

Contour forward flux sampling: Sampling rare events along multiple collective variables

Cite as: J. Chem. Phys. **150**, 024103 (2019); <https://doi.org/10.1063/1.5063358>

Submitted: 26 September 2018 . Accepted: 18 December 2018 . Published Online: 09 January 2019

Ryan S. DeFever , and Sapna Sarupria 



View Online



Export Citation



CrossMark

ARTICLES YOU MAY BE INTERESTED IN

[Automated design of collective variables using supervised machine learning](#)

The Journal of Chemical Physics **149**, 094106 (2018); <https://doi.org/10.1063/1.5029972>

[Perspective: Computational chemistry software and its advancement as illustrated through three grand challenge cases for molecular science](#)

The Journal of Chemical Physics **149**, 180901 (2018); <https://doi.org/10.1063/1.5052551>

[Perspective: Identification of collective variables and metastable states of protein dynamics](#)

The Journal of Chemical Physics **149**, 150901 (2018); <https://doi.org/10.1063/1.5049637>

The Journal
of Chemical Physics

Submit Today

The Emerging Investigators Special Collection and Awards
Recognizing the excellent work of early career researchers!

Contour forward flux sampling: Sampling rare events along multiple collective variables

Cite as: J. Chem. Phys. 150, 024103 (2019); doi: 10.1063/1.5063358

Submitted: 26 September 2018 • Accepted: 18 December 2018 •

Published Online: 9 January 2019



Ryan S. DeFever and Sapna Sarupria^{a)}

AFFILIATIONS

Department of Chemical and Biomolecular Engineering, Clemson University, Clemson, South Carolina 29634, USA

^{a)}Electronic mail: ssarupr@g.clemson.edu

ABSTRACT

Many rare event transitions involve multiple collective variables (CVs), and the most appropriate combination of CVs is generally unknown *a priori*. We thus introduce a new method, contour forward flux sampling (cFFS), to study rare events with multiple CVs simultaneously. cFFS places nonlinear interfaces on-the-fly from the collective progress of the simulations, without any prior knowledge of the energy landscape or appropriate combination of CVs. We demonstrate cFFS on analytical potential energy surfaces and a conformational change in alanine dipeptide.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5063358>

I. INTRODUCTION

Rare events remain uniquely challenging to study in molecular simulations.¹ These infrequent transitions between long-lived (meta)stable states are characterized by large differences between the time scales of the relevant physics (e.g., molecular vibrations, hydrogen bond lifetimes, etc.) and the time between events (often μ s to s). Exemplars include crystal nucleation,^{2–5} ion-pair dissociation in solution,^{6,7} conformational changes in biomolecules,^{8,9} and chemical reactions.¹⁰ Due to the prevalence and importance of rare events, several advanced sampling methods have been developed^{11–21} to estimate transition rate constants and sample unbiased trajectories connecting the stable states. However, even with increasing computational power some phenomena remain challenging to study and continued method development is required.

We present contour forward flux sampling (cFFS), a novel method to sample rare events with multiple collective variables²² (CVs) simultaneously. Building on forward flux sampling (FFS), cFFS leverages overall trajectory behavior to on-the-fly determine nonlinear interface placement in multiple CVs. FFS is a rare event sampling method that uses a series of non-overlapping interfaces to drive a system from an initial state A to final state B.^{15,16,23,24} Each interface is defined

by some value of an order parameter, λ , which changes monotonically from A to B. Straightforward simulation in A is used to estimate the flux, Φ_{A0} , from A to the first interface, λ_0 , and to collect a large number of first-crossing phase points at λ_0 . The designation of a phase point as a first-crossing point indicates that upon following the trajectory backwards in time from the point, one would reach λ_A before $\lambda > \lambda_0$. Several trajectories are initiated from each phase point collected at λ_0 (λ_i). Stochasticity from the dynamics or velocity perturbation at the start of each simulation ensures trajectory divergence. Trajectories returning to A are discarded, while those reaching the next interface, λ_1 (λ_{i+1}), are stored for the next iteration. This procedure is repeated for each interface until the boundary of B is reached, or the probability of advancing to the next interface, $P(\lambda_{i+1}|\lambda_i)$, plateaus to 1. The transition rate constant is calculated as $k_{AB} = \Phi_{A0} \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i)$, and transition paths from A to B are generated by connecting the partial paths backward from B to A. FFS has emerged as a popular choice for studying rare events in simulation because it is applicable to equilibrium and nonequilibrium systems, and implementation is algorithmically straightforward and embarrassingly parallel.

Despite its advantages, FFS has shortcomings. Assuming reasonable definitions for the boundaries of A and B, the rate

constant and transition path ensemble (TPE) computed with FFS are, in principle, independent of the order parameter used for the calculation.¹⁶ In practice, a poor choice of the order parameter is detrimental to the efficiency of FFS^{25,26} and can even lead to incorrect results.^{26,27} This arises when portions of λ_i which are important to the transition are sparingly sampled. More formally, imagine some coordinate ($\lambda^\perp$) orthogonal to λ . Challenges arise for FFS when there is poor overlap between the distribution of first-crossing phase points captured at λ_i , $\rho(\lambda^\perp|\lambda_i)$, and the probability of reaching λ_B from some point on λ_i , $P(\lambda_B|\lambda_i; \lambda^\perp)$.²⁶ There are two approaches to overcome this issue: (1) increase sampling to collect more phase points at problematic interface(s) or (2) improve the choice of the order parameter to increase overlap between the two distributions. The first approach yields more phase points everywhere along an interface but with sufficient sampling the paths spawned from phase points with a higher $P(\lambda_B|\lambda_i; \lambda^\perp)$ will come to dominate the eventual path ensemble, resulting in the correct rate constant and TPE. Unfortunately, the efficiency of FFS will still be poor. By contrast, the second approach increases the efficiency of FFS, meaning that FFS will converge to the correct rate constant and TPE with less sampling. Unfortunately, optimal order parameters are rarely known *a priori*. More often, one of the reasons for generating a path ensemble with a method such as FFS is to identify order parameters which best describe the transition.

Since sampling of all interfaces $i > 0$ in FFS depends on the phase points collected at λ_0 , methods have been proposed to optimize placement of and ensure adequate sampling of λ_0 .^{2,28,29} If the situation is not too dire, increasing the length of the basin simulation and collecting more phase points at λ_0 may provide a sufficient remedy. However, if overlap between $\rho(\lambda^\perp|\lambda_0)$ and $P(\lambda_B|\lambda_0; \lambda^\perp)$ is extremely small, this may be insufficient. Furthermore, the problem is not limited to λ_0 ; in principle, the distribution of phase points sampled at any λ_i could suffer from this problem. Poor overlap between $\rho(\lambda^\perp|\lambda_i)$ and $P(\lambda_B|\lambda_i; \lambda^\perp)$ becomes particularly problematic for systems with multiple transition tubes. There, a poor choice of the order parameter may result in some transition tubes becoming (artificially) favored over others. In the extreme, entire transition tubes can be missed by FFS.

A related situation worth mentioning is when $\rho(\lambda^\perp|\lambda_0)$ converges extremely slowly.^{2,29} If this is the problem, extending the basin simulations until convergence is achieved which will remedy the situation.² A greater number of phase points at λ_0 are not required; just phase points correctly sampled from the converged distribution are required.

The choice of the order parameter strongly affects the overlap between $\rho(\lambda^\perp|\lambda_i)$ and $P(\lambda_B|\lambda_i; \lambda^\perp)$. If the order parameter is the committor function, $P(\lambda_B|\lambda_i; \lambda^\perp)$ is constant with $\lambda^\perp$, thereby assuring good overlap between $\rho(\lambda^\perp|\lambda_i)$ and $P(\lambda_B|\lambda_i; \lambda^\perp)$.²⁶ Borrero and Escobedo thus devised a method to optimize the order parameter with a series of FFS simulations.³⁰ Though the approach yields improvements,⁹ it is challenging for systems which require extraordinary computational resources for even a single FFS run.^{4,5} Furthermore, some

processes are inherently multidimensional,^{6,31–33} and driving the transition along a single CV may not be ideal.

cFFS takes a different approach. We extend FFS to use multiple CVs on-the-fly. This allows researchers to test multiple CVs simultaneously and improves the chances of capturing important orthogonal coordinates within the set of CVs used to drive the transition. At each interface, cFFS identifies the next interface as a nonlinear combination of specified CVs on-the-fly from the behavior of simulations initiated from the previous interface. In doing so, cFFS also reveals the role of each CV through the entire transition. Only the combination of CVs must separate A and B, and so each CV need not monotonically change from A to B. If some CV is unimportant, this will be reflected by but not impede cFFS. These features offer substantial flexibility in CVs that can be used with cFFS. cFFS generates an estimate of the transition rate constant and a collection of $A \rightarrow B$ trajectories belonging to the TPE. We demonstrate cFFS with two CVs, but in principle it can be extended to three or more CVs.

In Sec. II, we explain cFFS. We proceed to demonstrate cFFS on several two-dimensional potential energy surfaces (PES) in Sec. III. In Sec. IV, we demonstrate cFFS with one position coordinate and one momentum coordinate, and in Sec. V, we test cFFS on a standard higher dimensional test case, a conformational transition in alanine dipeptide. Discussion and closing remarks are provided in Secs. VI and VII, respectively.

II. CONTOUR FORWARD FLUX SAMPLING

The central idea of cFFS is to allow the system to naturally evolve along multiple CVs to reveal how different CVs participate in the transition. This is achieved by placing the subsequent interface based on sampling initiated from the current interface. The FFS formalism can still be used to calculate the rate constant and TPE. Interface placement is designed such that the distribution of first-crossing points is uniform along the interface, ensuring that each interface is well-sampled everywhere within the chosen CVs.

The first step of cFFS is to run straightforward basin simulations in A to identify the bounds of A (λ_A) and the first interface (λ_0) and to collect phase points at λ_0 . The value of each CV in time, $\lambda(t)$, is calculated, where $\lambda \equiv \{\lambda^I, \lambda^{II}, \dots, \lambda^N\}$ is the set of CVs. CV space is discretized to create an N-D grid. The discretization size is selected such that the system rarely travels more than a single grid site in one time step. The discrete probability distribution, $P(\lambda)$, is calculated from the basin simulations. Grid sites exceeding a threshold probability are added to the set of sites describing A, $\mathbf{s}_A$. Regions of CV space which are not in $\mathbf{s}_A$ but completely surrounded by $\mathbf{s}_A$ are added to $\mathbf{s}_A$. λ_A is defined as the boundary between sites in $\mathbf{s}_A$ and those that are not. Trajectories exit A when they cross from a grid site in $\mathbf{s}_A$ to a grid site not in $\mathbf{s}_A$.

Several criteria are used to identify $\mathbf{s}_0$, the set defining λ_0 . $\mathbf{s}_0$ should (a) completely contain $\mathbf{s}_A$ so that λ_0 does not overlap

with or cross λ_A , (b) not create regions of CV space completely surrounded by $\mathbf{s}_0$, but not included in it, (c) not include sites in $\mathbf{s}_B$, the set of sites describing B, (d) be selected such that some desired number of phase points can be collected at λ_0 , and (e) be selected such that there is an equal flux of trajectories exiting $\mathbf{s}_0$ along the entire λ_0 interface. Criteria (e) is crucial as it ensures that cFFS does not bias the system to sample any one direction more readily than another. Further discussion is provided later. Once λ_A and λ_0 are defined the basin simulations are re-analyzed to calculate Φ_{A0} and collect phase points at λ_0 .

The remainder of cFFS proceeds as follows. Several trajectories are initiated from each phase point at λ_i ($\lambda_i = \lambda_0$ for the first iteration). Trajectories are terminated when they return to λ_A or reach a maximum number of steps. The set of sites defining λ_{i+1} , $\mathbf{s}_{i+1}$, is determined from the behavior of trajectories initiated at λ_i using analogous criteria to those described for determining λ_0 . Note that $\mathbf{s}_{i+1}$ must completely contain $\mathbf{s}_i$ to satisfy the effective positive flux (EPF) formalism.^{14,34} Once $\mathbf{s}_{i+1}$ is identified, trajectories are re-analyzed to determine if they cross λ_{i+1} (i.e., exit $\mathbf{s}_{i+1}$) before returning to A. For each trajectory that crosses λ_{i+1} , the phase point at the time step which the trajectory crosses λ_{i+1} is saved. Trajectories which fail to reach λ_{i+1} or return to A before the maximum number of steps are extended until they reach λ_{i+1} or return to A. The probability, $P(\lambda_{i+1}|\lambda_i)$, is calculated from the number of trajectories that reach λ_{i+1} before returning to A.

Eventually, sites in $\mathbf{s}_{i+1}$ will be adjacent to sites in $\mathbf{s}_B$. Trajectories initiated from λ_i can then reach λ_{i+1} , return to A, or proceed directly to B. This indicates the kinetic barrier has been surmounted, and thus cFFS is nearly

complete. Two probabilities are now calculated, $P(\lambda_{i+1}|\lambda_i)$ and $P(\lambda_B|\lambda_i)$. Our approach is to continue cFFS until $\mathbf{s}_{i+1}$ surrounds $\mathbf{s}_B$. At this point, i becomes the final interface, n . Trajectories initiated from λ_n are continued until they reach λ_B or return to λ_A to close the probabilities for the rate calculation. As with multi-state FFS,³⁵ the transition rate constant is calculated as

$$k_{AB} = \Phi_{A0} \sum_{j=0}^n P(\lambda_B|\lambda_j) \prod_{i=0}^{j-1} P(\lambda_{i+1}|\lambda_i). \quad (1)$$

The collection of trajectories comprising the TPE is constructed by connecting the partial paths backwards from B to A. Note that all trajectories *do not* have equal weight in the TPE. The relative weight of each trajectory is $w = 1/\prod_{i=0}^j k_i$, where j is the final interface crossed by a trajectory before reaching B and k_i is the number of trajectories initiated from each configuration at interface i .

III. DEMONSTRATION ON 2D POTENTIAL ENERGY SURFACES

We demonstrate cFFS with Langevin dynamics of a single particle on four 2D potential energy surfaces (PESs) with different topographical features [see Figs. 1(a)–1(d)]. PES-1 has a single transition tube which follows two monotonically increasing CVs. PES-2 has a single transition tube with hysteresis in the x coordinate. PES-3 and PES-4 both contain two transition tubes; the potential energy barriers are the same for the two tubes on PES-3 and different for the two tubes on PES-4. For each PES, we study $A \rightarrow B$ transitions with straightforward Langevin dynamics (SLD), FFS_{opt}, FFS_x, and cFFS. FFS_{opt} denotes FFS performed with the optimal linear

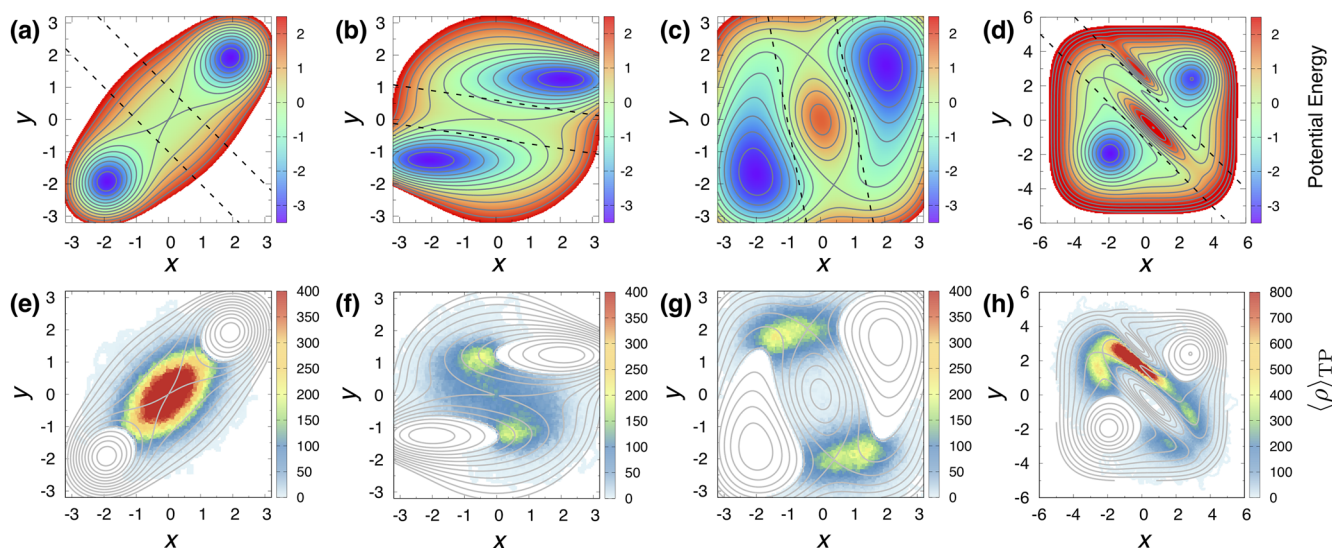


FIG. 1. (Top panels) PESs used to test cFFS: (a) PES-1, (b) PES-2, (c) PES-3, and (d) PES-4. Colors represent the potential energy. Contour lines are separated by 0.5 units. The region between the dashed lines was used to quantitatively compare $\rho(q|TP)$ between different methods. (Bottom panels) TPE sampling from SLD at $\beta = 2.5$ on (e) PES-1, (f) PES-2, (g) PES-3, and (h) PES-4.

combination of x and y (i.e., the order parameter orthogonal to the dividing surface of the PES), and FFS_x indicates FFS performed with x as the (suboptimal) order parameter. We stress that optimal order parameters are not known *a priori* for most realistic systems, and therefore FFS is generally performed with suboptimal order parameters. Further details of the PESs, Langevin dynamics, and FFS/cFFS parameters are provided in the [supplementary material](#).

A. Rate constants

$A \rightarrow B$ transition rate constants are reported in [Table I](#). Transitions were studied at $\beta = 2.5$ and $\beta = 5.0$ ($\beta = 1/k_B T$). The higher temperature ($\beta = 2.5$) enables rigorous comparison of TPE sampling with SLD, whereas the lower temperature ($\beta = 5.0$) provides a test at more challenging conditions. SLD rate constants are unbiased estimates. FFS_x provides accurate estimates of the rate constants at $\beta = 2.5$ but at $\beta = 5.0$ FFS_x underestimates the rate constants. This suggests that suboptimal order parameters perform worse as the barrier becomes larger relative to $k_B T$. We explain the breakdown of FFS_x by examining the TPE sampling below. FFS_{opt} and cFFS perform better. Rate constants from FFS_{opt} and cFFS both agree well with SLD at $\beta = 2.5$. At $\beta = 5.0$, FFS_{opt} underestimates rate constants for PES-2 and PES-3. By contrast, cFFS provides correct estimates of the rate constants for all four PESs at $\beta = 5.0$.

TABLE I. $A \rightarrow B$ transition rate constants for four 2D PESs. One standard deviation of the mean is reported in parentheses.

PES	$k_{AB} \times 10^5$ at $\beta = 2.5$			
	SLD	FFS_{opt}	FFS_x	cFFS
PES-1	2.9(0.2)	2.8(0.3)	3.1(0.9)	2.8(0.2)
PES-2	9.1(0.1)	7.9(0.9)	10.2(2.5)	8.8(0.7)
PES-3	2.6(0.3)	2.4(0.4)	2.3(0.6)	2.4(0.1)
PES-4	1.1(0.1)	1.0(0.1)	1.1(0.1)	1.0(0.1)

PES	$k_{AB} \times 10^9$ at $\beta = 5.0$			
	SLD	FFS_{opt}	FFS_x	cFFS
PES-1	5.4(1.2)	4.4(0.2)	3.1(0.2)	4.5(0.6)
PES-2	23.2(2.0)	18.0(3.5)	18.3(1.0)	21.9(2.3)
PES-3	6.4(1.3)	2.9(0.2)	2.5(0.2)	5.4(0.5)
PES-4	2.8(0.9)	1.9(0.4)	0.42(0.02)	2.6(0.1)

B. Transition path ensemble sampling

Though attaining the correct $A \rightarrow B$ rate constant is a crucial test of cFFS, it is also important that cFFS correctly samples the TPE. TPE sampling is calculated as $\langle \rho \rangle_{\text{TP}} = \langle n_{\text{visits}}/l^2 \rangle_{\text{TP}}$, where $\langle \dots \rangle_{\text{TP}}$ indicates an ensemble average over all transition paths and n_{visits} is the number of times a transition path visited each $l \times l$ region of space. For reference, TPE sampling from SLD at $\beta = 2.5$ is shown in the bottom panels of [Fig. 1](#).

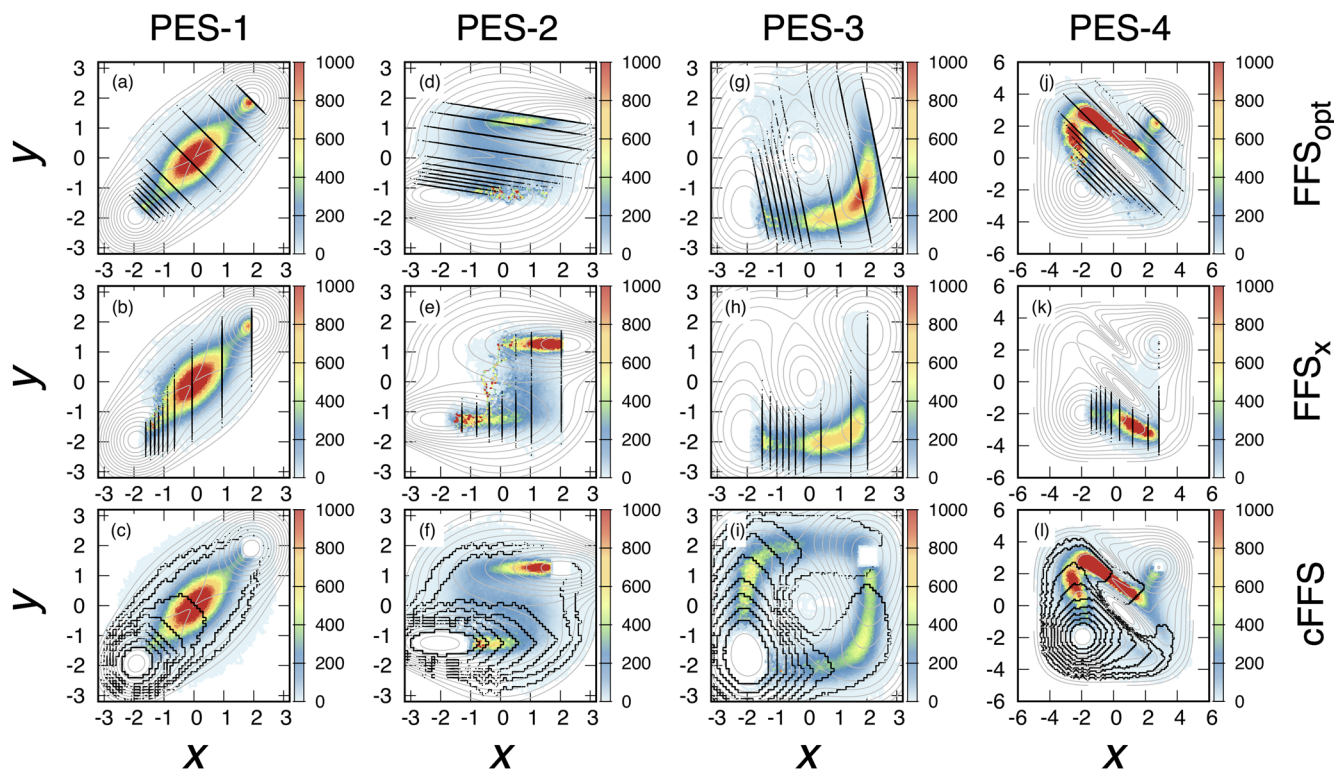


FIG. 2. Comparison of interface placement and TPE sampling generated with FFS_{opt} , FFS_x , and cFFS on PES-1–PES-4 at $\beta = 5.0$. PES contours are shown as gray lines. Configurations collected at each interface are shown with black points. TPE sampling represented by the heat map.

Figure 2 summarizes the behavior of FFS_{opt}, FFS_x, and cFFS on PES-1–PES-4 at $\beta = 5.0$. All methods result in qualitatively similar sampling for PES-1. The other surfaces proved more challenging for FFS_x and FFS_{opt}. By contrast, cFFS results in the qualitatively correct sampling for all four PESs. On PES-2, the hysteresis provides a challenge for FFS_x. Unlike FFS_{opt} and cFFS, FFS_x undersamples the $x < 0$ portion of the transition tube. On PES-3 and PES-4, the failure of FFS_x is even more stark; FFS_x only samples one of the two transition tubes. Even FFS_{opt} fails to sample both transition tubes equally on PES-3. PES-3 and PES-4 have two distinct transition tubes, and the minimum energy paths change direction from A to B. On PES-3, both transition tubes have the same potential energy barrier. However, one transition tube approaches the transition state from A with a gentler slope. Results from SLD at $\beta = 2.5$ in Fig. 1(c) indicate that both transitions should be equally traveled. cFFS reproduces this behavior at both $\beta = 2.5$ (Fig. S1 of the [supplementary material](#)) and the more challenging $\beta = 5.0$ [Fig. 2(i)]. At $\beta = 5.0$, FFS_x only samples a single transition tube [Fig. 2(h)]. Even FFS_{opt} struggles to sample both transition tubes equally on PES-3 [Fig. 2(g)]. The behavior of FFS_{opt} and FFS_x on PES-3 can be explained by the framework put forth in the introduction. In both cases, it is apparent that $\rho(\lambda^\pm|\lambda_0)$ sampled during the basin simulations only has good overlap with $P(\lambda_B|\lambda_0; \lambda^\pm)$ for one of the two transition tubes. The result is that FFS oversamples the tube with greater overlap, at the expense of the other transition tube. FFS sensitivity to the choice of the order parameter on PES-3 is further demonstrated in Fig. S2 of the [supplementary material](#). Though FFS will converge to the correct TPE in the limit of infinite sampling, as a practical matter FFS can lead to incorrect results. cFFS again performs well on PES-4, illustrating that cFFS is able to navigate a tortuous transition landscape with two transition tubes and unequal potential energy barriers.

Near the dividing surface (see Fig. 1), we quantitatively compare the TPE density of states, $\rho(q|TP)$, from SLD with that from FFS_{opt}, FFS_x, and cFFS using the Jensen-Shannon divergence.³⁶ We restrict our comparison to $\beta = 2.5$, where a large number of transitions can be generated with SLD, hence providing a robust reference. The results shown in Fig. 3 confirm qualitative conclusions from Fig. 2 ($\beta = 5.0$) and Fig. S1 of the [supplementary material](#) ($\beta = 2.5$). At $\beta = 2.5$, FFS_{opt} and cFFS perform similarly. For the simplest case (PES-1), FFS_x performs nearly as well as FFS_{opt} and cFFS. However, for the more complex surfaces, including the surface with hysteresis (PES-2) and surfaces with two transition tubes (PES-3, PES-4), FFS_x performs notably worse.

C. cFFS interface placement

Figure 2 also demonstrates cFFS interface placement. Interfaces are spaced further apart in directions that trajectories more readily advance and closer together in directions that trajectories struggle to advance. For these low-dimensional systems, interface locations adhere closely to the contours of the PESs. We strongly emphasize that no

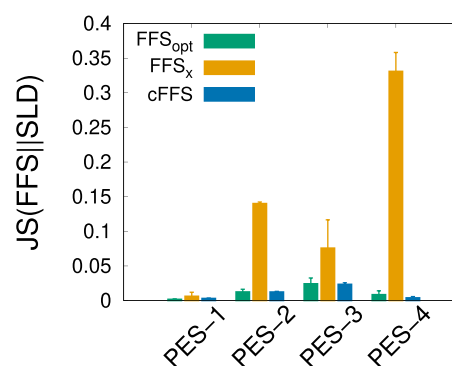


FIG. 3. Jensen-Shannon divergence between $\rho(q|TP)$ calculated with SLD and FFS_{opt}, FFS_x, and cFFS at $\beta = 2.5$. A value of zero indicates identical probability distributions, while a value of 1.0 indicates completely non-overlapping distributions. Error bars represent one standard deviation on the mean of three independent trials.

knowledge of the PES is employed; cFFS places the interface λ_{i+1} from the progress of trajectories initiated from λ_i alone.

If not done properly, performing FFS with multiple CVs simultaneously can bias the system to over-sample or under-sample regions of CV space. The amount of work performed by FFS is related to interface spacing (i.e., $\lambda_{i+1} - \lambda_i$), slope of the free energy landscape between λ_i and λ_{i+1} , and the number of trajectories initiated from λ_i . If the slope of the free energy landscape between two interfaces becomes steeper, λ_{i+1} is moved closer to λ_i or the number of trajectories initiated from λ_i is increased. Multiple CVs introduces a new prospect that unequal amounts of work are inserted along different CVs, biasing the system to over-sample in the direction that more work is inserted.

We introduced a condition of constant flux along an interface in cFFS interface placement to address this problem. The force exerted by the underlying free energy surface is proportional to $-dn_{\text{cross}}/d\lambda$, where n_{cross} is the number of trajectories crossing an interface placed at some value of λ . If n_{cross} changes more quickly with changing λ , then the underlying surface must have a steeper slope. Applying the differential definition of work, $dW = Fd\lambda$, and thus $dW \propto dn_{\text{cross}}$ and $W \propto n_{\text{cross}}$. Constant flux along the interface requires that all small sections of λ_{i+1} have approximately the same number of trajectories crossing them. This condition ensures that equal work is inserted everywhere along the interface (i.e., in all directions) and results in λ_{i+1} closer to λ_i in directions trajectories struggle to advance and farther from λ_i in directions trajectories readily advance. The fact that cFFS is able to reproduce the correct TPE symmetry for PES-3 and PES-4 provides strong evidence that the constant flux along the interface condition is correct.

In complex systems, the optimal order parameter is often expected to be a combination (linear or nonlinear) of multiple (suboptimal) order parameters. This combination is generally

nonintuitive and difficult to predict. As such, most applications of FFS use a suboptimal order parameter (e.g., FFS_x). On the four PESSs, cFFS successfully produces correct TPE sampling without knowing how x and y should be combined. Though x and y are part of the optimal order parameter, independently, x and y are suboptimal order parameters. This suggests that cFFS can outperform FFS when multiple suboptimal order parameters are known, but the optimal order parameter remains unknown. In addition, nonlinear combinations of CVs have increased degeneracy compared with linear combinations of CVs in creating reaction coordinates (i.e., optimal order parameters).³⁷ Since cFFS interfaces are arbitrarily complex combinations of the specified CVs, there may be substantial flexibility in selecting good CVs for cFFS. A variety of approaches have been proposed for identifying important CVs for rare event transitions.^{30,38–44} For example, recent work suggests that important CVs can be identified from local fluctuations in the (meta)stable basins.⁴⁵ We envision using such approaches to identify key CVs for cFFS.

IV. cFFS WITH A MOMENTUM COORDINATE

FFS is most often applied in the diffusive limit, and the CVs used as FFS order parameters are generally only functions of the atomic positions. In this section, we demonstrate cFFS on a simple analytical potential where momentum plays a key role during the transition. A previous study shows that FFS fails and under-predicts the transition rate constant when using a position-based order parameter alone.²⁶

Reference 26 tested several path sampling methods for a transition on a simple 1D analytical potential described by $V(r) = r^4 - 2r^2$. Of the tested methods, replica exchange transition interface sampling (RETIS) and partial path transition interface sampling (PPTIS) provided the best estimates to the reference effective positive flux (EPF) rate ($k_{AB}^{\text{EPF}} = 2.4 \pm 0.1 \times 10^{-7}$, $k_{AB}^{\text{RETIS}} = 2.8 \pm 0.7 \times 10^{-7}$, and $k_{AB}^{\text{PPTIS}} = 2.7 \pm 0.6 \times 10^{-7}$). FFS performed worst, underestimating the rate constant by 1–2 orders of magnitude depending on the length of the basin simulation. With a basin simulation of 4×10^6 steps, FFS produced a rate constant of $k_{AB}^{\text{FFS-short}} = 2.2 \pm 0.2 \times 10^{-9}$. When the basin simulation was extended to 10×10^6 steps, $k_{AB}^{\text{FFS-long}} = 1.2 \pm 0.1 \times 10^{-8}$. As explained in Ref. 26, the source of the systematic error in the rate constant was the lack of overlap between $\rho(\lambda^\perp|\lambda_0)$ and $P(\lambda_B|\lambda_0; \lambda^\perp)$. Successful transitions require large momentum when exiting the initial basin, and few to none of the trajectories captured at λ_0 had the requisite momentum. Even successful transition paths from FFS exited the initial state with lower momenta compared with other methods, resulting in a low estimate of the rate constant. This also resulted in the unphysical result that the momenta of transition paths from FFS were not symmetric about the barrier.

We perform cFFS with the above potential at identical conditions as in Ref. 26. The two variables for cFFS are the position (r) and momenta (p). The basin simulation is

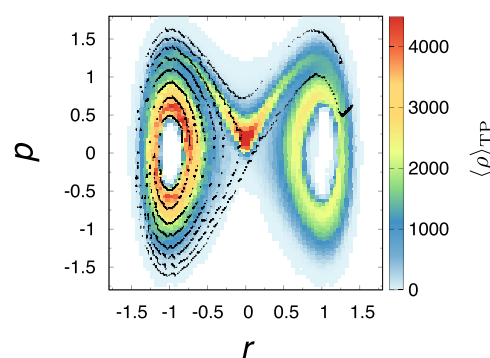


FIG. 4. cFFS on 1D potential with one position coordinate (r) and one momentum coordinate (p). Initial basin A is $r < 0$ minima, and final basin B is $r > 0$ minima. Configurations collected at each interface are shown as black points. Color map shows the TPE sampling.

performed with 4×10^6 steps. We place interfaces adaptively, collecting ~ 2000 configurations per interface. As in Ref. 26, we initiate 20 000 trajectories from each interface. cFFS resulted in shooting from 8 interfaces, compared with the 7 interfaces used in Ref. 26. The average rate constant from three cFFS trials was $k_{AB}^{\text{cFFS}} = 2.0 \pm 0.1 \times 10^{-7}$, slightly underestimating the EPF rate constant from Ref. 26. The TPE and configurations collected at each interface from cFFS are shown in Fig. 4. Paths exit the initial state orbiting the basin and acquiring more kinetic energy until they are able to escape. Their momenta then approaches zero as they cross through the transition state, before accelerating towards and orbiting into the final state. Consistent with theoretical expectations, the TPE generated by cFFS is symmetric about the barrier. We also tested cFFS with less sampling. Even with a twenty-fold reduction in sampling (1000 trajectories and 100 configurations per interface), the rate constant calculated with cFFS is $k_{AB}^{\text{cFFS}} = 2.6 \pm 0.7 \times 10^{-7}$ and the TPE remains symmetric about the barrier.

These results demonstrate the potential for using cFFS to study transitions with important momentum variables. Though the above test case represents an extremely simple analytical model, it demonstrates the advantages of cFFS in such scenarios. If an important momentum variable is known for a transition, cFFS allows the basins to be separated with a position coordinate while the momentum coordinate can be used to help drive the transition.

V. DEMONSTRATION ON ALANINE DIPEPTIDE

In keeping with tradition, we close by demonstrating cFFS on the C_{7ax} -to- C_{7eq} conformational change in alanine dipeptide in vacuum. Details of the simulations and cFFS are reported in the [supplementary material](#). ϕ and ψ backbone dihedral angles were used as CVs for cFFS. The progression of cFFS is shown in Fig. 5(a). Starting from the C_{7ax} basin centered near $\phi = 60^\circ$ and $\psi = -30^\circ$, cFFS drives the system to the C_{7eq} basin defined by $-94^\circ < \phi < -60^\circ$ and $12^\circ < \psi < 90^\circ$. The shape of the interfaces shows that ϕ plays the larger role

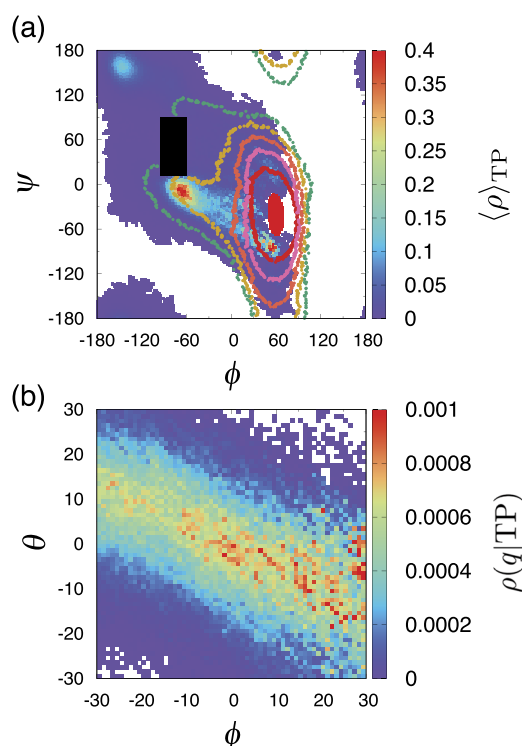


FIG. 5. cFFS for alanine dipeptide in vacuum. (a) Initial and final states are shown as red and black regions, respectively. Configurations collected at λ_0 , λ_1 , λ_2 , λ_3 , and λ_4 are reported as red, pink, salmon, gold, and green points, respectively. The color map represents the TPE sampling. ϕ and ψ angles are reported in degrees. (b) Correlation between ϕ and θ in the TPE. The color map represents the TPE density of states.

in the transition and reveals the location of the primary transition tube. The transition rate constant predicted by cFFS ($k_{AB}^{\text{cFFS}} = 5.0 \times 10^6 \text{ s}^{-1}$) compares favorably with straightforward simulation ($k_{AB}^{\text{SLD}} = 4.8 \times 10^6 \text{ s}^{-1}$). In Fig. 5(b), we show the relationship between ϕ and another dihedral angle, θ , in the TPE. It has been shown that θ is part of the reaction coordinate.^{17,46} cFFS captures the proper relationship between ϕ and θ even though θ is not one of the CVs used during cFFS.^{17,19}

VI. DISCUSSION

cFFS helps overcome a few challenges posed by FFS. cFFS allows one to try multiple CVs simultaneously. This is beneficial for systems where investigators have some *a priori* insight into the CVs that are expected to play a role in the transition, but a detailed analysis of the mechanism is missing and the best order parameter remains unknown. By using multiple CVs simultaneously and enforcing constant flux along an interface, the method can alleviate issues associated with poor overlap between $\rho(\lambda^\perp|\lambda_i)$ and $P(\lambda_B|\lambda_i; \lambda^\perp)$. Of course, it is possible that there are additional important orthogonal coordinates beyond the chosen CVs. This situation could pose sampling challenges for cFFS. Finally, we demonstrated cFFS with a combination

of momenta and position based coordinates. This may extend the practical applicability of FFS to more ballistic systems. FFS depends on stochasticity for trajectory divergence between subsequent interfaces, so it will still not be applicable in the limit of fully deterministic dynamics.

cFFS can in principle be extended to a large number of CVs. However, we surmise the method will not scale well to more than three or four CVs. In high dimensional space, the area through which trajectories can cross an interface will become exceedingly large. From a practical standpoint, this will make it difficult to maintain the constant flux condition. From an efficiency standpoint, most of each interface will drive the system towards regions of phase space which are irrelevant to the transition of interest. Even if successful transitions are generated, they will probably originate from a tiny subset of the phase points collected at λ_0 and thus be highly correlated. Challenges associated with scaling to large numbers of CVs are hardly limited to cFFS. A variety of advanced sampling methods, including nonequilibrium umbrella sampling^{47,48} and metadynamics⁴⁹ have come across similar problems. One solution is to collapse the reaction coordinate to a single dimension using a string-type approach.^{17,42,48} The string-type approach will prove difficult to implement in FFS without resorting to an iterative scheme requiring multiple FFS runs because each path ensemble in FFS is generated sequentially and there is no opportunity to relax the string. Moreover, the string-type approach could defeat one of the benefits of cFFS, which is that it enables exploration of transitions with multiple tubes.

Extending cFFS to large numbers of CVs will thus require alternative approaches. Dimensionality reduction techniques such as isomaps^{50,51} or diffusion maps^{52,53} could be employed to reduce a large number of CVs to two or three reduced coordinates which capture the largest spread in the data. In this manner, multiple transition tubes would hopefully be preserved^{33,51} within the reduced coordinates. Furthermore, several groups are actively working to combine machine learning and advanced sampling methods to identify important CVs on-the-fly.^{54–58} We are exploring if such methods or variations thereof can be incorporated with cFFS. One challenge to incorporate on-the-fly identification of reduced coordinates with FFS-type methods is again related to the sequential generation of ensembles. Sampling from the initial basin alone is unlikely to reveal reduced coordinates ideal for studying the transition. As FFS progresses, sampling from each interface ensemble will result in reduced coordinates which increasingly describe the transition. However, FFS requires that each ensemble be visited sequentially, and changing the definition of the reduced coordinates after each ensemble may cause substantial difficulty in maintaining this condition.

Studying rare events in simulations is an important and challenging problem that has spawned the development of many methods in the past decades. Here we restrict our comparison to two methods which use multiple CVs to sample and calculate rate constants for rare transitions with unbiased

dynamics in equilibrium or nonequilibrium systems. Vanden-Eijnden and Venturoli developed a method⁵⁹ that calculates the transition rate constants and transition paths from the steady state distribution under the boundary conditions that state A is a source and state B is a sink. The space between the stable states is tiled into enclosed Voronoi cells, and parallel simulations are performed in each cell. The steady state flux and probability distribution can be estimated from the time spent in each cell and exchange between cells. Like cFFS, the method is applicable to equilibrium as well as nonequilibrium systems and does not require that A and B be well separated in both variables. Since each parallel path is restricted to a single cell, the method may prove advantageous compared with cFFS for systems with metastable intermediates. The method does not provide direct access to dynamical transition paths, although, in principle, transition paths could probably be reconstructed with an extensive bookkeeping scheme. It is not immediately apparent which method would be better for systems with slowly decorrelating transition paths.

As mentioned in the introduction, Borrero and Escobedo³⁰ developed a method to optimize the FFS order parameter through a series of FFS runs. The approach uses committor information obtained from the prior FFS run to identify the best order parameter from a set of specified CVs. The procedure can be repeated until TPE sampling or the optimal order parameter converges. Like cFFS, the procedure in Ref. 30 allows FFS to be used in situations where there are a number of possible CVs. Since the FFS runs themselves are performed along a single order parameter (which may be a linear or nonlinear combination of multiple CVs), there is no limitation to the number of CVs which can be tested. For certain systems, this may represent a substantial advantage over cFFS, which in current form is practically limited to three or four CVs. Unfortunately, the method presented in Ref. 30 requires multiple (often expensive) FFS runs. Additionally, given the sensitivity of FFS sampling to the choice of the order parameter in the presence of multiple transition tubes, we suspect cFFS will perform better for such systems.

Finally, we would like to comment on the possibility of combining a cFFS-type approach with other path sampling methods. At the most basic level, cFFS divides CV space into a fine grid to help define regions of phase space and interfaces between those regions with arbitrary shape. In cFFS, criteria for boundary identification were selected to meet the needs of FFS – a minimum number of total first crossings and constant average flux along the interface to avoid biasing the system to proceed in one direction over another. It is easy to imagine modifying the boundary identification criteria for other applications. Within the family of FFS approaches, it may prove fruitful to combine the approach of Borrero and Escobedo³⁰ with a cFFS-type approach for interface definitions. This could allow interfaces with any arbitrary shape which could better reproduce the committor function. Transition interface sampling is less sensitive to the definition of the order parameter.^{25,26} However, a procedure has been proposed to optimize

interface placement given the order parameter.⁶⁰ This criterion for optimal interface placement could be combined with a cFFS-type approach for dividing the CV space for transition interface sampling.

VII. CONCLUDING REMARKS

We described cFFS, a method to sample rare event transitions along multiple CVs simultaneously. cFFS uses automated nonlinear interface placement and reveals on-the-fly the evolution of CVs during a transition. cFFS was tested with two CVs, but in principle, it can be extended to three or more. In practice, extending cFFS in current form to more than three or four CVs may prove challenging. The stable states only need to be separated in a combination of CVs, which may change nonmonotonically between the stable states. We introduced a criterion of constant flux along each interface to prevent biasing TPE. cFFS results in correct estimates of the transition rate constants and TPE sampling on several 2D PESs and the C_{7ax} -to- C_{7eq} transition in alanine dipeptide in vacuum. We additionally demonstrated cFFS on a 1D analytical potential using one position coordinate and one momenta coordinate. cFFS substantially improved upon FFS results on the same potential, where only the position coordinate was used as the order parameter. On the 2D PESs, cFFS performed particularly well for systems with hysteresis or multiple transition tubes. cFFS with multiple suboptimal order parameters consistently outperformed FFS with a single suboptimal order parameter. Since optimal order parameters are not known in most applications of FFS, cFFS with two or more suboptimal order parameters will be beneficial for studies of complex systems such as macromolecular conformational transitions and crystal nucleation.

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for details of Langevin dynamics, PESs, TPE sampling at $\beta = 2.5$, FFS/cFFS sampling on PES-3 at $\beta = 5.0$, and alanine dipeptide simulations.

ACKNOWLEDGMENTS

We thank the referees for thoughtful suggestions which substantially improved an earlier version of this work. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award No. DE-SC0015448. Clemson University is acknowledged for generous allotment of compute time on the Palmetto cluster.

REFERENCES

- ¹B. Peters, *Reaction Rate Theory and Rare Events* (Elsevier, 2017).
- ²Y. Bi and T. Li, *J. Phys. Chem. B* **118**, 13324 (2014).
- ³A. Haji-Akbari, R. S. DeFever, S. Sarupria, and P. G. Debenedetti, *Phys. Chem. Chem. Phys.* **16**, 25916 (2014).
- ⁴A. Haji-Akbari and P. G. Debenedetti, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 10582 (2015).
- ⁵R. S. DeFever and S. Sarupria, *J. Chem. Phys.* **147**, 204503 (2017).
- ⁶P. L. Geissler, C. Dellago, and D. Chandler, *J. Phys. Chem. B* **103**, 3706 (1999).

- ⁷E. Pluhařová, M. D. Baer, G. K. Schenter, P. Jungwirth, and C. J. Mundy, *J. Phys. Chem. B* **120**, 1749 (2015).
- ⁸J. Juraszek and P. G. Bolhuis, *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15859 (2006).
- ⁹C. Velez-Vega, E. E. Borrero, and F. A. Escobedo, *J. Chem. Phys.* **133**, 105103 (2010).
- ¹⁰M. Moqadam, A. Lervik, E. Riccardi, V. Venkatraman, B. K. Alsberg, and T. S. van Erp, *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4569 (2018).
- ¹¹C. H. Bennett, "Molecular dynamics and transition state theory: The simulation of infrequent events," in *Algorithms for Chemical Computations* (American Chemical Society, 1977), Chap. 4, pp. 63–97.
- ¹²C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).
- ¹³W. E, W. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).
- ¹⁴T. S. van Erp, D. Moroni, and P. G. Bolhuis, *J. Chem. Phys.* **118**, 7762 (2003).
- ¹⁵R. J. Allen, P. B. Warren, and P. R. ten Wolde, *Phys. Rev. Lett.* **94**, 018104 (2005).
- ¹⁶R. J. Allen, D. Frenkel, and P. R. ten Wolde, *J. Chem. Phys.* **124**, 024102 (2006).
- ¹⁷L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *J. Chem. Phys.* **125**, 024106 (2006).
- ¹⁸T. S. van Erp, *Phys. Rev. Lett.* **98**, 268301 (2007).
- ¹⁹P. Tiwary and M. Parrinello, *Phys. Rev. Lett.* **111**, 230602 (2013).
- ²⁰H. Jung, K. Okazaki, and G. Hummer, *J. Chem. Phys.* **147**, 152716 (2017).
- ²¹P. G. Bolhuis and C. Dellago, "Trajectory-based rare event simulations," in *Reviews in Computational Chemistry* (Wiley-Blackwell, 2010), Chap. 3, pp. 111–210.
- ²²In this work, a collective variable is a quantity that can be calculated from the configuration space coordinates of the system. An order parameter is a collective variable that can distinguish between states A and B.
- ²³R. J. Allen, C. Valeriani, and P. R. ten Wolde, *J. Phys.: Condens. Matter* **21**, 463102 (2009).
- ²⁴F. A. Escobedo, E. E. Borrero, and J. C. Araque, *J. Phys.: Condens. Matter* **21**, 333101 (2009).
- ²⁵T. S. van Erp, *J. Chem. Phys.* **125**, 174106 (2006).
- ²⁶T. S. van Erp, "Dynamical rare event simulation techniques for equilibrium and nonequilibrium systems," in *Kinetics and Thermodynamics of Multistep Nucleation and Self Assembly in Nanoscale Materials* (Wiley-Blackwell, 2012), Chap. 2, pp. 27–60.
- ²⁷J. Juraszek and P. G. Bolhuis, *Biophys. J.* **95**, 4246 (2008).
- ²⁸C. Velez-Vega, E. E. Borrero, and F. A. Escobedo, *J. Chem. Phys.* **130**, 225101 (2009).
- ²⁹G. C. Sosso, T. Li, D. Donadio, G. A. Tribello, and A. Michaelides, *J. Phys. Chem. Lett.* **7**, 2350 (2016).
- ³⁰E. E. Borrero and F. A. Escobedo, *J. Chem. Phys.* **127**, 164101 (2007).
- ³¹J. Russo and H. Tanaka, *J. Chem. Phys.* **145**, 211801 (2016).
- ³²M. A. Rohrdanz, W. Zheng, and C. Clementi, *Annu. Rev. Phys. Chem.* **64**, 295 (2013).
- ³³S. B. Kim, C. J. Dsilva, I. G. Kevrekidis, and P. G. Debenedetti, *J. Chem. Phys.* **142**, 085101 (2015).
- ³⁴T. S. Van Erp and P. G. Bolhuis, *J. Comput. Phys.* **205**, 157 (2005).
- ³⁵A. Vijaykumar, P. R. ten Wolde, and P. G. Bolhuis, *J. Chem. Phys.* **148**, 124109 (2018).
- ³⁶J. Lin, *IEEE Trans. Inf. Theory* **37**, 145 (1991).
- ³⁷C. Leitold, W. Lechner, and C. Dellago, *J. Phys.: Condens. Matter* **27**, 194126 (2015).
- ³⁸R. B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6732 (2005).
- ³⁹A. Ma and A. R. Dinner, *J. Phys. Chem. B* **109**, 6769 (2005).
- ⁴⁰B. Peters and B. L. Trout, *J. Chem. Phys.* **125**, 054108 (2006).
- ⁴¹A. Singer, R. Erban, I. G. Kevrekidis, and R. R. Coifman, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 16090 (2009).
- ⁴²W. Lechner, J. Rogal, J. Juraszek, B. Ensing, and P. G. Bolhuis, *J. Chem. Phys.* **133**, 174110 (2010).
- ⁴³W. Li and A. Ma, *Mol. Simul.* **40**, 784 (2014).
- ⁴⁴B. Peters, *Annu. Rev. Phys. Chem.* **67**, 669 (2016).
- ⁴⁵D. Mendels, G. Piccini, and M. Parrinello, *J. Phys. Chem. Lett.* **9**, 2776 (2018).
- ⁴⁶P. G. Bolhuis, C. Dellago, and D. Chandler, *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5877 (2000).
- ⁴⁷A. Warmflash, P. Bhimalapuram, and A. R. Dinner, *J. Chem. Phys.* **127**, 154112 (2007).
- ⁴⁸A. Dickson, A. Warmflash, and A. R. Dinner, *J. Chem. Phys.* **130**, 074104 (2009).
- ⁴⁹A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562 (2002).
- ⁵⁰J. B. Tenenbaum, V. De Silva, and J. C. Langford, *Science* **290**, 2319 (2000).
- ⁵¹P. Das, M. Moll, H. Stamati, L. E. Kaviraki, and C. Clementi, *Proc. Natl. Acad. Sci. USA* **103**, 9885 (2006).
- ⁵²R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker, *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7426 (2005).
- ⁵³A. L. Ferguson, A. Z. Panagiotopoulos, I. G. Kevrekidis, and P. G. Debenedetti, *Chem. Phys. Lett.* **509**, 1 (2011).
- ⁵⁴E. Chiavazzo, R. Covino, R. R. Coifman, C. W. Gear, A. S. Georgiou, G. Hummer, and I. G. Kevrekidis, *Proc. Natl. Acad. Sci. U. S. A.* **114**, E5494 (2017).
- ⁵⁵W. Chen and A. L. Ferguson, *J. Comput. Chem.* **39**, 2079 (2018).
- ⁵⁶J. M. L. Ribeiro, P. Bravo, Y. Wang, and P. Tiwary, *J. Chem. Phys.* **149**, 072301 (2018).
- ⁵⁷M. M. Sultan and V. S. Pande, *J. Chem. Phys.* **149**, 094106 (2018).
- ⁵⁸C. Wehmeyer and F. Noé, *J. Chem. Phys.* **148**, 241703 (2018).
- ⁵⁹E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **131**, 044120 (2009).
- ⁶⁰E. E. Borrero, M. Weinwurm, and C. Dellago, *J. Chem. Phys.* **134**, 244118 (2011).