure for conservative and non-conservative forces.

The free-energy surface of the coarse-grained reference system is shown in Fig. 4(a), where we project along two CVs: the end-to-end distance R_{14} and the dihedral angle φ. We use PCCA+ to identify metastable states—a method aimed at identifying coarse-grained states that preserve the slow timescales.[60] We further chose PCCA+ parameters, leading to metastable states that are both small and well separated. Three basins were identified, representing the helical states H, extended state E, and one intermediate state I. State H is associated with helical states located to the right of the middle free-energy barrier at $\varphi \approx 0.15\,\pi$. State I is an intermediate state at $\varphi \approx 0.4\,\pi$.

The driving along R_{14} can be casted to an additional attractive or repulsive interaction potential—leaving the system in equilibrium. On the other hand, we can also drive the peptide in a NESS along the periodic dihedral angle φ. The direction of driving will impact the dynamics because the free energy surface lacks the symmetry of the previous toy model. Thus, we can test the method for reweighting between equilibrium states, NESS, or from equilibrium to NESS and vice versa.

1. Equilibrium reweighting

Figure 4(b) shows the implied timescale analysis for the original force field and an applied driving along R_{14}. We choose a lag-time of $200\,\mathcal{T}$ to capture the two slowest processes of both systems, where $\mathcal{T} = 1$ fs. The second process is captured by the MSM and is virtually unaffected by the additional forces applied. In the following, we assume this process to remain unaffected by larger forces.
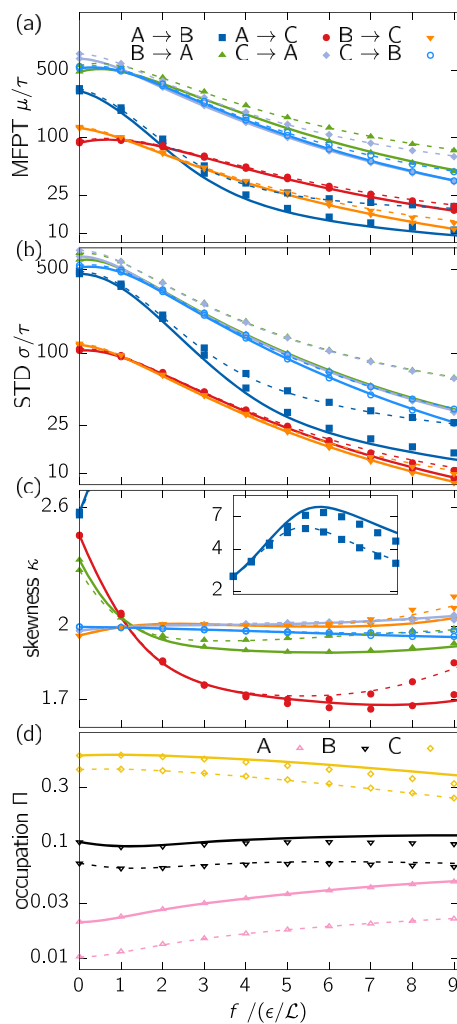
FIG. 3. (a)–(c) The first three moments of the FPTD for all six processes between metastable states under varying external force f. (d) The occupation probability of each metastable state. The dots represent the value measured from simulation. The line represents the reduced 1D equilibrium system continuously reweighted. The dashed lines represent the continuously reweighting processes of the underlying equilibrium processes in 2D space. The error bars are smaller than the points and lines.
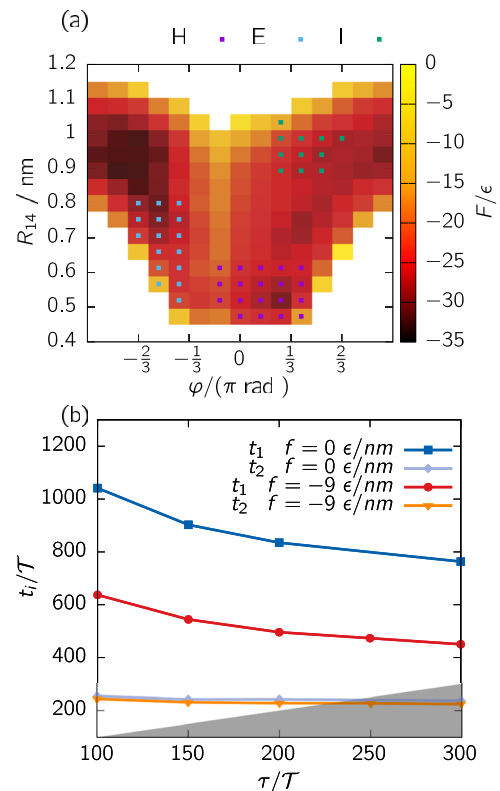

FIG. 4. (a) Free energy surface of tetra-alanine of the reference system. The metastable states are indicated by helical (H), extended (E), and intermediate (I). (b) Implied timescale analysis of the system defined by the reference force field ($f_R = 0$) and driven along the end-to-end distance with $f_R = -9 \frac{\epsilon}{nm}$. The shaded area marks the non-physical area where $t_i < \tau$.

Figures 5(a)–5(c) show the first three moments of the FPTD between the metastable states when driving along R_{14}. For the reference system at $f = 0$, we note the two fast processes I → E and H → E. The next two slower processes are H → I and E → I, and finally, the two slowest processes are both going to the helical state, E → H and I → H. Under driving, transitions to I slow down under an attractive end-to-end potential (i.e., negative forces) and speed up for a repulsive end-to-end potential (i.e., positive forces). The opposite happens for the processes going to H: An attractive end-to-end potential increases the speed of these processes. Transitions to the extended state E are comparatively unaffected by the driving. We

note that the STD behaves roughly proportional to the MFPT, while the skewness varies extremely weakly.

Looking at Fig. 5(d), increasingly repulsive R_{14} interactions lead to a stabilization of state I. On the other hand, this separation of the residues destabilizes both H and E, where the former decays more strongly.

The impact of the driving force on the MSM's implied timescales is shown in Fig. 5(e). The nature of the driving force retains the system in equilibrium so that path-dependent effects are not expected. In agreement with Fig. 4(b), the first timescale depends strongly on driving, while the second one is virtually unaffected.

Results on the three moments of the FPTD indicate that the transitions are recovered accurately. Minor deviations at large forces are rationalized by a significant change in the relevant populations: regions at large or small end-to-end distance become highly populated but may be insufficiently sampled in the reference system. These errors are mostly apparent for the higher-order moments. Overall, though, we report extremely encouraging results in terms of dynamical reweighting for a complex molecular system driven by a constant conservative force.
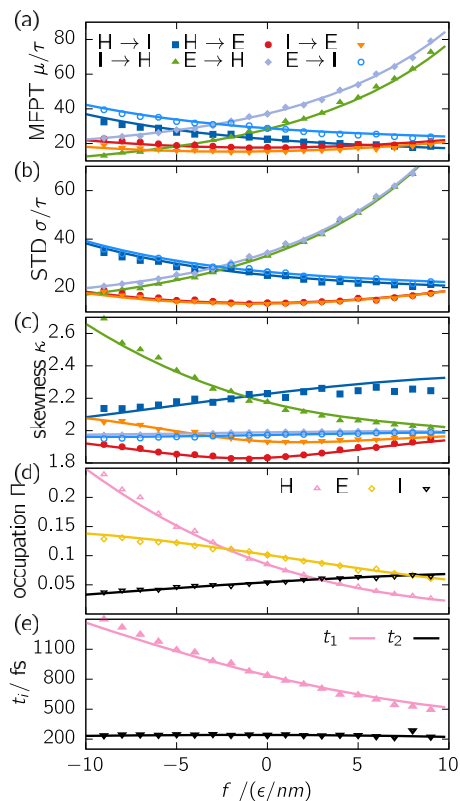
FIG. 5. (a)–(c) The first three moments of the FPTD for all six processes between metastable states under varying external force f along R_{14}. (d) The occupation probability of each metastable state. (e) The timescale of the two slowest processes covered by the MSM. The dots represent the value measured from simulation. The line represents the reference system continuously reweighted.
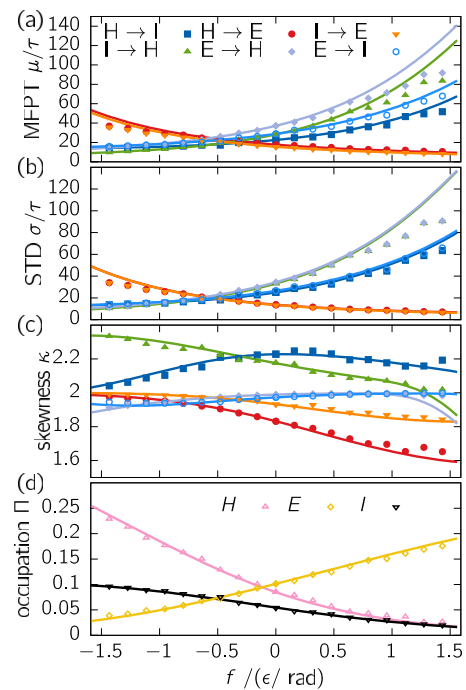

FIG. 6. (a)–(c) The first three moments of the FPTD for all six processes between metastable states in Fig. 4(a) under varying external force f along φ. (d) The occupation probability of each metastable state. The dots represent the value measured from simulation. The line represents the reference system continuously reweighted.

2. NESS reweighting

Figure 6 shows NESS driving along the dihedral φ in either direction. The dynamics of the system are largely dominated by its large free-energy barrier at $\varphi \approx -\frac{\pi}{6}$. Driving in the positive direction speeds up the processes H → E and I → E, while I → H slows down as it runs opposite to the driving force. On the other hand, H → I slows down, even though it runs along the external force. Most trajectories starting from H bypass I under heavy driving, leading, instead, directly to state E. This can be seen by the narrow, diagonal stripe below state I, which becomes more tightly populated. The trajectories find a direct path to the global basin (E) without hitting the intermediate state I, as can be seen in the transition density of H → E.

Overall agreement between direct simulations and reweighting is observed for the FPTD, especially up to $|f| < 1\frac{\epsilon}{\text{rad}}$. Similar to equilibrium reweighting, we find that the STD follows the behavior of the MFPT, and the skewness varies weakly. We observe some discrepancies at larger driving, notably for I → H and E → H. While they all increase, the simulation curves seem to reach a maximum.

Examination of the simulations shows that the process becomes faster by crossing over the free-energy barrier at $\varphi \approx -\frac{\pi}{6}$. Driving in the negative direction inverts the effect on the dynamics: Processes aligned with the force speed up, whereas opposing processes slow down. H → I does not follow this trend and instead accelerates. The trajectories that bypass the intermediate state under positive driving are now pushed into occupying state I. This is indicated by the increasing population of I under negative driving and depopulation under positive driving. The helical state population shows similar, but even stronger, behavior. The extended state, on the other hand, displays the opposite behavior. Here, again, the occupation probabilities are recovered accurately by the reweighting, even with small deviations in the dynamics at large driving.

3. Path dependence of entropy production

The observed deviations between simulation and reweighting at strong driving along the dihedral angle stem from the local entropy production. This key quantity is determined by both the external force and the set of paths connecting every pair of microstates. Finding the *shortest* connection between two microstates was straightforward for the toy-model system because they were separated by three or more barriers: Paths transitioning over a single barrier have much higher probabilities so that the other set of paths can be neglected.[20]

In tetra-alanine, driving along the R_{14}-direction did not lead to this issue because there are no periodic boundary conditions and the forces can be mapped to a potential. Transitions become path independent, and errors based on path dependence, thus, do not occur. Driving along the φ-direction, on the other hand, results in a NESS with periodic boundaries. It shows only one major barrier along the dihedral angle that dominates the dynamics. Choosing the appropriate path direction to feed into the local entropy production is more challenging here. For every jump in the Markov model, one has to determine if the underlying trajectory is aligned with, or directed against, the external force.

To shed light on path directions, we analyze the matrix of transition probabilities, as shown in Fig. 7(a) for the reference system. We fix a starting point, denoted by a green dot, and analyze the expected direction, given any final microstate. We expect all states to the right of the starting point (blue shaded area) to arise from trajectories going right, i.e., in the positive φ-direction. On the other hand, all states to the left of the starting point (green shaded area) arise from trajectories going left, i.e., negative φ-direction—taking periodic boundaries into account. A dividing mark (red dashed line) separates the two regions. The lag-time of the MSM is chosen to be small enough to avoid transition close to the divider, i.e., no transitions lead to ambiguity as to their likely direction. This disconnect in the transition matrix between left- and right-trajectories is associated with a discontinuity in the local entropy production. Figure 7(b) shows the transition matrix with a driving force $f = 1.4 \frac{\epsilon}{rad}$, initiated from the same starting point as before. The non-zero driving leads to

a change in transition probabilities, but the spatially long transitions are still forbidden. The discontinuity in local entropy production can be set in the same position. This is important because the target transition matrix is not known before reweighting. Having a gap at a similar position in the reference and target driving forces is essential for the reweighting algorithm.

Next, we illustrate the impact of an incorrect assignment of path direction. We displace the starting point directly to the left of the large, central barrier along φ [Fig. 7(c)]. Upon driving [Fig. 7(d)], the divider line cuts the transition matrix through a connected region, close to the intermediate state. As such, a discontinuity in the local entropy production will be present among the paths connecting this intermediate region. Left- and right-trajectories are no more well separated, leading to ambiguities. These issues directly result in the discrepancies observed in Fig. 6. They only materialize at strong driving; otherwise, the local entropy productions of these conflicting paths are negligible.

This analysis highlights the dependence of NESS reweighting by the choice of MSM. Upon reweighting from reference to target driving forces, the set of paths connecting two microstates should consist of similar sets of trajectories. Unfortunately, one cannot easily predict whether conditions for reweighting are met. The small gaps between the groups of trajectories in Figs. 7(a) and 7(b) are a warning signal. Models with three or more barriers, as constructed in the toy model, are less susceptible to this issue. A particle crossing a barrier is expected to take the shorter path over a single barrier and is unlikely to hop over two barriers within one lag-time.
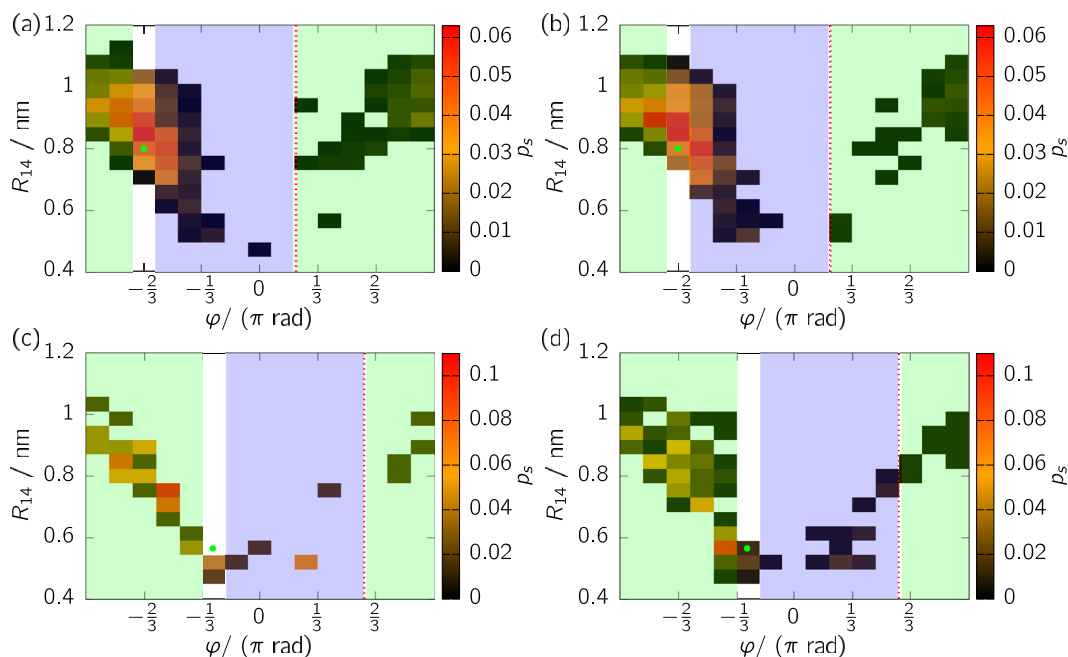
FIG. 7. Transition probabilities starting at the state marked by the green dot. [(a) and (c)] reference systems and [(b) and (d)] system driven along φ at $1.4 \frac{\epsilon}{rad}$. The red line represents the discontinuity in local entropy production, starting from the marked initial state. All states shaded green are connected to the starting state by trajectories going left; all states shaded blue are connected by a trajectories going right.

The tackling of larger systems should do away with these artifacts: both slower diffusion and more complex free-energy landscapes will remove directional ambiguity.

Clearly, these issues are brought about by the MSM construction of microtrajectories. Can we refine the MSM parameterization? Shorter lag-times would result in shorter trajectories and, thus, shorter jumps. Unfortunately, Fig. 4 shows that smaller lag-times show non-Markovian dynamics. Other options point at the role of CVs and microstate selection. We may select a different second collective variable (CV) when reweighting along the first. Such choices can have great impact and better represent the free-energy landscape. Alternatively, increasing the number of microstates does not allow us to decrease the lag-time. The microstates in the present model are discretized in equal size along the CVs. Advanced clustering techniques, such as k-means[61] or k-medoids,[62] help define more complex sets of microstates that could allow us to reduce the lag-time. Both the selection and clustering of the CVs influence how dynamics are described by the MSM. The connection of two microstates should be described by a unique bundle of paths. This means that the CVs and their separation in microstates should be chosen to reflect underlying kinetic distances of the system, as was formulated as a requirement for reweighting of dynamics in equilibrium by Wan *et al.*[17]

V. CONCLUSION

This study contributes to the sparse field of dynamical reweighting between non-equilibrium steady states. The presented method is based on Maximum Caliber (MaxCal) of Jaynes. It relies on an ensemble description of NESS by physical constraints—global balance and local entropy productions—and an efficient construction of microtrajectories by means of Markov state models (MSMs). Instead of being directly sampled, microtrajectories are *constructed* from the transition probability matrix, which robustly addresses issues of path sampling. On the other hand, an MSM description requires the use of appropriately chosen collective variables (CVs) that can describe the dynamics of the slow processes. Our initial description of NESS dynamical reweighting was based on the configurational variables of the system themselves. To scale up, this study presented an extension to CVs. The expression for the local entropy production was extended from individual forces to mean configurationally averaged forces. We tested the CV-based dynamical reweighting to both conservative and non-conservative forces, applied to both a toy model and a molecular system: a tetra-alanine peptide.

Strong agreement in both the static and dynamical properties is found overall. Discrepancies can be found at strong driving. We showed that they can be traced back to ambiguities in the direction of the constructed MSM-based paths. The periodic boundary conditions and relatively small landscape can lead to significant path contributions from both directions along the CV. Finding better CVs is an ever-present challenge[46] and is expected to systematically improve the MaxCal-based reweighting scheme presented here. Parallel avenues for improvements include the combination with established enhanced-sampling methods. For instance, while we herein reweight from a single state point, we expect the possibility to combine information from multiple state points, akin to

several optimal-estimator methods.[13,63–65] Furthermore, our current use of stochastic thermodynamics limits us to a single temperature reservoir. Extension of the method in this direction would open the door to more complex non-equilibrium systems, such as temperature gradients, or methodologies such as reverse non-equilibrium molecular dynamics.[66] We hope that the present methodology will help drive forward NESS reweighting for large molecular systems.

SUPPLEMENTARY MATERIAL

See the supplementary material for the mean-first-passage-time distributions and Chapman–Kolmogorov test of the 1D and 2D particle in a potential, as well as the stationary distribution of reactive trajectories for the peptide.

See the supplementary material at [URL] for additional information.

ACKNOWLEDGMENTS

APPENDIX A: CALIBER MAXIMIZATION

We consider the caliber

$$\mathcal{C} = -\sum_{i,j} \pi_i p_{ij} \ln \frac{p_{ij}}{q_{ij}} + \sum_i \mu_i \pi_i \left(\sum_j p_{ij} - 1 \right) + \zeta \left(\sum_i \pi_i - 1 \right)$$
$$+ \sum_j \nu_j \left(\sum_i \pi_i p_{ij} - \pi_j \right) + \sum_{ij} \pi_i \alpha_{ij} \left(\ln \left(\frac{p_{ij}}{p_{ji}} \right) - \Delta S_{ij} \right). \quad (A1)$$

The maximization with respect to the transition probabilities p_{ij} gives

$$0 = -\pi_i \ln \left(\frac{p_{ij}}{q_{ji}} \right) - \pi_i + \pi_i \mu_i + \pi_i \nu_j + \pi_i \frac{\alpha_{ij}}{p_{ij}} - \pi_j \frac{\alpha_{ji}}{p_{ij}}. \quad (A2)$$

Solving for p_{ij} with $\pi_i \neq 0$ yields

$$p_{ij} = q_{ij} \exp \left(-1 + \mu_i + \nu_j + \frac{\gamma_{ij}}{p_{ij}} \right), \quad (A3)$$

where $\gamma_{ij} = \alpha_{ij} - \frac{\pi_j}{\pi_i} \alpha_{ji}$ is used. Enforcing the local entropy productions explicitly by $\Delta S_{ij} = \ln \frac{p_{ij}}{p_{ji}}$ and after some algebra, one finds that

$$\frac{\gamma_{ij}}{p_{ij}} = w_{ij} \left(\Delta S_{ij} - \Delta S_{ij}^q - \mu_i + \nu_i + \mu_j - \nu_j \right), \quad (A4)$$

where $w_{ij} = 1 / \left(1 + \frac{\pi_i p_{ij}}{\pi_j p_{ji}} \right)$ and $\Delta S_{ij}^q = \ln \frac{q_{ij}}{q_{ji}}$ are used. This expression is set into Eq. (A3), and using $w_{ij} + w_{ji} = 1$, we find that

$$p_{ij} = q_{ij} \exp\left(-1 + w_{ji}\mu_i + w_{ij}\mu_j + w_{ji}\nu_j + w_{ij}\nu_i + w_{ij}(\Delta S_{ij} - \Delta S_{ij}^q)\right). \tag{A5}$$

The caliber maximization with respect to the stationary distribution gives

$$0 = -\sum_k p_{ik} \ln\left(\frac{p_{ik}}{q_{ik}}\right) + \mu_i \sum_k p_{ik} - \mu_i + \zeta - \nu_i$$
$$+ \sum_k \nu_k p_{ik} + \sum_k \alpha_{ik}\left(\ln\left(\frac{p_{ik}}{p_{ki}}\right) - \Delta S_{ik}\right). \tag{A6}$$

By combining with Eq. (A3) and making use of the probability conservation constraints, one finds a relation between the Lagrangian multipliers γ_{ij}, ν_i, and μ_i,

$$\mu_i + \nu_i = 1 + \zeta + \sum_k \gamma_{ik}. \tag{A7}$$

Enforcing the constraint $\sum_k p_{ik} = 1$ on Eq. (A5) results in a set of N equations, where N is the number of microstates. Combined with the set of N equations from Eq. (A7), there is a set of $2N$ coupled non-linear equations to be solved. To solve the problem, we assume that deviations from detailed balance are small: $\frac{\pi_i p_{ij}}{\pi_j p_{ji}} \approx 1$, resulting in $w_{ij} \approx \frac{1}{2}$. The approximation is applied to each Markovian jump individually such that the aggregate contributions to a microtrajectory may yield significant entropy productions. The approximation applied to Eq. (A5) yields

$$p_{ij} = q_{ij} \exp\left(\frac{1}{2}\left(-2 + \mu_i + \nu_j + \mu_j + \nu_i + \Delta S_{ij} - \Delta S_{ij}^q\right)\right). \tag{A8}$$

Using the result of Eq. (A7) and the definition $c_i = \sum_k \gamma_{ik}$, we obtain

$$p_{ij} = q_{ij} \exp\left(\zeta + \frac{1}{2}\left(c_i + c_j + \Delta S_{ij} - \Delta S_{ij}^q\right)\right). \tag{A9}$$

APPENDIX B: LOCAL ENTROPY PRODUCTION IN COLLECTIVE COORDINATES

To solve the reweighting equation, we need an expression for the relative local entropy production $\Delta S_{ij} - \Delta S_{ij}^q$ between target and reference states, the latter being indicated by superscript q. The indices i, j denote microstates that occur from discretizing the coordinates of the system of interest. Having access to the full set of coordinates allows us to analyze a trajectory $\boldsymbol{x}(t)$ and calculate the entropy production using

$$\Delta S[\boldsymbol{x}(t)] = \int dt \, \frac{\boldsymbol{F} \cdot \dot{\boldsymbol{x}}}{k_B T}, \tag{B1}$$

where $\dot{\boldsymbol{x}}$ is the velocity, $\boldsymbol{F}$ is the force, and T is the temperature.[38] Making use of numerically discretized trajectories, $\boldsymbol{x}(t) \approx \{\boldsymbol{x}_k\}$, $\Delta S(\{x_k\})$ is approximated between initial and target points, x_0 and x_T, respectively,

$$\Delta S[\{x_k\}] \approx \sum_d \sum_{t=1}^{T} \frac{\left(x_t^{(d)} - x_{t-1}^{(d)}\right)\left(F^{(d)}(\boldsymbol{x}_t) + F^{(d)}(\boldsymbol{x}_{t-1})\right)}{2k_B T}$$
$$\approx \frac{1}{2k_B T} \sum_d \sum_{t=0}^{T} x_t^{(d)}\left(F^{(d)}(\boldsymbol{x}_{t-1}) - F^{(d)}(\boldsymbol{x}_{t+1})\right), \tag{B2}$$

where Stratonovich integration is used[67] and d iterates over the configurational dimensions. The second approximation neglects end terms assuming long enough trajectories. We project this equation to D-dimensional collective variables $\boldsymbol{z} = \boldsymbol{M}(\boldsymbol{x})$, making use of a linear mapping operator $\boldsymbol{M}$. Analogous to structure-based coarse-graining, the local entropy production is transformed to CV space by a path-ensemble average,[47]

$$\Delta S[\boldsymbol{z}(t)] = \frac{\int \mathcal{D}[\boldsymbol{x}(t)]\delta(\boldsymbol{M}(\boldsymbol{x}(t)) - \boldsymbol{z}(t))\Delta S[\boldsymbol{x}(t)]}{\int \mathcal{D}[\boldsymbol{x}(t)]\delta(\boldsymbol{M}(\boldsymbol{x}(t)) - \boldsymbol{z}(t))},$$
$$\Delta S[\{\boldsymbol{z}_k\}] = \frac{\prod_{t=0}^{T} \int d\boldsymbol{x}_t \delta(\boldsymbol{M}(\boldsymbol{x}_t) - \boldsymbol{z}_t)\Delta S[\{\boldsymbol{x}_t\}]}{\prod_{t=0}^{T} \int d\boldsymbol{x}_t \delta(\boldsymbol{M}(\boldsymbol{x}_t) - \boldsymbol{z})}. \tag{B3}$$

Using the approximation in Eq. (B2), all integrals over $x_t^{(d)}$ can be performed separately and we find the entropy production in CV space,

$$\Delta S[\boldsymbol{z}(t)] = \frac{1}{2k_B T} \sum_d^D \sum_{t=0}^{T} z_t^{(d)}\left(F^{(d)}(\boldsymbol{z}_{t-1}) - F^{(d)}(\boldsymbol{z}_{t+1})\right), \tag{B4}$$

where $F^{(d)}(\boldsymbol{z}_t)$ are mean forces projected along dimension d evaluated at time t. The solution mentioned above requires to integrate along stochastic trajectories. We approximate the equation by ignoring fluctuations, which allows us to apply Riemann integration. By averaging over all existing pathways between two microstates later on, this approximation becomes exact because the fluctuations are a symmetric contribution to dynamics and do not contribute to entropy production. We express the forces through a conservative contribution derived from the potential of mean force, $G(\boldsymbol{z})$, and a non-conservative contribution, $\boldsymbol{f}$, resulting in $\boldsymbol{F} = -\frac{\partial G(\boldsymbol{z})}{\partial \boldsymbol{z}} + \boldsymbol{f}$. The non-conservative force is directed along the CVs. We find that

$$\Delta S[\{\boldsymbol{z}_t\}] = \frac{1}{k_B T} \int dt \sum_d^D \left(\frac{\partial G}{\partial z_d}\frac{\partial z_d}{\partial t} + f_d \frac{\partial z_d}{\partial t}\right)$$
$$= \frac{1}{k_B T}\left(\int dt \frac{dG}{dt} + \sum_d^D \left(\int dt f_d \frac{\partial z_d}{\partial t}\right)\right)$$
$$= \frac{G(\boldsymbol{z}_T) - G(\boldsymbol{z}_0)}{k_B T} + \frac{1}{k_B T}\sum_d^D\left(\int dt f_d \frac{\partial z_d}{\partial t}\right). \tag{B5}$$

Analogous to reweighting in full-configurational coordinates, two points in CV space can be connected along or against a constant external force. This can create ambiguity for periodic systems. By choosing the lag-time sufficiently small, one set of (long) trajectories has negligible weight compared to the other one. The expression for local entropy production thereby only depends on the initial and target points of the trajectory,

$$\Delta S(z_0, z_T) \approx \frac{G(z_T) - G(z_0) + f \cdot (z_0 - z_T)}{k_{\mathrm{B}} T}. \tag{B6}$$

We numerically estimate the *change* in the entropy production between reference (superscript "q") and target systems by

$$\Delta S(z_0, z_T) - \Delta S^q(z_0, z_T)$$
$$= \frac{1}{k_{\mathrm{B}} T} \Big[G(z_T) - G^q(z_T) - \big(G(z_0) - G^q(z_0)\big) + (z_T - z_0) \cdot (f - f^q) \Big] \tag{B7}$$

as an input for the reweighting formula in Eq. (A9). z_0 and z_T are chosen in the geometric center of each microstate. Because we restrict $G(z)$ to the equilibrium state, the solution of the caliber [Eq. (A6)] does not contain an explicit dependence of ΔS on the stationary distribution.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] J. I. Steinfeld, J. S. Francisco, and W. L. Hase, *Chemical Kinetics and Dynamics* (Prentice Hall Upper, Saddle River, NJ, 1999).

[2] S. Eberhard, G. Finazzi, and F.-A. Wollman, Annu. Rev. Genet. **42**, 463 (2008).

[3] A. U. Khan and M. Kasha, Proc. Natl. Acad. Sci. U. S. A. **80**, 1767 (1983).

[4] R. S. Knox, Biophys. J. **9**, 1351 (1969).

[5] A. Arkin, J. Ross, and H. H. McAdams, Genetics **149**, 1633 (1998).

[6] K. O. Chong, J.-R. Kim, J. Kim, S. Yoon, S. Kang, and K. An, Commun. Phys. **1**, 25 (2018).

[7] C. Cormick, T. Schaetz, and G. Morigi, New J. Phys. **13**, 043019 (2011).

[8] J. P. Dougherty, Philos. Trans. R. Soc., A **346**, 259 (1994).

[9] J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, Curr. Opin. Struct. Biol. **31**, 64 (2015).

[10] A. M. Ferrenberg and R. H. Swendsen, Comput. Phys. **3**, 101 (1989).

[11] C. A. F. De Oliveira, D. Hamelberg, and J. A. McCammon, J. Chem. Phys. **127**, 11B605 (2007).

[12] P. Tiwary and M. Parrinello, Phys. Rev. Lett. **111**, 230602 (2013).

[13] H. Wu, F. Paul, C. Wehmeyer, and F. Noé, Proc. Natl. Acad. Sci. U. S. A. **113**, E3221 (2016).

[14] J. F. Rudzinski, K. Kremer, and T. Bereau, J. Chem. Phys. **144**, 051102 (2016).

[15] L. S. Stelzl, A. Kells, E. Rosta, and G. Hummer, J. Chem. Theory Comput. **13**, 6328 (2017).

[16] L. Donati, C. Hartmann, and B. G. Keller, J. Chem. Phys. **146**, 244112 (2017).

[17] H. Wan, G. Zhou, and V. A. Voelz, J. Chem. Theory Comput. **12**, 5768 (2016).

[18] P. B. Warren and R. J. Allen, Mol. Phys. **116**, 3104 (2018).

[19] J. D. Russo, J. Copperman, and D. M. Zuckerman, preprint arXiv:2006.09451 (2020).

[20] M. Bause, T. Wittenstein, K. Kremer, and T. Bereau, Phys. Rev. E **100**, 060103 (2019).

[21] E. T. Jaynes, *Complex Systems—Operational Approaches in Neurobiology, Physics, and Computers* (Springer, 1985), pp. 254–269.

[22] P. D. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh, and K. A. Dill, J. Chem. Phys. **148**, 010901 (2018).

[23] P. D. Dixit, A. Jain, G. Stock, and K. A. Dill, J. Chem. Theory Comput. **11**, 5464 (2015).

[24] P. D. Dixit, E. Lyashenko, M. Niepel, and D. Vitkup, bioRxiv 137513 (2018).

[25] Z. F. Brotzakis, M. Vendruscolo, and P. Bolhuis, PNAS **118**(2), e2012423118 (2021).

[26] D. Meral, D. Provasi, and M. Filizola, J. Chem. Phys. **149**, 224101 (2018).

[27] K. Ghosh, P. D. Dixit, L. Agozzino, and K. A. Dill, Annu. Rev. Phys. Chem. **71**, 213 (2020).

[28] C. Maes, *Non-Dissipative Effects in Nonequilibrium Systems* (Springer, 2018).

[29] F. Noé and E. Rosta, J. Chem. Phys. **151**, 190401 (2019).

[30] B. E. Husic and V. S. Pande, J. Am. Chem. Soc. **140**, 2386 (2018).

[31] G. R. Bowman, V. S. Pande, and F. Noé, *An Introduction to Markov State Models and their Application to Long Timescale Molecular Simulation* (Springer Science & Business Media, 2013), Vol. 797.

[32] N. Plattner, S. Doerr, G. De Fabritiis, and F. Noé, Nat. Chem. **9**, 1005 (2017).

[33] J. W. Gibbs, *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundations of Thermodynamics* (C. Scribner's Sons, 1902).

[34] E. T. Jaynes, Annu. Rev. Phys. Chem. **31**, 579 (1980).

[35] L. Agozzino and K. A. Dill, Phys. Rev. E **100**, 010105 (2019).

[36] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).

[37] G. E. Crooks, J. Stat. Mech.: Theory Exp. **2011**, P07008.

[38] U. Seifert, Phys. Rev. Lett. **95**, 040602 (2005).

[39] P. D. Dixit and K. A. Dill, J. Chem. Theory Comput. **10**, 3002 (2014).

[40] M. A. Branch, T. F. Coleman, and Y. Li, SIAM J. Sci. Comput. **21**, 1 (1999).

[41] J. Tóbik, R. Martoňák, and V. Cambel, Phys. Rev. B **96**, 140413 (2017).

[42] R. Radhakrishnan and B. L. Trout, J. Am. Chem. Soc. **125**, 7743 (2003).

[43] P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. Natl. Acad. Sci. U. S. A. **97**, 5877 (2000).

[44] O. Valsson, P. Tiwary, and M. Parrinello, Annu. Rev. Phys. Chem. **67**, 159 (2016).

[45] F. Noé and C. Clementi, Curr. Opin. Struct. Biol. **43**, 141 (2017).

[46] M. A. Rohrdanz, W. Zheng, and C. Clementi, Annu. Rev. Phys. Chem. **64**, 295 (2013).

[47] W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008).

[48] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé, J. Chem. Phys. **134**, 174105 (2011).

[49] J. F. Rudzinski and W. G. Noid, J. Chem. Theory Comput. **11**, 1278 (2015).

[50] J. D. Halverson, T. Brandes, O. Lenz, A. Arnold, S. Bevc, V. Starchenko, K. Kremer, T. Stuehn, and D. Reith, Comput. Phys. Commun. **184**, 1129 (2013).

[51] J. F. Rudzinski and T. Bereau, Eur. Phys. J.: Spec. Top. **225**, 1373 (2016).

[52] T. Bereau and J. F. Rudzinski, Phys. Rev. Lett. **121**, 256002 (2018).

[53] P. Deuflhard and M. Weber, Linear Algebra Appl. **398**, 161 (2005).

[54] N. F. Polizzi, M. J. Therien, and D. N. Beratan, Isr. J. Chem. **56**, 816 (2016).

[55] E. Suárez, A. J. Pratt, L. T. Chong, and D. M. Zuckerman, Protein Sci. **25**, 67 (2016).

[56] P. K. Depa and J. K. Maranas, J. Chem. Phys. **123**, 094901 (2005).

[57] M. Guenza, Eur. Phys. J.: Spec. Top. **224**, 2177 (2015).

[58] J. F. Rudzinski, Computation **7**, 42 (2019).

[59] M. K. Meinel and F. Müller-Plathe, J. Chem. Theory Comput. **16**, 1411 (2020).

[60] S. Röblitz and M. Weber, Adv. Data Anal. Classif. **7**, 147 (2013).

[61] A. Likas, N. Vlassis, and J. J. Verbeek, Pattern Recognit. **36**, 451 (2003).

[62] H.-S. Park and C.-H. Jun, Expert Syst. Appl. **36**, 3336 (2009).

[63] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).

[64] M. R. Shirts and J. D. Chodera, J. Chem. Phys. **129**, 124105 (2008).

[65] A. S. Mey, H. Wu, and F. Noé, Phys. Rev. X **4**, 041018 (2014).

[66] F. Müller-Plathe, J. Chem. Phys. **106**, 6082 (1997).

[67] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, 1992), Vol. 1.