

## Exact milestoneing

Juan M. Bello-Rivas<sup>1</sup> and Ron Elber<sup>1,2</sup>

<sup>1</sup>*Department of Chemistry, Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712, USA*

<sup>2</sup>*Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA*

(Received 6 October 2014; accepted 11 February 2015; published online 2 March 2015)

A new theory and an exact computer algorithm for calculating kinetics and thermodynamic properties of a particle system are described. The algorithm avoids trapping in metastable states, which are typical challenges for Molecular Dynamics (MD) simulations on rough energy landscapes. It is based on the division of the full space into Voronoi cells. Prior knowledge or coarse sampling of space points provides the centers of the Voronoi cells. Short time trajectories are computed between the boundaries of the cells that we call milestones and are used to determine fluxes at the milestones. The flux function, an essential component of the new theory, provides a complete description of the statistical mechanics of the system at the resolution of the milestones. We illustrate the accuracy and efficiency of the exact Milestoning approach by comparing numerical results obtained on a model system using exact Milestoning with the results of long trajectories and with a solution of the corresponding Fokker-Planck equation. The theory uses an equation that resembles the approximate Milestoning method that was introduced in 2004 [A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**(23), 10880-10889 (2004)]. However, the current formulation is exact and is still significantly more efficient than straightforward MD simulations on the system studied. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4913399>]

### I. INTRODUCTION

The Molecular Dynamics (MD) method is a useful tool for studying properties of matter with atomically detailed simulations. MD makes it possible to connect microscopic structures and interactions with thermodynamics, kinetics, and mechanisms of molecular processes. Nevertheless, a significant limitation of these simulations is that of time scales. The fundamental numerical time step ( $\sim 10^{-15}$  s) is much shorter than many observation times in biophysics, for example, enzymatic reactions can take milliseconds and longer. This limitation makes MD simulations for these systems extremely expensive. While considerable progress in extending time scales of MD was made by improvements in specialized and general hardware,<sup>1</sup> significant limitations remain on the length of a single trajectory (microseconds for readily accessible machines) and on generating an ensemble of long trajectories necessary for estimating kinetics. Methods to speed up these calculations are desired.

Why are so many time steps required when the spatial reorganization of the molecules that we examine is frequently small? MD simulations can be long because a significant fraction of the time the system is located at metastable states or deep free energy minima. Nothing much happened while the system diffuses in the metastable state. Shortening the wait time at the metastable states while still retaining the correctness of the sampling and time scales is a prime motivation behind the exact Milestoning algorithm. We comment that in many biomolecular systems, the number of metastable states can be very large. In general, it is not sufficient to use a two state system (reactant and product) as an effective description

of complex dynamics. In many biophysical systems, it is necessary to consider a truly rough energy landscape with almost a continuum of temporal and spatial scales. A classic example is the study of myoglobin.<sup>2</sup> Appropriate technologies for such complex systems are desired.

Indeed, a number of different theoretical and algorithmic approaches aim to extend the time scale of simulations and produce trajectories probing slow kinetics. Notable methods are action-based approaches. In this class of techniques, trajectories of time scales much longer than temporal ranges accessible to MD are estimated.<sup>3</sup> Other approaches<sup>4</sup> are aimed at sampling trajectories that pass over a few significant energy barriers. Individual trajectories in the latter case are not long in time but are rare and therefore the average process is slow. On rough energy landscapes, in which we find numerous metastable states, broad distributions of barrier heights, and wide range of minimum depths, individual trajectories may be long in time. Technologies based on short and rare trajectories are difficult to use in a straightforward fashion in these systems. This is since individual trajectories between reactants and products can be long (the trajectories may be trapped for a long time in metastable states that are not the initial or the final states). Hence, kinetics may not be sampled properly by rare (and short) trajectory strategies. Methods like replica exchange transition interface sampling<sup>5</sup> aimed to enhance sampling by identifying the metastable states and focusing on the transitions between the metastable states using short trajectories. The number of these metastable states and the complexity of characterizing them can grow exponentially with the system size.<sup>6</sup> Studies of such systems are a significant challenge, and an efficient algorithm to enumerate them is not

known. It is desired to develop a technique that enhances the time scale of simulations and is less sensitive to the features of the underlying energy landscape.

Approaches like PPTIS (Partial Paths Transition Interface Sampling),<sup>7</sup> Weighted Ensemble (WE),<sup>8</sup> and Milestoning<sup>9</sup> aim to address the last problem and consider dynamics that can be a mix of diffusive (small barrier) and activated (large barrier) processes. Designing the method with rough energy landscapes in mind leads to technologies that are less impacted by the broad distributions of barriers and minima we frequently encounter in molecular biophysics<sup>10</sup> or the “small barrier” problem of material science.<sup>11</sup> Both PPTIS and Milestoning were introduced as highly efficient and approximate methods, while WE is in principle exact for *stochastic* dynamics. PPTIS and Milestoning exploit the use of short trajectory fragments to estimate a local kinetic operator. The operator is then used in probabilistic modeling of longer time scales.

The purpose of the present manuscript is to formulate and illustrate an exact Milestoning approach. It offers an exact description of the statistical mechanics of the system while retaining many of the useful features of the original approximate algorithm.<sup>12</sup> The theory was tested and illustrated extensively in the past (see, for instance, Refs. 13 and 14) and a review is available.<sup>12</sup> Nevertheless, since the theory and implementation have evolved considerably and are now exact, we repeat and redefine some of the concepts that were introduced earlier in the formulation of Faradjian and Elber<sup>9</sup> and Kirmizialtin and Elber.<sup>12</sup> The current formulation also builds on the Voronoi tessellation of Markovian Milestoning introduced by Venturoli and Vanden-Eijnden<sup>15</sup> and adjusted to avoid milestone crossing by Majek and Elber.<sup>16</sup> Finally, the idea of exact molecular dynamics trajectories through interfaces was discussed in a more limited fashion in earlier works by Warmflash, Bhimalapuram, and Dinner and in a follow up paper by Dickson, Warmflash, and Dinner.<sup>17,18</sup> It was also examined by Vanden-Eijnden and Venturoli in their trajectory tilting approach.<sup>19</sup> These studies did not offer a complete statistical mechanical view, which is offered by the approach of exact Milestoning.

Compared to weighted ensemble, another method that is statistically exact,<sup>8</sup> Milestoning is more flexible in the choice of the dynamics. The WE approach requires stochastic equations of motion to allow splitting and termination of trajectories. Milestoning uses stochastic considerations only to determine initial conditions and can use straightforward and deterministic equations such as Hamilton equations. The exact Milestoning algorithm can be applied to non-equilibrium cases.

This manuscript is organized as follows. In Sec. II, we motivate the study and present the general principles of exact Milestoning and the corresponding algorithms. More specifically, we start with the definition of the milestones (Sec. II A), continue to define the core equations for the probability of crossing a milestone and of the absolute flux (Sec. II B), derive the formula for the stationary flux, a core entity of the Milestoning theory (Sec. II C), discuss the algorithm to compute the stationary flux (Sec. II D), consider the probabilities of the states (Sec. II E), and finally examine moments of the distribution of the first passage time (Sec. II F).

In Sec. III, we consider a numerical example. In Sec. III A, we describe the simple two-dimensional energy surface that we use for illustration. The type of the dynamics (Brownian) is described in Sec. III B. The results and discussion are in Sec. IV. In Sec. IV A, we consider numerical calculations of the stationary flux and the MFPT (Mean First Passage Time) using exact Milestoning. In Sec. IV B, we solve the same problem using the Fokker-Planck equation (FPE). In Sec. IV C, we compare exact Milestoning and the FPE solutions. In Sec. IV D, we compare the exact Milestoning calculations to a calculation based on very long trajectories (the most straightforward approach for this type of problem in systems with a large number of degrees of freedom) and discuss accuracy and efficiency. In Sec. IV E, we examine the rate of convergence as a function of the initial guess and illustrate that starting from equilibrium distributions at the milestones is significantly more efficient than initiating the system at one milestone. The summary of the paper is in Sec. V.

## II. THEORY AND CALCULATIONS

### A. Definition and the choice of the milestones

A system of  $N$  particles at time  $t$  is fully characterized by a phase space vector,  $x(t)$ .

We are given a set of points (in phase or coordinate spaces),  $\{x_i\}_{i=1}^L$ , which we call anchors. These points guide the trajectory calculations and establish centers of Voronoi cells, which were introduced to Milestoning by Vanden-Eijnden and Venturoli.<sup>15</sup> A Voronoi cell  $j$  is defined as the set of all points  $x$  such that their distance to  $x_j$  is shorter than a distance to any other anchor,  $x_k$  ( $k \neq j$ ). A dividing surface between two cells,  $i$  and  $j$ , which we call a milestone, is the set of all points  $x$  such that their distances from  $i$  and  $j$  is the same and smaller than the distance to any other anchor. For example, if the anchor points form a straight line in a two-dimensional space, the milestones would be orthogonal lines exactly between the anchors. A trajectory may cross a milestone and transition between two Voronoi cells. *Crossing a milestone is the fundamental event in the theory of Milestoning and a state is defined by the last milestone that was crossed.* We denote milestones by  $M_\alpha$  where the index  $\alpha$  is a shortcut for a pair of anchors, say,  $ij$ . The total number of milestones is  $M$ .

How do we obtain the initial set of anchors? There are numerous ways of obtaining them. In the past, we have used our algorithms for reaction path calculations<sup>21,22</sup> to determine points along a reaction coordinate and to conduct Milestoning calculations in one dimension.<sup>14,23,24</sup> In other studies, we have used chemical intuition to determine relevant sample coordinates for protein folding (backbone torsions),<sup>25</sup> pick up structure from replica exchange equilibrium calculations,<sup>26</sup> or even use a field representation of material density.<sup>27</sup> Hence, the choice of anchors is highly flexible, and we rely on the user to provide a sensible sample of anchors that can (a) describe the process of interest as a sequence of transitions between cells associated with the anchors and that (b) the transitions between cells can be effectively sampled by short trajectories. While we have not emphasized this point in the past, the set of anchors can be adjusted dynamically. As we explore

more phase space points using trajectories between interfaces, we may discover new portions of phase space that require sampling and therefore new anchors can be generated on the fly.

There is also considerable flexibility in the definition of the distances mentioned above. In the present discussion, the distance between any two points,  $x_i$  and  $x_j$ , is simply  $d_{ij} = \sqrt{(x_i - x_j)^T (x_i - x_j)}$ . However, the distance can be defined also in coarse space. Let a vector of coordinates in coarse space be  $Q(x)$   $Q \in R^J$ , where  $J < 3N$ . The distance in coarse space is  $d_{Q,ij} = \sqrt{(Q_i - Q_j)^T (Q_i - Q_j)}$ . For example, when describing protein folding, the coordinates of the protein and not of the solvent are likely to be sufficient to describe the reaction. Of course, the dynamics is still conducted in the full space of  $x(t)$ .

## B. Definitions: Space of events, probability of events, flux, and kernel

### 1. Definitions

**Milestoning events:** The basic event of the statistical theory for kinetics and thermodynamics is the last crossing of a milestone (say,  $\alpha$ ) by a trajectory. The set of all events is  $\Omega$ . A trajectory such that the last milestone that it crossed is  $\alpha$  is said to be in a state  $\alpha$ .

$p_\alpha(x, t) \cdot dx$ : It is the probability that a trajectory *last* crossed a milestone  $\alpha$  at a phase space point  $x$  at time  $t$  in the time interval  $t' \in [0, t]$ , given that at time 0, the trajectory already crossed a milestone and was in  $\Omega$ .

Hence, no other milestone was crossed up to time  $t$  after milestone  $\alpha$  was crossed at  $t'$ . Any trajectory in the set can be assigned to an event of the set  $\Omega$  with a probability  $p_\alpha(x, t)$ . Since a trajectory must have crossed a milestone earlier, the sum of the probabilities of all events is one and can be written as  $\sum_{M_\alpha} \int p_\alpha(x, t) dx = 1$ . This normalization may raise a concern about time zero. That is, what is the milestone that was crossed when we start probing the process? We therefore set an initial condition that at time zero, the trajectory must have been in  $\Omega$  and we satisfy this condition by requiring that the trajectory crosses a milestone exactly at time 0 with a probability  $p_\alpha(x, t = 0)$ .

$q_\alpha(x, t) \cdot dt \cdot dx$ : It is the probability that a trajectory *last* crossed a milestone  $\alpha$  at a phase space point  $x$  between times  $t$  and  $t + dt$  given that at time zero the trajectory crossed already a milestone and therefore was in  $\Omega$ .

We call the function  $q_\alpha(x, t)$  the absolute flux or, in short, the flux. Note that the flux function as defined is always positive. Therefore, it differs from the usual definition of flux in continuum mechanics. Since the system is assumed ergodic, and since we setup initial conditions that imply crossing at time zero, a trajectory must have crossed the latest milestone at some time  $t$ . We can therefore normalize this probability as

$$\sum_{M_\alpha} \int_0^\infty dx \int dt \cdot q_\alpha(x, t) = 1.$$

$K_{\alpha\beta}(x', t'; x, t) \cdot dx \cdot dt$ : It is the probability of crossing milestone  $\beta$  between times  $t$  and  $t + dt$  ( $t > t'$ ) at phase space point  $x$  given that milestone  $\alpha$  was crossed at time  $t'$  at

phase space point  $x'$  and no other milestone was crossed between  $t$  and  $t'$ . It is also called the kernel and is normalized in a similar way to the flux function. We sum up over all crossing events and times given the earlier crossing event of milestone  $\alpha$  at phase space point  $x$  and time  $t$ . We have

$$\sum_{\beta \in M_\beta} \int dx' \int_t^\infty dt' K_{\alpha\beta}(x, t; x', t') = 1.$$

We are now ready to write an expression for  $p_\alpha(x, t)$ ,

$$p_\alpha(x, t) = \int_0^t q_\alpha(x, t') \left[ 1 - \sum_{\beta \in \bar{\alpha}} \int_{M_\beta} \int_{t'}^t K_{\alpha\beta}(x, t'; x', \tau) d\tau dx' \right] dt'. \quad (1)$$

Equation (1) means the following: we ask what is the probability of finding a trajectory at state  $\alpha$ , i.e., that at time  $t$ , the last milestone that the trajectory crossed was  $\alpha$ . Consider all crossing events into state  $\alpha$  at earlier times  $t' \in [0, t]$ . These crossing events may contribute to the probability of being at state  $\alpha$ . However, some of these events are lost from state  $\alpha$  between  $t'$  and  $t$  since they cross another milestone before time  $t$ . We therefore remove from these crossing events the losses to other states  $\beta$ . The notation  $\beta \in \bar{\alpha}$  means milestones  $\beta$  that can be crossed after milestone  $\alpha$  was crossed without passing any other milestone as an intermediate event. In other words, the milestones  $\beta \in \bar{\alpha}$  share a Voronoi cell with milestone  $\alpha$ . For example, if  $\alpha$  is an interface between cells  $i$  and  $j$ , then  $\beta$  is an interface between cells  $j$  and  $k$ .

Because Eq. (1) is formulated from a phase point  $x'$  to another phase point  $x$ , it is exact. In the past, we formulated an approximate version of Eq. (1) in which only a weight function of the milestone was computed.<sup>9</sup> The flux and the kernel still depend on the phase space points. We approximate the flux function as a canonical distribution conditioned on being at a milestone multiplied by a constant, and the kernel was applied on the sampled points exactly.<sup>12</sup> This approximation implies memory loss of trajectories between milestones and has proved useful in a number of applications (for recent applications, see Refs. 25 and 27–30). Nevertheless, the focus of the present paper is on exact applications. Equation (1) holds of course also for non-equilibrium processes.

The sequence of events of Eq. (1) is illustrated schematically in Figures 1 and 2.

While  $q_\alpha(x, t)$  seems like a flux function in the sense of continuum mechanics, it is not the direct derivative of  $p_\alpha(x, t)$ . The time derivative of the probability of being at state  $\alpha$  is

$$\frac{dp_\alpha(x, t)}{dt} = q_\alpha(x, t) - \sum_{\beta \in M_\beta} \int_0^t \int_{M_\beta} q_\alpha(x, t') K_{\alpha\beta}(x, t'; x', t) dt' \cdot dx'. \quad (2)$$

Equation (2) is a sum of a gain and a loss at time  $t$ . We gain from trajectories that cross milestone  $\alpha$  exactly at time  $t$  and lose trajectories that cross  $\alpha$  (enter state  $\alpha$ ) earlier and exactly at  $t$  leave state  $\alpha$  through another milestone  $\beta$ . The loss term

is an integral over earlier entry times and summation over all exit channels  $\beta$  and  $x'$ .

Note that Eq. (1) is not closed. We expressed  $p_\alpha(x, t)$  in terms of two other functions which at present are not known

( $q_\alpha(x, t)$  and  $K_{\alpha\beta}(x, t'; x', t)$ ). We therefore write another equation for the flux. In our kinetic model, no particles are lost during a transition. Hence, we write an equation stating the conservation of trajectories in terms of the flux function,

$$q_\alpha(x, t) \cdot dt = \begin{cases} p_\alpha(x, t=0) \cdot dt & t=0 \\ \left[ \sum_\beta \int_0^t \int_{M_\beta} q_\beta(x', t') K_{\beta\alpha}(x', t'; x, t) \cdot dx' \cdot dt' \right] \cdot dt & t>0 \end{cases}. \quad (3)$$

Equation (3) is at the center of the Milestoning theory and has the following physical meaning: we ask on the left side what is the probability that milestone  $\alpha$  was crossed between time  $t$  and time  $t + dt$  at phase space point  $x$ . On the right hand side we count the pathways in which this transition may happen. If we are at time zero, then there is an initial condition (top term) that counts the trajectories that cross milestone  $\alpha$  exactly at time 0. The lower term is for times that are different from zero. Crossing  $\alpha$  at time  $t$  is by a trajectory of a well-defined state. Hence at earlier time, it must have crossed another milestone ( $\beta$ ). The summation and integral in Eq. (3) is over all the trajectories that cross milestone  $\beta$  at earlier times  $t'$  and then continue to cross  $\alpha$  exactly at time  $t$ . This summation accounts for all the events of crossing of  $\alpha$  at time  $t$ . No other events in  $\Omega$  that contribute to this transition are possible.

We will now show that Eq. (3) for the flux function  $q_\alpha(x, t)$  can be solved exactly. The flux function is a core entity of Milestoning that enables the calculations of equilibrium and kinetic observables such as the free energy and the moments of the mean first passage time. It is therefore no wonder that

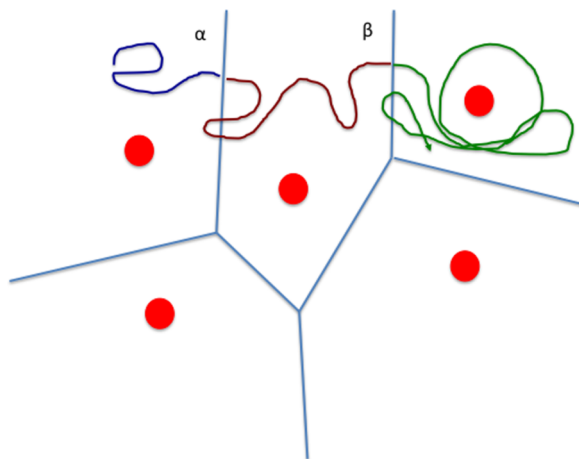


FIG. 1. A schematic drawing of Milestoning in two dimensions. The milestones are the blue lines between the anchors (red circles). A trajectory is shown as a curve that alternates colors when the last milestone crossed is modified. In the illustration, it crossed milestone  $\alpha$  and then  $\beta$ . See text for more details.

we spent considerable time on the theory of how to compute  $q_\alpha(x, t)$  and on an algorithm to extract it from MD trajectories.

### C. An equation for the stationary flux function, $q_{\alpha, \text{stat}}(x)$

To simplify the basic expression provided by Eq. (3), we consider stationary processes and stationary flux. A subset of stationary processes includes systems at equilibrium. We comment that even at equilibrium the flux is not necessarily zero. In the directional Milestoning picture proposed by Májek and Elber,<sup>16</sup> the flux from cell  $i$  to cell  $j$  is considered a different function from the flux from cell  $j$  to  $i$  and it is not zero for a system at equilibrium at a finite temperature.

We consider processes for which a stationary long time limit is well defined. Of course, it is not guaranteed that a stationary solution exists and one can imagine a non-equilibrium process with an oscillating external force for which the flux never becomes stationary. However, in the current formulation, we restrict ourselves to stationary, long-time processes. The goal of the current section is the derivation of an equation for the stationary flux  $q_{\alpha, \text{stat}}(x)$ .

Another restriction is the type of processes that we consider which is of processes homogeneous in time. For

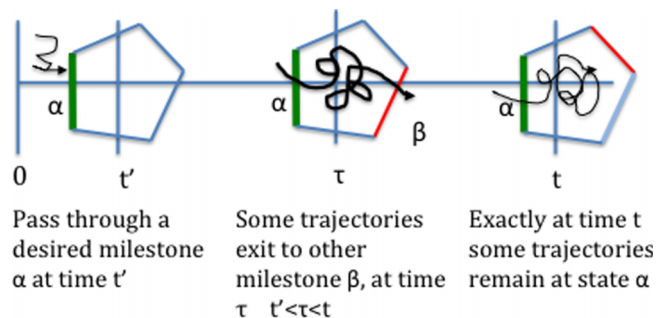


FIG. 2. A schematic representation of the sequence of events considered in the basic Milestoning equation (Eq. (1)). The first event is the passage through milestone  $\alpha$  at time  $t'$ . Then during propagation from time  $t'$  to  $t$ , some trajectories change their state by passing another milestone  $\beta$ . We consider only the fraction of trajectories that do not pass any other milestone up to time  $t$  as in state  $\alpha$ .



example, typical MD trajectories with time independent potentials are time homogeneous. In this case, the kernel does not depend on the absolute times but only on the time difference; hence, we write

$$K_{\beta\alpha}(x', t'; x, t) \rightarrow K_{\beta\alpha}(x'; x, t - t').$$

The Laplace transform is a convenient approach to handle Eqs. (1) and (3) because both equations are convolutions of time homogeneous processes. After the Laplace transform, the convolutions are reduced to algebraic products, which are simpler to manipulate. Moreover, it is easy to use Laplace transformed functions to derive the long time limit as we illustrate below.

We express the stationary flux in terms of the Laplace transform of the time dependent flux function. We define the Laplace transform of a function  $g(t)$  as  $\tilde{g}(u) = \int_{0^-}^{\infty} \exp(-ut) g(t) dt$ , where  $u$  is the Laplace variable, and the time integral starts from small negative values. For the flux function, we write

$$\tilde{q}_{\alpha}(x, u) = \int_{0^-}^{\infty} q_{\alpha}(x, t) \exp(-ut) dt. \quad (4)$$

Consider now a time  $\bar{t}$  that is finite but sufficiently long such that for all practical purposes, the flux  $q_{\alpha}(x, \bar{t})$  is time independent (stationary) and is equal to  $q_{\alpha, \text{stat}}(x)$ . We can separate Eq. (4) to two integrals

$$\begin{aligned} \tilde{q}_{\alpha}(x, u) &= \int_{0^-}^{\bar{t}} q_{\alpha}(x, t) \exp(-ut) dt \\ &\quad + \int_{\bar{t}}^{\infty} q_{\alpha, \text{stat}}(x) \exp(-ut) dt \\ &= \int_{0^-}^{\bar{t}} q_{\alpha}(x, t) \exp(-ut) dt \\ &\quad + q_{\alpha, \text{stat}}(x) \cdot \frac{1}{u} \exp(-u\bar{t}). \end{aligned}$$

Multiplying the above equation by the Laplace variable  $u$  and taking the limit of  $u \rightarrow 0$ , we have

$$\begin{aligned} \lim_{u \rightarrow 0} [u \cdot \tilde{q}_{\alpha}(x, u)] &= \lim_{u \rightarrow 0} \left[ u \cdot \int_{0^-}^{\bar{t}} q_{\alpha}(x, t) \exp(-ut) dt \right] \\ &\quad + \lim_{u \rightarrow 0} \left[ u \cdot q_{\alpha, \text{stat}}(x) \cdot \frac{1}{u} \exp(-u\bar{t}) \right], \quad (5) \\ \lim_{u \rightarrow 0} [u \cdot \tilde{q}_{\alpha}(x, u)] &= q_{\alpha, \text{stat}}(x). \end{aligned}$$

Similarly, we define the Laplace transform of the time homogeneous kernel

$$\tilde{K}_{\beta\alpha}(x'; x, u) = \int_{0^-}^{\infty} K_{\beta\alpha}(x'; x, \tau) \exp(-u\tau) d\tau. \quad (6)$$

Consider the Laplace transform of the kernel at the limit of  $u \rightarrow 0$ ,

$$\begin{aligned} \lim_{u \rightarrow 0} [\tilde{K}_{\beta\alpha}(x'; x, u)] &= \lim_{u \rightarrow 0} \left[ \int_{0^-}^{\infty} K_{\beta\alpha}(x'; x, \tau) \exp(-u\tau) d\tau \right] \\ &= \int_{0^-}^{\infty} K_{\beta\alpha}(x'; x, \tau) d\tau, \quad (7) \end{aligned}$$

$$\lim_{u \rightarrow 0} [\tilde{K}_{\beta\alpha}(x'; x, u)] \equiv K_{\beta\alpha}(x'; x),$$

Equation (7) has a simple physical interpretation as the probability that a trajectory will cross milestone  $\alpha$  at phase space point  $x$  for the first time given that it crossed milestone  $\beta$  at phase space point  $x'$  before. The actual time difference between the crossings no longer matters. For future reference, we also define the time independent kernel— $K_{\beta\alpha}(x'; x)$ .

Keeping Eq. (5) in mind, we now take the Laplace transform of Eq. (3), multiply the result by the Laplace variable  $u$ , and take the limit of  $u \rightarrow 0$ ,

$$\begin{aligned} \lim_{u \rightarrow 0} [u \cdot \tilde{q}_{\alpha}(x, u)] &= \lim_{u \rightarrow 0} [u \cdot p_{\alpha}(x, t = 0)] \\ &\quad + \lim_{u \rightarrow 0} \left[ \sum_{\beta} \int_{M_{\beta}} dx' \cdot u \cdot \tilde{q}_{\beta}(x', u) \tilde{K}_{\beta\alpha}(x'; x, u) \right]. \quad (8) \end{aligned}$$

The first term on the right hand side of Eq. (8) vanishes, as it should for initial conditions that have no impact on the stationary solution. We use Eqs. (5) and (7) to write

$$q_{\alpha, \text{stat}}(x) = \sum_{\beta} \int_{M_{\beta}} dx' q_{\beta, \text{stat}}(x') \cdot K_{\beta\alpha}(x'; x). \quad (9)$$

Equation (9) is a linear equation for the vector  $\mathbf{q}_{\text{stat}}$  with the milestone number as an index. The kernel is known exactly from short MD trajectories between milestones. Given a set of initial conditions at the interface (provided as a sample from the flux function), the kernel transformed these initial conditions by trajectory calculations into hitting points on nearby milestones. We can write Eq. (9) more compactly using bold-faced letters to denote vectors and matrices. We have

$$\mathbf{q}_{\text{stat}}^t = \mathbf{q}_{\text{stat}}^t \mathbf{K}. \quad (10)$$

Both the vector and the matrix are typically of high dimension. However, the kernel is known exactly and is a transition probability. This facilitates the computational algorithm, which is described next.

#### D. An algorithm to compute the stationary flux

Equation (10) is a linear problem on the unknown  $\mathbf{q}^t$  given the transition kernel (matrix)  $\mathbf{K}$ . The linear problem can be solved, in principle, by standard means. For example, the left eigenvector of the matrix  $\mathbf{K}$  with an eigenvalue of one is the desired solution for Eq. (10). All the absolute values of the eigenvalues of  $\mathbf{K}$  are between zero and one and the eigenvector with eigenvalue one is unique if the system is ergodic. This

also follows from the kernel being a transition matrix. It is therefore clear that multiplying  $\mathbf{q}'$  by  $\mathbf{K}$  numerous times will reduce the contributions of all the eigenvalues that are smaller than one and will retain the desired solution.<sup>31</sup>

We solve for the flux vector  $\mathbf{q}'$  Eq. (10) by power iterations.<sup>31</sup> The iterations are over flux vectors and are denoted by  $\{\mathbf{q}^{(n)t}\}_{n=0}^N$ , where  $N$  is the number of iterations used. The sequence of iterations is given by

$$\begin{cases} q_{\alpha}^{(0)}(x) \propto \exp(-\beta H(x)) & x \in M_{\alpha} \quad \forall \alpha \\ \mathbf{q}^{(n+1)t} = \mathbf{q}^{(n)t} \mathbf{K} & n = 0, 1, 2, \dots \end{cases} \quad (11)$$

The Boltzmann distribution follows from the NVT ensemble and is clearly not exactly the flux, hence the need for iterations. We generate a sample from the Boltzmann distribution by Monte Carlo (MC) or MD algorithms tailored for constant temperature simulations conditioned to be at milestone  $\alpha$ . A better initial guess can be obtained by computing forward and backward trajectories to the nearby interfaces. The phase space points, provided by the exact flux function, are sampled from a first hitting distribution. That is, integration backward in time of the phase space points yields trajectories that cross first a milestone different from the starting milestone (Figure 3). This picture was discussed first in the context of PPTIS.<sup>7</sup>

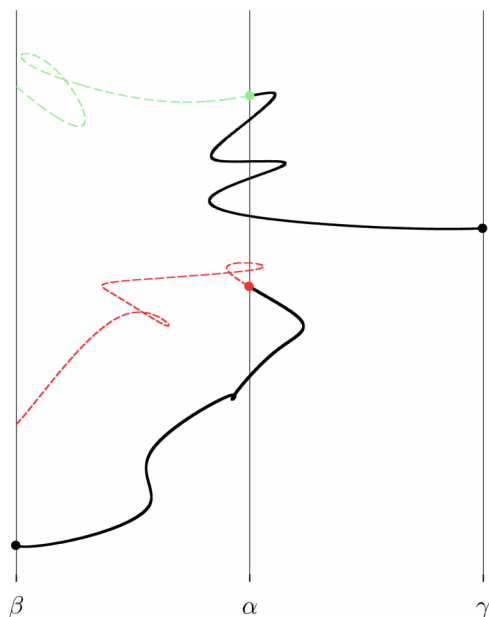


FIG. 3. A schematic drawing of trajectory fragments initiated at milestone  $\alpha$ . Backward trajectories in time (dashed lines) are used to assess if the trajectories are sampled from FHPD. The red dashed trajectory re-crosses the initial milestone  $\alpha$  before hitting milestone  $\beta$ . Therefore, it is not sampled from FHPD. The green trajectory hits milestone  $\beta$  before any other milestone and is sampled from FHPD. Note that the forward trajectories (black lines) are allowed to re-cross the initial milestone before they finally terminate at a nearby milestone. The forward trajectories illustrate the operation of the transition matrix  $\mathbf{K}$  that takes phase space points from one milestone and transmits them with certain probability to phase space points at another milestone. For Milestoning calculations, we also record the time length of the forward trajectory fragments (black lines). It is possible to exploit the identity of the terminating milestones of the backward trajectories (dashed lines) as well, but so far we have not done so. The final termination points of the trajectories (for example, the black circle at milestone  $\beta$ ) are used as starting points for the next iteration of the flux  $\mathbf{q}^{(n+1)}$ .

We implemented this condition for acceptance of trajectories in directional Milestoning<sup>16</sup> to increase the accuracy of the initial flux function avoiding the over counting of trajectories that cross the same interface multiple times. Each trajectory must be counted only once regardless of the number of times it crosses sequentially and repeatedly the same interface. Hawk and Makarov introduced a filter that extends to more than one milestone.<sup>32</sup>

Even this filter is not providing the exact flux function since we have no record of the fate of the trajectories beyond the nearby interfaces. It is possible that the distribution at a nearby interface is still inaccurate. The condition provided in Eq. (11) is global and it cannot be satisfied exactly by local considerations unless iterations are applied. In principle, the initial guess for the flux function should not affect the final results, however, the number of iterations that is required to obtain the correct answer is influenced by the initial condition. It makes sense to try to determine a best approximation to the flux function locally and to reduce the number of required (expensive) iterations.

The iterations have a simple physical interpretation. We initiate trajectories at a milestone and continue these trajectories to the nearby milestones. The next iteration uses the final phase space point of the previous terminating trajectories at a milestone as initial conditions and continues these trajectories to yet another layer of milestones. Hence, each of the iterations extends the time length of the trajectories. If a very large number of iterations are used, exact long trajectories are generated. Hence, it is not surprising that Milestoning becomes exact at the limit of a large number of power iterations. Of course, this particular scenario suggests no computational advantage compared to straightforward molecular dynamics. It only suggests a formal connection to exact calculations and an analysis method.<sup>29</sup> We expect, as we illustrate a number of times in the past, that if the system is close to equilibrium, rapid convergence to the correct flux can be achieved using good guesses (and only short trajectories).

The flux is not normalized in the usual sense. For convenience, we write

$$q_{\alpha, \text{stat}}(x) = w_{\alpha} f_{\alpha}(x), \quad (12)$$

where  $w_{\alpha}$  is an undetermined weight that defines the relative contributions of different milestones to the overall flux. This weight is a prime target of the computations since it is useful in various numerical variants of Milestoning. For example,  $w_{\alpha}$  determines the weights of edges for molecular kinetic networks.<sup>33</sup> It is the only term that carries information about the kinetics in the zero iteration, since the distribution within the milestone ( $f_{\alpha}^{(0)}(x)$ ) is of equilibrium. Hence, a physical interpretation of the zero iteration is of a system that is equilibrated within a milestone but not necessarily between milestones. The local equilibrium within a milestone is the “memory loss” approximation that was used in the first version of Milestoning.<sup>34</sup> It is assumed that trajectories spend enough time between the milestones such that the distribution on the terminating milestone is of local equilibrium. This separation of time scales can be achieved by proper selection of the interfaces. For example, if a picture of a single reaction

coordinate is appropriate and if the space normal to the reaction coordinate is small and local equilibrium in the milestone can be achieved rapidly, then the memory loss assumption is likely to be valid.<sup>14</sup> In a reduced coarse-grained description of the system, only the milestones and their weights remain.

We emphasize that the above considerations of “memory loss” are not required in exact Milestoning. In exact Milestoning, the iterations guarantee convergence to the exact answer regardless of the initial guess for the flux function. Poor initial guesses may cause however slower convergence rate.

The function  $f_\alpha(x)$  is normalized,

$$\int f_\alpha(x) dx = 1. \quad (13)$$

Note that the Eq. (10) determines the overall flux only up to a multiplying constant,  $\lambda$ , which we take to be positive. Hence, if  $(q_{\alpha,stat}(x) = w_\alpha f_\alpha(x))_{\alpha=1}^M$  is a solution of Eq. (10), so is the vector  $(\lambda \cdot w_\alpha f_\alpha(x))_{\alpha=1}^M$ .

The second line of Eq. (11) requires the calculations of a high dimensional integral over the phase space points  $x$ . Each phase space point is used to initiate a trajectory fragment that runs to termination, which is formally equivalent to the operation of  $\mathbf{K}$  on the initial phase space point. We calculate this integral by sampling in phase space. The sampling is conducted according to the flux  $\mathbf{q}^{(n)}$ . In the first iteration, the sample is determined by the canonical distribution. In the next iteration, we use points that were created by terminating trajectories at the milestones. Let the set of  $n$ th iteration phase space points at milestone  $\alpha$  be  $\{x_{\alpha i}^{(n)}\}_{i=1}^{L_\alpha}$ , where  $L_\alpha$  is the number of sampled points. The probability density associated with this set is

$$q_{\alpha,stat}^{(n)}(x) = \frac{w_\alpha^{(n)}}{L_\alpha} \sum_{i=1, \dots, L_\alpha} \delta(x - x_{\alpha i}^{(n)}) \quad (14)$$

which is equivalent to approximating the function  $f_\alpha(x)$  in Eq. (12) by a sum of Dirac's delta functions. Operating on this set yields a new set of phase space points using Eq. (11),

$$\begin{aligned} \frac{w_\alpha^{(n+1)}}{L_\alpha} \sum_{i=1, \dots, L_\alpha} \delta(x - x_{\alpha i}^{(n+1)}) \\ = \sum_\beta \frac{w_\beta^{(n)}}{L_\beta} \int \sum_{j=1, \dots, L_\beta} \delta(x' - x_{\beta j}^{(n)}) \mathbf{K}_{\beta\alpha}(x', x) dx', \\ w_\alpha^{(n+1)} = \sum_\beta w_\beta^{(n)} \\ \times \left[ \sum_{i=1, \dots, L_\alpha} \sum_{j=1, \dots, L_\beta} \mathbf{K}_{\beta\alpha}(x_{\beta j}^{(n)}, x_{\alpha i}^{(n+1)}) / (L_\alpha L_\beta) \right]. \end{aligned} \quad (15)$$

The second line is obtained by integration of the phase space variable  $x$  in the first line of Eq. (15). The final formula is a direct equation for the vector of weights of the milestones  $\mathbf{w}$ . At the limit of sufficient number of iterations, we have  $\mathbf{w}^{(n+1)} \rightarrow \mathbf{w}^{(n)}$ . It is convenient to re-write Eq. (15) as a linear equation for the vector of coefficients  $\mathbf{w}^{(n+1)}$ , which is consistent with the approximate variation of Milestoning.<sup>12</sup> Equation (15) is a power iteration of the type  $\mathbf{w}^{(n+1)t} = \mathbf{w}^{(n)t} \mathbf{K}$

and it converges to the dominant eigenvector of the matrix  $\mathbf{K}$ . It can also be cast as an eigenvector problem and, as such, one can use an eigenvalue solver with better convergence. Both equations converge to the same limit,

$$\begin{aligned} \frac{w_\alpha^{(n+1)}}{L_\alpha} \sum_{i=1, \dots, L_\alpha} \delta(x - x_{\alpha i}^{(n+1)}) \\ = \sum_\beta \frac{w_\beta^{(n+1)}}{L_\beta} \int \sum_{j=1, \dots, L_\beta} \delta(x' - x_{\beta j}^{(n)}) \mathbf{K}_{\beta\alpha}(x', x) dx', \\ w_\alpha^{(n+1)} = \sum_\beta w_\beta^{(n+1)} \\ \times \left[ \sum_{i=1, \dots, L_\alpha} \sum_{j=1, \dots, L_\beta} \mathbf{K}_{\beta\alpha}(x_{\beta j}^{(n)}, x_{\alpha i}^{(n+1)}) / (L_\alpha L_\beta) \right]. \end{aligned} \quad (16)$$

In practical applications, the length of the vector and the dimensionality of the transition matrix, once the phase space variables are integrated out, are manageable and so far have not exceeded tens of thousands. With the weight of the milestones calculated, the fluxes in Eq. (14) can be determined as well.

An important question is the error assessment and tests of the convergence of the iterations. In general, we use ensemble averages to test convergence. This is because we can estimate them with reasonable accuracy even for large systems. For example, we may consider  $w_\alpha \equiv \int_{M_\alpha} q_\alpha(x_\alpha) dx_\alpha$ . However, for the small and simple test case, we consider here it is possible to look at the results at greater details. We therefore introduce also the Rayleigh quotient  $r$  for the fluxes which is given by

$$r \equiv \frac{\langle \mathbf{q}^{(n+1)}, \mathbf{q}^{(n)} \rangle}{\langle \mathbf{q}^{(n)}, \mathbf{q}^{(n)} \rangle} = \frac{\langle \mathbf{q}^{(n)} \mathbf{K}, \mathbf{q}^{(n)} \rangle}{\langle \mathbf{q}^{(n)}, \mathbf{q}^{(n)} \rangle}, \quad (17)$$

where the  $\langle \cdot, \cdot \rangle$  is the  $L_2$  inner product. The Rayleigh quotient is bound between minus one and one since the kernel  $\mathbf{K}$  does not increase the length of a vector. It is equal to one if the results are converged. Our representation of the flux density uses Dirac delta functions.

The second convergence measure that we used is the relative error in the 1-norm,  $\Delta$ , between flux vectors of sequential iterations,

$$\Delta_\alpha = \left\| q_\alpha^{(n-1)} - q_\alpha^{(n)} \right\|_1 / \left\| q_\alpha^{(n)} \right\|_1. \quad (18)$$

The simulations converge when  $\Delta_\alpha = 0$ . These measures are useful also for conducting the iterations optimally. It is possible that fluxes at selected milestones vary slowly or converge rapidly. Therefore an adaptive iterative scheme is possible and likely to show efficiency gains. We illustrate this phenomenon in Sec. III.

We summarize below the algorithm for computing the stationary flux  $\mathbf{q}$ .

1. For each milestone  $M_\alpha$ , sample configurations,  $x_\alpha$ , from a known trial distribution, typically canonical  $q_\alpha(x) \propto \exp(-H(x)/kT)$   $x \in M_\alpha$ .

2. Compute forward trajectories from the points  $x_\alpha$  sampled from the current  $q_\alpha$  for all  $\alpha$ . We record the initial and final points of the trajectory fragments as well as the time

length. This operation samples the matrix product  $\mathbf{q}' \cdot \mathbf{K}$ . The time length of the trajectories is used in other calculations.

3. Estimate the total flux through the interface  $\alpha$  by summing up the points that terminate at  $\alpha$ . This operation is equivalent to the integration  $w_{\alpha}^{(n+1)} = \int_{M_{\alpha}} q_{\alpha}^{(n+1)}(x_{\alpha}) dx_{\alpha}$ .

4. Check convergence of the iterations: (1)  $|\langle \mathbf{qK}, \mathbf{q} \rangle / \langle \mathbf{q}, \mathbf{q} \rangle - 1| < \varepsilon$ , (2)  $\|q_{\alpha}^{(n-1)} - q_{\alpha}^{(n)}\|_1 / \|q_{\alpha}^{(n)}\|_1 \leq \varepsilon$ , or by (3) observables of interest. If the calculation has not converged, use the final phase space points of the trajectory fragment to obtain a new  $q_{\alpha}$  and go to 2.

The stationary flux function is the basic entity of the Milestoning theory and algorithm. However, it is not a common experimental observable. To make the connection to the more typical observables of equilibrium and kinetics, we show below how (with the help of the flux function) we can determine the probability of a state and its free energy (Sec. II E). We follow up with the calculations of the moments of the first passage time (Sec. II F).

We finally comment that the original version of Milestoning<sup>9</sup> is a simplified form of the exact Milestoning formulation. In the original version, no iterations were conducted and the trial distribution  $q_{\alpha}(x) \propto \exp(-H(x)/kT)$   $x \in M_{\alpha}$  was used in the equation for the weights of the milestones (Eq. (15)). There was no further refinement of the distribution within the milestone. Hence, the memory loss approximation was used.

## E. The stationary probability

We consider the probability,  $p_{\alpha}(x, t)$ , that  $M_{\alpha}$  is the last milestone that was passed at a phase space point  $x$  before time  $t$ . At stationary or equilibrium conditions, the dependence on time can be omitted and we use this probability to define a corresponding “free energy” as  $F_{\alpha}(x) = -(k_B T) \cdot \log(p_{\alpha}(x))$ , where the temperature,  $T$ , and the Boltzmann constant,  $k_B$ , enters the calculations through the initial conditions of the trajectories (see Sec. II D and Eq. (11)). As we have done for the flux, we consider the Laplace transform for the probability (Eq. (1)) to have

$$\tilde{p}_{\alpha}(x, u) = \tilde{q}_{\alpha}(x, u) \cdot \frac{1}{u} \left[ 1 - \sum_{\beta} \int_{M_{\beta}} dx' \tilde{K}_{\alpha\beta}(x, x', u) \right]. \quad (19)$$

We multiply Eq. (21) by the Laplace variable  $u$  and consider the limit in which it is approaching zero. This is the long time limit that gives the stationary solution of the probability (similar to the limit of the flux function in Eq. (5)),

$$\lim_{u \rightarrow 0} u \cdot \tilde{p}_{\alpha}(x, u) = p_{\alpha, \text{stat}}(x). \quad (20)$$

We elaborate on the expression on the right hand side of Eq. (19),

$$\begin{aligned} & \lim_{u \rightarrow 0} \frac{1}{u} \left[ 1 - \sum_{\beta} \int_{M_{\beta}} dx' \tilde{K}_{\alpha\beta}(x, x', u) \right] \\ &= \lim_{u \rightarrow 0} \frac{1}{u} \left[ 1 - \sum_{\beta} \int_{0-}^{\infty} \int_{M_{\beta}} dt \cdot dx' \cdot \exp(-ut) K_{\alpha\beta}(x, x', t) \right] \end{aligned}$$

$$\begin{aligned} &= \lim_{u \rightarrow 0} \frac{1}{u} \left[ 1 - \sum_{\beta} \int_{0-}^{\infty} \int_{M_{\beta}} dt \cdot dx' \cdot \{1 - ut\} K_{\alpha\beta}(x, x', t) \right] \\ &= \lim_{u \rightarrow 0} \frac{1}{u} \left[ 1 - \sum_{\beta} \int_{0-}^{\infty} dt \int_{M_{\beta}} dx' K_{\alpha\beta}(x, x', t) \right. \\ &\quad \left. + u \sum_{\beta} \int_{0-}^{\infty} dt \int_{M_{\beta}} dx' \cdot t \cdot K_{\alpha\beta}(x, x', t) \right] \\ &= \lim_{u \rightarrow 0} \frac{1}{u} \left[ 1 - 1 + u \sum_{\beta} \int_{0-}^{\infty} dt \int_{M_{\beta}} dx' \cdot t \cdot K_{\alpha\beta}(x, x', t) \right] \\ &= \sum_{\beta} \int_{0-}^{\infty} dt \int_{M_{\beta}} dx' \cdot t \cdot K_{\alpha\beta}(x, x', t) \\ &\equiv \langle t_{\alpha}(x) \rangle, \end{aligned} \quad (21)$$

where we denote the average lifetime of a position  $x_{\alpha}$  at milestone  $\alpha$  as  $\langle t_{\alpha}(x) \rangle$ . It is the average time that it takes a trajectory initiated at  $x_{\alpha}$  to reach a milestone different from  $\alpha$  for the first time. We now combine Eqs. (19) to (21) to obtain

$$\lim_{u \rightarrow 0} [\tilde{p}_{\alpha}(x, u)] = \lim_{u \rightarrow 0} \left[ \tilde{q}_{\alpha}(x, u) \cdot \frac{1}{u} \left[ 1 - \sum_{\beta} \int_{M_{\beta}} dx' \tilde{K}_{\alpha\beta}(x, x', u) \right] \right], \quad (22)$$

$$p_{\alpha, \text{stat}}(x) = q_{\alpha, \text{stat}}(x) \cdot \langle t_{\alpha}(x) \rangle.$$

The last expression is a remarkably simple result for the probability of being at state  $\alpha$ . It is a product of the stationary flux function to cross milestone  $\alpha$  multiplied by the average time that the crossing trajectory lives in this state (before crossing another milestone). We call this time—the lifetime of a milestone. Compared to the previous calculations of the flux function, we need to add only a single new entity (the milestone lifetime) that can be computed from the same trajectory fragments that we used to estimate the flux. If we sample  $L_{\alpha}$  trajectories at milestone  $\alpha$  and phase space point  $x$  with termination times at other milestones  $\{t_i\}_{i=1}^{L_{\alpha}}$ , the lifetime of milestone  $\alpha$  at phase space point  $x$  is  $\langle t_{\alpha}(x) \rangle = \frac{1}{L_{\alpha}} \sum_{i=1, \dots, L_{\alpha}} t_i$ . Hence, no new trajectories are required.

## F. The mean first passage time

To study kinetics, we focus on the first passage time. The first passage time is defined as the time it takes the system to reach for the first time a state  $f$  given that it started from a state  $i$ . The first passage time is a random variable that can be sampled with trajectory calculations or its distribution function can be computed. We prefer to calculate the moments of the distribution function of the first passage time. In particular, the first moment of this distribution,  $\langle \tau \rangle$ , or the MFPT is widely used to study kinetics. For reactants' population that decays exponentially in time, the MFPT is the inverse of the rate constant.



To model the first arrival to the final state, we consider a specific choice of the kernel with an absorbing boundary at milestone  $f$  which we call  $K_A$ . We set

$$K_{A,f\alpha}(x; x', t) = 0 \quad \forall \alpha. \quad (23)$$

The MFPT or  $\langle \tau \rangle$  is the time that it takes a trajectory to enter the absorbing state multiplied by the probability to enter the absorbing state at time  $t$ , averaged over all times. In other words, it is the first moment in time of the flux at milestone  $f$ ,

$$\begin{aligned} \langle \tau(x) \rangle &= \int_{0^-}^{\infty} t \cdot q_f(x, t) dt \text{ or } \langle \tau_f \rangle \\ &= \int_{M_f} dx_f \int_{0^-}^{\infty} dt \cdot t \cdot q_f(x, t). \end{aligned} \quad (24)$$

It is important to note that the flux as defined in Eq. (24) is no longer stationary. This is a result of the choice of the kernel made in Eq. (23). The kernel is no longer conserving and the flux is decaying as a function of time. This is necessary since Eq. (24) is diverging for a non-zero stationary flux.

We can write the integral over time as a derivative in Laplace space,

$$\begin{aligned} \int_{0^-}^{\infty} t \cdot q_f(x, t) dt &= \lim_{u \rightarrow 0} \int_{0^-}^{\infty} \exp(-ut) t \cdot q_f(x, t) dt \\ &= - \lim_{u \rightarrow 0} \frac{d}{du} \int_{0^-}^{\infty} \exp(-ut) q_f(x, t) dt \\ &= - \lim_{u \rightarrow 0} \frac{d}{du} \tilde{q}_f(x, u). \end{aligned} \quad (25)$$

We already discussed  $\tilde{q}_\alpha(x_\alpha, u)$  (see, for instance, Eq. (8)) which we could further exploit for our purpose here,

$$\begin{aligned} \tilde{q}_\alpha(x, u) &= p_\alpha(x, t=0) \\ &+ \sum_{\beta} \int_{M_\beta} dx' \cdot \tilde{q}_\beta(x', u) \tilde{K}_{A,\beta\alpha}(x', x_\alpha, u), \\ \frac{d\tilde{q}_\alpha(x, u)}{du} &= \sum_{\beta} \int_{M_\beta} dx' \left[ \frac{d\tilde{q}_\beta(x', u)}{du} \tilde{K}_{A,\beta\alpha}(x', x, u) \right. \\ &\quad \left. + \tilde{q}_\beta(x', u) \frac{d\tilde{K}_{A,\beta\alpha}(x', x, u)}{du} \right]. \end{aligned} \quad (26)$$

Or in a more compact matrix notation,

$$\begin{aligned} \tilde{\mathbf{q}}^t (\mathbf{I} - \tilde{\mathbf{K}}_A) &= \mathbf{p}^t(t=0), \\ \frac{d\tilde{\mathbf{q}}^t}{du} (\mathbf{I} - \tilde{\mathbf{K}}_A) - \tilde{\mathbf{q}}^t \frac{d\tilde{\mathbf{K}}_A}{du} &= 0. \end{aligned} \quad (27)$$

We determine the derivative of  $\tilde{\mathbf{q}}$  with respect to the Laplace variable in the compact notation

$$\begin{aligned} \frac{d\tilde{\mathbf{q}}^t}{du} &= \tilde{\mathbf{q}}^t \frac{d\tilde{\mathbf{K}}_A}{du} (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \\ &= \mathbf{p}^t(t=0) \cdot (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \frac{d\tilde{\mathbf{K}}_A}{du} (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1}. \end{aligned} \quad (28)$$

We define the matrix  $\mathbf{K}_A \equiv \lim_{u \rightarrow 0} \tilde{\mathbf{K}}_A = \int_{0^-}^{\infty} \mathbf{K}_A(t) dt$ . Hence, whenever we do not write explicit time dependence, we imply integration over time.

We also consider the matrix of local average transition times between the milestones,  $\mathbf{T}$ . This matrix is also minus the derivative of the Laplace transform of the matrix  $\tilde{\mathbf{K}}_A(u)$  at the limit of zero Laplace variables. It has a simple physical interpretation. It is the first moment of time of the probability transition matrix  $\mathbf{K}_A(t)$ ,

$$\begin{aligned} \mathbf{T} &= - \lim_{u \rightarrow 0} \frac{d\tilde{\mathbf{K}}_A(u)}{du} \\ &= - \lim_{u \rightarrow 0} \frac{d}{du} \left[ \int_{0^-}^{\infty} \exp(-ut) \cdot \mathbf{K}_A(t) dt \right] \\ &= \int_{0^-}^{\infty} t \cdot \mathbf{K}_A(t) dt. \end{aligned} \quad (29)$$

Note that  $\langle \tau \rangle = - \lim_{u \rightarrow 0} \frac{d\tilde{q}_f}{du}$  so the negative signs of  $\mathbf{T}$  and  $\langle \tau \rangle$  that emerge from the differentiation with respect to the Laplace variable cancel out to give

$$\langle \tilde{\tau} \rangle^t = \mathbf{p}^t(t=0) \cdot (\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{T} (\mathbf{I} - \mathbf{K}_A)^{-1}. \quad (30)$$

We note that the transition time is a function of the phase space points at the two interfaces. Interestingly, the explicit dependence on the absorbing milestone disappears in the final expression as is illustrated by further manipulations below. This is typical to Markovian processes in which the MFPT depends only on the starting and not the end point.<sup>9</sup>

The MFPT in Eq. (30) is a vector that contains information on all milestones. Here, we are interested only in the average time to reach for the first time the absorbing milestone  $f$ . Hence, we are interested in a single element of the vector  $(\langle \tilde{\tau} \rangle^t)_f$ . To obtain this single element, we multiply Eq. (32) from the right by a column vector  $\mathbf{e}_f(x_f)$  which is a unit vector in the direction of the absorbing milestone. The scalar product implies an integration over the phase space points of the absorbing milestone,  $x_f$ , to have

$$\langle \tilde{\tau} \rangle^t \cdot \mathbf{e}_f = \mathbf{p}^t(t=0) \cdot (\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{T} (\mathbf{I} - \mathbf{K}_A)^{-1} \cdot \mathbf{e}_f. \quad (31)$$

As stated in the definition the rows of the transition matrix  $\mathbf{K}$  (including the integration over the phase space points in the milestone) must add to one. The matrix  $\mathbf{K}_A$  is different since the elements of the row of the absorbing state in  $\mathbf{K}_A$ , and obviously their sum, are zero. As a result, the matrix  $\mathbf{K}_A$  has non-negative eigenvalues smaller than one. This implies that the inverse of  $(\mathbf{I} - \mathbf{K}_A)$  in Eq. (31) is well defined. The summation of the rows of  $\mathbf{K}_A$  can be written as

$$\mathbf{K}_A \cdot \mathbf{1} = \mathbf{1} - \mathbf{e}_f, \quad (32)$$

where  $\mathbf{1}$  is a column vector. Each of the elements of  $\mathbf{1}$  is equal 1. Using Eq. (32), we can write

$$(\mathbf{I} - \mathbf{K}_A) \mathbf{1} = \mathbf{e}_f, \mathbf{1} = (\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{e}_f. \quad (33)$$

Substituting in Eq. (31), we have

$$\langle \tilde{\tau} \rangle^t \cdot \mathbf{e}_f = \mathbf{p}^t(t=0) \cdot (\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{T} \cdot \mathbf{1}. \quad (34)$$

Consider the matrix vector product  $\mathbf{t} \equiv \mathbf{T} \cdot \mathbf{1}$ . The resulting elements of the column vector  $\mathbf{t}$  are the summation of the row of the matrix  $\mathbf{T}$ . For example,

$$\begin{aligned} t_\alpha(x) &= \sum_\beta \int_{M_\beta} T_{\alpha\beta}(x, x') dx' \\ &= \sum_\beta \int_0^\infty \int_{M_\beta} dt \cdot dx' \cdot t \cdot K_{A,\alpha\beta}(x, x', t). \end{aligned} \quad (35)$$

Equation (36) therefore takes its final compact form

$$\langle \vec{\tau} \rangle_f = \mathbf{p}^t(t=0) \cdot (\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{t}. \quad (36)$$

Equation (36) is central to the understanding of kinetics as the MFPT plays an important role in many theories, experiments, and simulations of molecular processes. Interestingly, if the kernel is replaced by a single exponential function,<sup>9</sup>  $K_{\alpha\beta} \propto \exp(-t/t_\alpha)$  which implies a Markov process,<sup>34,37</sup> we obtain the same expression as Eq. (36) where  $t_\alpha = (\mathbf{t})_\alpha$ .

The expression in Eq. (36) can be difficult to compute in exact Milestoning since the solution of a large linear system  $(\mathbf{I} - \mathbf{K}_A)$  or a geometric series expansion in  $\mathbf{K}_A$  is required. In the original version of Milestoning, this was of a smaller concern since the matrix depends only on the milestone index and not on the phase space point in the Milestone. In exact Milestoning, we retain the dependence on the phase space point. To find a simpler expression that depends only on the flux vector, we define a new kernel  $\mathbf{K}_C$  ( $C$  for cyclic) in which the flux absorbed at milestone  $f$  is instantaneously transported to the initiating milestone(s),

$$\mathbf{K}_A = \mathbf{K}_C - \mathbf{e}_f \mathbf{p}^t(t=0). \quad (37)$$

An intriguing relationship is stated below,

$$q_f \mathbf{p}^t(t=0) = \mathbf{q}^t (\mathbf{I} - \mathbf{K}_A). \quad (38)$$

Equation (38) is demonstrated in Eq. (39) in which we used the condition for a stationary solution  $\mathbf{q}^t = \mathbf{q}^t \mathbf{K}_C$  for a conserving  $\mathbf{K}$ ,

$$\begin{aligned} \mathbf{q}^t (\mathbf{I} - \mathbf{K}_A) &= \mathbf{q}^t - \mathbf{q}^t \mathbf{K}_A \\ &= \mathbf{q}^t - \mathbf{q}^t (\mathbf{K}_C - \mathbf{e}_f \mathbf{p}^t(t=0)) \\ &= \mathbf{q}^t - \mathbf{q}^t \mathbf{K}_C + (\mathbf{q}^t \cdot \mathbf{e}_f) \mathbf{p}^t(t=0) \\ &= \mathbf{q}^t - \mathbf{q}^t + q_f \mathbf{p}^t(t=0) \\ &= q_f \mathbf{p}^t(t=0). \end{aligned} \quad (39)$$

Multiplying Eq. (36) by  $q_f$ , we have

$$\begin{aligned} q_f \langle \vec{\tau} \rangle_f &= q_f \mathbf{p}^t(0) \cdot (\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{t} = \mathbf{q}^t (\mathbf{I} - \mathbf{K}_A) (\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{t}, \\ q_f \langle \vec{\tau} \rangle_f &= \mathbf{q}^t \cdot \mathbf{t}, \\ \langle \vec{\tau} \rangle_f &= (\mathbf{q}^t \cdot \mathbf{t}) / q_f. \end{aligned} \quad (40)$$

Equation (40) is a major result for the calculation of the MFPT since it includes only vectors. It is a product of the vector of fluxes and the vector of local life times of the milestones. Hence, the short trajectories described in the flux and stationary probability calculations are used again to compute the MFPT with no additional cost.

Equation (40) is similar to an expression derived by Reimann, Schmid, and Hänggi (RSH) for the MFPT for

non-Markovian stationary processes.<sup>38</sup> We discuss below the connection between their formula and Eq. (40).

We denote the alternative definition by  $\langle \tau_{RSH} \rangle$ ,

$$\langle \tau_{RSH} \rangle = \frac{\sum_\alpha \int_{M_\alpha} p_{\alpha,stat}(x) dx}{\int_{M_i} q_{i,stat}(x) dx}. \quad (41)$$

The flux  $q_i$  is through the  $i$ th milestone that feeds new trajectories into the system until a stationary flux is reached with the balancing of the absorbing boundary. The absorbing milestones are set in such a way that their own flux is matched by the flux into the source milestone (and hence the use of  $\mathbf{K}$  below). In Milestoning, we emphasize the use of fluxes and therefore write an adjusted expression, which is equivalent to Eq. (41). The sum in the denominator in our case is over the flux into the absorbing state, while RSH sum the flux to the initial state. However, because of the cyclic boundary conditions, these two fluxes are the same,

$$\langle \tau_{RSH} \rangle = \frac{\sum_\alpha \int_{M_\alpha} q_{\alpha,stat}(x) t_\alpha(x) dx}{\int_{M_f} q_{f,stat}(x) dx} \equiv (\mathbf{q} \cdot \mathbf{t}) / q_{f,stat}. \quad (42)$$

Further exploitation of Laplace transforms allows us to derive expressions for higher moments of the MFPT. For example, the second moment of the first passage time is

$$\langle \tau^2(x_f) \rangle = \int_{0^-}^{\infty} t^2 q_f(x_f, t) dt = \lim_{u \rightarrow 0} \frac{d^2 \tilde{q}_f(x_f, u)}{du^2}. \quad (43)$$

Based on the collection of equalities, we derived earlier

$$\begin{aligned} \tilde{\mathbf{q}}^t (\mathbf{I} - \tilde{\mathbf{K}}_A) &= \mathbf{p}^t(t=0), \quad \tilde{\mathbf{q}}^t = \mathbf{p}^t(t=0) (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1}, \\ \left( \frac{d\tilde{\mathbf{q}}}{du} \right)^t (\mathbf{I} - \tilde{\mathbf{K}}_A) - \tilde{\mathbf{q}}^t \frac{d}{du} \tilde{\mathbf{K}}_A &= 0, \\ \left( \frac{d\tilde{\mathbf{q}}}{du} \right)^t &= \mathbf{p}^t(t=0) (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \frac{d\tilde{\mathbf{K}}_A}{du} (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1}. \end{aligned}$$

Differentiating Eq. (28) with respect to the Laplace variable, we have

$$\begin{aligned} \left( \frac{d^2 \tilde{\mathbf{q}}}{du^2} \right)^t (\mathbf{I} - \tilde{\mathbf{K}}_A) - 2 \left( \frac{d\tilde{\mathbf{q}}}{du} \right)^t \frac{d\tilde{\mathbf{K}}_A}{du} - \tilde{\mathbf{q}}^t \frac{d^2 \tilde{\mathbf{K}}_A}{du^2} &= 0, \\ \left( \frac{d^2 \tilde{\mathbf{q}}}{du^2} \right)^t (\mathbf{I} - \tilde{\mathbf{K}}_A) &= 2 \mathbf{p}^t(t=0) (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \frac{d\tilde{\mathbf{K}}_A}{du} (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \frac{d\tilde{\mathbf{K}}_A}{du} \\ &\quad + \mathbf{p}^t(t=0) (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \frac{d^2 \tilde{\mathbf{K}}_A}{du^2}, \\ \left( \frac{d^2 \tilde{\mathbf{q}}}{du^2} \right)^t &= \mathbf{p}^t(t=0) (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \\ &\quad \times \left[ 2 \frac{d\tilde{\mathbf{K}}_A}{du} (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1} \frac{d\tilde{\mathbf{K}}_A}{du} + \frac{d^2 \tilde{\mathbf{K}}_A}{du^2} \right] (\mathbf{I} - \tilde{\mathbf{K}}_A)^{-1}. \end{aligned} \quad (44)$$

As in the calculation for the MFPT, we multiply from the right by the unit vector  $\mathbf{e}_f$ , and using the relationship  $(\mathbf{I} - \mathbf{K}_A)^{-1} \mathbf{e}_f = \mathbf{1}$  in the limit  $u \rightarrow 0$ , we have

$$\begin{aligned}\langle \tau^2 \rangle_f &= \mathbf{p}^t(t=0)(\mathbf{I} - \mathbf{K}_A)^{-1} [2\mathbf{T}(\mathbf{I} - \mathbf{K}_A)^{-1}\mathbf{t} + \mathbf{T}^{(2)}\mathbf{1}], 6pt \\ \langle \tau^2 \rangle_f &= (\mathbf{q}^t/q_f) [2\mathbf{T}(\mathbf{I} - \mathbf{K}_A)^{-1}\mathbf{t} + \mathbf{t}^{(2)}],\end{aligned}\quad (45)$$

where we define the second moment matrix and time

$$\mathbf{T}^{(2)} = \int_0^\infty t^2 \mathbf{K}_A(t) dt, \quad \mathbf{t}^{(2)} = \mathbf{T}^{(2)}\mathbf{1}. \quad (46)$$

This completes the illustrative derivation of the second moment of the first passage time.

Equation (9) for the stationary flux, Eq. (22) for the stationary probability, Eqs. (36) and (40) for the MFPT, and Eq. (45) for the average of the second moment of the first passage time are the main results of this section.

### III. AN ILLUSTRATION

#### A. A model with an entropy barrier

We consider the two-dimensional energy landscape discussed in Ref. 9. It consists of a double well potential with the basins connected through a narrow channel. This model for entropic barrier was used in the past to evaluate the quality of the Milestoning approximation.<sup>9</sup> Here, we use the same model to evaluate the impact of the iterations on the accuracy and efficiency of the exact Milestoning algorithm. The potential energy (see Figure 4) is

$$U(x, y) = x^6 + y^6 + \exp(-x/\sigma)^2 [1 - \exp(-y/\sigma)^2]. \quad (47)$$

The numerical results shown below were computed with  $\sigma = 1/10$ .

#### B. Model for the dynamics

For illustration purposes, we focus on a low dimensionality and a simple model that we can solve accurately. The two dimensional double well potential, which is described in III A, has this advantage. However, a complication of

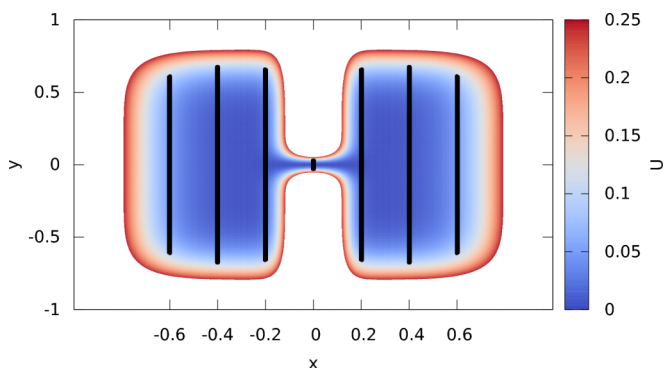


FIG. 4. A contour plot of the two-dimensional energy landscape that is used in the text to illustrate the exact Milestoning algorithm. We consider a transition from the well in the left to the well on the right. Also shown are the milestones (thick black lines). Note the narrow channel connecting the two wells. The channel does not include an energy barrier. Hence, the model considered is of an entropic barrier.

low dimensionality systems can be the lack of ergodicity. Milestoning requires ergodicity since the transition kernel must connect all states. To ensure ergodicity in the current test case, we use a stochastic model for the dynamics, overdamped Langevin

$$\frac{dz}{dt} = -\nabla U + R, \quad (48)$$

where  $z$  is a two dimensional vector  $z = (x, y)$  and  $R$  is a random force that follows the fluctuation dissipation theorem  $\langle R \rangle = 0$  and  $\langle R(t) R(t') \rangle = 2kT\delta(t - t')$ . Milestoning can be used on a diverse set of ODEs and SDEs, including the overdamped Langevin equation. To solve Eq. (48) and obtain the coordinates  $z$  as a function of time, we used the BAOAB algorithm.<sup>39</sup> The temperature,  $kT$ , was 0.025 in all calculations with the exception of the study of the MFPT in which it was 0.008. The lower temperature helped us establish clearer dynamic characteristics of an activated process. However, the lower temperature requires more expensive calculations and we prefer to do the other studies at higher temperatures. The time step was  $\Delta t = 10^{-4}$ . Straightforward long time trajectories from the initiating milestone to the last (absorbing) milestone were computed for comparison.

We consider a total of seven milestones defined as the hyperplanes (lines in this particular case),

$$M_i = \{(x, y) \in \mathbb{R}^2 \mid x = -0.6 + (i - 1)\Delta x\}, \quad (49)$$

where  $\Delta x = 0.2$ . The hyperplane  $M_1$  is the initial milestone, while  $M_7$  is the absorbing milestone.

### IV. RESULTS AND DISCUSSIONS

#### A. Stationary flux and mean first passage time from exact Milestoning

Fig. 5 shows the estimates of the stationary flux (integrated over  $y$  and hence the same as the Milestoning weights,  $w$ , defined in Eq. (12)) for each milestone as functions of the number of iterations. Note that the scale does not start from zero and the errors are highly non-uniform. The convergence is considerably more rapid far from the absorbing boundary that distorts the initial guess of an equilibrium distribution at the milestone. Milestones 1-4 are near convergence after 10 iterations. Near the absorbing boundary more iterations are required with a maximum number of about 30.

The estimated milestone lifetimes (the components of the vector  $t$ ) as the iterations proceed are shown in Figure 6. Note that since the last milestone is absorbing, we never obtain an estimate for  $t_7$ . The range of times that we obtain is narrow and rapid convergence to the asymptotic value is observed. In a number of cases, essentially a single iteration brings us within ten percent of the converged value. Interestingly, the convergence of the local mean first passage time is less influenced by the proximity to the absorbing boundary in contrast to the calculation of the stationary flux.

In Fig. 7 we show the overall mean first passage time as a function of the iteration number. The MFPT is computed according to Eq. (40). Note that the range of the MFPT values is narrow and it is possible to obtain a sound result with a

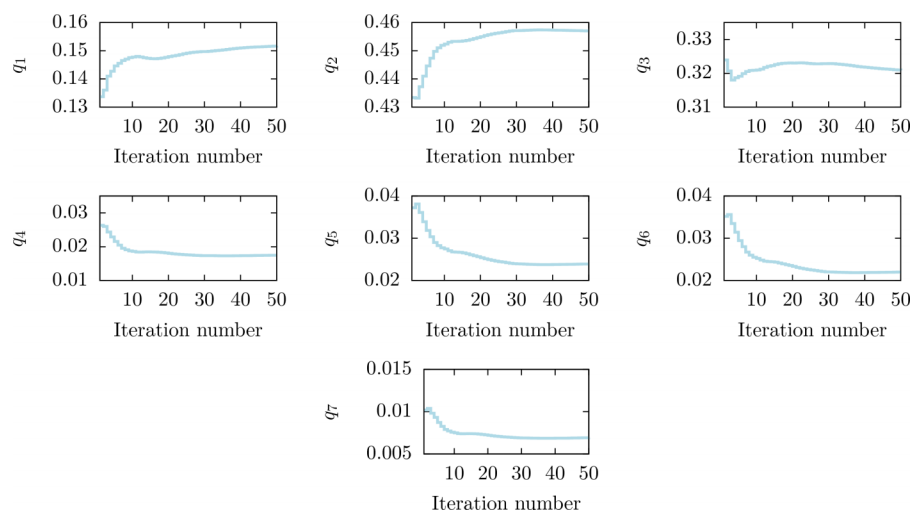


FIG. 5. Estimates for the components of the stationary flux vector at each iteration.

relatively small number of iterations. The inset in the figure focuses on the first 50 iterations. Results within 15% of the final value can be obtained by ten iterations.

The above calculations illustrate the convergence of the algorithm. It is useful to establish their correctness by conducting a study of the system using different computational means. We shall compare the ensembles of short trajectories used in Milestoning to long (uninterrupted) trajectories to verify the accuracy and efficiency of our procedure. This comparison is especially useful when considering the alternative of straightforward molecular dynamics simulations. However, the results of trajectory ensembles tend to be noisy, and similar to other stochastic sampling procedures, the averages converge relatively slowly with respect to the number of operations. It is therefore useful to compare the calculation to a significantly different computational approach with better convergence properties for the system at hand. The present test case is of low dimensionality. Systems of low dimensions can be studied accurately using numerical solutions of the Fokker-Planck equation. We describe a Fokker Planck solution of the model system and compare the results to Milestoning data below.

## B. The Fokker-Planck solution

Consider the planar strips

$$\Omega_i = \{(x, y)^2 \in \mathbb{R}^2 \mid x_i - \Delta x < x < x_i + \Delta x\}. \quad (50)$$

For each  $i = 1, \dots, 6$ , we see that the boundary  $\partial\Omega_i$  of the  $i$ th strip is the disjoint union of the two milestones  $M_{i-1}$  and  $M_{i+1}$ .

The *survival probability* of the system, denoted by  $S = S(x, y, t)$ , is the solution of the initial-boundary value problem,

$$\begin{cases} \frac{\partial S}{\partial t}(x, y, t) = \nabla \cdot J(x, y, t), & (x, y) \in \Omega_i, t > 0, \\ S(x, y, t) = 0, & (x, y) \in \partial\Omega_i, t \geq 0, \\ S(x, y, 0) = g_i(y)\delta(x - x_i), & (x, y) \in \Omega_i. \end{cases} \quad (51)$$

The flux  $J$  is given by

$$J(x, y, t) = \beta^{-1} \nabla S(x, y, t) + S(x, y, t) \nabla U(x, y) \quad (52)$$

and  $S_i$  is a density function. The Milestoning calculation is conducted between milestones  $i = 1$  and  $i = 7$ , the last milestone being absorbing. For the solution of the partial differential equation (PDE), which is conducted on a spatial grid, we add a reflective milestone  $M_0$  that precedes  $M_1$ .

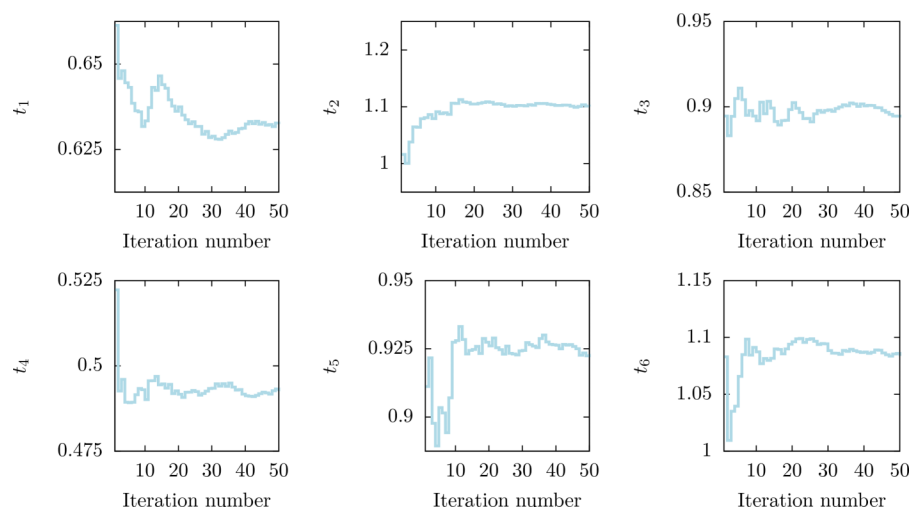


FIG. 6. Estimates for the local mean first passage times for the first six milestones. No estimate of the local MFPT can be provided for the final absorbing milestone (milestone 7).



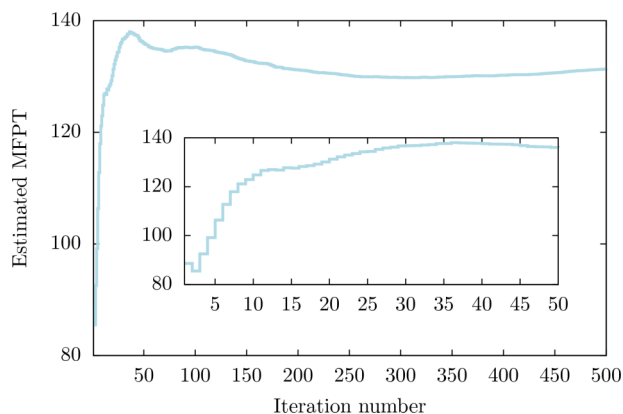


FIG. 7. Estimated (global) mean first passage time. The inset provides the evolution of the estimate during the first 50 iterations of the algorithm.

No such reflective boundary was added for the trajectory calculations since the potential energy itself prevents the trajectories from reaching  $M_0$ . Since the probability of reaching  $M_0$  is exceptionally small, the trajectory calculations in Milestoning and the solution of the Fokker-Planck equation on the grid are equivalent.

From the survival probability, we can obtain the density of the first hitting point distribution at the milestones  $M_{i\pm 1}$  by the formula

$$S_{i\pm 1}(y) = \int_0^\infty |J_1(x_{i\pm 1}, y, t)| dt. \quad (53)$$

The non-zero entries of the transition matrix  $K$  are then  $K_{M,1} = 1$  and

$$K_{i,i\pm 1} = \frac{\int_{-\infty}^\infty S_{i\pm 1}(y) dy}{\int_{-\infty}^\infty (S_{i-1}(y) + S_{i+1}(y)) dy} \quad (54)$$

for  $i = 1, \dots, M-1$ .

Hence, we can compare the transition matrix computed by the Brownian trajectories to the results of the partial

differential equation. Our initial condition is a stationary flux distribution (determined by the Milestoning calculations) at **only** one interface and we run this distribution in time until the density terminates at the nearby milestones. The fraction of density that is accumulated at the nearby milestones determines the elements of the transition matrix. The transition matrix answers the question: given that we start from a particular milestone, what is the probability that the system will reach a particular (other) milestone at any time?

Time evolutions of densities computed with the Fokker Planck equation with the conditions described above are shown in Figure 8.

### C. A comparison between the Fokker Planck and exact milestoning solutions

For comparison, we consider the transition probabilities averaged over time and over the coordinate in the milestone. In the Milestoning language, we have

$$K_{\alpha\beta} = \int_{M_\alpha, M_\beta} dx_\alpha dx_\beta \left[ \int_0^\infty dt \cdot K_{\alpha\beta}(x_\alpha, x_\beta, t) \right] \quad (55)$$

which is also the transition matrix used directly in the approximated version of the Milestoning theory.<sup>9</sup>

In Table I we compared the transition matrices,  $K_{\alpha\beta}$ , computed with exact Milestoning and by the Fokker Planck equation. It is obvious that the agreement is excellent.

In Table II we compare the weights of the stationary fluxes through the milestone (Eq. (15)). The agreement of the results from Milestoning and the Fokker-Planck equation is excellent. Of course, the stationary vectors are the eigenvectors of the transition matrix with the eigenvalue of one. Since the transition matrix is well reproduced, it is expected that the eigenvectors will be in agreement as well.

The lifetimes of the milestones measure how long it takes on average for a trajectory initiated on milestone  $\alpha$

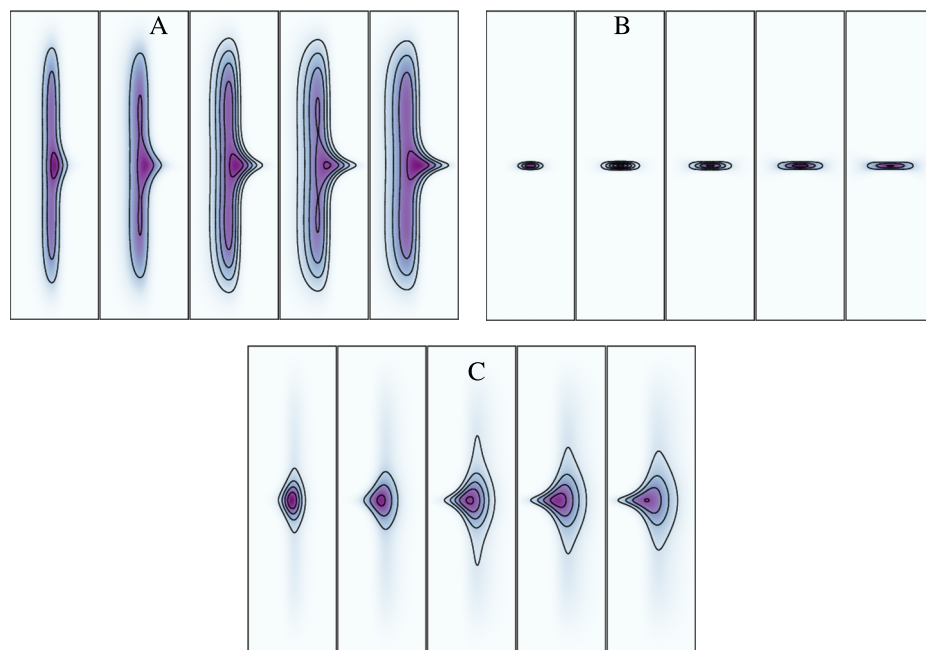


FIG. 8. The time evolution of the probability density of a milestone. The simulation starts with the density at a milestone and is propagated in time using a solver for the partial differential equation. The density is absorbed at the nearby milestones and the simulation continues until it disappears. The results are used to estimate the probability of a transition to a particular milestone and the decay time. In the figure, the time dependent density is renormalized each step for clarity. A is for milestone at  $x = -0.2$ , B for a milestone at  $x = 0$ , and C for a milestone at  $x = 0.2$ . Time progression is recorded from left to right. See text for more details.

TABLE I. The transition matrix as a function of the milestone index.

The transition matrix $K_{\alpha\beta}$ computed with exact Milestoning						
0	1	0	0	0	0	0
0.3186	0	0.6814	0	0	0	0
0	0.9491	0	0.0509	0	0	0
0	0	0.4958	0	0.5042	0	0
0	0	0	0.0810	0	0.919	0
0	0	0	0	0.6806	0	0.3194
1	0	0	0	0	0	0
The transition matrix $K_{\alpha\beta}$ computed with the Fokker Planck equation						
0	1	0	0	0	0	0
0.3197	0	0.6821	0	0	0	0
0	0.9492	0	0.0508	0	0	0
0	0	0.4996	0	0.5004	0	0
0	0	0	0.0848	0	0.9152	0
0	0	0	0	0.6818	0	0.3182
1	0	0	0	0	0	0

to reach (any) other milestone for the first time. Formally, it is defined as  $t_\alpha \equiv \sum_\beta \int_{0^-}^\infty \int_{M_\alpha} \int_{M_\beta} t \cdot K_{\alpha\beta}(x_\alpha, x_\beta, t) dx_\beta \cdot dx_\alpha \cdot dt$ . The lifetimes are compared in Table III. We remark that the milestone lifetime is a good estimator for the length of a trajectory that we will need to use in Milestoning and can be used to predict expected efficiency.

The overall MFPT, the time that it takes on average for a trajectory initiated at the first milestone to reach for the first time the absorbing milestone on the right, is 129.7525 in exact Milestoning and 129.4489 in the Fokker Planck solution. It is interesting that the MFPT is two orders of magnitude larger than the individual transition times, suggesting significantly higher efficiency in sampling a single transition event in Milestoning compared to a straightforward trajectory (in Milestoning, we only compute the trajectory fragments). Further evaluation of Milestoning efficiency is discussed in Sec. IV D.

A more detailed picture is obtained when we considered the lifetimes as a function of the position in the milestone and between specific pairs of milestones. These distributions are shown in Figure 9. Again the agreement between Fokker Planck (solid black line) and blue dots (Milestoning) is excellent.

In a Markovian process, the local transition times are distributed exponentially (without the delay time seen in all the distributions of Figure 9). The delays suggest that at the short time limit the Markovian assumption is violated. Indeed, we do not expect a master equation to be accurate at short times in the coarse space of the milestones.

TABLE II. The stationary flux as a function of the milestone index.

Exact Milestoning						
0.1524	0.4556	0.3195	0.0183	0.0246	0.0226	0.0072
Fokker-Planck equation						
0.1520	0.4558	0.3200	0.0183	0.0244	0.0223	0.0071

TABLE III. Comparing the lifetimes of Milestones.

Computed with exact Milestoning						
0.6304	1.0896	0.8985	0.4937	0.9261	1.0862	0
Computed with the Fokker-Planck equation						
0.6224	1.0666	0.8850	0.5009	0.9104	1.0638	0

Other useful indicators of the dynamics are the stationary first hitting point distributions (FHPD) which are shown in Figure 10. The FHPD are obtained in the Fokker Planck

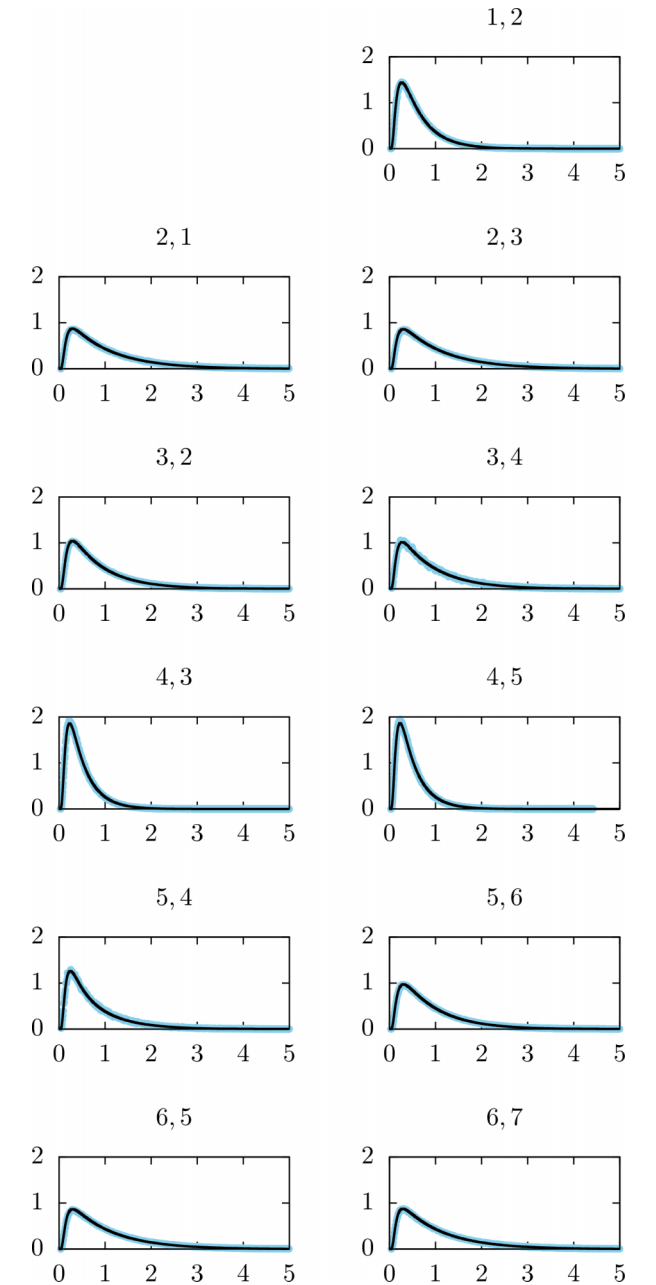


FIG. 9. The probability densities of local transition times between pairs of milestones as a function of the transition time are shown. The indices of the milestones are indicated at the top of the panels. The black solid line is the solution of the Fokker Planck equation and the exact Milestoning are the blue dots.

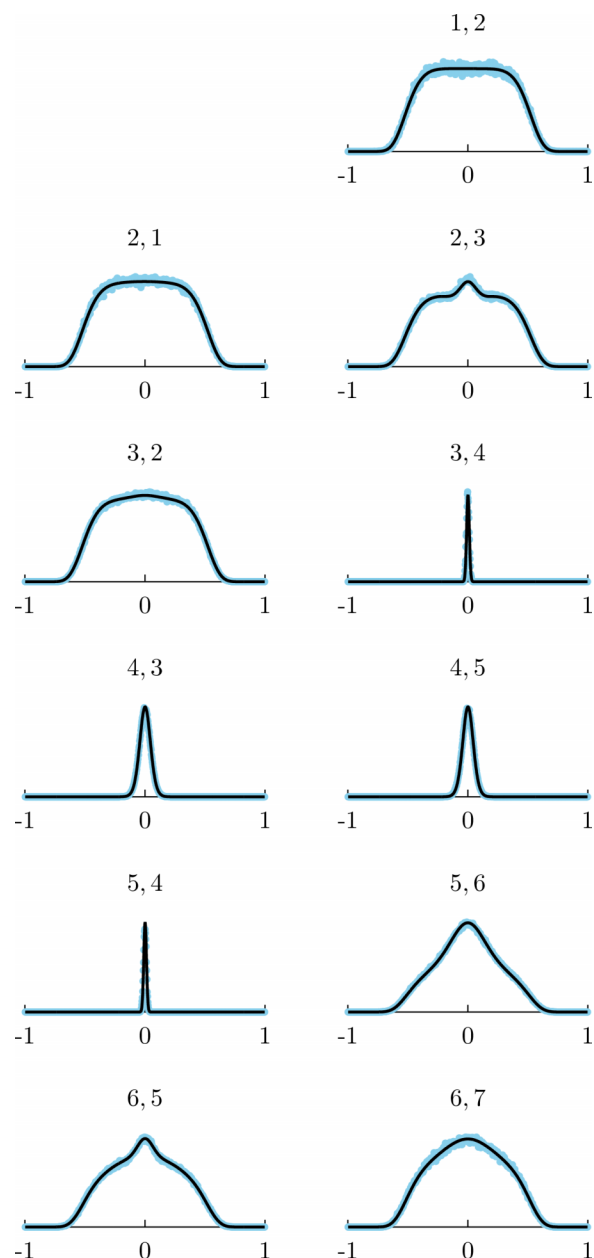


FIG. 10. First hitting distributions between pairs of milestones are shown as a function of the coordinate in the milestone line. The pairs are indicated at the top of the panel. The solid lines are the solutions of the Fokker Planck equation. The blue dots are the results of exact Milestoning calculations. The distributions are normalized to one. See text for more details.

equation by imposing absorbing boundaries at the nearby interfaces to the milestone on which trajectories were started.

Another measure of accuracy is the convergence of the stationary flux. The stationary flux at individual milestones, computed with exact Milestoning, must be time independent also under the Fokker Planck formulation. Hence, if we provide to the Fokker Planck equation, the stationary fluxes of exact Milestoning as initial conditions and we propagate the spatial distribution as a function of time no spatial changes should be observed.

To show the convergence to the stationary flux distribution, we ran a simulation of exact Milestoning for up to 30 iterations and then took the resulting estimate  $q$  of the

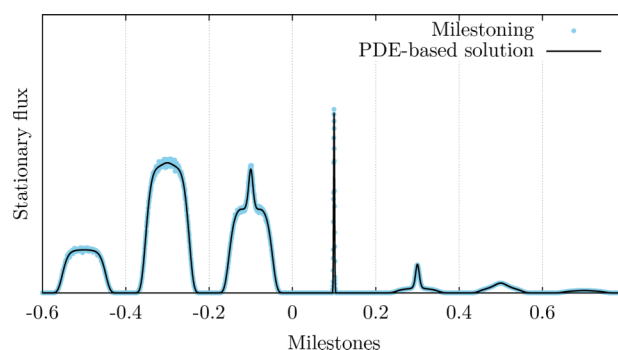


FIG. 11. Estimate for the stationary flux in the milestones compared with an additional iteration of exact Milestoning that solves the densities of the first hitting points by PDE methods. Note that the distributions are asymmetric with respect the origin at  $x$  since there is an absorbing boundary on the right.

stationary flux as the initial conditions for the Fokker Planck equation in Eq. (51),

$$g_i(y) = \frac{S_i(y)}{\int S_i(y) dy}$$

for the problems with  $i = 1, \dots, M$ . The resulting matrix  $K$  and first hitting distributions  $g_i$  can be used to obtain a new stationary flux vector  $q$ . This is akin to running an additional iteration of exact Milestoning where the first hitting distributions are obtained numerically by solving the partial differential equation for their densities. The estimate for the stationary flux obtained with Milestoning is indeed a fixed point of the transition operator as can be seen on Figure 11. The relative error in the  $L_1$ -norm between the estimated flux from Milestoning and the one coming from this additional iteration solving is equal to 0.0177.

We conclude that the accuracy obtained from the ensemble of trajectories in Milestoning is comparable to what we can get from the direct solution of the Fokker Planck equation.

In Sec. IV D, we compare the computational efficiency of exact Milestoning with long trajectories.

#### D. Comparison of Milestoning to long trajectories

A total of one hundred independent instances of each of the following two types of numerical experiments were conducted: uninterrupted trajectories from the first milestone on the left to the absorbing milestone on the right and Milestoning. For both types of experiments, we compute the mean first passage time to go from a point sampled from the canonical distribution conditioned to  $x = -0.6$  to any point at  $x = 0.6$ . We setup the system at a temperature of  $kT = 0.008$  and used a time step length of  $\Delta t = 10^{-4}$  for the numerical integration of the trajectory fragments. In the uninterrupted trajectories, each independent experiment accumulates its own set of estimates for the first passage time and outputs the *running average*. For the Milestoning calculations, we setup milestones at the lines  $x = -0.6, -0.55, -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.55$ , and  $0.6$ . From each of them we drew a total of 250 trajectory samples per iteration and we allowed the simulation to proceed for a total of 250 iterations.

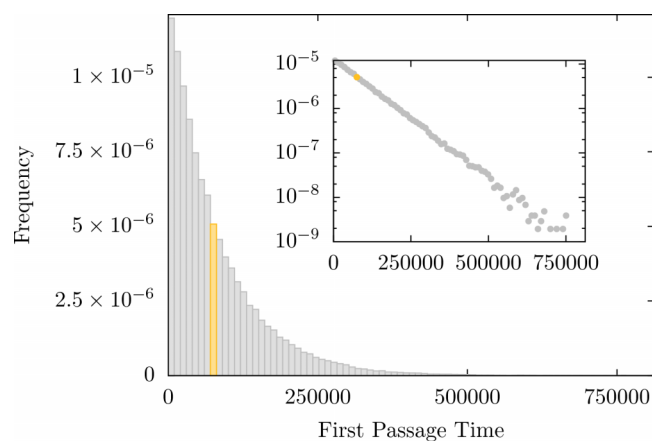


FIG. 12. Histogram of empirical distribution of first passage times. The bar highlighted in yellow is the numerical estimate of the MFPT. In the inset, we plot the same data on logarithmic scale to emphasize the exponential nature of the distribution.

Each of the 108 815 individual samples of the first passage time obtained by the uninterrupted trajectories was collected in order to compute an empirical distribution for the FPTs whose details are found in Figures 12 and 13.

To check the efficiency of the calculation, it is useful to examine running averages. We consider the MFPT as a function of the number of force evaluations. Every time step requires one force evaluation. In molecular dynamics simulations, the calculations of the forces are the most time consuming operation per step. The running average of the MFPT is shown in Figure 13.

Observe that there are very few samples of the running MFPT coming from uninterrupted trajectories below  $10^{10}$  force evaluations. Due to this lack of samples, the statistics are poor. However, the Milestoning runs sample significantly the MFPT for this range of force evaluations.

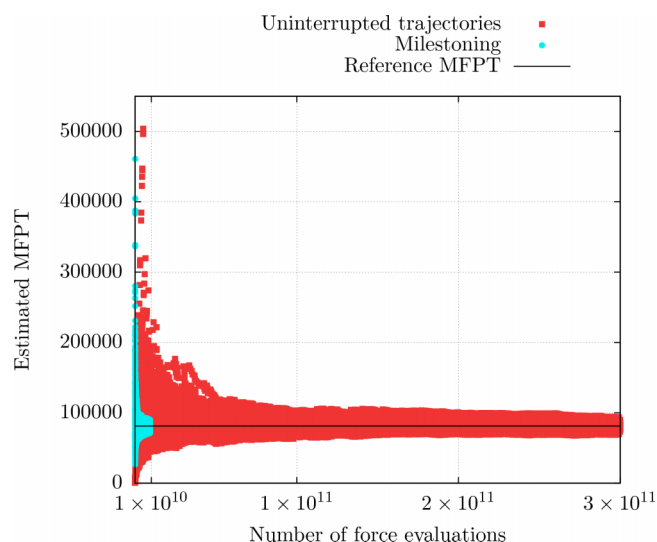


FIG. 13. Running averages coming from both types of experiments. The black line is the best numerical estimate for the MFPT. The distribution of the dots show the running average for the 100 repeats of the calculation either for long trajectories (red) or Milestoning (blue).

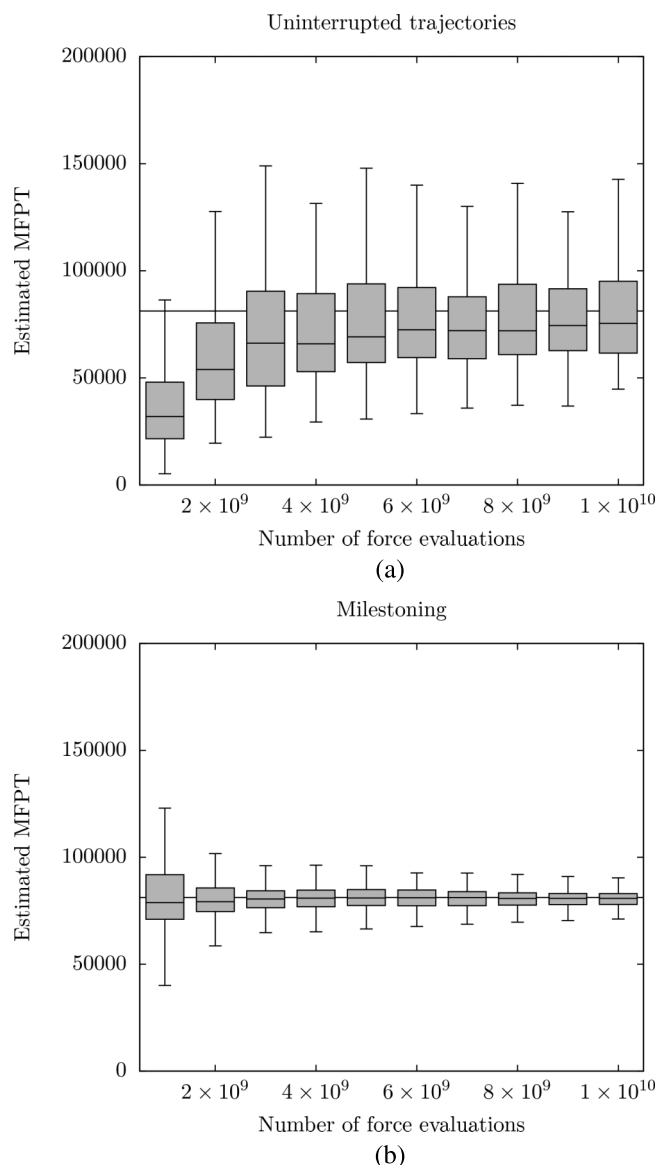


FIG. 14. (a) Box plots for MFPTs obtained from uninterrupted trajectories. The horizontal line is the best estimate for the MFPT. (b) Box plots for MFPTs obtained from Milestoning.

Further information about the error and the convergence is provided in Figures 14(a) and 14(b). These figures use box plot representations to illustrate the changes in the MFPT as a function of the number of force evaluations.

From the above plots, it is evident that Milestoning is significantly more efficient than straightforward molecular dynamics even in a case without an energy barrier (the barrier in our example is entropic).

### E. Dependence on the initial guess

Milestoning is an iterative solution of vector-matrix operation that determines an eigenvector with an eigenvalue one. The speed of convergence depends on the initial guess of the eigenvector. Here, we consider this dependence. In the first version of Milestoning,<sup>9</sup> the distribution at the interface was taken from equilibrium. We use the same initial condition



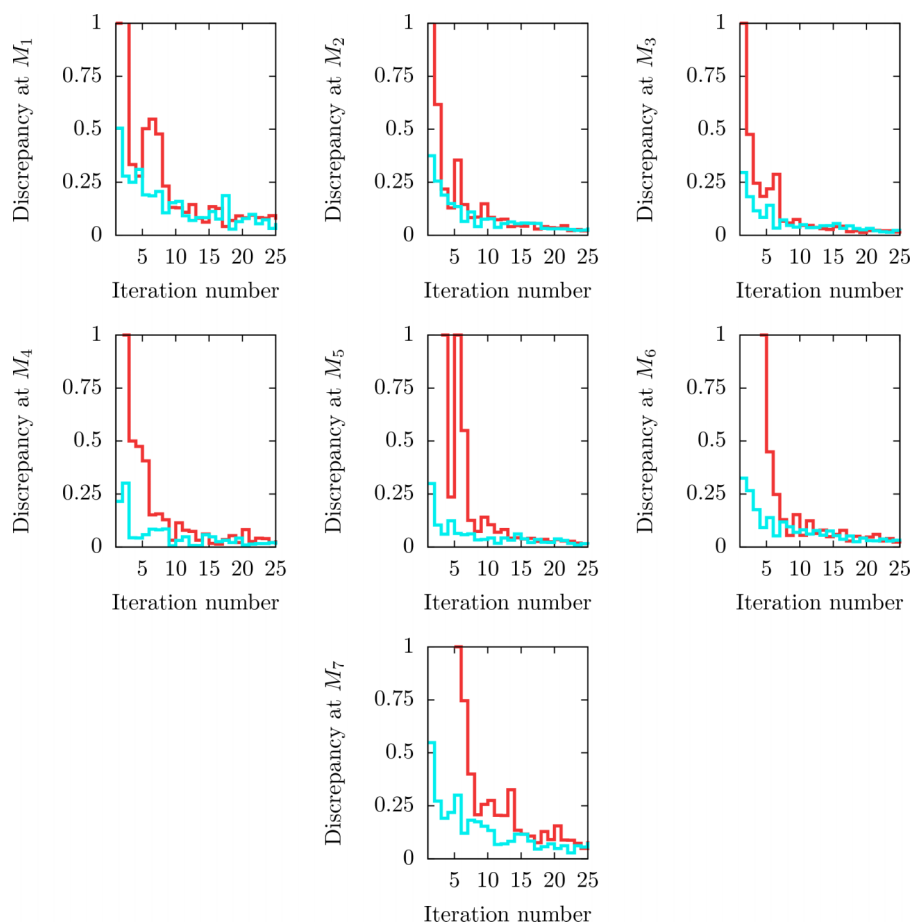


FIG. 15. Local convergence of each flux distribution. For each milestone, we use the error measurement  $\|q_i^n - q_i^{n-1}\|_\infty / \|q_i^n\|_\infty$  to highlight the maximum discrepancy between the current and the previously obtained flux distribution. The red line corresponds to the experiment in which all the trajectories are initially at the first milestone while the blue line shows the corresponding results for the experiment in which trajectories are initially sampled from all milestones according to the Boltzmann distribution.

here. Now we also consider an extreme case in which the whole population is concentrated at the first milestone and then propagated with iterations throughout the system. The last choice mimics the calculations in the Forward Flux Sampling (FFS) algorithm.<sup>40</sup>

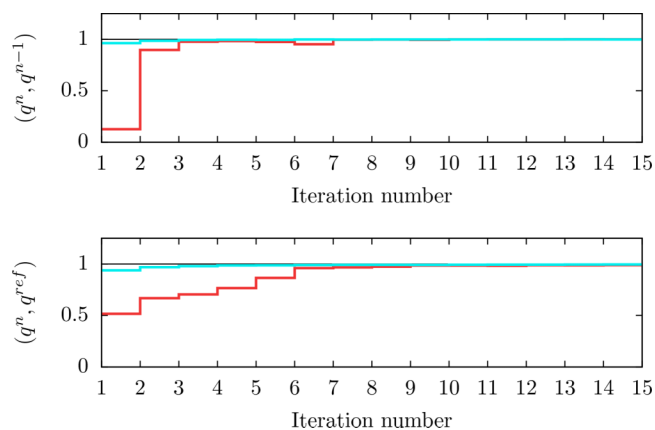


FIG. 16. Global convergence results using Rayleigh quotients. The figure on the top displays the inner products (appropriately normalized) of the flux distributions  $q^{n-1}$  and  $q^n$ . The graph on the bottom shows the inner product for the current iteration  $q^n$  and the reference flux  $q^{\text{ref}} = q^{500}$ . Again, the red color corresponds to the experiment where the initial trajectories are started only at the first milestone while the blue color is for the experiment where we initially run trajectories from every milestone assuming canonical distribution at each milestone.

We consider again the same set of seven milestones  $M_1, \dots, M_7$  that were used in previous experiments. We started a total of 100 trajectory fragments per (relevant) milestone per iteration for a total of 500 iterations.

In this case, we used reflecting boundary conditions at the last milestone. This yields a symmetric stationary flux distribution due to the last row of the transition matrix being  $K_{7,k} = \delta_{6,k}$ .

In Figure 15 we show the convergence of the flux as a function of the iteration number. It is obvious that 500 iterations were not required and that starting from the canonical distribution is a much faster algorithm. Nevertheless, the efficiency of the algorithm starting from a single initiating milestone is not too bad. It suggests that one could use the exact Milestoning procedure to compute non-equilibrium process in which the distributions at nearby milestones are built on the fly. The only input is the milestone we start from.

Global convergence is shown in Figure 16 illustrating again faster convergence of the thermally distributed “old Milestoning” initial guess.

## V. SUMMARY

We have introduced a new variant of Milestoning, which is exact (subject to converged sampling). This variant proposed a new way of computing kinetics and thermodynamics from trajectories using as a guideline a mesh of Voronoi cells.

Since short trajectories are computed over a limited spatial range, the algorithm avoids many of the ergodicity problems of straightforward MD and trapping in metastable states. Of course, this conceptual gain is not coming completely for free and some initial guesses and exploratory simulations are necessarily to identify plausible locations of the Voronoi cells. However, if such a sample is available (the center of the Voronoi cells, the anchors, can be obtained by high temperature trajectories, for example) the gain in efficiency and convergence is profound. Essentially, a complete picture of the statistical mechanics of the system is obtained at reduced cost.

The algorithm is based on iterative determination of stationary flux vectors at milestones. The stationary fluxes are elements of the eigenvector of the transition matrix with an eigenvalue one. The transition matrix (in contrast to the prior versions of Milestoning) is not computed explicitly. Instead, we consider the products of this matrix times a vector, which is the current iteration and estimate of the stationary flux. The vector-matrix products are conducted repeatedly to generate an iterative solution to the flux vector. The properties of the transition matrix guarantee convergence of the iteration process under mild conditions.

We illustrate that the algorithm is significantly more efficient than straightforward MD, which is not surprising since the earlier approximate versions of Milestoning were already much more efficient than MD. The older version of Milestoning is essentially a single iteration of the exact Milestoning procedure. The iterations make the exact version between 10 and 100 times slower in the current implementation. This still leaves ample speedup of exact Milestoning compared to straightforward MD. Because of the relatively short distances between milestones and the short duration of the trajectories, the calculations are not very sensitive to the structure of the energy landscape. This is because at short distances and times the variations in the energy landscape are small.

Also encouraging is the observation that the approximate version of Milestoning that uses only a single iteration is quite reasonable for the example presented and provides results that are not too far from the exact answer.

Nevertheless, we comment that the exact and full determination of the flux vector in Milestoning (if desired) remains a challenge. It is a very long vector of length of the number of milestones times the number of phase space points within a milestone. So, while we have an exact linear equation for it, an exact solution of the flux vector at high dimension is not possible. The more likely scenario is the use of sampling to compute averages and integrals that rely on  $\mathbf{q}$  for sampling in the milestone. For increased accuracy of estimates of the flux that goes beyond the canonical weight of the zero order iteration (Eq. (11)), trajectory fragments can be computed in the forward and backward direction like in directional Milestoning.<sup>16</sup> This is similar in spirit to the methods of PPTIS<sup>7</sup> and to the Milestoning with memory.<sup>32</sup> Also, similar to the approach taken in the Markov state model,<sup>41</sup> full trajectories from reactants to products, if available, can be analyzed for crossing events and the exact Milestoning theory can be used to extract kinetics and thermodynamics. For the

case of full exact trajectories, the sampling of the flux vector at the milestone is done from the exact distribution.

More can be done to increase the efficiency of the Milestoning algorithm and work in that direction is in progress. For example, the sampling in a milestone need not be uniform and it depends on the level of convergence achieved locally. The local sampling at milestones, which is trivially parallelizable, can be optimized to ensure better distributions of effort. Addition and subtraction of milestones can be automated to take into account newly discovered domains of the coarse variables. A useful work in that direction can be found in Ref. 42.

## ACKNOWLEDGMENTS

This research was supported by a grant from the NIH GM59796 and from the Welch Foundation Grant No. F-1783. R.E. thanks David Shalloway for teaching him Laplace transforms and Giovanni Ciccotti for many useful comments on the manuscript.

- <sup>1</sup>D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, C. McLeavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. B. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang, *Commun. ACM* **51**(7), 91-97 (2008).
- <sup>2</sup>R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, *Biochem.* **14**(24), 5355-5373 (1975).
- <sup>3</sup>R. Olender and R. Elber, *J. Chem. Phys.* **105**(20), 9299-9315 (1996).
- <sup>4</sup>P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291-318 (2002).
- <sup>5</sup>D. W. H. Swenson and P. G. Bolhuis, *J. Chem. Phys.* **141**(4), 044101 (2014).
- <sup>6</sup>R. Elber and M. Karplus, *Science* **235**(4786), 318-321 (1987).
- <sup>7</sup>D. Moroni, P. G. Bolhuis, and T. S. van Erp, *J. Chem. Phys.* **120**(9), 4055-4065 (2004).
- <sup>8</sup>B. W. Zhang, D. Jasnow, and D. M. Zuckerman, *J. Chem. Phys.* **132**(5), 054107 (2010).
- <sup>9</sup>A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**(23), 10880-10889 (2004).
- <sup>10</sup>R. Czermanski and R. Elber, *J. Chem. Phys.* **92**(9), 5580-5601 (1990).
- <sup>11</sup>K. A. Fichtorn and Y. Z. Lin, *J. Chem. Phys.* **138**(16), 164104 (2013).
- <sup>12</sup>S. Kirmizialtin and R. Elber, *J. Phys. Chem. A* **115**(23), 6137-6148 (2011).
- <sup>13</sup>S. M. Kreuzer, R. Elber, and T. J. Moon, *J. Phys. Chem. B* **116**(28), 8662-8691 (2012).
- <sup>14</sup>S. Kirmizialtin, V. Nguyen, K. A. Johnson, and R. Elber, *Structure* **20**(4), 618-627 (2012).
- <sup>15</sup>E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**(19), 194101 (2009).
- <sup>16</sup>P. Majek and R. Elber, *J. Chem. Theory Comput.* **6**(6), 1805-1817 (2010).
- <sup>17</sup>A. Dickson, A. Warmflash, and A. R. Dinner, *J. Chem. Phys.* **130**(7), 074104 (2009).
- <sup>18</sup>A. Warmflash, P. Bhimalapuram, and A. R. Dinner, *J. Chem. Phys.* **127**(15), 154112 (2007).
- <sup>19</sup>E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **131**(4), 044120 (2009).
- <sup>20</sup>R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II: Nonequilibrium Statistical Mechanics* (Springer Verlag, Berlin, 1978).
- <sup>21</sup>R. Czermanski and R. Elber, *Int. J. Quantum Chem.* **38**, 167-186 (1990).
- <sup>22</sup>R. Olender and R. Elber, *J. Mol. Struct.: THEOCHEM* **398**, 63-71 (1997).
- <sup>23</sup>R. Elber, *Biophys. J.* **92**(9), L85-L87 (2007).
- <sup>24</sup>R. Elber and A. West, *Proc. Natl. Acad. Sci. U. S. A.* **107**, 5001-5005 (2010).
- <sup>25</sup>S. Kreuzer and R. Elber, *Biophys. J.* **105**(4), 951-961 (2013).
- <sup>26</sup>G. S. Jas, W. A. Hegefeld, P. Majek, K. Kuczera, and R. Elber, *J. Phys. Chem. B* **116**(23), 6598-6610 (2012).
- <sup>27</sup>A. E. Cardenas and R. Elber, *J. Chem. Phys.* **141**(5), 054101 (2014).
- <sup>28</sup>A. Cardenas and R. Elber, *Mol. Phys.* (to be published).
- <sup>29</sup>S. M. Kreuzer, T. J. Moon, and R. Elber, *J. Chem. Phys.* **139**(12), 121902 (2013).
- <sup>30</sup>A. E. Cardenas and R. Elber, *Mol. Phys.* **111**(22-23), 3565-3578 (2013).

- <sup>31</sup>G. H. Golub and C. Van Loan, *Matrix Computation*, 4th ed. (Johns Hopkins University Press, Baltimore, 2012).
- <sup>32</sup>A. T. Hawk and D. E. Makarov, *J. Chem. Phys.* **135**(22), 224109 (2011).
- <sup>33</sup>S. Viswanath, S. M. Kreuzer, A. E. Cardenas, and R. Elber, *J. Chem. Phys.* **139**(17), 174105 (2013).
- <sup>34</sup>A. M. A. West, R. Elber, and D. Shalloway, *J. Chem. Phys.* **126**(14), 145104 (2007).
- <sup>35</sup>A. T. Hawk, *J. Chem. Phys.* **138**(15), 154105 (2013).
- <sup>36</sup>F. Nuske, B. G. Keller, G. Perez-Hernandez, A. Mey, and F. Noe, *J. Chem. Theory Comput.* **10**(4), 1739-1752 (2014).
- <sup>37</sup>E. Vanden Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, *J. Chem. Phys.* **129**(17), 174102 (2008).
- <sup>38</sup>P. Reimann, G. J. Schmid, and P. Hanggi, *Phys. Rev. E* **60**(1), R1-R4 (1999).
- <sup>39</sup>B. Leimkuhler and C. Matthews, *J. Chem. Phys.* **138**(17), 174102 (2013).
- <sup>40</sup>R. J. Allen, D. Frenkel, and P. R. ten Wolde, *J. Chem. Phys.* **124**(2), 024102 (2006).
- <sup>41</sup>C. Schutte, F. Noe, J. F. Lu, M. Sarich, and E. Vanden-Eijnden, *J. Chem. Phys.* **134**(20), 204105 (2011).
- <sup>42</sup>D. Shalloway and A. K. Faradjian, *J. Chem. Phys.* **124**(5), 054112 (2006).