Enhanced Sampling of Nonequilibrium Steady States

Alex Dickson and Aaron R. Dinner

Department of Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637; email: dinner@uchicago.edu

Annu. Rev. Phys. Chem. 2010. 61:441–59

The *Annual Review of Physical Chemistry* is online at physchem.annualreviews.org

This article's doi: 10.1146/annurev.physchem.012809.103433

0066-426X/10/0505-0441$20.00

Key Words

nonequilibrium umbrella sampling, forward flux sampling, transition rates, path sampling, string method

Abstract

We review recent progress in methods for accelerating the convergence of simulations of nonequilibrium systems, specifically nonequilibrium umbrella sampling (NEUS) and forward flux sampling (FFS). These methods account for statistics of dynamical paths between interfaces to enforce sampling of low probability regions of phase space for computing steady-state averages, including transition rates, for systems driven arbitrarily far from equilibrium. Recent advances in NEUS allow for efficient sampling of complex systems by focusing sampling in the vicinity of a one-dimensional manifold (string) that connects regions of interest in phase space; this procedure can be extended to the case of two strings that describe the forward and backward transition ensembles separately, which is useful, as they do not, in general, coincide. We recast FFS in the framework of NEUS to facilitate comparison of the two methods. We conclude by discussing selected applications of interest.

1. INTRODUCTION

Many systems of great experimental interest in the molecular sciences consume and dissipate energy through separate channels and are thus far from equilibrium. These include, but are not limited to, synthetic (1) and natural (2–4) molecular machines, polymers (5, 6) and colloids (7–12) under shear, and regulatory modules of living cells (13, 14). Recently, there has been renewal and dramatic expansion of the study of such systems owing to advances in both theory and experiment: in theory, the development of a statistical mechanics of trajectories (space-time objects) (15, 16), and, in experiment, the development of improved methods for detecting stochastic fluctuations in single molecules (17–20). In particular, these two strands of research have given rise to the appreciation that systems with well-defined steady states are constrained by certain symmetry relations (21). Simulations are playing an important complementary role to these advances by enabling the validation of theories and interpretation of experiments on complex systems of fundamental and practical interest.

A number of robust, well-established algorithms exist for simulating the dynamics of microscopically irreversible systems (see 22–25), and the choice between them depends on the process of interest. However, just as in simulations of complex microscopically reversible systems, relaxation to the steady state can often be prohibitively slow, especially when there are multiple (meta)stable states with bottlenecks between them in phase space. This problem can be overcome by methods that enhance the sampling of low probability states and reaction pathways but still enable the retrieval of the steady-state distribution associated with the original dynamics. Indeed, such methods are essential for obtaining information about dynamical bottlenecks (transition states), which are of central importance to elucidating mechanisms. However, almost all such existing algorithms rely on detailed balance (microscopic reversibility) and a priori knowledge of the distribution function (typically, Boltzmann weighting) (22, 23). By definition, these do not hold in systems in nonequilibrium steady states and limit cycles. New simulation paradigms are just now emerging to treat this important class of systems.

Although several algorithms for the enhanced sampling of nonequilibrium systems were suggested over the past decade, most assumed either detailed balance (26–28) (making them fundamentally flawed; see discussions in 29, 30) or a specific dynamics (31, 32) (making them useful for some problems but not others). However, the advances in theory mentioned above led to practical methods based on the statistics of trajectories for microscopically reversible systems (33–36) and in turn for microscopically irreversible ones. The latter include two general and powerful methods: nonequilibrium umbrella sampling (NEUS) (29, 30, 37) and forward flux sampling (FFS) (38–40). Both can be viewed as forms of umbrella sampling, in which a space of collective variables that characterize the system (order parameters) is covered (hence the name) by separate simulations to ensure uniform sampling, and then the results are combined with physical weighting. In NEUS, the sampling of different regions of phase space is conducted in parallel, with little communication between the regions. In FFS, the regions are treated serially, which has advantages and disadvantages.

Below we discuss these two methods in detail. To this end, we describe a common theoretical framework that can be generalized to any enhanced sampling method that works by simulating a process in a piecewise fashion (Section 2). In Section 3, we review the basic NEUS algorithm, as well as its extension to study transitions in spaces of many order parameters. Recent work that allows for the calculation of transition rates using NEUS is also reviewed. In Section 5, we show how FFS can be viewed as a NEUS simulation with overlapping regions. In Section 6, we exploit the common framework to compare the two methods, and discuss how the differences between them are likely to affect their ability to treat more demanding systems. We

close by discussing some applications and outstanding problems of interest in nonequilibrium systems.

2. THEORETICAL FRAMEWORK

Although a space-time analog of the free energy can be defined and calculated (15, 16, 41, 42), the steady-state probability distribution as a function of experimentally accessible order parameters provides the most direct connection to observables. In this sense, the steady-state probability distribution is the central quantity of interest for nonequilibrium systems. Formally, it can be defined as

$$P(\boldsymbol{z}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \delta(\theta[\boldsymbol{x}(t)] - \boldsymbol{z}) dt, \tag{1}$$

where θ is a mapping from each point in the full phase space, denoted by $\boldsymbol{x}$, to a point in the space of order parameters, and $\boldsymbol{z}$ is the point in the space of order parameters at which we want to evaluate the probability. In this review, we restrict our attention to systems for which the limit in Equation 1 exists and is independent of the choice of initial condition, $\boldsymbol{x}(0)$. The steady-state distribution is normalized such that

$$\int_{\mathbb{R}^m} P(\boldsymbol{z}) d\boldsymbol{z} = 1, \tag{2}$$

where $\mathbb{R}^m$ is the space formed by the m order parameters used to describe the system.

In practice, numerically estimating the steady-state distribution using Equation 1 can be prohibitive owing to trapping in phase space. To overcome this problem, one can break the space accessible to a process into smaller regions and enforce their even sampling. This way, instead of waiting for rare events to occur during a simulation, each stage of the event is simulated separately and the results are recombined. This can provide exponential gains in efficiency compared with conventional simulations. The key insight that allowed this strategy to be applied to microscopically irreversible systems was that one can use the statistics of trajectories to account properly for the flux into each region.

Let us consider a set of regions $\{\Omega_i\}$ that span $\mathbb{R}^m$, and are nonoverlapping in space. The special case of overlapping regions can be treated by creating an extended phase space in which the regions are nonoverlapping (as discussed below). The goal is to run trajectory segments that are restricted to each region and, in so doing, to obtain a steady-state distribution for each region:

$$P_{\Omega_i}(\boldsymbol{z}) = \lim_{T\to\infty} \frac{\int_0^T \delta(\theta[\boldsymbol{x}(t)] - \boldsymbol{z}) h_i(\theta[\boldsymbol{x}(t)]) dt}{\int_0^T h_i(\theta[\boldsymbol{x}(t)]) dt}, \tag{3}$$

where $h_i(\theta[\boldsymbol{x}(t)])$ is an indicator function equal to 1 if $\theta[\boldsymbol{x}(t)] \in \Omega_i$ and 0 otherwise. If we can devise a way to simulate the unbiased dynamics of a trajectory while confining it to the region Ω_i, Equation 3 reduces to

$$P_{\Omega_i}(\boldsymbol{z}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \delta(\theta[\boldsymbol{x}^{C_i}(t)] - \boldsymbol{z}) dt, \tag{4}$$

where $\boldsymbol{x}^{C_i}(t)$ is a point on a (discontinuous) trajectory that is confined to region Ω_i, but otherwise evolves according to unbiased dynamics. Given the set of regional distributions P_{Ω_i}, the complete steady-state distribution is reconstructed as

$$P(\boldsymbol{z}) = \sum_i P_{\Omega_i}(\boldsymbol{z}) \left[\lim_{T\to\infty} \frac{1}{T} \int_0^T h_i(\theta[\boldsymbol{x}(t)]) dt \right] = \sum_i P_{\Omega_i}(\boldsymbol{z}) W_i, \tag{5}$$

where we call W_i the weight of region Ω_i, and the property $\sum_i W_i = 1$ holds by definition.

The task of a piecewise sampling method is to compute these weights and regional steady-state distributions. To this end, such methods seek to conduct sampling within the regions in a manner that is consistent with the underlying dynamics, which, for irreversible systems, requires properly accounting for the incoming fluxes. To illustrate this point, we consider the simulation of a system in equilibrium. If a move is made that results in a walker leaving its region, the move can simply be rejected, the momenta reversed in sign, and the trajectory continued. This is an explicit implementation of the detailed balance condition, which guarantees that each exit point from a region is also an entry point with equal probability. In systems that are out of equilibrium, this condition no longer holds. For example, in a region in two-dimensional space that has a strong flow imposed from left to right, all the entry points will be on the left boundary, and all exit points will be on the right boundary. To conduct sampling in a physical manner, one must, upon an attempted exit through the right boundary, restart the walker on the left boundary. In general, to sample nonequilbrium systems in a physical manner, one must restart walkers according to their physically weighted flux input function.

3. NONEQUILIBRIUM UMBRELLA SAMPLING

The original motivation for developing NEUS (29) was to calculate higher-order correlation functions in discrete stochastic models of reaction networks (43, 44); such averages can have significant contributions from the tails of (monomodal) multivariate steady-state distributions of molecular copy numbers. Although FFS existed, it was poorly suited to this problem as it was designed for obtaining rates of transitions between stable states and limited to one order parameter. NEUS by contrast was designed to enable the restriction of copies of the system to arbitrary regions in the space of interest. The freedom to choose arbitrary regions in turn led to NEUS algorithms for efficient sampling of paths in spaces of many order parameters (30) and, in turn, separately focused sampling of forward and reverse pathways (37).

In equilibrium umbrella sampling simulations, the space is divided into a number of arbitrary regions, and separate copies of the system (i.e., walkers) are confined to those regions using a bias potential, which is often harmonic but need not be (22, 23, 45, 46). This approach is not directly applicable to nonequilibrium systems owing to their lack of detailed balance. Warmflash et al. (29) demonstrated that umbrella sampling simulations can be conducted in nonequilibrium systems if one explicitly accounts for the fluxes between regions by reinitializing walkers according to lists of entering phase-space points (the flux input functions) that are determined during the course of the simulation. This aspect of the algorithm is similar in spirit to a list-based procedure introduced by de Oliveira and Dickman for studying a quasi-stationary state in a nonequilibrium reaction-diffusion system with an absorbing phase (47, 48). Information on how the flux distributions for NEUS are computed can be found in previous work (29, 30, 37).

In the original algorithm, the weights of the regions used in Equation 5 are determined as follows. Each time a walker attempts to leave its region, an incremental amount of weight is transferred from its region (Ω_i) to the region to which it attempted to go (Ω_j):

$$-\Delta W_i = \Delta W_j = s\, W_i, \tag{6}$$

where Δ denotes an additive change, and s is a user-defined parameter chosen to optimize convergence. A slight modification of this equation is required if the simulations in each region are run asynchronously (see 29). This weight transfer scheme is motivated by considering a region Ω_i at steady state. For small s, the change in the weight of region Ω_i over a long time interval T is

$$\Delta W_i^s = s\,T \sum_j W_j^s \phi_{j\to i}^s - s\,T \sum_j W_i^s \phi_{i\to j}^s = s\,T\,(\text{Flux in} - \text{Flux out}) = 0, \tag{7}$$

where W_i^s is the weight of region Ω_i at steady state, and $\phi_{j\to i}^s$ is the steady-state (time-averaged) flux from region Ω_j to region Ω_i, equal to the number of trajectories leaving Ω_j for Ω_i per unit time. Importantly, the final equality results from the conservation of probability at steady state, and not a detailed balance criterion. Equation 7 shows that the correct steady-state values of the weights are a fixed point of the mapping in Equation 6. In Appendix A we make this argument more precise for Markovian systems; we show that this fixed point is the only one and that it is attractive. In addition to this local weight update scheme, two global weight update schemes have been proposed that use region-to-region transition statistics to adjust the weights of all regions simultaneously (37, 49). These both use flux balance conditions: Vanden-Eijnden & Venturoli (49) balance the total flux in and out of each region, and Dickson et al. (37) adjust the weights of the regions using the continuity of flux across each interface. Dickson et al. showed that their global weight balancing algorithm outperforms the local weight balancing algorithm early in a simulation but ultimately converges more slowly to the steady state.

3.1. Sampling Along a String

If the regions used for sampling are even discretizations of the order-parameter space (e.g., line segments in one dimension, rectangles in two dimensions, and so forth for higher dimensions), then the computational expenditure is roughly proportional to the volume of the order-parameter space, which scales exponentially with its dimension. Although it is often possible to capture the essential features of the dynamics of complex systems with only a few collective variables, their choice is often not obvious. A poor choice can actually slow the convergence of an umbrella sampling calculation because the remaining degrees of freedom can then be very slow to relax. This presents a dilemma: A many-dimensional order-parameter space is desirable to capture the essential dynamics of a transition, but it is too computationally expensive to cover this space evenly with regions.

One strategy is to use statistical approaches to determine a small number of collective variables as functions of the physical variables of the system, using estimates for the probabilities that trajectories commit to a given stable state for different starting configurations (50–54). Alternatively, because the formalism of NEUS introduced above allows for arbitrary regions, we can adapt ideas developed for the finite-temperature string method (36, 55–61) and use regions that are arranged along a one-dimensional string that winds its way through a many-dimensional space. The computational cost of the calculation scales linearly with the number of regions along the string and is now independent of the dimensionality of the order-parameter space. Although this strategy requires further analysis to ultimately identify dynamically meaningful physical variables, it is preferable from the perspective of sampling in that all the quantities used for optimization of the path are local in nature (i.e., are determined by short trajectory segments).

The idea is to define the regions in terms of the set of points (images) in order-parameter space that uniformly discretize the string (30). To this end, each point in the order-parameter space is associated with the closest image (according to a suitable distance metric, as discussed below). This partitioning of the space corresponds to tiling it with regions that are Voronoi polyhedra, each of which contains a walker. The string is updated periodically during the simulation by moving each image toward the average position of its walker over the last time interval. The individual image movements are typically complemented by a smoothing procedure that reduces the local curvature of the string, and a reparameterization step that spaces the images equally in arc length to ensure even sampling along the path. More information on these procedures can be found elsewhere (30, 37, 60). The string is updated until it stops changing to within the desired level of

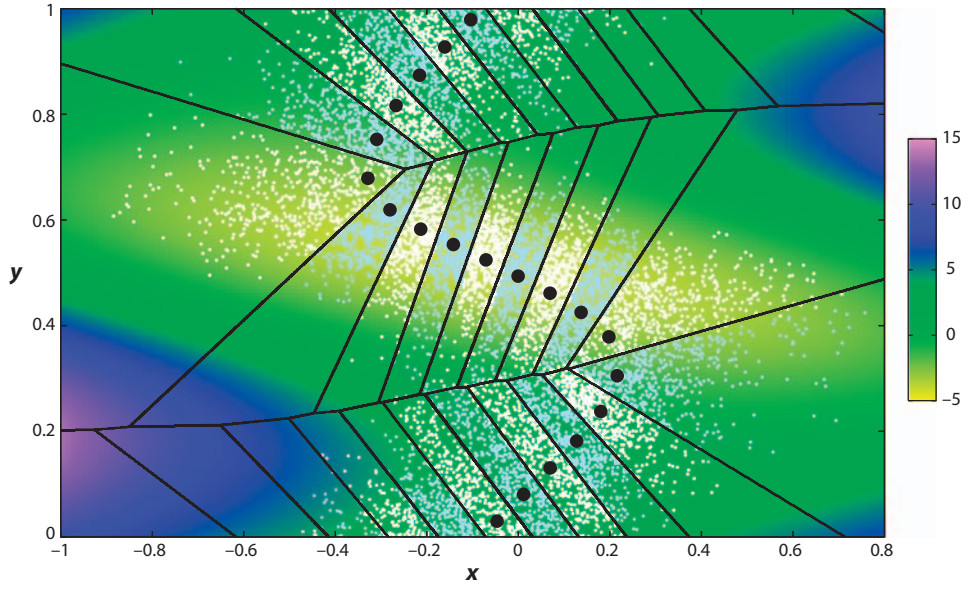

Figure 1

An example of a converged string obtained by nonequilibrium umbrella sampling. The system is defined by a two-dimensional potential surface indicated by the color map (the color scale bar is in units of kT). A constant upward force, in combination with a periodic y boundary, pushes the system out of equilibrium. A path of 20 images and its Voronoi polyhedra are superimposed. The images are shown as black dots. For each region, a random sample of 500 points is shown in white or light blue, alternating for clarity. Figure adapted from Reference 30.

precision. For example, **Figure 1** shows an example of a converged string, its images, and their Voronoi polyhedra for a two-dimensional potential.

Although these polyhedra are most easily visualized in two dimensions, their main advantage is that they are generalizable to higher-dimensional order-parameter spaces as well. **Figure 2** shows a string in a 64-dimensional order-parameter space for an Ising model under shear, in which the order parameters are defined by the average occupancy of spins inside each coarse-grained box. Indeed, the Voronoi-based approach is applicable to any system in which an appropriate distance metric can be found. In the special case in which curvilinear coordinates are used as order parameters, the distance in order-parameter space can be transformed into an appropriate Cartesian distance.

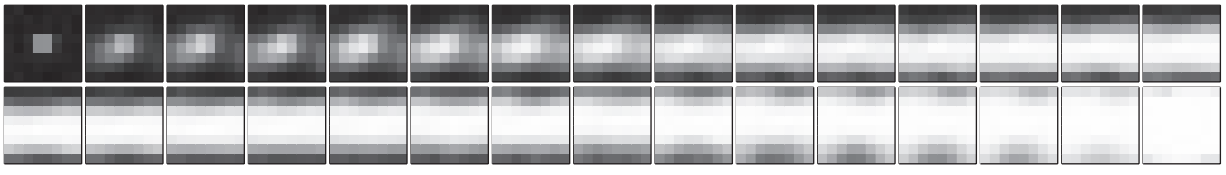

Figure 2

A string obtained by nonequilibrium umbrella sampling for the Ising model under shear transitioning from one stable state to another (30). The frames show configurations in the order-parameter space (images), which define the set of Voronoi sampling regions that confine the walkers. The Ising system here is composed of 16 × 16 spins, and the order parameters are defined as the average value of the spins inside a square box. In this case, there are 8 × 8 boxes; each box contains four spins. The string then winds through a 64-dimensional order-parameter space. Figure taken from Reference 30.

Methods to perform this operation have been described for equilibrium systems (57, 60), and although a similar treatment for nonequilibrium systems would be necessary to maintain the dynamical meaning of the path [for instance, as a path of maximum likelihood (32)], it remains to be shown whether such a treatment is necessary in general. Relatedly, one could determine the limit cycle (a time-ordered sequence of probability distributions) of systems subject to time-dependent external forces by extending the distance metric to include time (37). This possibility highlights the general fact that, when combining different types of variables, one must choose a weighting for their contributions to the distance metric. Nevertheless, given such a weighting, boundary crossings can be detected simply by direct evaluation of the distance metric, making the Voronoi-based approach generally applicable.

As discussed in more detail elsewhere (30, 55, 57), the efficacy of string-based methods relies on the ability of the order parameters to describe the transition sufficiently well and, relatedly, the most probable transition paths being largely confined to a tube in the space of order parameters. Because systems out of equilibrium do not obey detailed balance, the ensemble of transition paths connecting two regions A and B can comprise separate forward (from A to B) and backward (from B to A) pathways that do not necessarily overlap. For this reason, it can be useful to define two strings: one using statistics from forward trajectories and the other using statistics from backward trajectories. This can be achieved in a straightforward fashion by performing a NEUS simulation in an extended phase space (37). If we let m be the number of order parameters that compose the space, and let trajectories have the additional property that $x_{m+1}(t) = 0$ if the trajectory last visited A, and $x_{m+1}(t) = 1$ if the trajectory last visited B, then trajectories can be defined in the extended phase space $\mathbb{R}^{m+} = \mathbb{R}^m \cup \{0, 1\}$. To employ the Voronoi-based procedure, we define the distance between two points $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^{m+}$ as

$$\|\boldsymbol{x} - \boldsymbol{y}\| = \begin{cases} \infty & \text{if } x_{m+1} \neq y_{m+1} \\ \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} & \text{otherwise} \end{cases} \tag{8}$$

such that points that originated from different basins are infinitely far apart. In this way, the regions are defined such that they are wholly in either $\mathbb{R}^m \cup 0$ or $\mathbb{R}^m \cup 1$, and two overlapping sets of regions in $\mathbb{R}^m$ can be seen as a single set of nonoverlapping regions in $\mathbb{R}^{m+}$. Boundary crossings occur in the same way as before, and no modifications to the algorithm itself are needed. Below, we show how rate constants for the forward and reverse transitions can be calculated using this approach.

This technique was applied to a system of circadian oscillations in which the two strings describe a cyclic pathway in a 22-dimensional order-parameter space. The order parameters in this system are defined by the copy number of each species in the reaction network that gives rise to the oscillations (37, 62, 63). Circadian oscillations provide an excellent test case for NEUS because there is a well-defined limit cycle in a large order-parameter space. Given an arbitrary initial path, NEUS is able to find the limit cycle that is formed by connecting together the forward and backward strings (see **Figure 3*a***). **Figure 3*b*** shows the extent of sampling in each region once the string has converged. There is significant overlap between the sets of points from different regions because the image shows only a two-dimensional projection of the pathway. An unconstrained trajectory is shown in **Figure 3*c*** for comparison.

4. RATE CALCULATIONS

A common goal of many enhanced sampling methods is to describe rare transitions in phase space between two stable basins (33, 34, 36, 38, 64). It is possible to easily obtain the rate of transition between two basins by separating trajectories based on their basin of origin. This has always been

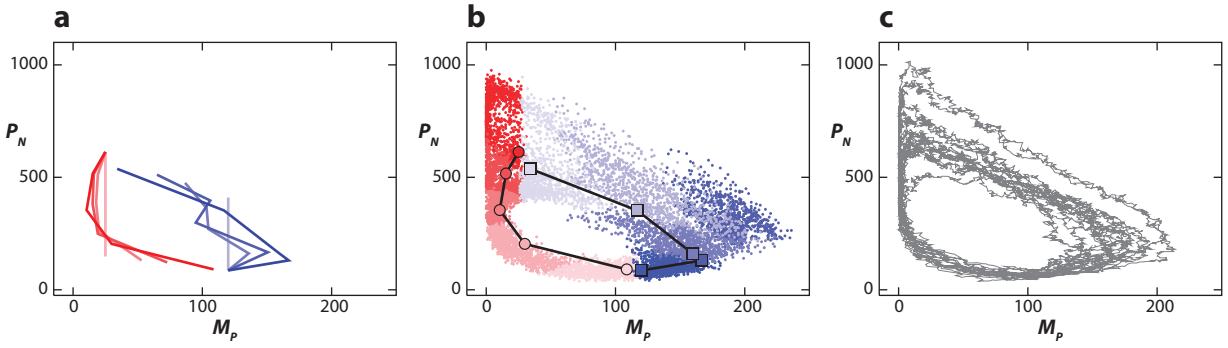

Figure 3

Sampling of circadian oscillations using nonequilibrium umbrella sampling. (*a*) The convergence of the forward (*red*) and backward (*blue*) strings from their initial configuration (*lightest*) to their final configuration (*darkest*). Because there are no stable basins in this system, the first image of each string was held fixed (top-most point in the forward string, and the bottom-most point in the backward string) to prevent the collapse of the string. Here, basin A is defined as the region where $M_P < 25$ and basin B is defined as the region where $M_P > 120$. (*b*) A projection of a random sampling of 1000 points obtained for each region. The strings used in this panel are the final strings from panel *a*. Each image in the string is marked by a symbol that is colored the same as the points sampled in the corresponding region. (*c*) An unconstrained trajectory obtained using a straightforward application of the Gillespie algorithm (96). All projections shown agree with previous results (62), and more information on the circadian model can be found elsewhere (37, 62). Panels *a* and *c* adapted from Reference 37.

at the heart of FFS rate calculations (38–40, 65), and a practical method for obtaining rates for NEUS was recently introduced (37, 49). Here we describe the theoretical underpinnings. We consider two basins A and B and define two sets of regions, $\mathcal{S}_A = \{\Omega_i^A\}$ and $\mathcal{S}_B = \{\Omega_i^B\}$, where a trajectory is in $\mathcal{S}_A$ if the trajectory was most recently in basin A, and a trajectory is in $\mathcal{S}_B$ if it was most recently in basin B. We note that $\mathcal{S}_A = \mathbb{R}^m \cup 0$ and $\mathcal{S}_B = \mathbb{R}^m \cup 1$ in the notation of the previous section. By definition, the rate constant for the transition from A to B is

$$k_{AB} = \frac{\overline{\Phi}_{B|\mathcal{S}_A}}{P_{\mathcal{S}_A}}, \tag{9}$$

where $\bar{\Phi}_{B|\mathcal{S}_A}$ is the time-averaged flux from $\mathcal{S}_A$ into the B basin and

$$P_{\mathcal{S}_A} = \sum_{i \in \mathcal{S}_A} W_i \tag{10}$$

is the probability of being in $\mathcal{S}_A$. In piecewise simulations, the total flux into B from all regions in $\mathcal{S}_A$ can be obtained from

$$\overline{\Phi}_{B|\mathcal{S}_A} = \sum_{j \in \mathcal{S}_A} \frac{W_j N_{jB}}{T_j}, \tag{11}$$

where N_{jB} is the number of transitions from region j to basin B over a given sampling period, and T_j is the time elapsed in region j during that sampling period. This method to calculate the rate is common to all piecewise sampling methods that separate trajectories based on their basin of origin, including NEUS and FFS, as shown below.

5. FORWARD FLUX SAMPLING

FFS was originally introduced to compute unidirectional rates for nonequilibrium processes [i.e., to calculate k_{AB} using a simulation in $\mathcal{S}_A$ (38)]. The method is an extension of interface-based

methods introduced for equilibrium systems such as milestoning (64, 66, 67) and (partial-path) transition interface sampling (34, 68, 69), but it is applicable to nonequilibrium processes as it does not make any assumptions about the distribution of trajectories on the interfaces, or assume detailed balance in any way. It is most similar to the weighted ensemble method (70, 71), which is much more broadly applicable than the reversible Brownian dynamics with which it was first employed. The idea is to calculate the rate (Equation 9) through a factorization of $\bar{\Phi}_{B|\mathcal{S}_A}$:

$$k_{AB} = \frac{\overline{\Phi}_{B|\mathcal{S}_A}}{P_{\mathcal{S}_A}} = \frac{\overline{\Phi}_{0|\mathcal{S}_A}}{P_{\mathcal{S}_A}} P(\lambda_B|\lambda_0) = \frac{\overline{\Phi}_{0|\mathcal{S}_A}}{P_{\mathcal{S}_A}} \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i), \tag{12}$$

where $\{\lambda_i\}$ is a set of nonintersecting interfaces between basins A and B, $\lambda_0 = \lambda_A$ is the boundary of basin A, and $\lambda_n = \lambda_B$ is the boundary of basin B. $\bar{\Phi}_{0|\mathcal{S}_A}$ is the flux across λ_0 in $\mathcal{S}_A$ in the forward direction, and $P(\lambda_{i+1}|\lambda_i)$ is the probability that a trajectory in $\mathcal{S}_A$ that crosses λ_i for the first time will subsequently cross λ_{i+1} before returning to λ_0. Since its original introduction (38), the algorithm has been modified and extended in a number of ways (39, 40, 53, 65, 72, 73). Of most relevance to this review, Valeriani et al. (65) showed that steady-state distributions can be obtained by combining the results of separate forward and backward FFS simulations.

To more clearly compare FFS with NEUS, we now recast FFS in the umbrella sampling framework. Below, we show that using a set of interfaces is equivalent to a piecewise simulation using a series of overlapping regions, and that Valeriani et al.'s (65) procedure, in effect, calculates the weight and flux input function for each region. The theory presented here is equivalent to that presented by Valeriani et al. (65), except that we relax the assumption that there are only two basins (i.e., we show how to allow for the presence of intermediates). To this end, we consider an FFS simulation in $\mathcal{S}_A$. The regions are defined using the series of interfaces $\lambda_0, \lambda_1, \ldots, \lambda_n$ as follows. If we associate with each trajectory a variable $\nu(t)$ that records the index of the last interface reached by the trajectory, we can define an extended space similar to that introduced in Section 3.1, by adding the maximum value of ν on the interval $[0, t]$ ($\nu_{\max}(t)$), to the order-parameter space. In this way, a trajectory is in region i at time t if $\nu_{\max}(t) = i$. An illustration of these regions is provided in **Figure 4**. The regions overlap in the original order parameter, so the boundary crossings are determined completely by $\nu_{\max}(t)$.

With the regions so defined, the flux entry points for each region all come from the same direction, and simulations can be performed in a sequential fashion. First, an unrestricted simulation in $\mathcal{S}_A$ is run to obtain M_0 starting points on λ_0, by registering outward crossings of λ_0. In this way, the quantity $\bar{\Phi}_{0|\mathcal{S}_A}/P_{\mathcal{S}_A}$ is obtained as M_0/T, where T is the total amount of time it took to register the M_0 points. The crossing points are then used as the flux input function for the walker that is restricted to region 0. If the walker attempts to exit region 0 by either re-entering A or reaching the next interface (λ_1), then it is reinitialized at one of the M_0 crossing points, randomly chosen with equal probability. The points of attempts to move to the next region are saved and used to form the flux input function for the next region. New trajectories continue to be initialized until the transition probability $P(\lambda_1|\lambda_0)$ has been determined to the desired accuracy. This process is repeated for each region, in sequence, up to λ_{n-1}. Importantly, to exit its region, a walker must either reach the next interface in the sequence, or return all the way to the first interface, λ_0. To obtain the full steady-state distribution, a symmetric set of simulations is conducted in $\mathcal{S}_B$.

The flux of trajectories entering region i through interface λ_i is given by $\overline{\Phi}_{0|A} P(\lambda_i|\lambda_0)$ (for regions in $\mathcal{S}_A$). The probability of an unconstrained trajectory being found in region Ω_i (i.e., W_i)

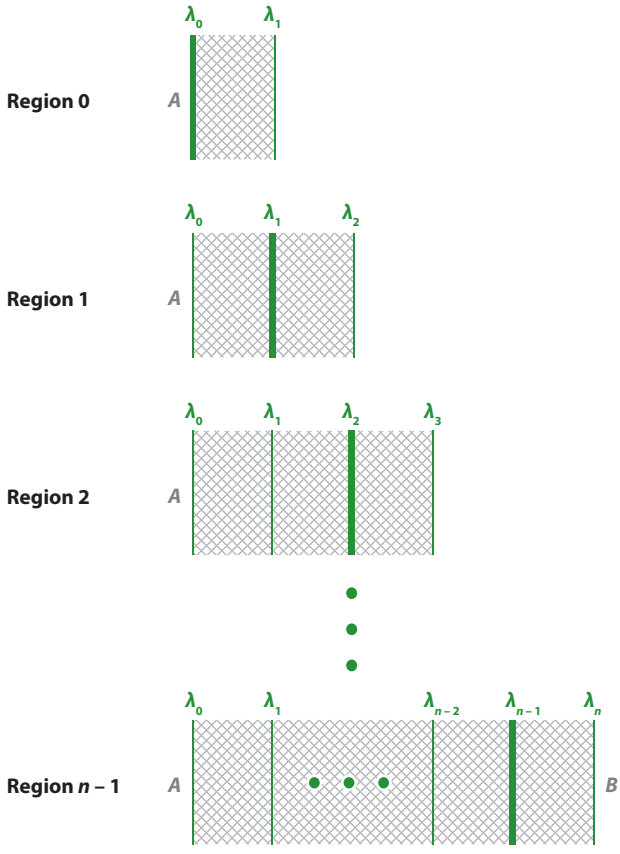

Figure 4

Recasting forward flux sampling. The effective nonequilibrium umbrella sampling regions formed by sequentially traversing the set of interfaces $\{\lambda_i\}$ are shown schematically. For a trajectory to be in region i, the interface i must be the interface with the highest index crossed by the trajectory (i.e., the trajectory must have crossed interface i, but not interface $i+1$). For each region, the interface with the highest index crossed by the trajectory is shown in bold. These regions contain only trajectories that last originated in basin A. A similar set of regions, with trajectories that last originated in basin B, is also needed to calculate the steady-state distribution.

is then

$$\overline{\Phi}_{0|A} P(\lambda_i|\lambda_0)\overline{\Delta T}_i, \tag{13}$$

where $\overline{\Delta T}_i$ is the average time length of a trajectory in region Ω_i. This gives the weights of all regions in $\mathcal{S}_A$, and the weights in $\mathcal{S}_B$ are similarly determined. To obtain $P_{\mathcal{S}_A}$ and $P_{\mathcal{S}_B}$, which are required to calculate Equation 13, we note that the flux from $\mathcal{S}_A$ to $\mathcal{S}_B$ must equal the flux from $\mathcal{S}_B$ to $\mathcal{S}_A$ ($\overline{\Phi}_{B|\mathcal{S}_A} = \overline{\Phi}_{A|\mathcal{S}_B}$) at steady state, even in the absence of detailed balance; it follows that

$$P_{\mathcal{S}_A} k_{AB} = P_{\mathcal{S}_B} k_{BA}. \tag{14}$$

Furthermore,

$$P_{\mathcal{S}_B} + P_{\mathcal{S}_A} = 1 \tag{15}$$

by the definition of $\mathcal{S}_A$ and $\mathcal{S}_B$. Equations 14 and 15 together determine the probabilities using the rates calculated by Equation 12. Importantly, Equation 14 is more general than Valeriani et al.'s

(65) equation 9:

$$p_A k_{AB} = p_B k_{BA}, \tag{16}$$

where p_A and p_B are the probabilities of being in basins A and B, respectively. Equation 16 is valid only for the limiting case of a two-state system where $p_A + p_B = 1$, in which case A and $\mathcal{S}_A$ are identical. Once the weights of each region are obtained by Equation 13, the steady-state distribution on $\mathbb{R}^m \backslash (A \cup B)$ can be calculated using Equation 5. We show in Appendix B that using conditional probabilities to calculate the rate, as in Equation 12, is equivalent to using the weights of the FFS regions (Equation 13) in the expression for the rate given in Equation 11.

6. COMPARING NONEQUILIBRIUM UMBRELLA SAMPLING AND FORWARD FLUX SAMPLING

With NEUS and FFS both expressed in the umbrella sampling formalism, it is easy to see the similarities and differences between the two methods. Both can be used to calculate rates and steady-state distributions for nonequilibrium processes using separate simulations in $\mathcal{S}_A$ and $\mathcal{S}_B$. In both methods, an order-parameter space is discretized into regions, and restricted simulations are performed in each region. The weight of each region is determined to calculate the full steady-state distribution, and proper sampling is ensured by reinitializing trajectories upon attempted exits according to a measured distribution of flux entry points. From a practical standpoint, both methods can be used with large-scale distributed computing architectures because they are both based on the integration of many (usually on the order of tens of thousands) short trajectory segments. In other words, they are almost trivially parallelizable in that each trajectory segment can be sent to a different processor.

The differences between the two algorithms stem from the nature of the regions used. Because the conditional probabilities in the factorization in Equation 12 are only meaningfully defined for nonintersecting interfaces, FFS is intrinsically limited to a single order parameter (which of course can be a complicated function of other variables). In cases in which a single order parameter that describes the transition well can be identified, the simplicity of FFS can be advantageous, especially if one is only interested in one direction of a transition. There is a minimum of overhead associated with the algorithm, and, once the flux input function for a region is known, simulations in it can be run independently. Of course, this requires convergence of the conditional probability for transition in the previous region, and there is no way to remedy incomplete sampling of the flux input functions at earlier interfaces (30, 74, 75). As is shown in **Figure 4**, the regions used in FFS also grow larger with increasing index. In the presence of intermediates, trajectories can go backward toward λ_0 and get caught in local minima. In these cases, long trajectory segments must be integrated before the trajectory exits the region, which results in inefficient sampling (37).

In NEUS, the weights and flux input functions are improved iteratively. Although this aspect of the algorithm contributes to the computational cost, it also enables the use of arbitrary regions and dynamic update of the sampling. In particular, as discussed in Section 3.1, given a large set of candidate physical variables that describe a transition of interest and a corresponding distance metric, one can optimize strings that focus sampling around the forward and backward transition path ensembles. This feature is expected to be advantageous in complex systems, where reaction coordinates are not often obvious. Relatedly, NEUS outperforms FFS in the presence of intermediates because the regions can be made arbitrarily small to avoid the integration of long trajectory segments.

7. APPLICATIONS

To set these ideas in context, we review recent applications and sketch areas ripe for enhanced sampling of nonequilibrium steady states.

Nucleation has been extensively studied because the size of the structured cluster typically provides a good handle on the dynamics. For an Ising model of crystal nucleation, Allen et al. (76) observed a peak in the rate of ordering as the probability of shifting the rows of the lattice (which represents the shear rate) varied. The authors suggested that the nonmonotonic dependence resulted from the competition between the tendency of the shear to drive like spins together and its tendency to break apart clusters. Dickson et al. (30) studied this model as a vehicle for investigating alternative coarse-graining schemes. FFS has also been used to study reversible nucleation, including crystal nucleation in molten NaCl (77), droplet coalescence in oil/water/surfactant systems (78), homogeneous bubble nucleation in a Lennard-Jones fluid (79), and lattice protein folding (53, 72, 73, 80). It will be of interest to extend this line of studies to examine the effects of shear on materials with amorphous microscopic structures (7–12), in which jamming can make internal rearrangements rare, and enhanced sampling methods can thus make a dramatic difference.

Polymers under the influence of external fields make up another area with significant activity. FFS has been used to study the translocation of a coarse-grained model of a polymer through a pore (39), as well as reversal inside a pore (81). We have been developing phenomenological models to interpret single-molecule measurements of a large RNA folding and unfolding in response to periodic changes in the concentration of magnesium ions in solution (82, 83). An interesting suggestion that emerged from this work is that the system has very slow underdamped modes. One possible source for such modes is a cyclic motion resulting from the interplay between the relaxation of the polymer and the flow in the microfluidic channel used to change the buffer (84), and we are using NEUS to examine this idea as well as the extent to which the flow perturbs the system away from the reversible dynamics. Given that several cell-surface receptors and cytoskeletal components appear to be shear sensitive (85, 86), we anticipate that these studies will be just the start of characterizing biomolecular conformational change in flow. The enhanced sampling methods should also enable studies of ion channels under strong electromotive forces, to examine in particular whether coupling between the ions and the protein at high fields contributes to rectification (asymmetric conductance) (87, 88).

As more data about the molecular interactions involved in cellular signaling and gene regulation are obtained, it will become increasingly important to analyze the dynamics of networks of such interactions. These systems are inherently far from equilibrium because they constantly consume nucleotide triphosphates (ATP and GTP), and enhanced sampling methods are required to bridge the gap between the timescale of fluctuations in protein activities (minutes) and the timescale of transitions between cell fates (days). Already, a model of a genetic toggle switch has emerged as the "hydrogen atom" of enhanced sampling of stochastic regulatory networks (38, 39, 53, 65, 72, 89). In this model, two proteins, A and B, can each homodimerize and then bind to an operon. The operon can only bind one dimer at a time. When a dimer of A (B) is bound to the operon, it represses transcription of the gene for B (A); there is no concomitant effect on the expression of A (B). As a result of these dynamics, the system has two stable states: one with abundant A and scarce B and the other with the opposite situation. Switching between the two stable states is rare (38). Despite its apparent simplicity, this system provides a challenging test. The bistability not only makes convergence of the relative populations of the stable states slow because there are few events connecting them, but also leads to a hysteresis in which the flow between stable states takes a different path through phase space in each direction (38). This line of study has been recently extended to more realistic models (90). Looking forward, carcinogenesis (91, 92) and cellular

reprogramming (93) can be framed as microscopically irreversible transition path problems, and we anticipate that the enhanced sampling methods described here will play an important role in addressing these and similar problems.

APPENDIX A: THE NONEQUILIBRIUM UMBRELLA SAMPLING WEIGHT TRANSFER PROCEDURE

In this appendix, we analyze Markovian systems with finite numbers of regions in their order-parameter spaces to show formally how the NEUS algorithm converges. To this end, consider a weight-space $\mathbb{W} = [0, 1]^m$ that contains all possible configurations of the m weight variables (one for each region). The transitions can be described by the master equation $\hat{X}^{t+1} = \boldsymbol{B}\hat{X}^t$, where $\hat{X}^t$ is a vector describing the population of each region at time t, $\boldsymbol{B}$ is a transition matrix defined such that b_{ij} is the probability of transitioning from state j to state i in the time interval δt, and $b_{ii} = 1 - \sum_{j\neq i} b_{ji}$ is the probability of remaining in state i for the interval δt. We assume the $\hat{X}$ vectors satisfy the normalization $\sum_{i=1}^{m} \hat{X}_i = 1$. If a steady state exists, and the system is ergodic, we can define $\hat{W}^f = \boldsymbol{B}^\infty \hat{X}$ for any $\hat{X}$, where $\hat{W}^f$ is the vector of steady-state weights. It follows that $\hat{W}^f = \boldsymbol{B}\hat{W}^f$, i.e., that $\hat{W}^f$ is an eigenvector of $\boldsymbol{B}$ with eigenvalue 1.

A.1. Defining a Propagator in Weight-Space

We can approximate the weight update step in Equation 6 by a propagator in $\mathbb{W}$. Over a given weight update period T (which is long compared with δt), the weight of region Ω_i will change according to Equation 7, which can be rearranged to get

$$W_i^{t+T} = W_i^t \left(1 - sT \sum_{\substack{j=1 \\ j\neq i}}^{m} \phi_{ji}\right) + \sum_{\substack{j=1 \\ j\neq i}}^{m} W_j^t (sT\phi_{ij}), \tag{17}$$

where ϕ_{ji} is the average flux of trajectories over the time interval $(t, t + T)$ from region Ω_i to region Ω_j. Making the assumption that $\phi_{ji}\delta t = b_{ji}$ (which we discuss below), the m equations represented by Equation 17 can be combined as $\hat{W}^{t+T} = \boldsymbol{\Gamma}\hat{W}^t$, where

$$\boldsymbol{\Gamma} = \begin{pmatrix} 1 - sT\sum_{\substack{j=1 \\ j\neq 1}}^{m} B_{j1} & sTB_{12} & sTB_{13} & \cdots & sTB_{1m} \\ sTB_{21} & 1 - sT\sum_{\substack{j=1 \\ j\neq 2}}^{m} B_{j2} & sTB_{23} & \cdots & sTB_{2m} \\ sTB_{31} & sTB_{32} & 1 - sT\sum_{\substack{j=1 \\ j\neq 3}}^{m} B_{j3} & \cdots & sTB_{3m} \\ \vdots & \vdots & & \ddots & \\ sTB_{m1} & sTB_{m2} & sTB_{m3} & \cdots & 1 - sT\sum_{\substack{j=1 \\ j\neq m}}^{m} B_{jm} \end{pmatrix} \tag{18}$$

can be seen as a propagator in weight-space. In summary, the matrix $\boldsymbol{\Gamma}$ propagates the weights over an update period T, whereas the matrix $\boldsymbol{B}$ propagates the system over a single time step.

A.2. Fixed Points of the Propagator

The fixed points of the propagator are given by $\boldsymbol{\Gamma}\hat{X}^f = \hat{X}^f$ and hence are eigenvectors of $\boldsymbol{\Gamma}$ with eigenvalue 1. Recognizing that $\boldsymbol{\Gamma} = (1 - sT)\boldsymbol{I} + sT\boldsymbol{B}$, it is easy to show that X^f is also an eigenvector of $\boldsymbol{B}$ with eigenvalue 1:

$$\boldsymbol{B}\hat{X}^f = \frac{1}{sT}\left[\boldsymbol{\Gamma} - (1 - sT)\boldsymbol{I}\right]\hat{X}^f = \frac{1}{sT}\left[1 - (1 - sT)\right]\hat{X}^f = \hat{X}^f. \tag{19}$$

The Perron-Frobenius theorem for positive stochastic matrices guarantees that for each matrix, this is the highest eigenvalue and that it is nondegenerate (94, 95). Therefore, $\hat{X}^f$ is equal to $\hat{W}^f$, the steady-state weight vector.

To show that $\hat{W}^f$ is a stable fixed point under $\mathbf{\Gamma}$, we consider $\Gamma^n(\hat{W}^f + \hat{\varepsilon})$ where $\hat{\varepsilon}$ is an arbitrary perturbation orthogonal to the fixed point. $\hat{\varepsilon}$ can be expressed as a linear combination of the eigenvectors of $\boldsymbol{B}$:

$$\hat{\varepsilon} = \sum_{i=1}^{m} c_i \hat{E}_i. \tag{20}$$

We now have

$$\lim_{n\to\infty} \mathbf{\Gamma}^n(\hat{W}^f + \hat{\varepsilon}) = \hat{W}^f + \lim_{n\to\infty} \sum_{i=1}^{m} c_i e_i^n \hat{E}_i = \hat{W}^f. \tag{21}$$

The eigenvalues and eigenvectors of $\boldsymbol{B}$ are given by e_i and $\hat{E}_i$ respectively, and the last step was performed using $-1 < e_i < 1$ for all $\hat{E}_i \neq \hat{W}^f$, and $c_i = 0$ for $\hat{E}_i = \hat{W}^f$.

The use of $\phi_{ji}\delta t = b_{ji}$ above assumes that the observed flux between regions over a given time interval is equal to the steady-state flux. This makes the mapping of the weights Markovian, when in practice the observed flux is subject to fluctuations in time and the current state of the evolving input distributions. Although it remains to incorporate these dynamics into the analytical framework, the above discussion provides an intuitive understanding of the algorithm.

APPENDIX B: EQUIVALENCE OF FORWARD FLUX SAMPLING RATE CALCULATION METHODS

Here we show that calculation of the rate through Equation 11 using the weights given by Equation 13 is equivalent to the standard FFS rate calculation using conditional probabilities (Equation 12). We consider the calculation of k_{AB} using a simulation in $\mathcal{S}_A$. Because the interfaces used to define the regions are nonintersecting, only one region $(n-1)$ can contribute to the flux from $\mathcal{S}_A$ into region B. The expression for the rate using Equation 11 then reduces to

$$k_{AB} = \frac{1}{P_{\mathcal{S}_A}} \left[\overline{\Phi}_{0|A} \left(\prod_{j=0}^{n-2} P(\lambda_{j+1}|\lambda_j) \right) \overline{\Delta T}_{n-1} \right] \frac{N_{(n-1)B}}{T_{n-1}}, \tag{22}$$

where the quantity in the square brackets is the weight of region $n-1$. We recall that $\overline{\Delta T}_{n-1}$ is the average time length of a trajectory segment in region $n-1$, $N_{(n-1)B}$ is the number of trajectories that go from region $n-1$ to B in a given sampling period, and T_{n-1} is the time elapsed in region $n-1$ during that sampling period. The quantity $N_{(n-1)B}$ must be equal to $P(\lambda_n|\lambda_{n-1})N_{\text{tot}}$, where N_{tot} is the total number of trajectory segments sampled in region $n-1$. By multiplying both sides of the equation by $\overline{\Delta T}_{n-1}$, it follows that

$$\overline{\Delta T}_{n-1} N_{(n-1)B} = P(\lambda_n|\lambda_{n-1}) T_{n-1}. \tag{23}$$

Substituting Equation 23 into Equation 22, we get

$$k_{AB} = \frac{\overline{\Phi}_{0|A}}{P_{\mathcal{S}_A}} \left(\prod_{j=0}^{n-1} P(\lambda_{j+1}|\lambda_j) \right), \tag{24}$$

which is the standard expression for rate calculation with FFS. This demonstrates the equivalence of the two approaches.

SUMMARY POINTS

1. Two general methods have emerged for enhanced sampling of steady-state properties of systems far from equilibrium: nonequilibrium umbrella sampling and forward flux sampling.
2. Both methods work by accumulating statistics for short trajectory segments. In this review, we introduce a common framework for describing algorithms of this nature. Within this framework, transition rates are obtained by using an extended order parameter space that separates forward and backward paths.
3. There are advantages and disadvantages to each method. In brief, nonequilibrium umbrella sampling provides greater flexibility than forward flux sampling at the expense of greater computational overhead.
4. Applications include nucleation under shear, polymer conformational change in the presence of external forces, and switching between states of bistable gene regulatory systems. Treating more complex processes remains an outstanding goal.
5. Although a number of analytical results now exist for nonequilibrium umbrella sampling, a better theoretical understanding of convergence properties could aid in choosing between competing implementations; to this end, it will be interesting to explore connections to other advances in the statistical mechanics of systems far from equilibrium (e.g., large deviation theory and fluctuation theorems).

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We wish to thank Michael Hagan, Aryeh Warmflash, and Mark Maienschein-Cline for critical readings of the manuscript, as well as Martin Tchernookov and Shannon Stewman for useful discussions. This work was supported by the National Science Foundation and the Natural Sciences and Engineering Research Council.

LITERATURE CITED

1. Browne WR, Feringa BL. 2006. Making molecular machines work. *Nat. Nanotechnol.* 1:25–35
2. Herbert KM, Greenleaf WJ, Block SM. 2008. Single-molecule studies of RNA polymerase: motoring along. *Annu. Rev. Biochem.* 77:149–76
3. Block SM. 1996. Fifty ways to love your lever: myosin motors. *Cell* 87:151–57
4. Karplus M, Gao YQ. 2004. Biomolecular motors: the F-1-ATPase paradigm. *Curr. Opin. Struct. Biol.* 14:250–59
5. Webster MA, Yeomans JM. 2005. Modeling a tethered polymer in Poiseuille flow. *J. Chem. Phys.* 122:164903

6. Arya G, Rottler J, Panagiotopoulos AZ, Sorolovitz DJ, Chaikin PM. 2005. Shear ordering in thin films of spherical block copolymer. *Langmuir* 21:11518–27
7. Barnes HA. 1989. Shear-thickening ("dilatancy") in suspensions of nonaggregating solid particles dispersed in Newtonian liquids. *J. Rheol.* 33:329–66
8. Brown E, Jaeger HM. 2009. Dynamic jamming point for shear thickening suspensions. *Phys. Rev. Lett.* 103:086001
9. Melrose JR, Ball RC. 2004. "Contact networks" in continuously shear thickening colloids. *J. Rheol.* 48:961–78
10. Brady JF, Bossis G. 1985. The rheology of concentrated suspensions of spheres in simple shear flow by numerical simulation. *J. Fluid Mech.* 155:105–29
11. Melrose JR, van Vliet JH, Ball RC. 1996. Continuous shear thickening and colloid surfaces. *Phys. Rev. Lett.* 77:4660–63
12. Grebenkov DS, Ciamarra MP, Nicodemi M, Coniglio A. 2008. Flow, ordering, and jamming of sheared granular suspensions. *Phys. Rev. Lett.* 100:078001
13. Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440:545–50
14. Laslo P, Spooner CJ, Warmflash A, Lancki DW, Lee H-J, et al. 2006. Multilineage transcriptional priming and stabilization of alternate hematopoietic cell fates. *Cell* 126:755–66
15. **Lecomte V, Appert-Rolland C, van Wijland E. 2007. Thermodynamic formalism for systems with Markov dynamics. *J. Stat. Phys.* 127:51–106**
16. Hedges LO, Jack RL, Garrahan JP, Chandler D. 2009. Dynamic order-disorder in atomistic models of structural glass formers. *Science* 323:1309–13
17. Selvin PR. 1995. Fluorescence resonance energy-transfer. *Methods Enzymol.* 246:300–34
18. Weiss S. 1999. Fluorescence spectroscopy of single biomolecules. *Science* 283:1676–83
19. Ha T. 2001. Single molecule fluorescence resonance energy transfer. *Methods* 5:78–86
20. Neuman KC, Block SM. 2004. Optical trapping. *Rev. Sci. Instrum.* 75:2787–809
21. Evans DJ, Morriss G. 2008. *Statistical Mechanics of Nonequilibrium Liquids*. New York: Cambridge Univ. Press. 2nd ed.
22. Allen MP, Tildesley DJ. 1987. *Computer Simulation of Liquids*. New York: Oxford Univ. Press
23. Frenkel D, Smit B. 2002. *Understanding Molecular Simulation: From Algorithms to Applications*. London: Academic
24. Todd BD, Daivis PJ. 2005. Homogeneous non-equilibrium molecular dynamics simulations of viscous flow: techniques and applications. *Mol. Sim.* 33:189–229
25. Gillespie DT. 2007. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* 58:35–55
26. Bandrivskyy A, Beri S, Luchinsky DG, Mannella R, McClintock PVE. 2003. Fast Monte Carlo simulations and singularities in the probability distributions of nonequilibrium systems. *Phys. Rev. Lett.* 90:210201
27. Blaak R, Auer S, Frenkel D, Löwen H. 2004. Crystal nuclation of colloidal suspensions under shear. *Phys. Rev. Lett.* 93:068303
28. Blaak R, Lowen H. 2005. Umbrella sampling in non-equilibrium computer simulations. *Comput. Phys. Commun.* 169:64–68
29. **Warmflash A, Bhimalapuram P, Dinner AR. 2007. Umbrella sampling for nonequilibrium processes. *J. Chem. Phys.* 127:154112**
30. Dickson A, Warmflash A, Dinner AR. 2009. Nonequilibrium umbrella sampling in spaces of many order parameters. *J. Chem. Phys.* 130:074104
31. Crooks GE, Chandler D. 2001. Efficient transition path sampling for nonequilibrium stochastic dynamics. *Phys. Rev. E* 64:026109
32. Heymann M, Vanden-Eijnden E. 2008. Pathways of maximum likelihood for rare events in nonequilibrium systems: application to nucleation in the presence of shear. *Phys. Rev. Lett.* 100:140601
33. **Dellago C, Bolhuis PG, Csajka FS, Chandler D. 1998. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* 108:1964–77**
34. **van Erp TS, Moroni D, Bolhuis PG. 2003. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* 118:7762–74**

15. Develops an analogy between equilibrium and nonequilibrium statistical mechanics in detail, and this leads to a free-energy-like quantity for dynamics known as the large deviation function; symmetries within this formalism correspond to fluctuation theorems, which have received much attention recently.

29. Introduced the NEUS method (which was then extended in Refs. 30, 37, and 49).

33. Transition path sampling (introduced here, and reviewed in Ref. 35) was the first practical and widely used computational approach based on the idea that dynamics and associated quantities like rates are determined by the statistics of an ensemble of trajectories.

34. Showed how the efficiency of the rate calculations within the transition-path-sampling framework could be significantly increased by directly estimating fluxes through a series of interfaces separating reactants and products (see also Refs. 68 and 69).

35. Dellago C, Bolhuis PG. 2009. Transition path sampling and other advanced simulation techniques for rare events. In *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, pp. 167–233. Vol. 221, Adv. Polymer Sci. Ser. New York: Springer.
36. E W, Vanden-Eijnden E, Ren W. 2005. Finite temperature string method for the study of rare events. *J. Phys. Chem. B* 109:6688–93
37. Dickson A, Warmflash A, Dinner AR. 2009. Separating forward and backward pathways in nonequilibrium umbrella sampling. *J. Chem. Phys.* 131:154104
38. Allen RJ, Warren PB, ten Wolde PR. 2005. Sampling rare switching events in biochemical networks. *Phys. Rev. Lett.* 94:018104
39. Allen RJ, Frenkel D, ten Wolde PR. 2006. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *J. Chem. Phys.* 124:024102
40. Allen RJ, Frenkel D, ten Wolde PR. 2006. Forward flux sampling-type schemes for simulating rare events: efficiency analysis. *J. Chem. Phys.* 124:194111
41. Giardina C, Kurchan J, Peliti L. 2006. Direct evaluation of large-deviation functions. *Phys. Rev. Lett.* 96:120603
42. Tchernookov M, Dinner AR. 2009. A list-based algorithm for evaluation of large deviation functions. *J. Stat. Mech.* Manuscript submitted
43. Warmflash A, Dinner AR. 2008. Signatures of combinatorial regulation in intrinsic biological noise. *Proc. Natl. Acad. Sci. USA* 105:17262–67
44. Maienschein-Cline M, Warmflash A, Dinner AR. 2009. Defining cooperativity in gene regulation locally through intrinsic noise. *IET Syst. Biol.* Manuscript submitted
45. Valleau JP, Torrie GM. 1977. Non-physical sampling distributions in Monte-Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* 23:187–99
46. Chandler D. 1987. *Introduction to Modern Statistical Mechanics*. New York: Oxford Univ. Press
47. de Oliviera MM, Dickman R. 2005. How to simulate the quasistationary state. *Phys. Rev. E.* 71:016129
48. Tchernookov M, Warmflash A, Dinner AR. 2009. Critical behavior for a model with catalyzed autoamplification. *J. Chem. Phys.* 130:134906
49. Vanden-Eijnden E, Venturoli M. 2009. Exact rate calculations by trajectory parallelization and tilting. *J. Chem. Phys.* 131:044120
50. Ma A, Dinner AR. 2005. Automatic method for identifying reaction coordinates in complex systems. *J. Phys. Chem. B* 109:6769–79
51. Peters B, Trout BL. 2006. Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.* 125:054108
52. Peters B, Beckham GT, Trout BL. 2007. Extensions to the likelihood maximization approach for finding reaction coordinates. *J. Chem. Phys.* 127:034109
53. Borrero EE, Escobedo FA. 2007. Reaction coordinates and transition pathways of rare events via forward flux sampling. *J. Chem. Phys.* 127:164101
54. Hu J, Ma A, Dinner AR. 2008. A two-step nucleotide-flipping mechanism enables kinetic discrimination of DNA lesions by AGT. *Proc. Natl. Acad. Sci. USA* 105:4615–20
55. Ren W, Vanden-Eijnden E, Maragakis P, E W. 2005. Transition pathways in complex systems: application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.* 123:134109
56. E W, Ren W, Vanden-Eijnden E. 2005. Transition pathways in complex systems: reaction coordinates, isocommittor surfaces and transition tubes. *Chem. Phys. Lett.* 413:242–47
57. Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. 2006. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* 125:024106
58. Miller TF, Vanden-Eijnden E, Chandler D. 2007. Solvent coarse-graining and the string method applied to the hydrophobic collapse of a hydrated chain. *Proc. Natl. Acad. Sci. USA* 104:14559–64
59. Venturoli M, Vanden-Eijnden E, Ciccotti G. 2009. Kinetics of phase transitions in two dimensional Ising models studied with the string method. *J. Math. Chem.* 45:188–222

38. Introduced the FFS method (which was then extended in Refs. 39, 40, 65, and 76).

45. Showed how one could enforce sampling of low probability regions and still recover the correct physical weighting of states described by order parameters. Most enhanced sampling methods for systems in equilibrium can be viewed as forms of umbrella sampling.

50. Introduced the idea that statistical methods could be used to efficiently relate physical properties of a system to the commitor, the probability that dynamics trajectories initiated from a configuration with random momenta reach one stable state before another. The commitor has become the standard means for characterizing reactants, transition states, and products of reactions in stochastic systems, and it is computationally costly to evaluate. A key part of this study was thus that the commitor values were determined only once for a set of representative structures, in contrast to earlier trial-and-error approaches to interpreting path sampling simulations.

60. Vanden-Eijnden E, Venturoli M. 2009. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* 130:194103

61. Pan AC, Roux B. 2008. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Phys. Chem.* 129:064107

62. Gonze D, Halloy J, Goldbeter A. 2002. Deterministic versus stochastic models for circadian rhythms. *J. Biol. Phys.* 28:637–53

63. Gonze D, Halloy J, Goldbeter A. 2002. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci. USA* 99:673–78

64. Faradjian AK, Elber R. 2004. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.* 120:10880

65. Valeriani C, Allen RJ, Morelli MJ, Frenkel D, ten Wolde PR. 2007. Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling. *J. Chem. Phys.* 127:114109

66. Shalloway D, Faradjian AK. 2006. Efficient computation of the first passage time distribution of the generalized master equation by steady state relaxation. *J. Chem. Phys.* 124:054112

67. West AMA, Elber R, Shalloway D. 2007. Extending molecular dynamics time scales with milestoning: example of complex kinetics in a solvated peptide. *J. Chem. Phys.* 122:145104

68. Moroni D, Bolhuis PG, van Erp TS. 2004. Rate constants for diffusive processes by partial path sampling. *J. Chem. Phys.* 120:4055–65

69. van Erp TS, Bolhuis PG. 2005. Elaborating transition interface sampling methods. *J. Comput. Phys.* 205:157–81

70. Huber GA, Kim S. 1996. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* 70:97–110

71. Zhang BW, Jasnow D, Zuckerman DM. 2007. Efficient and verified simulation of a path ensemble for conformational change in a united-residue model of calmodulin. *Proc. Natl. Acad. Sci. USA* 104:18043–48

72. Borrero EE, Escobedo FA. 2008. Optimizing the sampling and staging for simulations of rare events via forward flux sampling schemes. *J. Chem. Phys.* 129:024115

73. Borrero EE, Escobedo FA. 2009. Simulating the kinetics and thermodynamics of transitions via forward flux/umbrella sampling. *J. Phys. Chem. B.* 113:6434–45

74. Sear RP. 2008. Nucleation in the presence of slow microscopic dynamics. *J. Chem. Phys.* 128:214513

75. Juraszek J, Bolhuis PG. 2008. Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys. J.* 95:4246–57

76. Allen RJ, Valeriani C, Tanase-Nicola S, ten Wolde PR, Frenkel D. 2008. Homogeneous nucleation under shear in a two-dimensional Ising model: cluster growth, coalescence, and breakup. *J. Chem. Phys.* 129:134704

77. Valeriani C, Sanz E, Frenkel D. 2005. Rate of homogeneous crystal nucleation in molten NaCl. *J. Chem. Phys.* 122:194501

78. Rekvig L, Frenkel D. 2007. Molecular simulations of droplet coalescence in oil/water/surfactant systems. *J. Chem. Phys.* 127:134701

79. Wang Z, Valeriani C, Frenkel D. 2009. Homogeneous bubble nucleation driven by local hot spots: a molecular dynamics study. *J. Phys. Chem. B* 113:3776–84

80. Borrero EE, Escobedo FA. 2006. Folding kinetics of a lattice protein via a forward flux sampling approach. *J. Chem. Phys.* 125:164904

81. Huang L, Makarov DE. 2008. The rate constant of polymer reversal inside a pore. *J. Chem. Phys.* 128:114903

82. Li Y, Zhao T, Bhimalapuram P, Dinner AR. 2008. How the nature of an observation changes single-trajectory entropies. *J. Chem. Phys.* 128:074102

83. Li Y, Qu XH, Ma A, Smith GJ, Scherer NF, Dinner AR. 2009. Models of single-molecule experiments with periodic perturbations reveal hidden dynamics in RNA folding. *J. Phys. Chem. B* 113:7579–90

84. Delgado-Buscaglioni R. 2006. Cyclic motion of a grafted polymer under shear flow. *Phys. Rev. Lett.* 96:088303

85. Marshall BT, Long M, Piper JW, Yago T, McEver RP, Zhu C. 2003. Direct observation of catch bonds involving cell-adhesion molecules. *Nature* 423:190–93

60. The finite-temperature string method (also described in Refs. 36, 55, 57–59) is a path sampling approach built around the commitor rather than the ensemble of reactive trajectories (as in Ref. 33). The Voronoi tessellation scheme introduced in this paper is the basis for the string form of NEUS in Ref. 30.

67. Combines interface sampling (in the form developed in Ref. 64) with a generalized master equation formalism (Ref. 66) to obtain rates as well as higher moments of the first passage time distribution of a reaction.

70. To the best of our knowledge, the first interface sampling method. It is similar to FFS in practice. The method failed to gain popularity in the absence of the rate calculation framework developed in Refs. 33, 34, and 38.

86. Wang Y-L. 2007. Flux at focal adhesions: slippage clutch, mechanical gauge, or signal depot. *Sci. STKE* 2007:pe10
87. Cheng MH, Cascio M, Coalson RD. 2007. Homology modeling and molecular dynamics simulations of the α1 glycine receptor reveals different states of the channel. *Proteins* 68:581–93
88. Cheng MH, Coalson RD, Cascio M, Kurnikova M. 2008. Computational prediction of ion permeation characteristics in the glycine receptor modified by photo-sensitive compounds. *J. Comput. Aided Mol. Des.* 22:563–70
89. Morelli MJ, Tanase-Nicola S, Allen RJ, ten Wolde PR. 2008. Reaction coordinates for the flipping of genetic switches. *Biophys. J.* 94:3413–23
90. Morelli MJ, ten Wolde PR, Allen RJ. 2009. DNA looping provides stability and robustness to the bacteriophage λ switch. *Proc. Natl. Acad. Sci. USA* 106:8101–6
91. Warren PB. 2009. Cells, cancer, and rare events: homeostatic metastability in stochastic non-linear dynamics models of skin cell proliferation. *Phys. Rev. E* 80:030903
92. Anderson ARA, Quaranta V. 2008. Integrative mathematical oncology. *Nat. Rev. Cancer* 8:227–34
93. Warmflash A, Dinner AR. 2009. Modeling gene regulatory networks for cell fate specification. In *Mechanics of Cellular Systems and Processes*, ed. M Zaman, pp. 121–54. Cambridge, UK: Cambridge Univ. Press
94. Horn RA, Johnson CR. 1990. *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press
95. Graham A. 1987. *Nonnegative Matrices and Applicable Topics in Linear Algebra*. New York: Ellis Horwood Ltd.
96. Gillespie DT. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–61

Annual Review of
Physical Chemistry
Volume 61, 2010

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Physical Chemistry* articles may be found at http://physchem.annualreviews.org/errata.shtml