

On the assumptions underlying milestoning

Eric Vanden-Eijnden,^{1,a)} Maddalena Venturoli,^{1,b)} Giovanni Ciccotti,^{2,c)} and Ron Elber^{3,d)}

¹*Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA*

²*INFN and Dipartimento di Fisica, Università di Roma La Sapienza, Piazzale Aldo Moro, 2, 00185 Roma, Italy*

³*Department of Chemistry and Biochemistry, and Institute of Computational Engineering and Sciences (ICES), University of Texas, Austin, Texas 78712, USA*

(Received 6 March 2008; accepted 17 September 2008; published online 5 November 2008)

Milestoning is a procedure to compute the time evolution of complicated processes such as barrier crossing events or long diffusive transitions between predefined states. Milestoning reduces the dynamics to transition events between intermediates (the milestones) and computes the local kinetic information to describe these transitions via short molecular dynamics (MD) runs between the milestones. The procedure relies on the ability to reinitialize MD trajectories on the milestones to get the right kinetic information about the transitions. It also rests on the assumptions that the transition events between successive milestones and the time lags between these transitions are statistically independent. In this paper, we analyze the validity of these assumptions. We show that sets of optimal milestones exist, i.e., sets such that successive transitions are indeed statistically independent. The proof of this claim relies on the results of transition path theory and uses the isocommittor surfaces of the reaction as milestones. For systems in the overdamped limit, we also obtain the probability distribution to reinitialize the MD trajectories on the milestones, and we discuss why this distribution is not available in closed form for systems with inertia. We explain why the time lags between transitions are not statistically independent even for optimal milestones, but we show that working with such milestones allows one to compute mean first passage times between milestones exactly. Finally, we discuss some practical implications of our results and we compare milestoning with Markov state models in view of our findings. © 2008 American Institute of Physics. [DOI: 10.1063/1.2996509]

I. INTRODUCTION

Milestoning, introduced in Ref. 1 and further developed in Refs. 2–5, is a procedure to analyze the progress in time of complicated processes such as barrier crossing events between predefined states or slow diffusive motion from an initial state toward a target state. Milestoning coarse grains the temporal and spatial description of the system by reducing the dynamics into a succession of independent transition events between intermediates (the milestones), which are hypersurfaces in the configuration space of the system. The necessary information about the probability of occurrence of these transitions and the time lags between them is obtained by collecting local kinetic information from straightforward and short trajectories between the milestones. By reducing the dynamics this way, milestoning allows one to simulate the overall kinetics of the system for time scales which are much beyond what is accessible by standard molecular dynamics (MD) simulations. Milestoning, however, rests on the assumptions that such a reduction of the dynamics to transitions between milestones is legitimate and that it is possible to retrieve the right kinetic information using short simula-

tions between these milestones. The aim of this paper is to analyze under which conditions these assumptions are valid.

To introduce the issues that this question brings up and recall the details of the milestoning procedure, it is useful to consider a thought experiment in which one is given an infinitely long unbiased trajectory of a MD system which, assuming ergodicity, visits a given set of milestones infinitely often. Denoting these milestones by S_1, S_2, \dots, S_N and by $\mathbf{x}(t)$ with $t \geq 0$ the instantaneous position of the system along the trajectory, let $S_{i_0}, S_{i_1}, S_{i_2}, S_{i_3}, \dots$ with $i_k \in \{1, 2, \dots, N\}$, $k = 0, 1, 2, \dots$ be the infinite sequence of successive milestones that $\mathbf{x}(t)$ crosses [thus we do not count recrossings of the same milestone until another milestone has been crossed in the meantime, i.e., $i_k \neq i_{k-1}$ for all $k = 1, 2, \dots$ (see Fig. 1 for an illustration)]. Let also $0 \leq t_0 < t_1 < t_2 < \dots$ be the times at which these crossings occur, i.e., t_0 is the first time ≥ 0 such that $\mathbf{x}(t_0) \in S_{i_0}$ and for $k = 1, 2, \dots$, t_k is the first time such that $\mathbf{x}(t_k) \in S_{i_k}$ after the trajectory crossed $S_{i_{k-1}}$.

The sequence $\{(S_{i_0}, t_0), (S_{i_1}, t_1), \dots\} \equiv \{(S_{i_k}, t_k)\}_{k=0,1,\dots}$ gives a coarse-grained picture of the MD trajectory, but one which, with a suitable choice of milestones, will still contain useful information about the kinetics of this trajectory. The main objective of the milestoning procedure is to generate this sequence and to analyze its statistical properties in a way which is computationally cheaper than running a long MD

^{a)}Electronic mail: eve2@cims.nyu.edu.

^{b)}Electronic mail: mventuro@cims.nyu.edu.

^{c)}Electronic mail: giovanni.ciccotti@romal.infn.it.

^{d)}Electronic mail: ron@ices.utexas.edu.

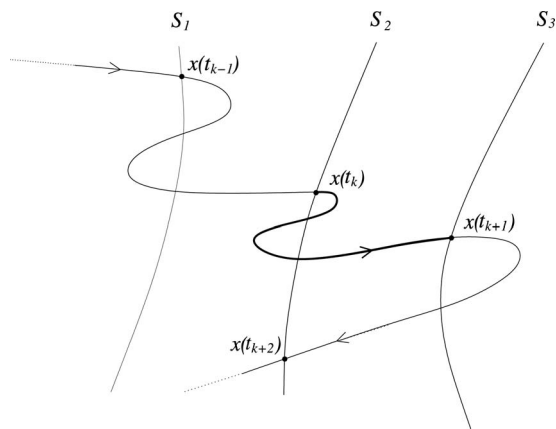


FIG. 1. Schematic of a piece of a long ergodic trajectory crossing a set of three milestones: S_1 , S_2 , and S_3 . In this example, $S_{i_{k-1}}=S_1$, $S_{i_k}=S_2$, $S_{i_{k+1}}=S_3$, and $S_{i_{k+2}}=S_2$. The part of the trajectory highlighted in bold contributes to one event counted in p_{23} and the time $t_{k+1}-t_k$ contributes to the statistics of $f_{23}(s)$. The figure also shows the previous transition event from S_1 to S_2 , which contributes to p_{12} with the time t_k-t_{k-1} contributing to $f_{12}(s)$, and the next one from S_3 to S_2 , which contributes to p_{32} with the time $t_{k+2}-t_{k+1}$ contributing to $f_{32}(s)$.

trajectory and waiting that it visits all the milestones as in the thought experiment above to gather its statistics.

To see how this objective is realized and the conditions this entails, let us introduce the following two statistical quantities which play a central role in milestoneing. First, by going back to the sequence $\{(S_{i_k}, t_k)\}_{k=0,1,\dots}$ obtained from the long ergodic MD trajectory and recording the proportion of times along this sequence that S_j is the milestone crossed next after S_i , we can define the probability p_{ij} that this event occurs,⁶

$$p_{ij} = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \delta_{i,i_{k-1}} \delta_{j,i_k}}{\sum_{k=1}^n \delta_{i,i_{k-1}}} = \left(\begin{array}{l} \text{probability that } S_j \text{ is the milestone} \\ \text{crossed next after } S_i \end{array} \right), \quad (1)$$

where $\delta_{i,j}=1$ if $i=j$ and $\delta_{i,j}=0$ otherwise. (Thus, by definition $p_{ii}=0$ and $\sum_{j=1}^N p_{ij}=1$ for all $i=1, \dots, N$.) Second, by recording the time it takes for the trajectory to go to S_j next after S_i each time this event happens, we can define the probability density $f_{ij}(s)$ of the lag time between these crossings conditional on them occurring: for any pair i, j such that $p_{ij} \neq 0$, $f_{ij}(s)$ is such that

$$\int_0^{\Delta t} f_{ij}(s) ds = \lim_{n \rightarrow \infty} \frac{\sum_{k=1}^n \delta_{i,i_{k-1}} \delta_{j,i_k} \mathbf{1}_{[0, \Delta t]}(t_k - t_{k-1})}{\sum_{k=1}^n \delta_{i,i_{k-1}} \delta_{j,i_k}} = \left(\begin{array}{l} \text{probability that the trajectory takes} \\ \text{less than } \Delta t \text{ to reach } S_j \text{ after crossing} \\ S_i \text{ given that } S_j \text{ is the milestone} \\ \text{crossed next after } S_i \end{array} \right) \quad (2)$$

where $\mathbf{1}_{[0, \Delta t]}(s)=1$ if $s \in [0, \Delta t]$ and $\mathbf{1}_{[0, \Delta t]}(s)=0$ otherwise. [Thus, $f_{ij}(s)$ is only defined if $p_{ij} \neq 0$ and in this case $\int_0^{\infty} f_{ij}(s) ds = 1$.]

As discussed below, the quantities (1) and (2) can, in principle, be computed more efficiently than via a long MD trajectory. Before explaining how to do so, however, one should wonder how useful these quantities are. Indeed, in general (i.e., for an arbitrary set of milestones), Eqs. (1) and (2) contain only partial information about the statistical properties of the sequence $\{(S_{i_k}, t_k)\}_{k=0,1,\dots}$ [for instance, we could be interested in the probability that S_j is crossed after S_i and S_k after S_j , which is not given by $p_{ij}p_{jk}$ in general, or the average time it takes to go from S_i to S_j , which is not expressible in terms of p_{ij} and $f_{ij}(s)$ alone in general]. However, *all* the statistical properties of the sequence $\{(S_{i_k}, t_k)\}_{k=0,1,\dots}$ could be deduced from Eqs. (1) and (2) alone if the following two properties were true. The first makes an assumption about the way successive transitions between milestones occur.

(P1) The probability that after crossing S_{j_0} , the system successively visits n milestones in the sequence $S_{j_1}, S_{j_2}, S_{j_3}, \dots, S_{j_n}$ is for any $n \in \mathbb{N}$ given by

$$p_{j_0 j_1} p_{j_1 j_2} p_{j_2 j_3} \cdots p_{j_{n-1} j_n}, \quad (3)$$

where the probability is again defined by recording the proportion of times this sequence of transitions is observed along the sequence $\{S_{i_k}\}_{k=0,1,2,\dots}$.

This property guarantees that the statistical properties of the sequence $\{S_{i_k}\}_{k=0,1,\dots}$ are identical to those of a discrete-time Markov process (here the ‘‘time’’ is the index k) with transition probability p_{ij} . The second property makes an assumption about the statistics of the lag times between crossings of successive milestones.

(P2) Denoting by $t_{j_0 j_1}, t_{j_1 j_2}, \dots, t_{j_{n-1} j_n}$ the lag times between successive crossings of the milestones along any sequence $S_{j_0}, S_{j_1}, S_{j_2}, \dots, S_{j_n}$, these lag times are statistically independent, i.e., if we compute the joint probability density function that $t_{j_0 j_1} = s_1, t_{j_1 j_2} = s_2, \dots, t_{j_{n-1} j_n} = s_n$, for any $n \in \mathbb{N}$ it reduces to the product

$$f_{j_0 j_1}(s_1) f_{j_1 j_2}(s_2) \cdots f_{j_{n-1} j_n}(s_n). \quad (4)$$

Together with (P1), this property guarantees that the statistical properties of the sequence $\{(S_{i_k}, t_k)\}_{k=0,1,\dots}$ are those of a discrete-time Markov process (where the time stands again for the index k , whereas the physical time t_k is part of the state variables of the chain⁷) with transition probability

$$\left(\begin{array}{l} \text{probability that } S_{i_{k+1}} = S_j \text{ and } t_{k+1} \leq \\ t + \Delta t \text{ given that } S_{i_k} = S_i \text{ and } t_k = t \end{array} \right) = p_{ij} \int_0^{\Delta t} f_{ij}(s) ds. \quad (5)$$

This says that given that milestone S_i is crossed at time t_k , with probability p_{ij} , the next milestone to be crossed is S_j and given that this event occurs, it happens at time $t_{k+1} = t_k + t_{ij}$, where t_{ij} is a random time distributed according to Eq. (2).

The key assumption made in the milestoning procedure is that properties (P1) and (P2) do hold true, thereby guaranteeing that the statistical properties of the sequence $\{(S_{i_k}, t_k)\}_{k=0,1,\dots}$ obtained from the MD trajectory are identical to those of the Markov chain with transition probability Eq. (5). This chain is fully specified if one knows the quantities in Eqs. (1) and (2) and it can be analyzed using the tools developed in Refs. 1–5. Of course, this still leaves us with the important question of how to compute Eqs. (1) and (2) in practice. A key component of milestoning is to do so by reinitializing the MD simulation on each milestone S_i . In principle, this reinitialization must be done using the probability density of the position at which a long ergodic trajectory first hits milestone S_i after crossing another one [i.e., the density of the points $\mathbf{x}(t_k)$ with k such that $i_k \equiv i$], but since this density is not known *a priori*, we are left with the question:

What is the expression of the first hitting point probability density on the milestones that should be used to reinitialize the short MD runs in order to compute Eqs. (1) and (2)?

If we can answer this question, then Eqs. (1) and (2) can be computed by performing relatively short MD runs until the trajectories initiated on a given milestone reach another milestone. This allows one to bypass the need to wait that the system actually reaches this milestone on its own (i.e., the need to use a long unbiased trajectory) and, in effect, it is the reason why milestoning offers a computational gain over direct MD. Indeed, with an appropriate choice of the milestones, the individual transition times between them can be made relatively short even though overall transition times between milestones which are far apart can be very long.

This finally brings us back to the aims of this paper. Our main objectives are to analyze the validity of properties (P1) and (P2) and to derive an expression for the first hitting point probability density that must be used to reinitialize the trajectories on the milestones. In Sec. II we will first analyze systems in the overdamped limit and show that there exist sets of milestones for which property (P1) is satisfied exactly. Property (P2), in general, will not hold exactly for this set of milestones, but we will show that this property is, in fact, not necessary to compute exactly certain quantities such as mean first passage times [in order to do such calculations exactly, only property (P1) is required]. For this reason, we shall refer to the milestones in sets such that property (P1) holds as “optimal milestones.” Using the results from transition path theory (TPT),^{8–10} we will show that these sets of optimal milestones involve the isosurfaces of the so-called committor function. We also derive the probability density of first hitting of these surfaces, from which one needs to initiate trajectories to compute Eq. (1) in the milestoning procedure. These results are illustrated in Sec. III via simple examples. In Sec. IV we then consider systems with inertia, for which the situation is more complex. Optimal milestoning and the exact calculation of mean first passage time are, in principle, possible using again the isosurfaces of the committor function, but this function now depends on both the system position and momentum. In addition, for reasons that we will

explain, an explicit expression for the probability density of first hitting of these surfaces is not available for systems with inertia, which complicates matters even more. How to deal with these issues, in practice, is explained in Sec. V where, among other things, we discuss how to identify optimal milestones in practice by combining milestoning with the string method.^{11–16} Finally, a few concluding remarks are given in Sec. VI where we compare milestoning with other strategies that use short runs between intermediates to deduce the overall kinetics. These include, in particular, Markov state models (MSMs) (e.g., Refs. 17–25), transition interface sampling^{26,27} (TIS) and forward flux sampling (FFS).²⁸ For the reader’s convenience, some background material on TPT (in particular on the committor function) and some technical derivations are deferred to two appendixes.

II. OPTIMAL MILESTONING IN THE OVERDAMPED LIMIT

In this section, we focus on systems governed by the overdamped (Smoluchowski) equation

$$\gamma \dot{\mathbf{x}}(t) = -\nabla V(\mathbf{x}(t)) + \sqrt{2\beta^{-1}} \boldsymbol{\eta}(t). \quad (6)$$

Here $\mathbf{x}(t) \in \mathbb{R}^d$ denotes the position of the system, γ is the friction coefficient, $V(\mathbf{x})$ is the potential, β is the inverse temperature, and $\boldsymbol{\eta}(t)$ is a white noise, i.e., a Gaussian process with mean zero and covariance $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$. The case of systems with inertia governed by the Langevin equation will be considered later in Sec. IV.

In Sec. II A we first establish that optimal sets of milestones which satisfy property (P1) exactly exist and involve the isosurfaces of the committor function. Then in Sec. II B we derive the probability density of first hitting points on these optimal milestones. Finally, in Sec. II C we discuss why, in general, property (P2) does not hold exactly even for optimal milestones, but we show that mean first passage times can nonetheless be computed exactly with these milestones.

A. Isocommittor surfaces as optimal milestones

Successive transitions between milestones are not independent, in general, and property (P1) does not always hold because typically, the position where the trajectory crossed milestone S_i influences whether it will cross next S_j rather than another milestone. In other words, the probability that $\mathbf{x}(t_k) \in S_j$ given that $\mathbf{x}(t_{k-1}) \in S_i$ depends on where $\mathbf{x}(t_{k-1})$ is located on S_i [if we knew $\mathbf{x}(t_{k-1})$, we could deduce the probability that $\mathbf{x}(t_k) \in S_j$ since $\mathbf{x}(t)$ is a Markov process, but the exact location of $\mathbf{x}(t_{k-1})$ is precisely what milestoning aims at coarse graining]. This issue, however, obviously disappears and property (P1) will hold if one can find sets of milestones such that the following property holds:

(P1’) the probability to reach first S_j from S_i is the same from any starting point on S_i . The main result of this section is to show that such sets of optimal milestones satisfying property (P1’) and, hence, (P1), exist.

To see why this is the case, suppose that we start with a first and a last milestone in a set of N milestones, i.e., two nonintersecting but otherwise arbitrary smooth dividing sur-

faces in configuration space, respectively, denoted by S_1 and S_N . We wish to complete this set with S_2, S_3, \dots, S_{N-1} in such a way that property (P1') hold exactly for this set. To do so, let us introduce the following function:

$$q(\mathbf{x}_0) = \left(\begin{array}{l} \text{probability that } \mathbf{x}(t) \text{ reaches first } S_N \\ \text{rather than } S_1 \text{ given that } \mathbf{x}(0) = \mathbf{x}_0 \end{array} \right). \quad (7)$$

Here and below, the probability is defined with respect to the realizations of the noise in Eq. (6). The function $q(\mathbf{x})$ is the so-called committor function introduced in TPT to describe the reaction between a reactant state $A \subset \mathbb{R}^d$ and a product state $B \subset \mathbb{R}^d$, which are here the two sets whose respective boundaries are S_1 and S_N and leave out the region in between S_1 and S_N (note that since S_1 and S_N can be chosen arbitrarily, in the present context the labels ‘‘reactant’’ and ‘‘product’’ are somewhat void of content). As we recall in Appendix A, the committor function $q(\mathbf{x})$ is the solution of the partial differential equation given in Eq. (A7) and the isosurfaces of $q(\mathbf{x})$ foliate the region in between S_1 and S_N (i.e., they are nonintersecting smooth dividing surfaces in this region). The key idea behind optimal milestoneing is to use some of these surfaces as milestones. Specifically, letting $z_1=0 < z_2 < z_3 < \dots < z_N=1$, we define

$$S_i = \{\mathbf{x}: q(\mathbf{x}) = z_i\}, \quad i = 1, \dots, N. \quad (8)$$

This definition is consistent with S_1 and S_N since $q(\mathbf{x})=0$ on S_1 and $q(\mathbf{x})=1$ on S_N by construction, and we claim that these surfaces are optimal milestones. To see this, look at the reaction from the extended reactant state at the left of S_{i-1} , i.e., the set $\{\mathbf{x}: q(\mathbf{x}) \leq z_{i-1}\}$, to the extended product state at the right of S_{i+1} , i.e., the set $\{\mathbf{x}: q(\mathbf{x}) \geq z_{i+1}\}$, and note that the surface $S_i = \{\mathbf{x}: q(\mathbf{x}) = z_i\}$ also is an isocommittor surface for this reaction. Indeed, the committor function for this reaction is related to the original $q(\mathbf{x})$ as $(q(\mathbf{x}) - z_{i-1}) / (z_{i+1} - z_{i-1})$ since, by construction, this function satisfies Eq. (A7) with the correct boundary conditions that it be 0 on S_{i-1} and 1 on S_{i+1} . Since $q(\mathbf{x}) = z_i$ on S_i by definition, this shows that the probability to reach first S_{i+1} rather than S_{i-1} starting from any point on S_i is equal to $(z_i - z_{i-1}) / (z_{i+1} - z_{i-1})$ for any $i = 2, \dots, N-1$ (obviously the probability to reach first S_2 from S_1 and the one to reach first S_{N-1} from S_N are both 1). This implies that properties (P1') and (P1) are indeed satisfied and gives the following expression for Eq. (1): $p_{12} = p_{NN-1} = 1$ and, for $i=2, \dots, N-1$,

$$p_{ij} = \begin{cases} \frac{z_{i+1} - z_i}{z_{i+1} - z_{i-1}} & \text{if } j = i - 1 \\ \frac{z_i - z_{i-1}}{z_{i+1} - z_{i-1}} & \text{if } j = i + 1 \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

The conclusions above rely on the properties of the exact $q(\mathbf{x})$, which we can only hope to compute from Eq. (A7) in systems of low dimensions. We note however that there are algorithms (see, e.g., Refs. 12, 15, 29, and 30) to effectively approximate the committor surfaces in systems with much higher dimensions under suitable assumptions (more on this in Sec. VI below).

B. First hitting point density and the issue of reinitialization

As explained earlier, using optimal milestones such that property (P1) holds exactly is not enough to make milestoneing practical. Indeed, since we typically do not have at our disposal an infinitely long ergodic trajectory, we also need to know how to reinitiate MD trajectories from the milestones to compute the lag-time density (2) [which is the only thing left to compute with optimal milestones since Eq. (1) is given by Eq. (9) for these milestones]. This is the issue that we discuss next.

Denote by $\rho_{ij}(\mathbf{x})$ the probability density of the first hitting point on S_j given that the infinitely long ergodic trajectory generated by Eq. (6) crossed S_i last before S_j . As demonstrated in Appendix B, a key property of the optimal milestones defined in Eq. (8) is that $\rho_{ij}(\mathbf{x})$ depends only on j and not on i , i.e., we have $\rho_{ij}(\mathbf{x}) \equiv \rho_j(\mathbf{x})$ with $\rho_j(\mathbf{x})$ given by³¹

$$\rho_j(\mathbf{x}) = Z_j^{-1} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})|. \quad (10)$$

Here Z_j is a normalization factor given by the surface integral

$$Z_j = \int_{S_j} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})| d\sigma_{S_j}(\mathbf{x}), \quad (11)$$

where $d\sigma_{S_j}(\mathbf{x})$ denotes the surface element on S_j (i.e., the Lebesgue measure on this surface). Because property (P1) holds with optimal milestones, one can compute the lag-time density (2) by reinitializing MD trajectories on these milestones, and the correct probability density to perform this reinitialization on S_j is $\rho_j(\mathbf{x})$.

The proof that Eq. (10) is the correct density of first hitting point on the optimal milestones is somewhat technical, and we defer it to Appendix B. Here let us simply stress that Eq. (10) is *not* the equilibrium density on S_j , which would be proportional to $e^{-\beta V(\mathbf{x})}$. The extra factor $|\nabla q(\mathbf{x})|$, which appears in Eq. (10), can be understood within the framework of TPT upon noting that the probability current of reactive trajectories is proportional to $e^{-\beta V(\mathbf{x})} \nabla q(\mathbf{x})$, and so the probability flux induced by this current through the surface S_j is precisely given by Eq. (11). The first hitting point density is related to this probability flux, and this is consistent with Eq. (10) rather than with the equilibrium density. How different the density (10) is with respect to the equilibrium one will be illustrated via examples in Sec. III. In Sec. IV we also discuss how to deal with Eq. (10) in practice.

C. Exact calculation of mean first passage times

As we have mentioned earlier, property (P2) is independent of (P1). In particular, even the set of optimal milestones made of isocommittor surfaces introduced in Sec. II A will not, in general, be such that the lag times between successive transitions between milestones are statistically independent (why this is the case will be illustrated via an example in Sec. III). This means that the reduction of the dynamics used in milestoneing will not be exact in general. However, optimal milestones allow one to compute exactly certain important quantities, such as the mean first passage time from any

milestone to any other one. Next we discuss why this is the case and give an explicit expression for these mean first passage times.

Consider the various instances when a long ergodic trajectory visits the milestones $S_{j_0}, S_{j_1}, \dots, S_{j_n}$ successively in this sequence. The duration of any one of such sequences is given by a sum of times, $t_{j_0j_1} + t_{j_1j_2} + \dots + t_{j_{n-1}j_n}$, each of which is the lag between the crossing times of two successive milestones, $t_{j_kj_{k+1}} = t_{k+1} - t_k$ with $\mathbf{x}(t_k) \in S_{j_k}$. If one builds the statistics of these times, one will observe that, in general, they are correlated, i.e., rather than Eq. (2) what one needs to describe their statistical properties in full is their joint probability density

$$\begin{aligned} & f_{j_0j_1 \dots j_n}(s_1, s_2, \dots, s_n) \\ &= \left(\begin{array}{l} \text{probability density that } t_1 - t_0 = s_1, \\ t_2 - t_1 = s_2, \dots, \text{ and } t_n - t_{n-1} = \\ s_n, \text{ given that } \mathbf{x}(t_k) \in S_{j_k} \text{ for } k= \\ 0, \dots, n \end{array} \right) \\ &\neq f_{j_0j_1}(s_1) f_{j_1j_2}(s_2) \dots f_{j_{n-1}j_n}(s_n). \end{aligned} \quad (12)$$

However, regardless of whether the times $t_{j_0j_1}, \dots, t_{j_{n-1}j_n}$ are statistically independent or not, Eq. (2) remains the marginal of the density (12) for any $t_{j_kj_{k+1}}$, i.e., for any $k=1, \dots, n$ we have

$$\begin{aligned} & f_{j_{k-1}j_k}(s_k) \\ &= \int f_{j_0j_1 \dots j_n}(s_1, s_2, \dots, s_n) ds_1 \dots ds_{k-1} ds_{k+1} \dots ds_n. \end{aligned} \quad (13)$$

As a result, the average duration of such sequence is given exactly by

$$\tau_{j_0j_1} + \tau_{j_1j_2} + \dots + \tau_{j_{n-1}j_n}, \quad (14)$$

where τ_{ij} is the first moment of Eq. (2),

$$\tau_{ij} = \int_0^\infty s f_{ij}(s) ds. \quad (15)$$

In essence, evaluating the mean first passage time amounts to summing up the mean duration of sequences of transitions between milestones weighed by their probability of occurrence [an explicit expression for the mean passage time is given as the solution of Eq. (21) below]. As a result, what is required to compute these times is Eq. (15) and, as long as one uses optimal milestones [in order that property (P1) holds and to get the correct Eq. (9)] and one reinitializes the trajectories according to Eq. (10) [to compute the first moment (15) correctly], it does not matter whether property (P2) holds or not. This is good news also because measuring the first moment τ_{ij} rather than the full density $f_{ij}(s)$ is easier.

Let us now give an explicit expression for the mean first passage time to a given milestone. Without loss of generality, we can assume that this milestone is the last one S_N since we can always make it so by relabeling the milestones index. Let us define

$$T_i = \left(\begin{array}{l} \text{mean first passage time from} \\ \text{milestone } S_i \text{ to milestone } S_N \end{array} \right). \quad (16)$$

Thus $T_i > 0$ for $i \neq N$ and $T_N = 0$. To compute this mean first passage time, we need to sum up the average lengths in time of all the sequences leading to S_N weighed by their probability of occurrence. This is easy to do with the help of a little trick which consists in introducing a modified p_{ij} , denoted as \hat{p}_{ij} and defined as

$$\hat{p}_{ij} = \begin{cases} p_{ij} & \text{if } i \neq N \\ \delta_{iN} & \text{if } i = N. \end{cases} \quad (17)$$

The modified \hat{p}_{ij} turns the milestone S_N into a so-called cemetery state: unlike with the original p_{ij} , with the new \hat{p}_{ij} the process stays at milestone S_N forever if it ever reaches this milestone. Therefore the average length in time of a path starting from S_i and moving from milestones to milestones for up to n transitions but staying in S_N if ever it reaches this state can be expressed as

$$\begin{aligned} T_i^{(n)} &= \sum_{j_1, \dots, j_n} (\tau_{ij_1} + \tau_{j_1j_2} + \dots + \tau_{j_{n-1}j_n}) \\ &\times \hat{p}_{ij_1} \hat{p}_{j_1j_2} \dots \hat{p}_{j_{n-1}j_n}, \end{aligned} \quad (18)$$

where τ_{ij} is the first moment of $f_{ij}(s)$ defined in Eq. (15) and we have set $\tau_{ii} = 0$: since the only nonzero \hat{p}_{ii} is $\hat{p}_{NN} = 1$, only τ_{NN} enters the sum in Eq. (18), and the choice of $\tau_{NN} = 0$ guarantees that the length in time of a path stops growing as soon as it reaches S_N . The mean first passage time T_i is the limit of Eq. (18) as $n \rightarrow \infty$, $T_i = \lim_{n \rightarrow \infty} T_i^{(n)}$ since this then takes into account paths of all possible length to reach S_N . Thus expression (18) confirms that the mean first passage time to a given milestones will be exact if property (P1) is satisfied (which ensures that p_{ij} and, hence, \hat{p}_{ij} contains the exact statistical information about probability of transitions between milestones), but regardless of whether (P2) holds or not.

To compute the limit of Eq. (18) as $n \rightarrow \infty$, let us introduce some notations and denote by $\mathbf{T}^{(n)} = (T_1^{(n)}, T_2^{(n)}, \dots, T_n^{(n)})^T$ the vector with entries $T_i^{(n)}$, by $\mathbf{1} = (1, 1, \dots, 1)^T$ the unit vector in \mathbb{R}^N , by \hat{P} the matrix with entries \hat{p}_{ij} , and by Q the matrix with entries $\tau_{ij} \hat{p}_{ij}$. Then, Eq. (18) can be expressed as

$$\mathbf{T}^{(n)} = \sum_{m=1}^n \hat{P}^{n-m} Q \hat{P}^{m-1} \mathbf{1} = \sum_{m=1}^n \hat{P}^{n-m} Q \mathbf{1}, \quad (19)$$

where we used $\hat{P} \mathbf{1} = \mathbf{1}$ which follows from \hat{P} being a transition probability matrix to get the second equality. Left multiplying by \hat{P} , we obtain the following recurrence relation for $\mathbf{T}^{(n)}$

$$\mathbf{T}^{(n+1)} = \hat{P} \mathbf{T}^{(n)} + \hat{\tau} \quad (20)$$

where $\hat{\tau} = Q \mathbf{1}$: explicitly $\hat{\tau}_i = \tau_i$ for $i \neq N$ and $\hat{\tau}_N = 0$ where $\tau_i = \sum_{j=1}^N p_{ij} \tau_{ij}$. We can now take the limit as $n \rightarrow \infty$ to obtain

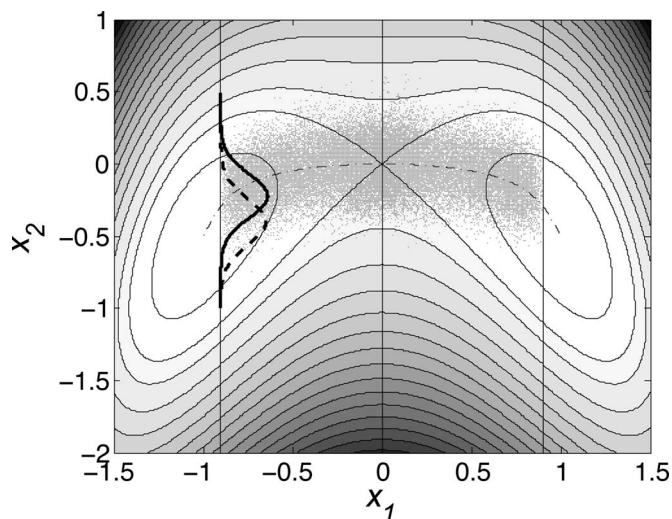


FIG. 2. Contour plot of the potential (23) with the three milestones $S_1 = \{x_1 = -0.9\}$, $S_2 = \{x_1 = 0\}$, and $S_3 = \{x_1 = 0.9\}$ (S_1 and S_3 are the boundaries of the reactant set $\{x_1 < 0.9\}$ and product set $\{x_2 > -0.9\}$ where $q(x) = 0$ and $q(x) = 1$, respectively, and S_2 is the isocommittor surface $q(x) = 0.5$ for this reaction). The minimum energy path is also shown (dot-dashed line). The gray dots are snapshots every $\delta t = 0.05$ along 10^3 trajectories starting from points distributed on S_2 according to Eq. (10), and the predicted density (10) of first hitting points on S_1 (thick black line) is compared to the equilibrium density on S_1 (dashed line).

$$(\text{Id} - \hat{P})\mathbf{T} = \hat{\tau}, \quad (21)$$

where $\mathbf{T} = (T_1, T_2, \dots, T_N)^T$ is the vector with entries T_i and Id is the identity matrix. This equation may seem problematic because $\det(\text{Id} - \hat{P}) = 0$ since $\text{Id} - \hat{P}$ has a zero eigenvalue associated with the eigenvector $\mathbf{1}$. However, since by definition the N th component of the vectors $(\text{Id} - \hat{P})\mathbf{T}$ and $\hat{\tau}$ are both zero, we can remove the last equation in the set [Eq. (21)], which leads to a regular set of $N-1$ equations for the $N-1$ unknown T_1, \dots, T_{N-1} which can easily be solved (and we already know that $T_N = 0$ by definition). Equation (21) agrees with the one for \mathbf{T} derived in Refs. 2 and 3 using different methods.

We conclude this section with a comment. From Eq. (21), the only quantities required to compute mean first passage times are p_{ij} and the mean time τ_i . As a result it does not matter what the actual shape of $f_{ij}(s)$ is, and we could approximate this function by

$$\tilde{f}_{ij}(s) = \tau_i^{-1} \exp(-s/\tau_i). \quad (22)$$

without changing Eq. (21). Since this approximation amounts to assuming that the dynamics of the transitions between successive milestones is a continuous-time Markov process,⁷ it was referred to as *Markovian milestoneing* in Ref. 3. In general, the approximation of $f_{ij}(s)$ by Eq. (22) is invalid, but we stress again that, when used with optimal milestones, it gives the correct mean first passage times. We will come back to this observation in Sec. VI when we compare milestoneing to MSMs.

III. ILLUSTRATIVE EXAMPLES

Let us now illustrate the results in Sec. II on a few simple examples. In Fig. 2, we analyze a two dimensional

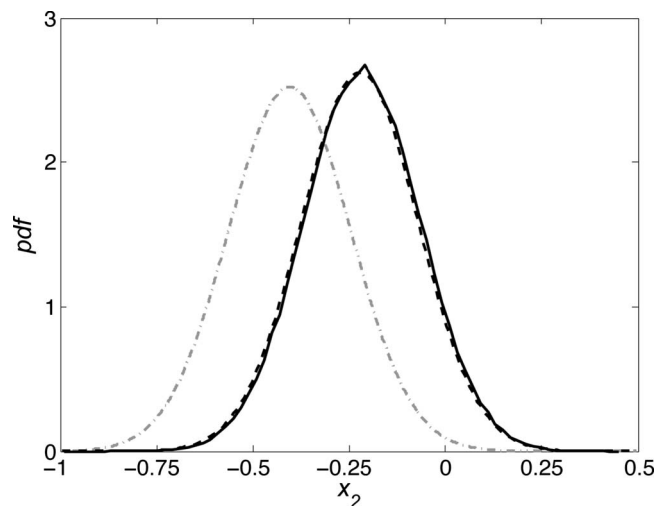


FIG. 3. Comparison of the probability density of first hitting points obtained by binning the location where 10^5 trajectories started from points in S_2 distributed according to Eq. (10) hit S_1 (black solid curve) with the density [Eq. (10)] (black dashed curve) and with the equilibrium density (gray dot-dashed curve). We also computed the first hitting point density using a long unbiased trajectory and, up to statistical errors, it coincides with the solid curve shown in the figure.

example with potential

$$V(x_1, x_2) = \frac{1}{4}(1 - x_1^2)^2 + \frac{1}{2}\left(x_2 + \frac{1}{2}x_1^2\right)^2 \quad (23)$$

in a situation where $\beta = 40$ and $\gamma = 1$ in Eq. (6). First we consider an extreme case in which we only take three milestones, $S_1 = \{x_1 = -0.9\}$, $S_2 = \{x_1 = 0\}$, and $S_3 = \{x_1 = 0.9\}$. Thus S_1 and S_3 mark the boundary of the reactant set $A = \{x_1 < -0.9\}$ and the product set $B = \{x_1 > 0.9\}$ [i.e., S_1 is also the isocommittor surface where $q(x) = 0$, while S_3 is the one where $q(x) = 1$], and by symmetry, S_2 is the isocommittor surface where $q(x) = 0.5$ for this reaction. These three surfaces are shown in Fig. 2 atop the contour plot of the potential (23). Also shown as gray dots are snapshots every $\delta t = 0.05$ of 10^3 trajectories generated from S_2 according to Eq. (10) in which we used a numerical approximation of $\nabla q(x)$ obtained by solving Eq. (A7) by finite elements (see Ref. 10 for details). The predicted density (10) of hitting points on S_1 is also shown (thick black line) and compared to the equilibrium density on S_1 , i.e., the density proportional to $e^{-\beta V(-0.9, x_2)}$ (dashed line). Clearly, these two densities are different, and Eq. (10) rather than the equilibrium density is the correct first hitting point density. This was also confirmed by computing this density directly from a long MD trajectory (not shown). The difference between Eq. (10) and the equilibrium density is also apparent from Fig. 3 where we compare the density of hitting points obtained by binning the location where 10^5 trajectories hit S_1 (black solid curve), with Eq. (10) (black dashed curve) and the equilibrium density (gray dot-dashed curve).

It should be stressed that working with three milestones only is an extreme case, and it is the reason why the density (10) on S_1 and S_3 is so different than the equilibrium one. Indeed, S_1 and S_3 are the boundary of sets A and B which are arbitrary. As a result, we should not expect Eq. (10) to agree with the equilibrium density on S_1 and S_3 . If, however, we

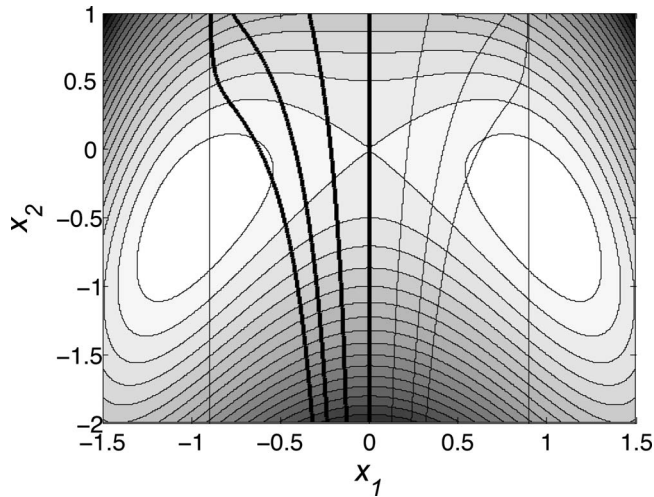


FIG. 4. Contour plot of the potential (23) with superimposed isocommittor surfaces used as milestones: from left to right, these surfaces are $q(x)=0$, $q(x)=0.001$, $q(x)=0.01$, $q(x)=0.1$, $q(x)=0.5$, $q(x)=0.9$, $q(x)=0.99$, $q(x)=0.999$, and $q(x)=1$.

include more milestones, then one can expect that the ones away from the boundaries of A and B will bend in the right direction in such a way that Eq. (10) and the equilibrium density may become more alike if $|\nabla q(x)| \approx cst$ in the region in $q(x)=z_j$, where $e^{-\beta V(x)}$ is peaked. This approximation will be valid if one can locally approximate the isocommittor surfaces by planes and neglect curvature effects along the reaction channel. A situation where this approximation is valid is illustrated in Figs. 4 and 5. In general, however, the extra factor $|\nabla q(x)|$ in Eq. (10) compared to the equilibrium density restricted to S_j will matter (as shown, e.g., in our second example below). How to account for it will be discussed some more in Sec. VI.

As a second illustration, let us consider an example

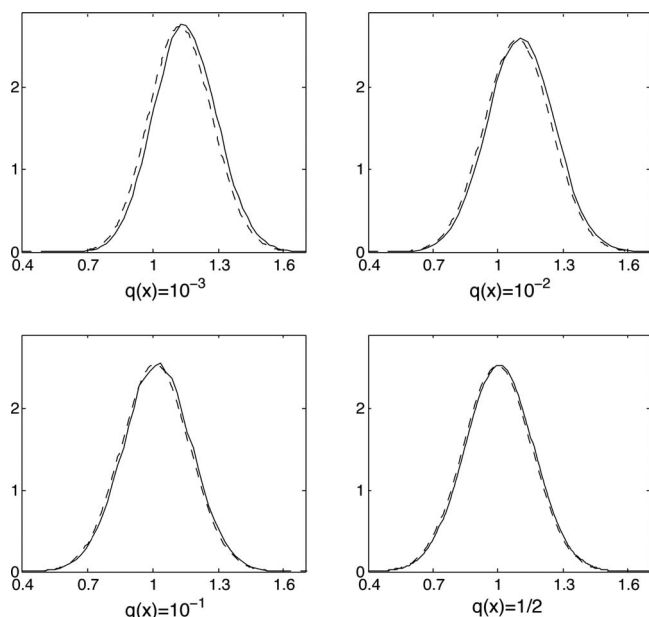


FIG. 5. Comparison of the probability density (10) (solid line) on the four milestones shown as thick lines in Fig. 4 with the equilibrium probability density (dashed line). The probability densities are plotted as functions of the arc length along the milestones.

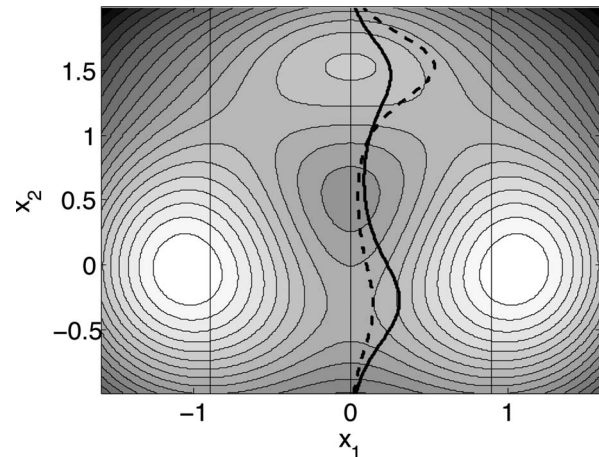


FIG. 6. Contour plot of the three-hole potential. We use three milestones, shown as vertical lines and corresponding to (from left to right) $q(x)=0$, $q(x)=0.5$, and $q(x)=1$. The density (10) (thick solid line) on the surface $q(x)=0.5$ shows that the lower channel is the preferred one, even though the equilibrium density (thick dashed line) is peaked in the upper channel.

originally proposed in Ref. 32 in which there are two reaction channels. The potential is shown in Fig. 6 and has been analyzed in detail in Ref. 10. This example shows why the time lags between transitions can be correlated even though these transitions are not. Indeed, because there are two channels (the upper and the lower ones), and the upper one contains a dynamical trap around the shallow minimum centered at $(0, 1.5)$, the time lags between transitions in the upper channel tend to be longer than between the ones in the lower one. Hence, a long lag between transition is more likely to be followed by a long one (indicating that the process goes through the upper channel), and a short lag by a short one (indicating that the process goes through the lower channel). Therefore, this is a situation where property (P1) holds but (P2) does not. It should also be stressed that this is an example where the factor $|\nabla q(x)|$ matters again. This is because there are two channels for the reaction, i.e., the Boltzmann factor $e^{-\beta V(x)}$ is peaked in two different regions in $q(x)=z_j$, and even though $|\nabla q(x)|$ is approximately constant in these two regions, these constants are different. This is illustrated in Fig. 6. Using the equilibrium density on $q(x)=0.5$ would suggest that the preferred channel is the upper one, whereas using Eq. (10) clearly shows that the lower one is preferred. This effect should be accounted for in the milestoning procedure. We come back to this point in Sec. VI.

IV. OPTIMAL MILESTONING IN SYSTEMS WITH INERTIA

In this section, we discuss what happens in systems with inertia, e.g., when we replace the overdamped Eq. (6) by the Langevin equation

$$\dot{x}(t) = v(t), \quad (24)$$

$$m\dot{v}(t) = -\nabla V(x(t)) - \gamma v(t) + \sqrt{2\beta^{-1}\gamma}\eta(t),$$

where $v(t)$ denotes the velocity, m the mass matrix, and γ the friction coefficient. Let us revisit the results of Sec. II in the context of Eq. (24). Much of the discussion below can be

generalized, at least formally, if the Langevin terms $-\gamma\mathbf{v}(t) + \sqrt{2\beta^{-1}}\boldsymbol{\eta}(t)$ in Eq. (24) are replaced by another thermal bath, possibly consistent with another ensemble than NVT .

A. Isocommittor surfaces as optimal milestones

Given any two nonintersecting smooth dividing surfaces in phase-space (\mathbf{x}, \mathbf{v}) , henceforth denoted as \mathcal{S}_1 and \mathcal{S}_N , the committor function associated with these surfaces can be defined as [cf. Eq. (7)]

$$q(\mathbf{x}_0, \mathbf{v}_0) = \left(\begin{array}{l} \text{probability that } (\mathbf{x}(t), \mathbf{v}(t)) \\ \text{reaches first } \mathcal{S}_N \text{ rather than } \mathcal{S}_1 \\ \text{given that } (\mathbf{x}(0), \mathbf{v}(0)) = (\mathbf{x}_0, \mathbf{v}_0) \end{array} \right). \quad (25)$$

As recalled in Appendix A, $q(\mathbf{x}, \mathbf{v})$ satisfies the partial differential Eq. (A9), and the isosurfaces of this function are nonintersecting smooth dividing surfaces foliating the region between \mathcal{S}_1 and \mathcal{S}_N . As a result, we can take as milestones the generalization of Eq. (8), i.e.,

$$\mathcal{S}_i = \{(\mathbf{x}, \mathbf{v}) : q(\mathbf{x}, \mathbf{v}) = z_i\}, \quad i = 1, \dots, N, \quad (26)$$

where $z_1=0 < z_2 < z_3 < \dots < z_N=1$, and these milestones will satisfy properties (P1) and (P1') exactly with the transition matrix p_{ij} given in Eq. (9). This shows that optimal milestoneing is possible, at least in principle, even for systems governed by Eq. (24).

There are, however, two caveats with this result. The first is that the optimal milestones are now nontrivial dividing surfaces in phase space rather than the simpler lift up in phase space of surfaces defined primarily in configuration space. The second, discussed next in Sec. IV B, is that, unlike in the case of systems in the overdamped limit, the first hitting point density associated with these optimal milestones is not available in closed form. (Practical considerations regarding both these issues are given in Sec. V.)

B. First hitting point density and the issue of reinitialization

To understand the issue which arises with the Langevin Eq. (24) when one tries to derive the first hitting point density for this equation, it is useful to recall some results of TPT. TPT can be applied to Eq. (24), but in this case the statistical properties of the reactive trajectories depend on two committor functions: the forward one, which was defined in Eq. (25), and the backward one. The forward committor function gives the probability to reach first a product rather than a reactant state starting from point (\mathbf{x}, \mathbf{v}) ; the backward committor function gives the probability to arrive at (\mathbf{x}, \mathbf{v}) coming last from a reactant state rather than a product state. Unlike in the overdamped case, the backward committor function is not simply one minus the forward one: rather it is given by $1 - q(\mathbf{x}, -\mathbf{v})$, where $q(\mathbf{x}, \mathbf{v})$ is the forward committor function defined in Eq. (25).

As discussed in Sec. IV A, optimal milestones having property (P1') and, hence, (P1) can be defined using the isosurfaces of the forward committor function. This is consistent with properties (P1) and (P1') dealing with what happens after the process leaves a milestones, i.e., forward in

time. However, when one tries to derive an explicit expression for the first hitting density of the milestones, one deals with the way the process reaches these surfaces, i.e., one must look backward in time. As a result, while it is possible to give an explicit expression for the first hitting point density of milestones defined as isosurfaces of the backward committor function (see Appendix B), such an explicit expression is not available for the optimal milestones defined in Eq. (26). To avoid confusions, let us stress that this negative conclusion does not mean that milestoneing (or even optimal milestoneing) is not viable in systems with inertia such as the ones governed by Eq. (24). Simply, it means that the exact probability density to reinitiate the MD trajectories on the milestones is not available in closed form, and one needs either to resort to approximations or find ways to bypass the need of this density altogether to perform milestoneing. These practical issues are discussed further in Sec. V.

C. Exact calculation of mean first passage times

Provided that one uses the optimal milestones defined in Eq. (26) and one uses the proper density to reinitialize the MD trajectories on these milestones to compute the first moment (15), the results given in Sec. II C apply to the present situation, i.e., Eq. (21) remains the exact equation for the mean first passage time from any milestone to the last.

V. SOME PRACTICAL CONSIDERATIONS

To make the results presented in this paper practical, two main issues must be addressed. The first is how to identify the isosurfaces of the committor function to be used as optimal milestones. The second is how to reinitialize the MD trajectories on these surfaces to compute the density (2) or its first moment (15).

The first issue suggests to combine milestoneing with the string method,^{11–16} which is a technique to calculate the committor function and its isosurfaces from scratch, at least in situations where the reaction channels are localized either in the original Cartesian space or when viewed in suitable collective variables. The string method can be applied either to the overdamped Eq. (6) or to the Langevin Eq. (24). In the latter case, however, it makes the approximation that the committor function $q(\mathbf{x}, \mathbf{v})$ can be approximated by a function of \mathbf{x} only, $q(\mathbf{x}, \mathbf{v}) \approx q(\mathbf{x})$ or, more generally, by a function of some suitable collective variables $\boldsymbol{\theta}(\mathbf{x}) = (\theta_1(\mathbf{x}), \theta_2(\mathbf{x}), \dots, \theta_N(\mathbf{x}))^T$ such as dihedral angles and bond distances: $q(\mathbf{x}, \mathbf{v}) \approx Q(\boldsymbol{\theta}(\mathbf{x}))$ [this second approximation can also be used for the overdamped Eq. (6)]. The output of the method is one or more curves, the hyperplanes perpendicular to which are local approximations near the curve of the isosurfaces of the committor function (see Refs. 5 and 16 for details). The accuracy of these approximations depends on the validity of both the assumptions above: that the reaction channels are localized and that the committor function can be approximated by a function of \mathbf{x} only. The latter assumption is difficult to justify on general grounds, but we note that a sufficient condition for it to be valid is that the overdamped Eq. (6) be a good approximation of Eq. (24) on longer time scales where the velocities have had time to decorrelate. This

suggests to use milestones that are separated enough so that the typical travel time between milestones will be longer than the velocity decorrelation time. For instance, it was observed empirically in Ref. 3 on the example of solvated alanine dipeptide that the milestones must be sufficiently separated to allow for accurate rate determination. This limits somewhat the efficiency gain achievable by milestoneing, but may not be a serious issue in practice since the velocity decorrelates in MD simulations on the subpicosecond time scale while we are interested in time scales that are at least a million times longer. For instance, termination times of picoseconds between milestones yield computational speed up of about 1000 for Scapharca hemoglobin.⁴

Considering now the issue of how to reinitialize the MD trajectories on the milestones, for overdamped systems at least, the results of the string method can also be used to estimate the extra factor $|\nabla q(\mathbf{x})|$ to be included in Eq. (10). The problem, however, is that such an approximation can be rather poor (approximating the gradient of a function is typically harder than approximating the function itself) and, in addition, this procedure is not an option for systems with inertia for which the equivalent of Eq. (10) is not available in closed form. One way to get around these difficulties in overdamped systems is to assume that $|\nabla q(\mathbf{x})|$ is approximately constant on the isosurfaces of $q(\mathbf{x})$. As was illustrated in Sec. III via examples, this approximation is sometimes valid but not always. Another way, which is applicable also for systems with inertia, is to use milestones sufficiently separated so that decorrelation effects along the trajectory make the influence of the density used to initiate the trajectories on the milestones less stringent. This is the procedure which was used in the original milestoneing papers,¹⁻⁴ where MD trajectories are generated on the milestones using the equilibrium density restricted to these milestones. Yet another way, applicable again both to overdamped systems and to those with inertia, is to come up with a sampling procedure that bypasses the need of the first hitting point density altogether. Such a sampling procedure will be presented in Ref. 5 and we refer the reader to this reference for details.

VI. CONCLUDING REMARKS

In this paper, we have analyzed the assumptions underlying milestoneing, namely, properties (P1) and (P2), whose specific roles and importance are different. Property (P1) states that transitions between successive milestones are statistically independent and one of the main results in this paper was to establish that property (P1) is satisfied exactly provided that one uses isocommittor surfaces as milestones. Property (P2), on the other hand, does not hold exactly in general, i.e., it is not true that the time lags between successive transitions between milestones are statistically independent. However, we showed that the importance of this property is less than that of (P1). In particular, it was established that mean first passage times from any milestone to any other one can be calculated exactly if one uses milestones satisfying property (P1) and computes the first moment (15) exactly, even if (P2) does not hold. This leads to what we called optimal milestoneing. Optimal milestoneing requires one to use

isocommittor surfaces as milestones, which in practice suggests to combine milestoneing with the string method, as mentioned in Sec. V. Optimal milestoneing also requires one to know the probability density of hitting points on the milestones. This density is the one to use to reinitiate the trajectory on the milestones in order to compute the key quantities (1) and (2). For systems in the overdamped limit, we derived an explicit expression for this density. We also explained why such a density is not available for the optimal milestones associated with system with inertia such as the ones governed by the Langevin Eq. (24) but, as discussed in Sec. V, there are various ways to address this issue in practice.

The fact that, in principle at least, optimal milestoneing allows one to compute exactly quantities such as mean first passage times is, we believe, a great advantage of the method. It should be stressed, however, that milestoneing can remain useful even in situations where property (P1) is not satisfied exactly. In fact, the way milestoneing was originally implemented in Refs. 1-5 is not optimal in the sense above, and yet it was demonstrated in this paper that the method does give accurate predictions of the long-time kinetics of various complicated MD systems. While it is difficult to give a precise assessment of why this is the case, this seems to be related to the use of well-separated milestones. The separation must be such that the trajectory decorrelates as it travels between successive milestones in such a way that properties (P1) and (P2) are approximately satisfied even though the milestones are not optimal. Clearly, using optimal milestones should remain the preferred solution, but it shows that things do not necessarily go wrong if this option is not available.

We now conclude this paper with quick comparisons of (i) milestoneing versus MSMs, and (ii) milestoneing versus TIS and FFS.

A. Milestoneing versus MSM

It is useful to compare milestoneing with MSMs in light of our results. MSMs have recently become popular to analyze the kinetics of complicated processes by gathering information about these processes either via short MD runs, or sometimes even in situations where the thought experiment described in the Introduction is realized, i.e., one actually has a very long MD trajectory at one's disposal. The basic idea in MSM is to reduce the dynamics to a Markov jump process between the appropriately chosen states in such a way that this reduced dynamics be consistent with the actual one. Thus, at first sight, MSM seems quite similar in spirit to the Markovian milestoneing procedure described at the end of Sec. II C. There is, however, an important difference, namely, that the states used in MSM typically form a partition of phase space, whereas those used in Markovian milestoneing are hypersurfaces in phase space (e.g., the states in MSM could be the regions between the milestones rather than the milestones themselves). A consequence of this is that it is hard to justify when MSMs of this type are valid and, in particular, it is not obvious how to choose the time lag at which to compute the transition probability matrix from the MD data in these models. There have been recent studies to address this issue (see, e.g., Refs. 23 and 24), but

to the best of our knowledge, there are no rigorous results as to when MSMs are accurate. This situation is to be contrasted with milestoning which, provided that optimal milestones are being used, permits to compute exactly mean first passage times. We believe that this is an important advantage of milestoning (Markovian or not) over standard MSM that makes the former preferable to the latter for the analysis of the kinetics of complicated MD processes.

B. Milestoning versus TIS and FFS

Finally, our results show that milestoning is an interesting alternative to techniques such as TIS^{26,27} or FFS.²⁸ TIS and FFS allow one to compute exactly the flux of reactive trajectories through a given set of interfaces and thereby the mean first passages between these interfaces. The interfaces used in TIS and FFS are, in principle, arbitrary (i.e., they do not need to be the isocommittor surfaces as in optimal milestoning). In practice, however, the efficiency of these methods will drop if one does not use appropriate surfaces. Ideally, one would like to choose these surfaces so as to optimize the computational gain offered by TIS or FFS, but how to do so is not clear. The isocommittor surfaces seem again a natural choice but, to the best of our knowledge, there is no rigorous proof that this choice is the best one. In addition, TIS and FFS are not as straightforward to parallelize as milestoning and as a result they will typically offer lesser computational gains. It would be interesting to investigate these questions further but we will leave this for future work.

ACKNOWLEDGMENTS

We are grateful to Philipp Metzner for providing us with the committor function data in the examples of Sec. III. We also thank the referees for their thoughtful comments that led us to rethink the results of this paper and enlarge its scope. Part of this work was performed while the authors were visiting the Erwin Schrödinger Institute (ESI) in Vienna whose support is gratefully acknowledged. This work was also partially supported by NIH under Grant No. GM59796, NSF under Grant Nos. DMS02-09959, DMS02-39625, and DMS07-08140, and ONR under Grant No. N00014-04-1-0565.

APPENDIX A: THE COMMITTOR EQUATION

Here, we give the equations for the committor function associated with the overdamped Eq. (6) and the Langevin Eq. (24). This is a standard material which can be found in any textbook on stochastic processes, e.g., Sec. 5.7.B in Ref. 33. We start by deriving the equation for the committor function associated with Eq. (6). If $\mathbf{x}(t)$ satisfies Eq. (6), given a function $u(\mathbf{x})$, Itô's change in variable formula (e.g., Sec. 3.3 in Ref. 33) asserts that

$$\dot{u}(\mathbf{x}(t)) = (Lu)(\mathbf{x}(t)) + \sqrt{2(\beta\gamma)^{-1}} \nabla u(\mathbf{x}(t)) \cdot \boldsymbol{\eta}(t). \quad (\text{A1})$$

Here L denotes the infinitesimal generator of the diffusion associated with Eq. (6), i.e.,

$$(Lu)(\mathbf{x}) = -\gamma^{-1} \nabla V(\mathbf{x}) \cdot \nabla u(\mathbf{x}) + (\beta\gamma)^{-1} \Delta u(\mathbf{x}), \quad (\text{A2})$$

where Δ denotes the Laplacian. Suppose that we consider Eq. (6) in a domain $\Omega \subset \mathbb{R}^d$, and denote by τ the first exit time of Ω , i.e.,

$$\tau = \min(t: \mathbf{x}(t) \in \partial\Omega), \quad (\text{A3})$$

where $\partial\Omega$ denotes the boundary of the set Ω . Integrating Eq. (A1) up to time τ and taking the expectation using the property that the noise term has zero expectation, we obtain

$$\langle u(\mathbf{x}(\tau)) \rangle = u(\mathbf{x}_0) + \left\langle \int_0^\tau (Lu)(\mathbf{x}(t)) dt \right\rangle, \quad (\text{A4})$$

where $\mathbf{x}_0 = \mathbf{x}(0)$ is the initial condition of Eq. (6). So far, u was an arbitrary function. Now, let us assume that it is the solution of

$$(Lu)(\mathbf{x}) = 0 \quad \text{if } \mathbf{x} \in \Omega, \quad (\text{A5})$$

$$u(\mathbf{x}) = f(\mathbf{x}) \quad \text{if } \mathbf{x} \in \partial\Omega,$$

where $f(\mathbf{x})$ is some given function defined on the boundary of Ω . Then Eq. (A4) reduces to

$$\langle f(\mathbf{x}(\tau)) \rangle = u(\mathbf{x}_0), \quad (\text{A6})$$

where we used that $\mathbf{x}(\tau) \in \partial\Omega$, and so $u(\mathbf{x}(\tau)) = f(\mathbf{x}(\tau))$ from the boundary condition in Eq. (A5). Equation (A6) is known as (an instance of) Feynman–Kac formula which gives an expression for the solution of the elliptic partial differential Eq. (A5) in terms of an expectation involving the solution of the stochastic differential Eq. (6). An interesting special case is when Ω is the whole configuration space minus the union of sets A and B , $\Omega = \mathbb{R}^d \setminus (A \cup B)$, and one sets $f(\mathbf{x}) = 0$ if $\mathbf{x} \in \partial A$ and $f(\mathbf{x}) = 1$ if $\mathbf{x} \in \partial B$. Denoting by $q(\mathbf{x})$ instead of $u(\mathbf{x})$ the solution of Eq. (A5) in this special case, one sees that Eq. (A5) reduces to

$$\begin{aligned} (Lq)(\mathbf{x}) &= 0 \quad \text{if } \mathbf{x} \in \mathbb{R}^d \setminus (A \cup B), \\ q(\mathbf{x}) &= 1 \quad \text{if } \mathbf{x} \in \partial B \\ q(\mathbf{x}) &= 0 \quad \text{if } \mathbf{x} \in \partial A, \end{aligned} \quad (\text{A7})$$

and, from Eq. (A6), the solution of this equation gives the probability that $\mathbf{x}(\tau)$ belongs to ∂B rather than ∂A , i.e., it is the committor function defined in Eq. (7) provided that we set $\partial A = S_1$ and $\partial B = S_N$. Note that, using the time-reversal symmetry of the overdamped equation, $q(\mathbf{x})$ also is the probability to arrive at \mathbf{x} coming last from B rather than A , i.e., $1 - q(\mathbf{x})$ is the backward committor function.

The case of the Langevin Eq. (24) can be treated similarly, the main difference is that all calculations must now be done in phase space rather than configuration space. The generator of the diffusion associated with Eq. (24) is [cf. Eq. (A2)]

$$\begin{aligned} (\mathcal{L}u)(\mathbf{x}, \mathbf{v}) &= -\mathbf{v} \cdot \nabla_{\mathbf{x}} u(\mathbf{x}, \mathbf{v}) - \nabla V(\mathbf{x}) \cdot \Delta_{\mathbf{v}} u(\mathbf{x}, \mathbf{v}) \\ &\quad - \gamma \mathbf{v} \cdot \nabla_{\mathbf{v}} u(\mathbf{x}, \mathbf{v}) + \beta^{-1} \gamma \Delta_{\mathbf{v}} u(\mathbf{x}, \mathbf{v}) \end{aligned} \quad (\text{A8})$$

and the equation for the committor function is [cf. (A7)]

$$\begin{aligned}
(\mathcal{L}q)(\mathbf{x}, \mathbf{v}) &= 0 \quad \text{if } (\mathbf{x}, \mathbf{v}) \in \mathbb{R}^{2d} \setminus (\mathcal{A} \cup \mathcal{B}) \\
q(\mathbf{x}, \mathbf{v}) &= 1 \quad \text{if } \mathbf{x} \in \partial\mathcal{B} \\
q(\mathbf{x}, \mathbf{v}) &= 0 \quad \text{if } \mathbf{x} \in \partial\mathcal{A},
\end{aligned} \tag{A9}$$

where \mathcal{A} and \mathcal{B} are sets in phase space defining the reactant and the product states, respectively, and $\partial\mathcal{A}$ and $\partial\mathcal{B}$ are the boundaries of these sets. The only additional caveat with Eq. (A9) compared to Eq. (A7) is that the local vector normal to the boundaries $\partial\mathcal{A}$ and $\partial\mathcal{B}$ must have a nonzero velocity component everywhere on $\partial\mathcal{A}$ and $\partial\mathcal{B}$ except possibly on a set of zero Lebesgue measure on these sets. This condition is required in order to be able to impose the Dirichlet boundary conditions in Eq. (A9) because the Laplacian term in Eq. (A9) involves only \mathbf{v} .

Finally, we note that the function $q(\mathbf{x}, \mathbf{v})$ is the forward committor function, i.e., it gives the probability to first reach \mathcal{B} rather than \mathcal{A} forward in time starting from point (\mathbf{x}, \mathbf{v}) . The backward committor function, giving the probability to arrive at point (\mathbf{x}, \mathbf{v}) coming last from \mathcal{A} rather than \mathcal{B} is given by $1 - q(\mathbf{x}, -\mathbf{v})$ [and not simply $1 - q(\mathbf{x}, \mathbf{v})$] as a result of the symmetry by time reversal of the Langevin Eq. (24).

APPENDIX B: FIRST HITTING POINT DENSITY

Here we show that the first hitting point density on milestone S_j is given by Eq. (10). Specifically, we show that if a solution of Eq. (6) is initially distributed on S_j according to $\rho_j(\mathbf{x})$ (consistent with this density being assumed to be the one of the first hitting point on this surface), then the probability density of the first hitting point on S_{j+1} (S_{j-1}) conditional on the process hitting this surface first is given by $\rho_{j+1}(\mathbf{x})$ [$\rho_{j-1}(\mathbf{x})$]. Since this argument can be repeated to all the milestones, this is consistent with Eq. (10) being the first hitting point density on milestone S_j .

To get the probability density of the first hitting point on S_{j+1} and S_{j-1} , we consider the probability density at time t of the solution of Eq. (6) in the region Ω_j between S_{j-1} and S_{j+1} , i.e., $\Omega_j = \{\mathbf{x} : z_{j-1} < q(\mathbf{x}) < z_{j+1}\}$ (if $j=1, N$, we simply set $\Omega_1 = \{\mathbf{x} : q(\mathbf{x}) < z_2\}$ and $\Omega_N = \{\mathbf{x} : z_{N-1} < q(\mathbf{x})\}$). Denoting by $p_t(\mathbf{x})$ the probability density that the particle be at \mathbf{x} at time t , $p_t(\mathbf{x})$ satisfies

$$\partial_t p_t(\mathbf{x}) = \gamma^{-1} \nabla \cdot (\nabla V(\mathbf{x}) p_t(\mathbf{x}) + \beta^{-1} \nabla p_t(\mathbf{x})) \tag{B1}$$

for $\mathbf{x} \in \Omega_j$, and $p_t(\mathbf{x}) = 0$ if $\mathbf{x} \in S_{j-1}$ or $\mathbf{x} \in S_{j+1}$. As initial condition for Eq. (B1), we take the distribution in \mathbb{R}^d whose density on S_j is $\rho_j(\mathbf{x})$,³¹

$$p_{t=0}(\mathbf{x}) = Z_j^{-1} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})|^2 \delta(q(\mathbf{x}) - z_j). \tag{B2}$$

We also set absorbing boundary conditions at S_{j-1} and S_{j+1} , i.e., $p_t(\mathbf{x}) = 0$ if $\mathbf{x} \in S_{j-1}$ or $\mathbf{x} \in S_{j+1}$. Due to the presence of these absorbing boundaries, we have that $p_t(\mathbf{x}) \rightarrow 0$ for $\mathbf{x} \in \Omega_j$ as $t \rightarrow \infty$, and this decay is associated with escape events by the boundaries S_{j-1} and S_{j+1} . Writing Eq. (B1) as

$$\partial_t p_t(\mathbf{x}) = -\nabla \cdot J_t(\mathbf{x}), \tag{B3}$$

where $J_t(\mathbf{x})$ is the probability current

$$J_t(\mathbf{x}) = -\gamma^{-1} \nabla V(\mathbf{x}) p_t(\mathbf{x}) - (\gamma\beta)^{-1} \nabla p_t(\mathbf{x}), \tag{B4}$$

we see that the probability flux density at point $\mathbf{x} \in S_{j-1} \cup S_{j+1}$ and time t is given by

$$\gamma^{-1} \hat{n}(\mathbf{x}) \cdot (\nabla V(\mathbf{x}) p_t(\mathbf{x}) + \beta^{-1} \nabla p_t(\mathbf{x})), \tag{B5}$$

where $\hat{n}(\mathbf{x})$ denotes the unit vector pointing outward Ω_j at $\mathbf{x} \in S_{j-1} \cup S_{j+1}$: explicitly, $\hat{n}(\mathbf{x}) = \nabla q(\mathbf{x}) / |\nabla q(\mathbf{x})|$ for $\mathbf{x} \in S_{j+1}$ and $\hat{n}(\mathbf{x}) = -\nabla q(\mathbf{x}) / |\nabla q(\mathbf{x})|$ for $\mathbf{x} \in S_{j-1}$. We are interested in the time-integrated probability flux density at the point $\mathbf{x} \in S_{j-1} \cup S_{j+1}$ since this time-integrated flux density gives the overall probability density that the particle escapes by this point (the hitting density). To calculate this flux density, let

$$P(\mathbf{x}) = \int_0^\infty p_t(\mathbf{x}) dt. \tag{B6}$$

Using the initial condition (B2) as well as $\lim_{t \rightarrow \infty} p_t(\mathbf{x}) = 0$, by integrating Eq. (B1) in time from 0 to ∞ , we arrive at the following equation for $P(\mathbf{x})$:

$$-\rho_j(\mathbf{x}) = \gamma^{-1} \nabla \cdot (\nabla V(\mathbf{x}) P(\mathbf{x}) + \beta^{-1} \nabla P(\mathbf{x})) \tag{B7}$$

for $\mathbf{x} \in \Omega_j$, and $P(\mathbf{x}) = 0$ if $\mathbf{x} \in S_{j-1}$ or $\mathbf{x} \in S_{j+1}$. We claim that the solution of Eq. (B7) is

$$P(\mathbf{x}) = \begin{cases} C_L (q(\mathbf{x}) - z_{j-1}) e^{-\beta V(\mathbf{x})} & \text{if } z_{j-1} \leq q(\mathbf{x}) < z_j \\ C_R (z_{j+1} - q(\mathbf{x})) e^{-\beta V(\mathbf{x})} & \text{if } z_j \leq q(\mathbf{x}) \leq z_{j+1}, \end{cases} \tag{B8}$$

where

$$\begin{aligned}
C_L &= \gamma Z_j^{-1} \frac{z_{j+1} - z_j}{z_{j+1} - z_{j-1}}, \\
C_R &= \gamma Z_j^{-1} \frac{z_j - z_{j-1}}{z_{j+1} - z_{j-1}}.
\end{aligned} \tag{B9}$$

Using Eq. (A7), it is easy to see that Eq. (B8) solves Eq. (B7) in the region $\Omega_j \setminus S_j$ in which $\rho_j(\mathbf{x}) = 0$, that it satisfies the right boundary condition on S_{j+1} and S_{j-1} , and that it is continuous (though not differentiable) on S_j . Thus it simply remains to establish that Eq. (B8) also satisfies Eq. (B7) on S_j in some appropriate sense. Specifically, we need to show that if we take $\delta > 0$ and an arbitrary test function $f(\mathbf{x})$, multiply Eq. (B7) by $f(\mathbf{x})$, integrate the resulting equation on the region $z_j - \delta \leq q(\mathbf{x}) \leq z_j + \delta$, and finally let $\delta \rightarrow 0$, we get an equation which is satisfied with $P(\mathbf{x})$ given by Eq. (B8). Performing this calculation and using integration by parts at the right-hand side, the result is

$$\begin{aligned}
& -Z_j^{-1} \int_{S_j} f(\mathbf{x}) e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})| d\sigma_{S_j}(\mathbf{x}) \\
&= -\gamma^{-1} (C_L + C_R) \int_{S_j} f(\mathbf{x}) e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})| d\sigma_{S_j}(\mathbf{x}),
\end{aligned} \tag{B10}$$

which is indeed satisfied for every $f(\mathbf{x})$ in view of Eq. (B9). Having established that Eq. (B8) is the solution of Eq. (B7), we can calculate the time-integrated flux at $\mathbf{x} \in S_{j-1} \cup S_{j+1}$ which gives the probability density that the particle exits at

this point. Time-integrating Eq. (B5) and using Eq. (B6), this hitting density can be expressed in terms of $P(\mathbf{x})$ as

$$\gamma^{-1} \hat{n}(\mathbf{x}) \cdot (\nabla V(\mathbf{x})P(\mathbf{x}) + \beta^{-1} \nabla P(\mathbf{x})), \quad (\text{B11})$$

which, using Eq. (B8), can be expressed, on S_{j-1} and S_{j+1} , as

$$\rho_L(\mathbf{x}) = \gamma^{-1} C_L e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})| \quad \text{if } \mathbf{x} \in S_{j-1}$$

$$\rho_R(\mathbf{x}) = \gamma^{-1} C_R e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})| \quad \text{if } \mathbf{x} \in S_{j+1}, \quad (\text{B12})$$

which, according to Eq. (10), are indeed proportional to $\rho_{j-1}(\mathbf{x})$ and $\rho_{j+1}(\mathbf{x})$, respectively. In fact, the integrals of $\rho_L(\mathbf{x})$ on S_{j-1} and of $\rho_R(\mathbf{x})$ on S_{j+1} give the probabilities that the particle exits by S_{j-1} or S_{j+1} , respectively,

$$\int_{\mathbb{R}^d} \rho_L(\mathbf{x}) d\mathbf{x} = \frac{z_{j+1} - z_j}{z_{j+1} - z_{j-1}} \equiv p_{jj-1},$$

$$\int_{\mathbb{R}^d} \rho_R(\mathbf{x}) d\mathbf{x} = \frac{z_j - z_{j-1}}{z_{j+1} - z_{j-1}} \equiv p_{jj+1}, \quad (\text{B13})$$

where we used Eqs. (11) and (B9) as well as the property (shown below) that for any $j, k=1, \dots, N$, we have

$$\int_{S_j} |\nabla q(\mathbf{x})| e^{-\beta V(\mathbf{x})} d\sigma_{S_j}(\mathbf{x}) = \int_{S_k} |\nabla q(\mathbf{x})| e^{-\beta V(\mathbf{x})} d\sigma_{S_k}(\mathbf{x}). \quad (\text{B14})$$

Eq. (B13) is consistent with Eq. (9), and we see that $\rho_{j-1}(\mathbf{x}) = \rho_L(\mathbf{x})/p_{jj-1}$ and $\rho_{j+1}(\mathbf{x}) = \rho_R(\mathbf{x})/p_{jj+1}$. This shows that the hitting probability on S_{j-1} conditional on exiting through S_{j-1} is indeed $\rho_{j-1}(\mathbf{x})$, and similarly for S_{j+1} .

Finally to show that Eq. (B14) holds, consider

$$\begin{aligned} I(z) &= \int_{q(\mathbf{x})=z} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})| d\sigma_{\{q(\mathbf{x})=z\}} \\ &= \int_{\mathbb{R}^d} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})|^2 \delta(q(\mathbf{x}) - z) d\mathbf{x}. \end{aligned} \quad (\text{B15})$$

We claim that $I(z)$ is independent of z , which implies Eq. (B14). To see this, use

$$\begin{aligned} \frac{dI(z)}{dz} &= - \int_{\mathbb{R}^d} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})|^2 \delta'(q(\mathbf{x}) - z) d\mathbf{x} \\ &= - \int_{\mathbb{R}^d} e^{-\beta V(\mathbf{x})} \nabla q(\mathbf{x}) \cdot \nabla \delta(q(\mathbf{x}) - z) d\mathbf{x} \\ &= \beta \gamma \int_{\mathbb{R}^d} e^{-\beta V(\mathbf{x})} (Lq)(\mathbf{x}) \delta(q(\mathbf{x}) - z) d\mathbf{x} = 0, \end{aligned} \quad (\text{B16})$$

where we integrated by parts to get the third equality and we used Eq. (A2) to get the fourth.

Similar calculations can be done for the Langevin Eq. (24) but, as explained in Sec. IV B, we can only get a closed form expression for the first hitting point density if we use as milestones the isosurfaces of the backward committor function rather than the forward one. For economy, let us omit these calculations since they are similar to the one presented

above and simply state the result: the first hitting point density of the surface $S_j^b = \{\mathbf{x}, \mathbf{v} : q(\mathbf{x}, -\mathbf{v}) = z_j\}$ [not to be confused with Eq. (26)] is given by

$$\varrho_j(\mathbf{x}, \mathbf{v}) = \mathcal{Z}_j^{-1} |\nabla_{\mathbf{v}} q(\mathbf{x}, -\mathbf{v})| e^{-\beta H(\mathbf{x}, \mathbf{v})}. \quad (\text{B17})$$

Here \mathcal{Z}_j is the following normalization factor:

$$\mathcal{Z}_j = \int_{S_j^b} |\nabla_{\mathbf{v}} q(\mathbf{x}, -\mathbf{v})| e^{-\beta H(\mathbf{x}, \mathbf{v})} d\sigma_{S_j^b}(\mathbf{x}, \mathbf{v}), \quad (\text{B18})$$

where $d\sigma_{S_j^b}(\mathbf{x}, \mathbf{v})$ denotes the surface element on S_j^b , and $H(\mathbf{x}, \mathbf{v})$ is the Hamiltonian of the system.

¹A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**, 10880 (2004).

²D. Shalloway and A. K. Faradjian, *J. Chem. Phys.* **124**, 054112 (2006).

³A. M. A. West, R. Elber, and D. Shalloway, *J. Chem. Phys.* **126**, 145104 (2007).

⁴R. Elber, *Biophys. J.* **92**, L85 (2007).

⁵E. Vanden-Eijnden, M. Venturoli, and R. Elber, "Markovian milestone with Voronoi tessellations," *J. Chem. Phys.* (2008) (in preparation).

⁶Instead of taking a single long ergodic trajectory, we can also consider an ensemble of these, e.g., by considering the ensemble of solutions of Eq. (6) or (24) generated using different realizations of the noise and initial conditions drawn from the canonical distribution. The definitions (1) and (2) for p_{ij} and $f_{ij}(s)$, which involve time averages, remain valid in this case, but, alternatively, we can also define these quantities using appropriate averages with respect to the ensemble of trajectories.

⁷To avoid confusions, note that assuming that the sequence (S_k, t_k) is Markov with a discrete time index $k=0, 1, 2, \dots$ does *not* imply that the successive crossings of the milestones can be described as a continuous-time Markov process whose state-space is the index of the last milestone crossed, say $i(t)$. Indeed this would require that Eq. (2) is of the form $f_{ij}(s) = \tau_i^{-1} \exp(-s/\tau_i)$ for some $\tau_i > 0$. This is assumed in Markovian milestone (see the remark at the end of Sec. II C), but not in milestoneing.

⁸W. E and E. Vanden-Eijnden, *J. Stat. Phys.* **123**, 503 (2006).

⁹E. Vanden-Eijnden, in *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer, Berlin, 2006), Vol. 1, pp. 439–478.

¹⁰P. Metzner, C. Schütte, and E. Vanden-Eijnden, *J. Chem. Phys.* **125**, 084110 (2006).

¹¹W. E, W. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).

¹²W. E, W. Ren, and E. Vanden-Eijnden, *Chem. Phys. Lett.* **413**, 242 (2005).

¹³W. E, W. Ren, and E. Vanden-Eijnden, *J. Phys. Chem. B* **109**, 6688 (2005).

¹⁴W. E, E. Vanden-Eijnden, and P. Maragakis, *J. Chem. Phys.* **123**, 134109 (2005).

¹⁵L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *J. Chem. Phys.* **125**, 024106 (2006).

¹⁶E. Vanden-Eijnden and M. Venturoli, "Revisiting the Finite Temperature String Method for the Calculation of Reaction Tubes and Free Energies," *J. Chem. Phys.* (to be published).

¹⁷D. Shalloway, *J. Chem. Phys.* **105**, 9986 (1996).

¹⁸A. Ulitsky and D. Shalloway, *J. Chem. Phys.* **109**, 1670 (1998).

¹⁹G. Hummer and I. Kevrekidis, *J. Chem. Phys.* **118**, 10762 (2003).

²⁰W. Swope, J. Pitera, and F. Suits, *J. Phys. Chem. B* **108**, 6571 (2004).

²¹W. C. Swope, J. W. Pitera, F. Suits, M. C. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubski, T. J. C. Ward, Y. Zhestkov, and R. J. Zhou, *J. Phys. Chem. B* **108**, 6582 (2004).

²²S. Krivov and M. Karplus, *J. Chem. Phys.* **117**, 10894 (2004).

²³F. Noe, I. Horenko, C. Schütte, and J. Smith, *J. Chem. Phys.* **126**, 155102 (2007).

²⁴N. Buchete and G. Hummer, *J. Phys. Chem. B* **112**, 6057 (2008).

²⁵J. Chodera, N. Singhal, W. C. Swope, V. S. Pande, and K. A. Dill, *J. Chem. Phys.* **126**, 155101 (2007).

²⁶D. Moroni, T. S. van Erp, and P. G. Bolhuis, *Physica A* **340**, 395 (2004).

²⁷D. Moroni, P. G. Bolhuis, and T. S. van Erp, *J. Chem. Phys.* **120**, 4055 (2004).

²⁸R. J. Allen, D. Frenkel, and P. R. ten Wolde, *J. Chem. Phys.* **124**, 024102 (2006).

²⁹ A. Ma and A. R. Dinner, *J. Phys. Chem. B* **109**, 6769 (2005).

³⁰ B. Peters and B. L. Trout, *J. Chem. Phys.* **125**, 054108 (2006).

³¹ Note that Eq. (10) is a probability density function with respect to the Lebesgue measure (i.e., the surface element) $d\sigma_{S_j}(\mathbf{x})$ on S_j . Thus the support of the associated probability distribution is S_j . Formally, this distribution can be written in \mathbb{R}^d as $Z_j^{-1} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})|^2 \delta(q(\mathbf{x}) - z_j) d\mathbf{x}$, where $\delta(q(\mathbf{x}) - z_j)$ denotes the Dirac delta function. This representation

can be justified from the identity $\int_{\mathbb{R}^d} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})|^2 \delta(q(\mathbf{x}) - z_j) f(\mathbf{x}) d\mathbf{x} = \int_{S_j} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})| f(\mathbf{x}) d\sigma_{S_j}(\mathbf{x})$ valid for any function $f(\mathbf{x})$ such that the surface integral exists.

³² S. Huo and J. E. Straub, *J. Chem. Phys.* **107**, 5000 (1997).

³³ I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Graduate Texts in Mathematics Vol. 113, 2nd ed. (Springer-Verlag, New York, 1991).