MULTISCALE MODEL. SIMUL.
Vol. 14, No. 1, pp. 301–322

A MATHEMATICAL FRAMEWORK FOR EXACT MILESTONING*

DAVID ARISTOFF†, JUAN M. BELLO-RIVAS‡, AND RON ELBER§

Abstract. We give a mathematical framework for exact milestoning, a recently introduced algorithm for mapping a continuous time stochastic process into a Markov chain or semi-Markov process that can be efficiently simulated and analyzed. We generalize the setting of exact milestoning and give explicit error bounds for the error in the milestoning equation for mean first passage times.

Key words. accelerated molecular dynamics, long-time dynamics, stationary distribution, semi-Markov processes

AMS subject classifications. 82C21, 82C80, 37A30

DOI. 10.1137/15M102157X

1. Introduction.

Molecular dynamics (MD) simulations, in which classical equations of motions are solved for molecular systems of significant complexity, have proven useful for interpreting and understanding many chemical and biological phenomena (for textbooks, see [47, 24, 3]). However, a significant limitation of MD is that of time scales. Many molecular processes of interest occur on time scales significantly longer than the temporal scales accessible to straightforward simulations. For example, permeation of molecules through membranes can take hours [9], while MD is usually restricted to a scale of microseconds. One approach to extending simulation times is to use faster hardware [48, 49, 45]. Other approaches focus on developing theories and algorithms for long-time phenomena. Most of the emphasis has been on methodologies for activated processes with a single dominant barrier, as in transition path sampling [16, 7, 15]. Approaches for dynamics on rough energy landscapes, and for more general and/or diffusive dynamics, have also been developed [41, 46, 14, 50, 43]. The techniques of exact milestoning [5] and milestoning [22] belong to the last category. They are theories and algorithms for accelerating trajectory calculations of kinetics and thermodynamics in complex molecular systems. The acceleration is based on the use of a large number of short trajectories instead of complete trajectories between reactants and products (Figure 1). The simulation of short trajectories is trivial to implement in parallel, making the formulation efficient to use on modern computing resources. Moreover, the use of short trajectories makes it possible to enhance sampling of improbable but important events by initiating the short trajectories near bottlenecks of reactions. A challenge is how to start the short trajectories, and how to analyze the result to obtain correct long-time behavior.

While milestoning is an approximate procedure, it shares the same philosophy and core algorithm as the exact milestoning approach. In both algorithms the phase

*Received by the editors May 14, 2015; accepted for publication (in revised form) December 16, 2015; published electronically March 3, 2016. This research was supported in part by the NIH (grant GM59796) and by the Welch Foundation (grant F-1783).

http://www.siam.org/journals/mms/14-1/M102157.html

†Department of Mathematics, Colorado State University, Fort Collins, CO 80523 (aristoff@rams.colostate.edu). This author's work was partially supported by the National Science Foundation via award NSF-DMS-1522398.

‡Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712 (jmb@ices.utexas.edu).

§Institute for Computational Engineering and Sciences, and Department of Chemistry, University of Texas at Austin, Austin, TX 78712 (ron@ices.utexas.edu).

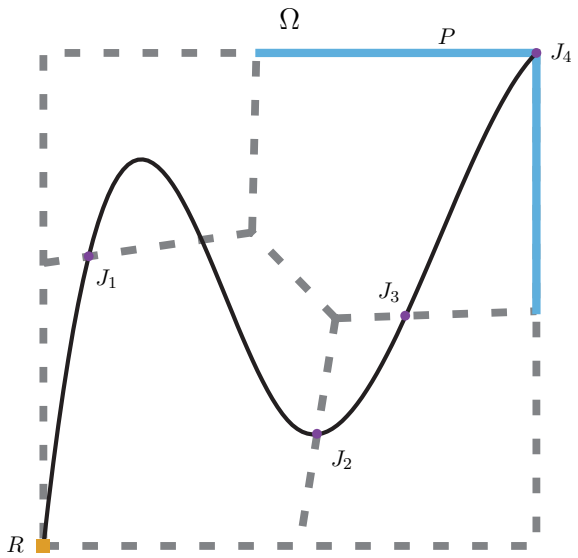

FIG. 1. *Representation of the state space* Ω *and the milestones. Each milestone is one of the line segments traced by dashed grey lines. The reactant state* R *is highlighted as a square dot in the bottom-left corner, while the product state* P *is comprised of the two line segments shown in blue in the upper-right corner. A particular realization of a long trajectory appears as a continuous black line, and the corresponding values of* (J_n) *are marked with dots.*

space Ω is partitioned by hypersurfaces, which we call milestones $M \subset \Omega$, into cells. The short trajectories are initiated on milestones and are terminated the first time they reach a neighboring milestone (Figure 1). The short trajectories can be simulated in parallel.

Milestoning uses an approximate distribution for the initial conditions of the trajectories at the hypersurfaces. The results are then analyzed within the milestoning theory. The approximation is typically the (normalized) canonical distribution restricted to the milestone interface M. In exact milestoning the distribution of hitting points at the interface is estimated numerically by iteratively computing trajectory fragments between milestones. In a straightforward implementation of the iterations (see also [5]) the final phase points of trajectories that were terminated on one milestone are continued until they hit another milestone. This type of trajectory continuation procedure is also used in nonequilibrium umbrella sampling (NEUS) [55] and trajectory tilting [53]. The continuation does not mean that full trajectories from reactants to products are computed. The calculations stop when the stationary distribution at the interface, or observables of interest, converge. In practice, and depending of course on the initial guess, the calculation ends significantly earlier than does computation of complete trajectories from R to P. The fast convergence of the iterations leads to significant computational savings.

A number of other algorithms build on the use of short trajectories to estimate long-time kinetics by "patching" these short trajectories at milestones or interfaces. These technologies include the weighted ensemble (WE) [57, 32], transition interface sampling (TIS) [52], partial path transition interface sampling (PPTIS) [43], forward flux sampling (FFS) [2], NEUS [55], trajectory tilting [53], and boxed molecular dy-

namics (BMD) [25]. Some of these techniques are similar; however, many subtle differences remain. Some of the differences are as follows. WE is the only method that makes it necessary to use stochastic dynamics. The trajectory sampling in NEUS, trajectory tilting, and exact milestoning is similar, even though the theories are quite different. Exact milestoning allows for the calculations of all the moments of the first passage time [5], a result which is not available for other technologies. BMD, milestoning, and PPTIS are approximate methods leading to greater efficiency. TIS, PPTIS, FFS, and BMD are focused on one-dimensional reaction coordinates. Other technologies (e.g., WE, milestoning, NEUS, and trajectory tilting) focus on a space of one or several coarse variables.

Hence, the overall scopes of these techniques differ significantly, which makes direct comparison between them less obvious. We have compared in the past the accuracy and efficiency of the methods of milestoning and exact milestoning with that of forward flux [5, 10]. Forward flux is one of the closest algorithms (in one dimension) to milestoning and exact milestoning. Numerous examples of kinetics of molecular systems studied with milestoning have been published [9, 11, 36, 34, 38, 21, 20, 56, 19]. We have also discussed extensively the features of alternate technologies that exploit trajectory fragments [56, 40].

The milestoning theory has not yet been subject to rigorous mathematical analysis, which is the goal of the present manuscript. In this manuscript we show that the exact milestoning method can be derived and analyzed in the framework of probability theory. The result is a useful link between physical intuition and a more formal approach. Readers that are interested in the efficiency of the algorithm on concrete examples, and comparison to other technologies, are referred to the sources mentioned in the previous paragraph.

This article is organized as follows. In section 2, we describe the setting for exact milestoning and introduce notation used throughout. In section 3, we show existence of and convergence to a stationary flux under very general conditions. In section 4 we precisely state the exact milestoning algorithm [5]. In section 5, we establish conditions under which convergence to the stationary flux is consistent in the presence of numerical error (Lemma 5.2 and Theorem 5.3), and we give a natural upper bound for the numerical error arising in exact milestoning (Theorem 5.4). Finally, in section 6 we consider some instructive examples.

2. Setup and notation.

2.1. The dynamics and MFPT. In milestoning we spatially coarse-grain a dynamics (X_t). The basic idea is to stop and start trajectories on certain interfaces, called milestones, and then reconstruct functions of (X_t) using these short trajectories, which can be efficiently simulated in parallel. We assume here that the dynamics is stochastic and focus on using milestoning for the efficient computation of *mean first passage times* (MFPTs) of (X_t), although similar ideas can be used to compute other nonequilibrium quantities.

To make our arguments we need some assumptions on (X_t). We let (X_t) be a time homogeneous strong Markov process with càdlàg paths taking values in a standard Borel space Ω. These assumptions allow us to stop and restart (X_t) on the milestones without knowing its history. In applications, (X_t) may be Langevin or overdamped Langevin dynamics, and Ω may be a subset of Euclidean space.

We write $\mathbb{P}$, $\mathbb{E}$ for all probability measures and expectations, with superscripts $\mathbb{P}^x$ (resp., $\mathbb{P}^\xi$) to indicate a starting point x (resp., distribution ξ). The symbol $\sim$ will indicate equality in probability law. We will use the words *distribution* and

probability measure interchangeably. Total variation norm will be denoted by $\|\cdot\|_{TV}$. Our analysis below will mostly take place in an idealized setting where we assume infinite sampling on the milestones. In this setting, distributions are smooth (if state space is continuous), and the total variation norm is the appropriate one.

Recall that we are interested in computing the MFPT of (X_t) from a reactant set R to a product set P. Throughout we consider fixed disjoint product and reactant sets $P, R \subset \Omega$. When R is not a single point, we will start (X_t) from a fixed probability measure ρ on R. If R is a single point, $\rho = \delta_R$, the delta distribution at R. As discussed above, milestoning allows for an efficient computation of the MFPT of (X_t) to P, starting at ρ. It is useful to think of P as a sink and R as a source for (X_t). More precisely, *we assume that when* (X_t) *reaches* P*, it immediately restarts on* R *according to* ρ. Obviously, this assumption has no effect on the MFPT to P. It will be useful, however, for computational and theoretical considerations.

Many of the results below follow from well-known theorems in probability theory. However, because of the special source-sink structure of (X_t), simpler proofs are often available, and we include them for clarity and completeness.

2.2. The milestones and semi-Markov viewpoint. We write $M \subset \Omega$ for the space of milestones used for parallelizing the computation of the MFPT. Each point $x \in M$ belongs to a *milestone* $M_x \subset M$. Thus, M is the union of all the milestones. We assume there are finitely many milestones, each of which is a closed set. Moreover, we demand that (X_t) pass through the intersection of two milestones with probability 0—thus, (X_t) can cross only one milestone at a time. This can be accomplished for Langevin or overdamped Langevin dynamics by taking the milestones to be codimension 1 with pairwise intersections of codimension 2 or larger; see Figure 1. The sets P and R will be two of the milestones. We always start (X_t) on M.

By following the sequence of milestones crossed by (X_t), we obtain a sequence of points (J_n) in M. See Figure 1. We now describe (J_n) more precisely. Let θ_n be the nth milestone crossing time for (X_t), defined recursively by $\theta_0 = 0$ and

$$\text{if } X_{\theta_n} = x, \text{ then } \theta_{n+1} := \inf\{t > \theta_n \,:\, X_t \in M_y \text{ for some } M_y \neq M_x\}.$$

Note that by a milestone crossing we mean a crossing of a milestone different from the previous one. The sequence of milestone crossing points is $J_n = X_{\theta_n}$.

We show now that (X_t) can be partially reconstructed from (J_n) and (θ_n). Let (Y_t) be defined[1] by setting $Y_t = J_n$ whenever $\theta_n \leq t < \theta_{n+1}$. Then (X_t) and (Y_t) agree at each milestone crossing time $t = \theta_n$, $n = 0, 1, 2, \ldots$, and (Y_t) is obtained from (X_t) by throwing away the path of (X_t) between milestone crossings, keeping only the endpoints. It follows that (X_t) and (Y_t) have the same MFPT to P. Thus, for our purposes it is enough to study (Y_t). We note that (Y_t), like (X_t), immediately restarts at ρ upon reaching P.

By our assumptions above, (J_n) is a Markov chain on M, and (Y_t) is a semi-Markov process on M, meaning it has the Markov property at crossing times. We write $K(x, dy)$ for the transition kernel of (J_n). Thus, if the initial distribution of (J_n) is $J_0 \sim \xi$, then the distribution at time n is $\mathbb{P}^\xi(J_n \in \cdot) = \xi K^n$. We also write $\xi K^n f := \mathbb{E}^\xi[f(J_n)]$ and $\xi f := \int_M f(x)\, \xi(dx)$ for suitable functions f.

[1]When (Y_t) has a probability density, it corresponds to the density $p(x, t)$ from [5] for the last milestone point passed.

The following notation will be needed. For $x \in M$, define local first passage times

$$\tau_M^x = \inf\{t > 0 : Y_t \in M_y \text{ for some } M_y \neq M_x\}.$$

Thus, τ_M^x is the first time for (Y_t) to cross some milestone other than M_x, starting at $Y_0 = x$. In particular, if $X_{\theta_{n-1}} = x$, then $\theta_n \sim \theta_{n-1} + \tau_M^x$. We also define τ_P to be the first time to cross P, and σ_P the number of crossings before reaching P:

$$\tau_P = \inf\{t > 0 : Y_t \in P\}, \qquad \sigma_P = \min\{n \geq 0 : J_n \in P\}.$$

We are interested in $\mathbb{E}^\rho[\tau_P]$, the MFPT from ρ to P.

3. Invariant measure and MFPT.

3.1. Stationary distribution on the milestones. The MFPT will be estimated via short trajectories between milestones. An important ingredient is the correct starting distribution for these trajectories. Exact milestoning makes use of a *stationary flux* of (X_t) on the milestones, which corresponds[2] to the stationary distribution μ of (J_n). It is worth noting that milestoning can also be made exact by choosing milestones as isocommittor surfaces [54]. The advantage of the formulation here is that the milestones can be arbitrary.

Some assumption is required to guarantee the existence of a stationary flux. We adopt the following sufficient condition, which we assume holds throughout:

$$\mathbb{E}^\xi[\tau_P] \text{ and } \mathbb{E}^\xi[\sigma_P] \text{ are finite for all probability measures } \xi \text{ on } M.$$

This ensures that (Y_t) reaches P in finite expected time and does not have an infinite number of milestone crossings in finite time. The condition can be readily verified in the standard settings for milestoning discussed above. Using this assumption and the source-sink structure of the dynamics—namely, that (Y_t) immediately restarts at ρ upon reaching P—we show in Theorem 3.1 below that μ exists.

THEOREM 3.1. *(J_n) has an invariant probability measure μ defined by*

$$\mu(\cdot) := \mathbb{E}^\rho\left[\sum_{n=0}^{\sigma_P} \mathbb{1}_{\{J_n \in \cdot\}}\right] \mathbb{E}^\rho[\sigma_P + 1]^{-1},$$

where $\mathbb{1}_{\{J_n \in C\}} = 1$ if $J_n \in C$ and otherwise $\mathbb{1}_{\{J_n \in C\}} = 0$.

Proof. Define $\nu(\cdot) = \mathbb{E}^\rho\left[\sum_{n=0}^{\sigma_P} \mathbb{1}_{\{J_n \in \cdot\}}\right]$, and observe that

$$\nu(\cdot) = \sum_{n=0}^{\infty}\sum_{m=0}^{n} \mathbb{P}^\rho(J_m \in \cdot \,|\, \sigma_P = n)\mathbb{P}^\rho(\sigma_P = n) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_n \in \cdot,\, \sigma_P \geq n).$$

If $C \cap R = \emptyset$, by bounded convergence,

$$\int_M \nu(dx)K(x,C) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_{n+1} \in C,\, \sigma_P \geq n) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_n \in C,\, \sigma_P \geq n) = \nu(C),$$

[2] Our μ is the same as the appropriately normalized stationary flux q in other milestoning papers. We use μ instead of q to emphasize that here it is a probability measure, not a density.

where the second equality uses $\mathbb{P}^\rho(J_0 \in C) = 0$ and $J_{n+1} \notin R \Rightarrow \sigma_P \neq n$. If $C \subset R$,

$$\int_M \nu(dx)K(x,C) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_{n+1} \in C,\, \sigma_P = n) + \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_{n+1} \in C,\, \sigma_P \geq n+1)$$
$$= \rho(C) - \mathbb{P}^\rho(J_0 \in C, \sigma_P \geq 0) + \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_n \in C,\, \sigma_P \geq n) = \nu(C). \quad \square$$

We will show below that (J_n) converges to μ under appropriate conditions. In that case μ is unique, and we will call μ the *stationary distribution* of (J_n). A successful application of exact milestoning will require some technique for sampling μ. The algorithm we present (Algorithm 1 below) is based on convergence of the distribution of (J_n) to μ. We demonstrate two types of convergence in Theorems 3.4 and 3.5 below.

It is worth noting that the proof of Theorem 3.1 leads to a representation of μ as a Neumann series, given in Corollary 3.2 below. The Neumann series is written in terms of the *transient kernel*

$$\bar{K}(x,dy) = \begin{cases} K(x,dy), & x \notin P, \\ 0, & x \in P. \end{cases} \tag{3.1}$$

$\bar{K}(x,dy)$ corresponds to a version of (J_n) that is absorbed (killed) on P.

COROLLARY 3.2. *We have*

$$\lim_{n\to\infty} \left\| \nu(M)^{-1} \sum_{i=0}^{n-1} \rho \bar{K}^i - \mu \right\|_{TV} = 0. \tag{3.2}$$

Proof. Recall that $\mu = \nu/\nu(M)$, where

$$\nu(\cdot) = \sum_{n=0}^{\infty} \mathbb{P}^\rho(J_n \in \cdot,\, \sigma_P \geq n) = \sum_{n=0}^{\infty} \rho \bar{K}^n.$$

Moreover,

$$\sup_{|f|\leq 1} \left| \nu(M)^{-1} \sum_{i=0}^{n-1} \rho \bar{K}^i f - \mu f \right| \leq \nu(M)^{-1} \sum_{i=n}^{\infty} \mathbb{P}^\rho(\sigma_P \geq i), \tag{3.3}$$

and the right-hand side of (3.3) is summable since by assumption $\mathbb{E}^\rho[\sigma_P] < \infty$. $\square$

3.2. Milestoning equation for the MFPT. Equipped with an invariant measure μ, we are now able to state the milestoning equation (3.4) for the MFPT. In exact milestoning, this equation is used to efficiently compute the MFPT. The algorithm is based on two principles: first, many trajectories can be simulated in parallel to estimate τ_M^x for various x; and second, the stationary distribution μ can be efficiently estimated through a technique based on power iteration. See the right-hand side of (3.4) below.

The gain in efficiency comes from the fact that the trajectories used to estimate τ_M^x are much shorter than trajectories from R to P. Whether we can efficiently sample μ may depend somewhat on whether we have a good initial guess. When (X_t) is Langevin dynamics, we have found that in some cases the canonical Gibbs distribution is a sufficiently good guess. See [5] and [6] for details and discussion.

THEOREM 3.3. *Let μ be defined as above. Then $\mu(P) > 0$ and*

$$\mu(P)\mathbb{E}^\rho[\tau_P] = \int_M \mu(dx)\mathbb{E}^x[\tau_M^x] := \mathbb{E}^\mu[\tau_M]. \tag{3.4}$$

Proof. The assumption $\mathbb{E}^\rho[\sigma_P] < \infty$ shows that $\mu(P) > 0$. For any $x \in M$,

$$\begin{aligned}\mathbb{E}^x[\tau_P] &= \int_M \mathbb{E}^x\left[\tau_P \,|\, Y_{\tau_M^x} = y\right] K(x, dy)\\ &= \int_M \mathbb{E}^x\left[\tau_M^x \,|\, Y_{\tau_M^x} = y\right] K(x, dy) + \int_{M\setminus P} \mathbb{E}^x\left[\tau_P - \tau_M^x \,|\, Y_{\tau_M^x} = y\right] K(x, dy)\\ &= \mathbb{E}^x[\tau_M^x] + \int_{M\setminus P} \mathbb{E}^y[\tau_P] K(x, dy).\end{aligned}$$

Thus,

$$\mathbb{E}^\mu[\tau_P] = \mathbb{E}^\mu[\tau_M] + \int_{M\setminus P}\int_M \mu(dx)\mathbb{E}^y[\tau_P]K(x, dy) = \mathbb{E}^\mu[\tau_M] + \int_{M\setminus P} \mu(dy)\mathbb{E}^y[\tau_P],$$

and so

$$\mathbb{E}^\mu[\tau_M] = \int_P \mu(dy)\mathbb{E}^y[\tau_P] = \mu(P)\mathbb{E}^\rho[\tau_P]. \qquad \square$$

In section 4 below we present the exact milestoning algorithm (Algorithm 1) recently used in [5] and [6]. The algorithm uses a technique which combines coarse-graining and power iteration to sample μ. Consistency of power iteration algorithms is justified via Theorem 3.4 below. Though we emphasize that there is a range of possibilities for sampling μ (for example, algorithms based on (3.2) or (3.7) below), we note that Algorithm 1 was shown to be efficient for computing the MFPT in the entropic barrier example of [5] and the random energy landscapes example of [6].

3.3. Convergence to stationarity. In this section we justify the consistency of power iteration–based methods for sampling μ by showing that ξK^n converges to μ in the total variation norm as $n \to \infty$. The theorem requires an extra assumption—aperiodicity of (J_n).

THEOREM 3.4. *Suppose that (J_n) is aperiodic in the following sense:*

$$g.c.d.\,\{n \geq 1 \,:\, \mathbb{P}^\rho(\sigma_P = n - 1) > 0\} = 1, \tag{3.5}$$

where g.c.d. is the greatest common divisor. Then for all probability measures ξ on M,

$$\lim_{n\to\infty} \|\mathbb{P}^\xi(J_n \in \cdot) - \mu\|_{TV} \equiv \lim_{n\to\infty} \|\xi K^n - \mu\|_{TV} = 0. \tag{3.6}$$

In particular, μ is unique.

Proof. We use a simple coupling argument. Let (H_n) be an independent copy of (J_n), and let $J_0 \sim \xi$ and $H_0 \sim \mu$. For $n \geq 0$, let S_n (resp., T_n) be the times at which (J_n) (resp., (H_n)) hit P for the $(n+1)$st time. Then $S_{n+1} - S_n$, $n \geq 0$, are i.i.d. random variables with finite expected value and nonlattice distribution, and $(S_{n+1} - S_n)_{n\geq 0} \sim (T_{n+1} - T_n)_{n\geq 0}$. It follows that $(S_n - T_n)_{n\geq 0}$ is a mean zero random walk with nonlattice step distribution. Thus, its first time to hit 0 is finite almost surely. So

$$\zeta := \inf\{n \geq 0 \,:\, J_n \in P,\, H_n \in P\}$$

obeys $\mathbb{P}(\zeta \geq n) \to 0$ as $n \to \infty$. Note that $J_n \sim H_n$ whenever $\zeta < n$. Thus

$$|\mathbb{P}^\xi(J_n \in C) - \mathbb{P}^\mu(H_n \in C)| \leq 2\mathbb{P}(\zeta \geq n).$$

Since μ is stationary for (H_n) we have $\mathbb{P}^\mu(H_n \in C) = \mu(C)$. Now

$$\|\mathbb{P}^\xi(J_n \in \cdot) - \mu\|_{TV} = \sup_{C \subset M} |\mathbb{P}^\xi(J_n \in C) - \mu(C)| \leq 2\mathbb{P}(\zeta \geq n),$$

which establishes the convergence result. To see uniqueness, suppose ξ is another invariant probability measure for (J_n); then the last display becomes $\|\xi - \mu\|_{TV} \leq 2\mathbb{P}(\zeta \geq n)$. Letting $n \to \infty$ shows that $\xi \sim \mu$. ∎

We now consider a class of problems where there is a smooth one-dimensional *reaction coordinate* $\psi : \Omega \to [0,1]$ tracking progress of (X_t) from R to P. In this case $\psi|_R \equiv 0$, $\psi|_P \equiv 1$, the milestones $M_1, \ldots, M_m$ are disjoint level sets of ψ, and $R = M_1$, $P = M_m$. Moreover (J_n) can only hop between neighboring milestones, unless $J_n \in P$. That is, if $J_n \in M_i$ for $i \notin \{1, m\}$, then $J_{n+1} \in M_{i-1}$ or $J_n \in M_{i+1}$; if $J_n \in M_1$, then $J_{n+1} \in M_2$; and if $J_n \in M_m$, then $J_{n+1} \in M_1$. Suppose that if $J_n \in M_i$ for $i \notin \{1, m\}$, then $J_{n+1} \in M_{i-1}$ with probability in $(0,1)$. Then the aperiodicity assumption (3.5) is satisfied if and only if m is odd. This is due to the fact that, if $J_0 \in M_1$, then $J_{m-1} \in M_m$ and $J_{m+1} \in M_m$ with positive probability, and m and $m+2$ are coprime when m is odd. On the other hand, if m is even, then the conclusion of Theorem 3.4 cannot hold. To see this, let m be even and suppose that J_0 is supported in an odd-indexed milestone. Then J_{2n} is always supported on an odd-indexed milestone, while J_{2n+1} is always supported on an even-indexed milestone.

Theorem 3.4 requires aperiodicity of (J_n). Even when (J_n) is not aperiodic, it converges in a time-averaged sense, as we show in Theorem 3.5 below. This means that problems in sampling μ arising from aperiodicity can be managed by averaging over time.

THEOREM 3.5. *Let $J_0 \sim \xi$, with ξ a probability measure on M. For bounded measurable $f : M \to \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(J_i) \overset{\text{a.s.}}{=} \int_M f \, d\mu \equiv \mu f. \tag{3.7}$$

Proof. Let S_n be the time at which (J_n) hits P for the $(n+1)$st time, and define

$$f_n = \sum_{i=S_n+1}^{S_{n+1}} f(J_i).$$

Note that f_n, $n \geq 0$, are i.i.d. Let $k(n) = \max\{k : S_k \leq n\}$, and write

$$\frac{1}{n} \sum_{i=0}^{n-1} f(J_i) = \frac{1}{n} \sum_{i=0}^{S_0} f(J_i) + \frac{1}{n} \sum_{i=0}^{k(n)-1} f_i + \frac{1}{n} \sum_{i=S_{k(n)}+1}^{n} f(J_i).$$

Since (J_n) hits P in finite time a.s., $n - S_{k(n)}$ and S_0 are finite a.s. Thus,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(J_i) \overset{\text{a.s.}}{=} \lim_{n\to\infty} \frac{k(n)}{n} \frac{1}{k(n)} \sum_{i=0}^{k(n)-1} f_i.$$

Notice that $R_n := S_{n+1} - S_n$, $n \geq 0$, are i.i.d. with finite expectation and

$$\frac{R_0 + \cdots + R_{k(n)-1}}{k(n)} \leq \frac{n - S_0}{k(n)} \leq \frac{R_0 + \cdots + R_{k(n)}}{k(n)}.$$

By the previous two displays and the law of large numbers,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(J_i) \overset{\text{a.s.}}{=} \frac{\mathbb{E}[f_0]}{\mathbb{E}[R_0]} = \mathbb{E}^\rho[\sigma_P + 1]^{-1} \sum_{i=0}^{\sigma_P} \mathbb{E}^\rho[f(J_i)] = \mu f. \qquad \square$$

It is worth noting that an additional aperiodicity condition leads to a limit for the distribution of Y_t. More precisely, suppose that (3.5) holds and, for each $x \in M \setminus P$ and $y \in M$, $\mathbb{P}^x(\tau_1 \in \cdot \,|\, J_1 = y)$ is nonlattice. Then for any $C \subset M$ and μ-a.e. x,

$$\lim_{t\to\infty} \mathbb{P}^x(Y_t \in C) = \frac{\int_C \mu(dy)\mathbb{E}^y[\tau_M^y]}{\int_M \mu(dy)\mathbb{E}^y[\tau_M^y]}. \tag{3.8}$$

See [4] for details[3] and a proof.

4. Exact milestoning algorithm. We now describe in detail an algorithm for sampling μ and the MFPT $\mathbb{E}^\rho[\tau_P]$, used successfully in [5] and [6]. We assume throughout this section that the conclusion of Theorem 3.4 holds. Let ξ be an initial guess for μ. (If (X_t) is Brownian or Langevin dynamics, we take ξ to be the canonical Gibbs distribution.) We write M_i, $i = 1, \ldots, m$, for the distinct milestones, so that $M = \bigcup_{i=1}^m M_i$. The algorithm will produce approximations

$$\xi \equiv \mu^{(0)}, \mu^{(1)}, \mu^{(2)}, \ldots$$

of μ. Let $\mu_i^{(n)}$ be the *nonnormalized* restriction of $\mu^{(n)}$ to M_i, and define

$$\mathbb{E}^{\mu_i^{(n)}}[\tau_M] := \mu^{(n)}(M_i)^{-1} \int_{M_i} \mu_i^{(n)}(dx)\mathbb{E}^x[\tau_M^x].$$

For $C \subset M_j$ we will also use the notation

$$a_{ij}^{(n)}(C) = \mu^{(n-1)}(M_i)^{-1} \int_{M_i} \mu_i^{(n-1)}(dx)\, K(x, C).$$

Below we think of $a_{ij}^{(n)}$ and $\mu_i^{(n)}$ as either distributions or densities. The $a_{ij}^{(n)}$ are obtained from trajectory fragments between milestone crossings. A simple Monte Carlo scheme for estimating these distributions is as follows. Let $x_1, \ldots, x_L$ be i.i.d. samples from the distribution $\mu_i^{(n-1)}/\mu^{(n-1)}(M_i)$. Starting at each $x_\ell \in M_i$, simulate (X_t) until it crosses the next milestone, say at the point $y_\ell \in M_j$. If we idealize by assuming that the simulation of (X_t) is done exactly, then by Chebyshev's inequality,

$$\mathbb{P}\left(\left| a_{ij}^{(n)}(C) - \int_C \frac{1}{L} \sum_{\ell=1}^{L} \delta_{y_\ell}(dy) \right| > \epsilon \right) \leq \frac{a_{ij}^{(n)}(C) - a_{ij}^{(n)}(C)^2}{L\epsilon^2},$$

[3] When the right-hand side of (3.8) has a density, it is the same as the stationary probability density $p(x)$ in [5] for the last milestone point passed.

where δ_y is the Dirac delta distribution at y. We therefore write, for $y \in M$,

$$a_{ij}^{(n)}(y) \approx \frac{1}{L} \sum_{\ell=1}^{L} \tilde{\delta}_{y_\ell}(y), \tag{4.1}$$

where $\tilde{\delta}_{y_\ell}$ is either some suitable approximation to the identity at y_ℓ or simply a delta function at y_ℓ. Thus, in Algorithm 1 we think of $a_{ij}^{(n)}$ and $\mu_i^{(n-1)}$ as either densities (in the former case) or distributions (in the latter). The local mean first passage times (i.e., the times between successive milestone crossings) are approximated by the sample means

$$\mathbb{E}^{\mu_i^{(n-1)}}[\tau_M] \approx \frac{1}{L} \sum_{\ell=1}^{L} \tau_M^{x_\ell}.$$

It is important to realize that we do not need to store the full coordinates of each y_ℓ in memory. Instead, it suffices to use a data-structure that keeps track of the pairs (y_ℓ, M_j). The actual coordinates of each point can be written to disk and read from it as needed.

Algorithm 1. Exact milestoning algorithm.

Input: Milestones $M = \bigcup_{j=1}^{m} M_j$, initial guess ξ, and tolerance $\varepsilon > 0$ for the absolute error in the MFPT.

Output: Estimates for μ, local MFPTs $\mathbb{E}^{\mu}[\tau_M]$, and overall MFPT $\mathbb{E}^{\rho}[\tau_P]$.

```
μ^(0) ← ξ.
T^(0) ← +∞.
for all n = 1, 2, ... do
  for i = 1 to m do
    Estimate a_ij^(n) and E^{μ_i^(n-1)}[τ_M].
    A_ij^(n) ← a_ij^(n)(M_j).
  end for
  Solve w^T A = w^T (with A = (A_ij^(n)) ∈ R_{≥0}^{m×m} and w = (w_1, ..., w_m) ∈ R_{≥0}^m).
  for j = 1 to m do
    μ_j^(n) ← Σ_{i=1}^m w_i a_ij^(n).
  end for
  Normalize μ^(n).
  T^(n) ← μ(P)^{-1} E^{μ^(n-1)}[τ_M].
  if |T^(n) − T^(n-1)| < ε then
    break
  end if
end for
return (μ^(n), E^{μ^(n-1)}[τ_M], T^(n)).
```

The eigenvalue problem in Algorithm 1 involves a stochastic matrix $\mathtt{A} \in \mathbb{R}_{\geq 0}^{m \times m}$ that is sparse. Indeed, the ith row corresponds to milestone M_i and may have only as many nonzero entries as the number of neighboring milestones M_j. In practice, to solve the eigenvalue problem we can use efficient and accurate Krylov subspace solvers [28] such as Arnoldi iteration [39] to obtain $\mathbf{w}$ without computing all the other eigenvectors.

In Algorithm 1, if $\mathbf{w}_i := \mu_i^{(n-1)}(M_i)$ is used instead of the solution $\mathbf{w}$ to $\mathbf{w}^{\mathtt{T}}\mathtt{A} = \mathbf{w}^{\mathtt{T}}$, then the algorithm approximates μ by simple power iteration, $\mu^{(n)} = \xi K^n$. The reason

for defining the weights as the solution to $\mathbf{w}^{\mathrm{T}}\mathtt{A} = \mathbf{w}^{\mathrm{T}}$ is practical: we have found that it gives faster convergence of the iterations, at no apparent cost to accuracy. It can be seen as a version of power iteration that uses coarse-graining. See [5, 6] for applications of the algorithm in exact milestoning, and [28, 39] for related discussions.

Finally, we mention the fact that pseudorandom number generators (PRNGs) can only produce a finite quantity of pseudorandom numbers. Once the maximum is reached, the generators may silently reuse the previous random numbers in the same order. It has been noted [13] that this phenomenon leads to unphysical artifacts in simulations. The simplest approach to properly using PRNGs (and avoiding the aforementioned artifacts altogether) consists of reseeding the generator from time to time, obtaining the new seeds from high-quality entropy sources such as those available in modern computer hardware (see [17, 33] for more details).

5. Error analysis.

5.1. Stationary distribution error. In practice, due to time discretization error, we cannot generate trajectories exactly according to the transition kernel K. Instead, we can generate trajectories according to a numerical approximation K_ϵ. We investigate here whether such schemes are consistent, that is, whether powers of K_ϵ of K converge to a distribution $\mu_\epsilon \approx \mu$. We emphasize that, even though we account for time discretization here, we still assume infinite sampling, and thus for a given $x \in M$, $K_\epsilon(x, dy)$ may be a continuous distribution. See section 5.2 below for related remarks and a discussion of how time discretization errors affect the exact milestoning estimate of the MFPT.

The following theorem, restated from [23], establishes consistency of power iteration–based schemes when K_ϵ is sufficiently close to K and (J_n) is geometrically ergodic. We give natural conditions for geometric ergodicity in Lemma 5.2 and Theorem 5.3 below.

THEOREM 5.1. *Suppose that(J_n) is geometrically ergodic: there exists $\kappa \in (0, 1)$ such that*

$$\sup_{x \in M} ||\delta_x K^n - \mu||_{TV} = O(\kappa^n).$$

Let $\{K_\epsilon\}$ be a family of stochastic kernels with $K_0 = K$, assumed to act continuously on the Banach space of bounded measurable functions on M with the sup norm, such that

$$\lim_{\epsilon \to 0} \sup_{|f| \le 1} \|K_\epsilon f - Kf\|_\infty = 0. \tag{5.1}$$

Then for each $\hat{\kappa} \in (\kappa, 1)$ there is a $\delta > 0$ such that, for each $\epsilon \in [0, \delta)$, K_ϵ has a unique invariant probability measure μ_ϵ, and

$$\sup_{\epsilon < \delta} \sup_{x \in M} \|\delta_x K_\epsilon^n - \mu_\epsilon\|_{TV} = O(\hat{\kappa}^n),$$
$$\lim_{\epsilon \to 0} \|\mu_\epsilon - \mu\|_{TV} = 0.$$

We now give two sufficient conditions for the geometric ergodicity of (J_n). The first condition is a uniform lower bound on the probability of reaching P in N steps (Lemma 5.2). The second is a strong Feller condition (Theorem 5.3). The latter is a very natural condition and is easy to verify in some cases, for instance when (X_t) is a nondegenerate diffusion and the milestones are sufficiently regular.

LEMMA 5.2. *Suppose that there exist* $\lambda \in (0,1)$ *and* $N \in \mathbb{N}$ *such that, for all* $x \in M$, $\mathbb{P}^x(J_{N-1} \in P) \geq \lambda > 0$. *Then* (J_n) *is geometrically ergodic:*

$$\sup_{x \in M} ||\delta_x K^n - \mu||_{TV} \leq \lambda^{-1}(1-\lambda)^{\lfloor n/N \rfloor}.$$

Proof. Let $\xi_1, \xi_2 \in \mathcal{P}$, consider the signed measure $\xi = \xi_1 - \xi_2$, and compute

$$\begin{aligned}
\|\xi_1 K^N - \xi_2 K^N\|_{TV} &= \sup_{|f| \leq 1} \left| \int_M \int_M \xi(dy) K^N(y, dz) f(z) \right| \\
&= \sup_{|f| \leq 1} \left| \int_M \int_M \xi(dy) \left(K^N(y, dz) - \lambda \rho(dz) \right) f(z) \right| \\
&= \sup_{|f| \leq 1} \left| \int_M \xi(dy) \int_M \left(K^N(y, dz) - \lambda \rho(dz) \right) f(z) \right| \\
&\leq (1-\lambda) \sup_{|f| \leq 1} \left| \int_M \xi(dy) f(y) \right| = (1-\lambda) \|\xi_1 - \xi_2\|_{TV}.
\end{aligned}$$

The last line uses the fact that $K^N(y, dz) - \lambda\rho(dz)$ is a positive measure. This shows that K^N is a contraction mapping on $\mathcal{P}$ with contraction constant $(1-\lambda)$. Observe also that $\|\xi_1 K - \xi_2 K\|_{TV} \leq \|\xi_1 - \xi_2\|_{TV}$. The result now follows from the contraction mapping theorem. See, for instance, Theorem 6.40 of [18]. ∎

Note that the λ in Lemma 5.2 is a quantity that can be estimated, at least in principle, by running trajectories of (X_t) starting at x which cross $N-1$ milestones before reaching P. However, this is likely impractical for the same reason direct estimation of the MFPT is impractical—the trajectories would be too long. One alternative would be to compute the probability $P^i(J^{\mathtt{A}}_{N-1} \in P)$ for the Markov chain $(J^{\mathtt{A}}_n)$ on $\{1, \ldots, m\}$ with transition matrix $\mathtt{A}$, and use the minimum over $i \in \{1, \ldots, m\}$ as a proxy for λ. Even without a practical way to estimate λ, we believe the characterization of Lemma 5.2 is useful for understanding the convergence rate.

Lemma 5.2 leads to the following condition for geometric ergodicity of (J_n).

THEOREM 5.3. *Suppose that* M *is compact and* (J_n) *is a strong Feller chain which is aperiodic in the sense of* (3.5). *Then* (J_n) *is geometrically ergodic.*

Proof. Let $\epsilon \in (0, \mu(P))$. By Theorem 3.4, for each $x \in M$ there is an $N_x \in \mathbb{N}$ such that $\mathbb{P}^x(J_n \in P) \geq \epsilon$ for all $n \geq N_x$. Because (J_n) is strong Feller, the map $x \to \mathbb{P}^x(J_n \in P)$ is continuous. By the compactness of M, it follows that for any $\lambda \in (0, \epsilon)$ there is an $N \in \mathbb{N}$ such that $\mathbb{P}^y(J_n \in P) \geq \lambda$ for all $y \in M$ and $n \geq N-1$. Lemma 5.2 now yields the result. ∎

5.2. MFPT error. As discussed above, (3.4) can be used to estimate the MFPT $\mathbb{E}^\rho[\tau_P]$ by sampling μ and local MFPTs τ^x_M. The error in this estimate has two sources. First, in general we only have an approximation $\tilde{\mu}$ of μ. The second source of error is in the sampling of τ^x_M, due to the fact that we can only simulate a time discrete version $(\tilde{X}_{n\delta t})$ of (X_t). In Theorem 5.4 below we give an explicit formula for the numerical error of the MFPT in terms of these two sources. We first need the following notation. Let $\tilde{\tau}^x_M$ be the minimum of all $n\,\delta t > 0$ such that the line segment between $\tilde{X}_{n\delta t}$ and $\tilde{X}_{(n+1)\delta t}$ intersects M_y for some $M_y \neq M_x$, and define

$$\mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M] := \int_M \tilde{\mu}(dx) \mathbb{E}^x[\tilde{\tau}^x_M].$$

Theorem 5.4 below gives an expression for the error in the *original* milestoning as well as in exact milestoning.

THEOREM 5.4. *There exists a nonnegative function ϕ such that*

$$\begin{aligned}\left|\mathbb{E}^{\rho}[\tau_P] - \tilde{\mu}(P)^{-1}\mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]\right| \le c_1 \left|\mu(P)^{-1} - \tilde{\mu}(P)^{-1}\right| \\ + \tilde{\mu}(P)^{-1}\left(c_2\|\mu - \tilde{\mu}\|_{TV} + \phi(\delta t)\right),\end{aligned} \tag{5.2}$$

where

$$c_1 := \mathbb{E}^{\mu}[\tau_M], \qquad c_2 := \sup_{x\in M} \mathbb{E}^{x}[\tau_M^x].$$

Proof. Note that

$$\begin{aligned}|\mathbb{E}^{\rho}[\tau_P] - \tilde{\mu}(P)^{-1}\mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]| &= |\mu(P)^{-1}\mathbb{E}^{\mu}[\tau_M] - \tilde{\mu}(P)^{-1}\mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]| \\ &\le |\mu(P)^{-1}\mathbb{E}^{\mu}[\tau_M] - \tilde{\mu}(P)^{-1}\mathbb{E}^{\mu}[\tau_M]| \\ &\quad + |\tilde{\mu}(P)^{-1}\mathbb{E}^{\mu}[\tau_M] - \tilde{\mu}(P)^{-1}\mathbb{E}^{\tilde{\mu}}[\tau_M]| \\ &\quad + |\tilde{\mu}(P)^{-1}\mathbb{E}^{\tilde{\mu}}[\tau_M] - \tilde{\mu}(P)^{-1}\mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]|,\end{aligned}$$

where we have written $\mathbb{E}^{\tilde{\mu}}[\tau_M] := \int_M \tilde{\mu}(dx)\mathbb{E}^x[\tau_M^x]$. We may write

$$\phi(\delta t) = |\mathbb{E}^{\tilde{\mu}}[\tau_M] - \mathbb{E}^{\tilde{\mu}}[\tilde{\tau}_M]|$$

for the term depending only on time stepping error. Note that

$$\begin{aligned}|\mathbb{E}^{\mu}[\tau_M] - \mathbb{E}^{\tilde{\mu}}[\tau_M]| &= \left|\int_M \mu(dx)\mathbb{E}^x[\tau_M^x] - \int_M \tilde{\mu}(dx)\mathbb{E}^x[\tau_M^x]\right| \\ &\le \left(\sup_{x\in M}\mathbb{E}^x[\tau_M^x]\right)\|\mu - \tilde{\mu}\|_{TV}.\end{aligned}$$

Combining the last three expressions yields the result. ∎

Recall that in the above we have ignored errors from finite sampling. We now discuss the implications of those errors. In the original milestoning, $\tilde{\mu}$ is the canonical Gibbs distribution on the milestones. In that setting, we can typically sample independently from $\tilde{\mu}$ on the milestones. Thus, the central limit theorem implies that the *true* error in the milestoning approximation of $\mathbb{E}^{\rho}[\tau_P]$ is bounded above with high probability by the right-hand side of (5.2) plus a constant times $1/\sqrt{N}$, where N is the number of samples. An analogous argument applies to exact milestoning if $\tilde{\mu}$ is sampled by simple power iteration. For our coarse-grained version of power iteration in Algorithm 1, however, we obtain samples of $\tilde{\mu}$ which are not independent, and thus a more detailed analysis would be required to determine the additional error from finite sampling.

We do not analyze the time discretization error $\phi(\delta t)$ and instead refer the reader to [27] and references therein. Here we simply remark that, if (X_t) is a diffusion process, then under certain smoothness assumptions on the drift and diffusion coefficients of (X_t) and on M, we have $\phi(\delta t) = \theta(\sqrt{\delta t})$ when $(X_{n\delta t})$ is the standard Euler time discretization with time step δt. See [26] for details and proof. See also [8, 31] for numerical schemes that mitigate time discretization error in the MFPTs.

6. Illustrative examples. In this section we discuss two examples of milestoning to illustrate the method.

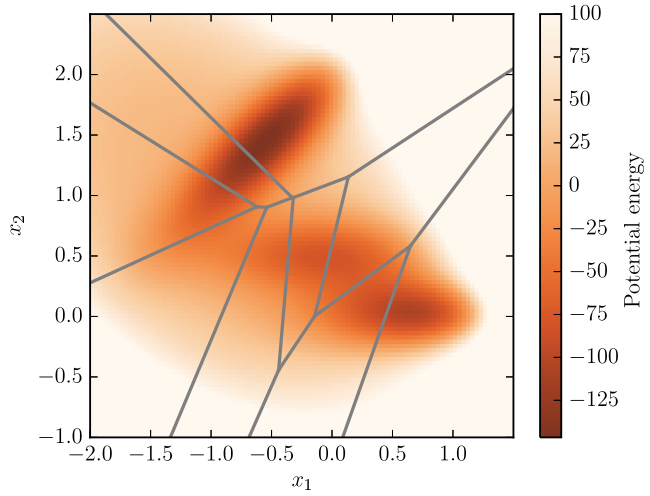

FIG. 2. *Graph of the Müller–Brown potential energy function. The milestones are shown as the overlaid line segments.*

We consider the solution, (X_t), of the Brownian dynamics equation,

$$\begin{cases} dX_t = -\nabla U(X_t)\,dt + \sqrt{2\beta^{-1}}\,dB_t, \\ X_0 \sim \rho, \end{cases} \tag{6.1}$$

where $U\colon \Omega \to \mathbb{R}$ is a smooth potential energy function, $\beta > 0$ is the inverse temperature, and (B_t) is a standard Brownian motion.

6.1. Müller–Brown potential. We begin with a system characterized by the Müller–Brown potential [44]. The energy function $U\colon \Omega \subset \mathbb{R}^2 \to \mathbb{R}$ is given by the formula (see also the corresponding energy landscape in Figure 2)

$$\begin{aligned} U(x_1, x_2) = &-200\,\mathrm{e}^{-(x_1-1)^2-10\,{x_2}^2} - 100\,\mathrm{e}^{-{x_1}^2-10\left(x_2-\frac{1}{2}\right)^2} \\ &- 170\,\mathrm{e}^{-\frac{13}{2}\left(x_1+\frac{1}{2}\right)^2+11\left(x_1+\frac{1}{2}\right)\left(x_2-\frac{3}{2}\right)-\frac{13}{2}\left(x_2-\frac{3}{2}\right)^2} \\ &+ 15\,\mathrm{e}^{\frac{7}{10}(x_1+1)^2+\frac{3}{5}(x_1+1)(x_2-1)+\frac{7}{10}(x_2-1)^2}. \end{aligned}$$

This system is a commonly used benchmark for numerical methods for obtaining reaction rates.

We chose to partition Ω using a Voronoi tessellation (displayed in Figures 2 and 3) generated from a set of points gathered by the method of locally updated planes [51]. However, any other set of points could have been chosen (as we shall discuss in the next example). Figure 2 also shows our choice of reactant and product milestones.

For the numerical experiments to be detailed below, we solve the stochastic differential equation in (6.1) using the Euler–Maruyama scheme [42] with a time step length $\Delta t = 10^{-5}$ at a temperature determined by $\beta^{-1} = 5$. We use the number of force evaluations as a measure of the computational cost of our methods, and we note that the Euler–Maruyama method requires one force evaluation per time step.

We compared two types of experiments that we now describe. The first experiment consists of running Brownian dynamics trajectories started at the reactant

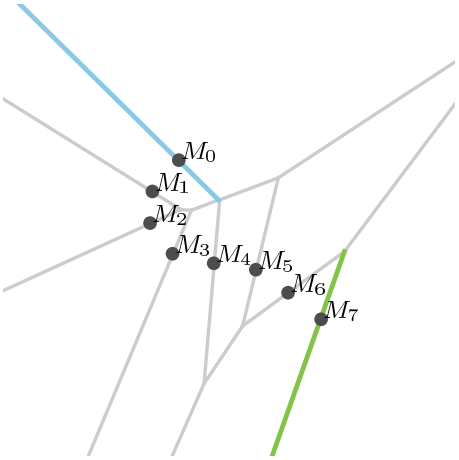

FIG. 3. *Milestones represented as line segments. Some of the milestones are labeled, and the reactant is colored in green (lower right) while the product is shown in blue (upper left).*

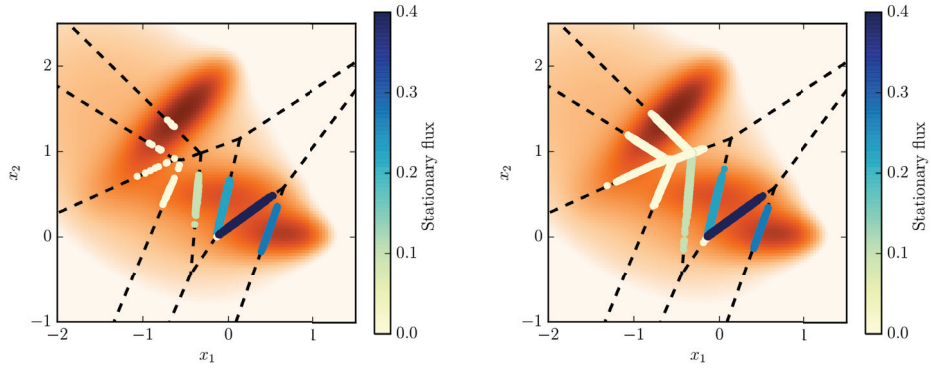

FIG. 4. *Phase space points in the empirical distributions of the stationary flux μ for two types of simulations: long trajectories using straightforward Brownian dynamics (left) and exact milestoning (right). Despite the fact that the two types of simulations involved comparable amounts of computational effort, we see that the sampling in exact milestoning is much more exhaustive.*

milestone until they reach the product milestone. As soon as a trajectory reaches the product, we initiate a new trajectory from the reactant milestone and so on. We refer to these as *long* trajectories. In the second experiment we run exact milestoning, starting the first iteration with exactly one phase space point at each of the milestones along the reaction path. Next, we run ten *short* trajectories per milestone per iteration. These short trajectories start at each milestone and stop whenever they reach any neighboring milestone, as described in section 4.

Despite allowing the long trajectories to go on for approximately 2.5×10^9 force evaluations, only seven reach the product milestone. This leads to a poor approximation of the MFPT. By contrast, running the exact milestoning method for approximately 2×10^9 force evaluations, we obtain good estimates of the stationary distribution μ and the local MFPTs.

The values of $\mu(M_i)$ are displayed in Figures 4 and 5. The empirical distributions

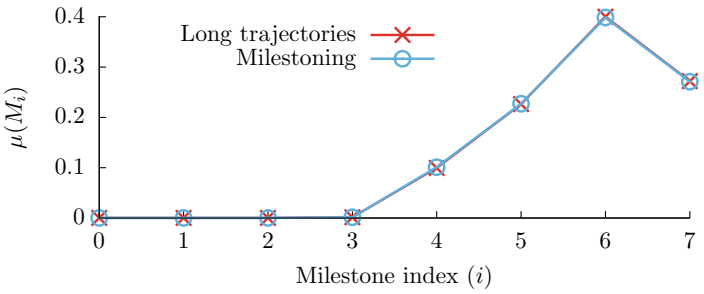

FIG. 5. *Values of $\mu(M_i)$ at some of the milestones in the Müller–Brown potential. The values correspond to the long trajectories (in red) and to exact milestoning (in blue), as discussed in section* 6.1. *Not shown are the milestones other than M_i for $i = 0, \ldots, 7$, where the sampling obtained from the long trajectories is insufficient for comparison.*

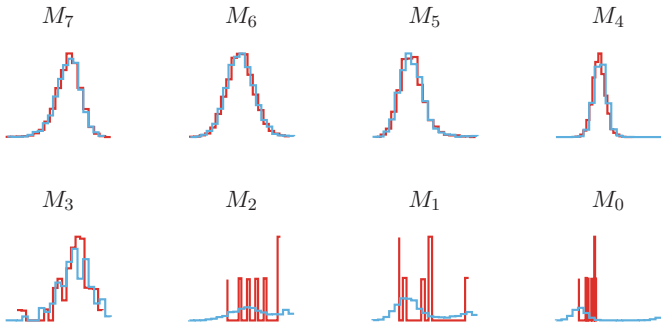

FIG. 6. *Empirical distributions of the stationary flux obtained by long trajectories (in red) and exact milestoning (in blue) corresponding to the system in section* 6.1. *Notice that the sampling of the long trajectories is very sparse at the milestones close to the product.*

corresponding to μ on some of the milestones are shown in Figure 6.

Figures 4 and 5 illustrate the nonequilibrium nature of exact milestoning. The stationary distribution that we compute differs noticeably from the equilibrium (canonical) distribution. Recall from Figures 2 and 3 that the reactant milestone, M_7, is located at the lower-right minimum, while the product milestone, M_0, is at the global minimum in the upper-left side of the graph. With this in mind, we see that trajectories initiated at M_7 arrive at the intermediate minimum located close to the center of the graph, and many of those trajectories return to the lower minimum, crossing M_7 again. This results in high values of $\mu(M_6)$ at the transition state, while the density at M_0 (the global minimum) is significantly lower. Equilibrium considerations, which are inappropriate here, would suggest that most of the stationary density (and the stationary probability) is concentrated at the global minimum and that the weight at the transition state would be small.

6.2. Rough energy landscape. In this case, we present an example of milestoning on the torus $\Omega = \mathbb{R}^2/\mathbb{Z}^2$. For our computations, we consider a uniformly spaced mesh of milestones with fixed product and reactant sets P, R; see Figure 7.

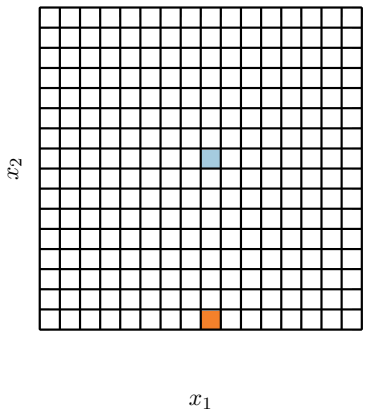

FIG. 7. *Diagram showing the reactant state (lower shaded square) and the product state (upper shaded square) within the set of all milestones for the example in section* 6.2. *Each milestone is an edge of one of the small squares in the diagram. (The total number of milestones has been decreased to enhance visibility.)*

We model a rough energy landscape by a potential energy function of the form

$$U(x_1, x_2) = \mathrm{Re} \sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} z_{k_1,k_2}\, \mathrm{e}^{2\pi \mathrm{i}(k_1 x_1 + k_2 x_2)}, \tag{6.2}$$

where Re denotes the real part of a complex number and $N \in \mathbb{N}$ is a constant that tunes the ruggedness of the potential. Each coefficient $z_{k_1,k_2} = a_{k_1,k_2} + \mathrm{i}\, b_{k_1,k_2} \in \mathbb{C}$ is determined by the random variables a_{k_1,k_2} and b_{k_1,k_2}, which are distributed according to

$$\begin{cases} c, & \text{with probability } \frac{1}{2}, \\ 0, & \text{with probability } \frac{1}{2}, \end{cases}$$

with c itself being a uniform random variable in the interval $(-1, 1)$. Since N is fixed, a particular realization of the coefficients specified above completely determines the potential energy function U. The graph of the canonical density of a potential energy of the form discussed above is shown in Figure 8. Notice that this class of energy functions generalizes the model for rough landscapes introduced by [58] and that similar potential energy functions have been used to model Wigner glasses [1].

We carry out exact milestoning in this example by solving boundary value problems, as described in [6]. The resulting stationary density obtained after convergence is shown in Figure 9.

It is interesting to note that it has been argued [54] that an optimal choice of milestones would consist of using the level sets (also called *isocommittors*) of the committor function. These surfaces (see Figure 10) are typically hard to compute in practice, which makes the use of exact milestoning more appealing, as its results are independent of how the milestones are set up.

7. Conclusions. The main goal of this manuscript is to present a rigorous mathematical derivation, based on probability theory, of exact milestoning. While the theory of exact milestoning and accompanying numerical examples were discussed

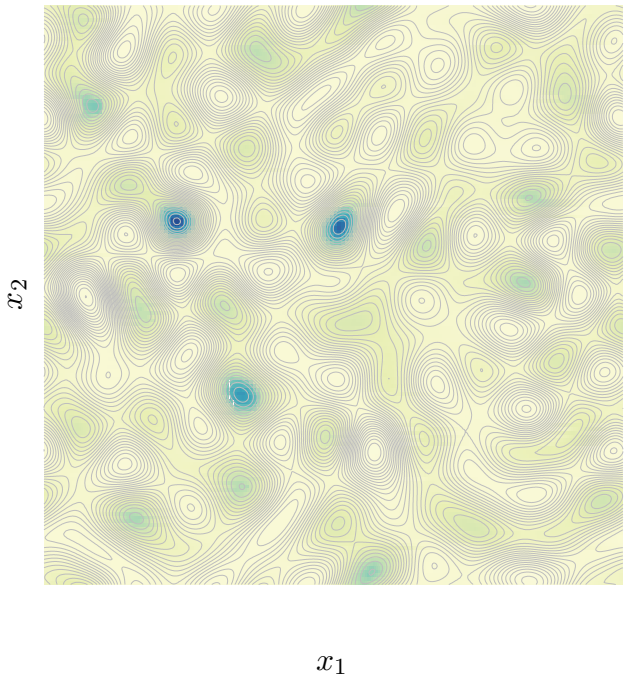

FIG. 8. *Graph of the density of the canonical distribution corresponding to a rough energy landscape with* $N = 7$ *at temperature* $\beta^{-1} = 1$.

elsewhere [5, 6], the mathematical formulation in the earlier paper was not as rigorous as in this manuscript. Once this formulation is established, it opens the way for further communication between chemical physicists and mathematicians, and it bridges the gap between the communities for further development of an important tool for computer simulation.

Exact milestoning belongs to a class of enhanced sampling methods for the calculation of kinetics. Approaches most closely related to milestoning are NEUS [55] and trajectory tilting [53]. The way in which trajectories are sampled is similar in all of these methods; however, the theoretical frameworks are different. For example, milestoning allows the calculation of all the moments of the first passage time (FPT) distribution [5], and hence better estimates of the FPT can be constructed, a result that was not reported for other methods.

It is not necessary in milestoning to establish or rely on metastability to estimate the average transition time. From this perspective the method is different from another exact approach—transition path sampling [16]— that exploits the short duration of rare trajectories between metastable states. Exact milestoning makes the sampled trajectories short by sampling trajectory fragments between boundaries of phase space cells or milestones. The statistics of short trajectories between milestones make it possible to investigate wide ranges of types of energy landscapes, which may be corrugated or not, as illustrated in the two examples in this paper. We have shown here that exact milestoning is both accurate and highly efficient.

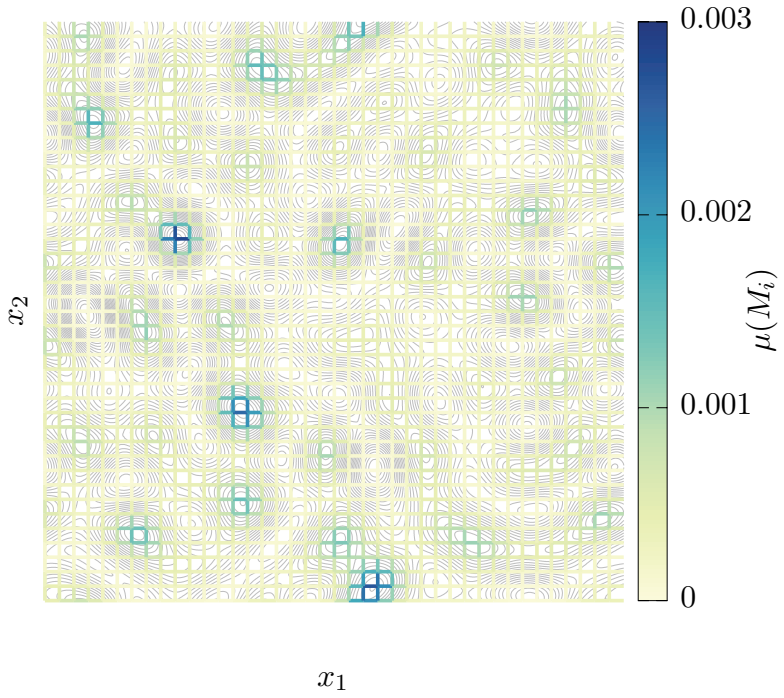

FIG. 9. *Stationary density μ on the rough energy landscape of section 6.2. The contour lines are the level sets of U. There are $2 \times 40 \times 40$ total milestones (shown as the segments in the overlaid grid).*

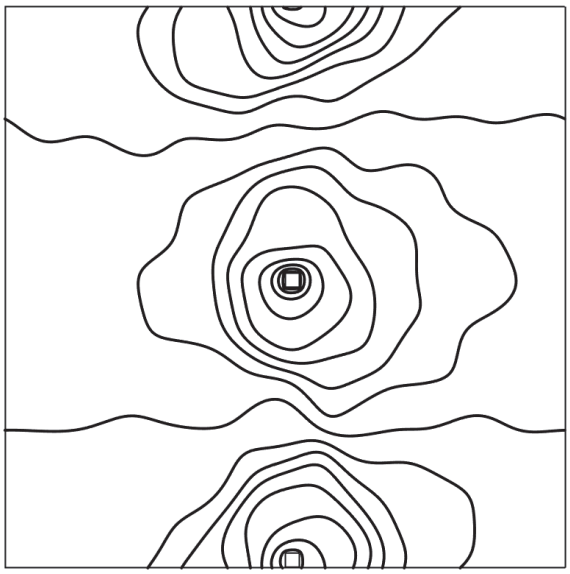

FIG. 10. *Isocommittor surfaces for the rough energy landscape discussed in section 6.2.*

It is important to emphasize that the choice of the milestones in exact milestoning is arbitrary from a formal viewpoint. Efficiency considerations suggest that it is beneficial to select them following two criteria: (i) the milestones should be sufficiently close in the kinetic sense to make the trajectories short, and (ii) milestones should be chosen to make the number of iterations as small as possible. For example, the number of iterations can be small if the system is close to equilibrium and the milestones are expressed in a space of slow variables. Then an initial choice of the canonical distribution is quite accurate.

We also note that the method of milestoning that was broadly used in the past (e.g., [35]) is approximate and assumes local equilibrium within the milestones. While corrections and further refinements were proposed [40, 30, 29], these approximations cannot be made exact and are similar in spirit to the local equilibrium and lag time approximations of Markov state models [14]. Nevertheless, these approximations can be accurate with a proper choice of coarse variables. These types of approximations are very useful as the system grows in complexity and size and exact calculations become prohibitively expensive. Milestoning made it possible to investigate kinetics of enzymes [37] and transport through membranes [12] in agreement with experimental observations. These are systems of tens to hundreds of thousands of particles, and with time scales of milliseconds. It will be of considerable interest to re-evaluate these approximations for large systems with the method of exact milestoning. As the efficiency of exact milestoning increases with faster hardware, we are breaking scale barriers that were not previously accessible to atomically detailed simulations.

REFERENCES

[1] S. Akhanjee and J. Rudnick, *Disorder induced transition into a one-dimensional Wigner glass*, Phys. Rev. Lett., 99 (2007), 236403 (see also arXiv:0706.4462).

[2] R. J. Allen, D. Frenkel, and P. R. ten Wolde, *Forward flux sampling-type schemes for simulating rare events: Efficiency analysis*, J. Chem. Phys., 124 (2006), 463102.

[3] P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford Science Publications, Clarendon Press, Oxford, UK, 1989.

[4] G. Alsmeyer, *On the Markov renewal theorem*, Stoch. Process. Appl., 50 (1994), pp. 37–56.

[5] J. M. Bello-Rivas and R. Elber, *Exact milestoning*, J. Chem. Phys., 142 (2015), 094102.

[6] J. M. Bello-Rivas and R. Elber, *Simulations of thermodynamics and kinetics on rough energy landscapes with milestoning*, J. Comput. Chem., (2015), to appear.

[7] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Transition path sampling: Throwing ropes over rough mountain passes, in the dark*, Annu. Rev. Phys. Chem., 53 (2002), pp. 291–318.

[8] B. Bouchard, S. Geiss, and E. Gobet, *First Time to Exit of a Continuous Itô Process: General Moment Estimates and L_1-convergence Rate for Discrete Time Approximations*, preprint, arXiv:1307.4247, 2013.

[9] A. E. Cardenas, G. S. Jas, K. Y. DeLeon, W. A. Hegefeld, K. Kuczera, and R. Elber, *Unassisted transport of n-Acetyl-l-tryptophanamide through membrane: Experiment and simulation of kinetics*, J. Phys. Chem. B, 116 (2012), pp. 2739–2750.

[10] A. E. Cardenas and R. Elber, *Enhancing the capacity of molecular dynamics simulations with trajectory fragments*, in Innovation in Biomolecular Modeling and Simulation, T. Schlick, ed., Royal Society of Chemistry, London, 2012, Vol. 1, pp. 117–137.

[11] A. E. Cardenas and R. Elber, *Computational study of peptide permeation through membrane: Searching for hidden slow variables*, Molec. Phys., 111 (2013), pp. 3565–3578.

[12] A. E. Cardenas, R. Shrestha, L. J. Webb, and R. Elber, *Membrane permeation of a peptide: It is better to be positive*, J. Phys. Chem. B, 119 (2015), pp. 6412–6420.

[13] D. S. Cerutti, R. Duke, P. L. Freddolino, H. Fan, and T. P. Lybrand, *A vulnerability in popular molecular dynamics packages concerning Langevin and Andersen dynamics*, J. Chem. Theory Comput., 4 (2008), pp. 1669–1680.

[14] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, *Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics*, J. Chem. Phys., 126 (2007), 155101.

[15] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, *Transition path sampling and the calculation of rate constants*, J. Chem. Phys., 108 (1998), pp. 1964–1977.
[16] C. Dellago, P. G. Bolhuis, and P. L. Geissler, *Transition Path Sampling*, Adv. Chem. Phys. 123, John Wiley & Sons, New York, 2002.
[17] Y. Dodis, A. Shamir, N. Stephens-Davidowitz, and D. Wichs, *How to eat your entropy and have it too—Optimal recovery strategies for compromised RNGs*, in Advances in Cryptology CRYPTO 2014 SE-3, Lecture Notes in Comput. Sci. 8617, J. Garay and R. Gennaro, eds., Springer, Berlin, Heidelberg, 2014, pp. 37–54.
[18] R. Douc, E. Moulines, and D. Stoffer, *Nonlinear Time Series: Theory, Methods and Applications with R Examples*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, Abingdon, UK, 2014.
[19] R. Elber, *A milestoning study of the kinetics of an allosteric transition: Atomically detailed simulations of deoxy Scapharca hemoglobin*, Biophys. J., 92 (2007), pp. L85–L87.
[20] R. Elber, K. Kuczera, and G. S. Jas, *The kinetics of helix unfolding: Molecular dynamics simulations with milestoning*, J. Phys. Chem. A, 113 (2009), pp. 7461–7473.
[21] R. Elber and A. West, *Atomically detailed simulation of the recovery stroke in myosin by milestoning*, Proc. Natl. Acad. Sci. USA, 107 (2010), pp. 5001–5005.
[22] A. K. Faradjian and R. Elber, *Computing time scales from reaction coordinates by milestoning*, J. Chem. Phys., 120 (2004), pp. 10880–10889.
[23] D. Ferré, L. Hervé, and J. Ledoux, *Regular perturbation of V-geometrically ergodic Markov chains*, J. Appl. Probab., 50 (2013), pp. 184–194.
[24] D. Frenkel and B. Smit, eds., *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed., Academic Press, San Diego, 2002.
[25] D. R. Glowacki, E. Paci, and D. V. Shalashilin, *Boxed molecular dynamics: Decorrelation time scales and the kinetic master equation*, J. Chem. Theory Comput., 7 (2011), pp. 1244–1252.
[26] E. Gobet and S. Menozzi, *Exact approximation rate of killed hypoelliptic diffusions using the discrete Euler scheme*, Stoch. Process. Appl., 112 (2004), pp. 201–223.
[27] E. Gobet and S. Menozzi, *Stopped diffusion processes: Boundary corrections and overshoot*, Stoch. Process. Appl., 120 (2010), pp. 130–162 (see also arXiv:0706.4042).
[28] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed., Johns Hopkins Stud. Math. Sci., Johns Hopkins University Press, Baltimore, MD, 2013.
[29] A. T. Hawk, *Milestoning with coarse memory*, J. Chem. Phys., 138 (2013), 154105.
[30] A. T. Hawk and D. E. Makarov, *Milestoning with transition memory*, J. Chem. Phys., 135 (2011), 224109.
[31] D. J. Higham, X. Mao, M. Roj, Q. Song, and G. Yin, *Mean exit times and the multilevel Monte Carlo method*, SIAM/ASA J. Uncertain. Quantif., 1 (2013), pp. 2–18.
[32] G. A. Huber and S. Kim, *Weighted-ensemble Brownian dynamics simulations for protein association reactions*, Biophys. J., 70 (1996), pp. 97–110.
[33] Intel Corporation, *Intel® Digital Random Number Generator (DRNG) Software Implementation Guide*, 2014; available online from https://software.intel.com/en-us/articles/intel-digital-random-number-generator-drng-software-implementation-guide.
[34] G. S. Jas, W. A. Hegefeld, P. Májek, K. Kuczera, and R. Elber, *Experiments and comprehensive simulations of the formation of a helical turn*, J. Phys. Chem. B, 116 (2012), pp. 6598–6610.
[35] S. Kirmizialtin and R. Elber, *Revisiting and computing reaction coordinates with directional milestoning*, J. Phys. Chem. A, 115 (2011), pp. 6137–6148.
[36] S. Kirmizialtin, V. Nguyen, K. A. Johnson, and R. Elber, *How conformational dynamics of DNA polymerase select correct substrates: Experiments and simulations*, Structure, 20 (2012), pp. 618–627.
[37] S. Kirmizialtin, K. A. Johnson, and R. Elber, *Enzyme selectivity of HIV reverse transcriptase: Conformations, ligands, and free energy partition*, J. Phys. Chem. B, 119 (2015), pp. 11513–11526.
[38] S. M. Kreuzer, R. Elber, and T. J. Moon, *Early events in helix unfolding under external forces: A milestoning analysis*, J. Phys. Chem. B, 116 (2012), pp. 8662–8691.
[39] R. B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide: Large-scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, Software Environ. Tools 6, SIAM, Philadelphia, 1998.
[40] P. Májek and R. Elber, *Milestoning without a reaction coordinate*, J. Chem. Theory Comput., 6 (2010), pp. 1805–1817.
[41] P. Metzner, C. Schütte, and E. Vanden-Eijnden, *Transition path theory for Markov jump processes*, Multiscale Model. Simul., 7 (2009), pp. 1192–1219; http://doi.org/10.1137/070699500.

[42] G. N. Milstein and M. V. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Sci. Comput., Springer, Berlin, Heidelberg, 2004.
[43] D. Moroni, P. G. Bolhuis, and T. S. Van Erp, *Rate constants for diffusive processes by partial path sampling*, J. Chem. Phys., 120 (2004), pp. 4055–4065 (see also arXiv:0310466).
[44] K. Müller and L. D. Brown, *Location of saddle points and minimum energy paths by a constrained simplex optimization procedure*, Theoret. Chim. Acta, 53 (1979), pp. 75–93.
[45] A. P. Ruymgaart, A. E. Cardenas, and R. Elber, *MOIL-opt: Energy-conserving molecular dynamics on a GPU/CPU system*, J. Chem. Theory Comput., 7 (2011), pp. 3072–3082.
[46] M. Sarich, F. Noé, and C. Schütte, *On the approximation quality of Markov state models*, Multiscale Model. Simul., 8 (2010), pp. 1154–1177; http://doi.org/10.1137/090764049.
[47] T. Schlick, *Molecular Modeling and Simulation: An Interdisciplinary Guide*, Springer, New York, 2010; http://doi.org/10.1007/978-1-4419-6351-2.
[48] D. E. Shaw, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Lerardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. M. Deneroff, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S. C. Wang, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, and K. J. Bowers, *Anton, A special-purpose machine for molecular dynamics simulation*, Comm. ACM, (2008); http://doi.org/10.1145/1364782.1364802.
[49] J. E. Stone, J. C. Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco, and K. Schulten, *Accelerating molecular modeling applications with graphics processors*, J. Comput. Chem., 28 (2007), pp. 2618–2640.
[50] D. W. H. Swenson and P. G. Bolhuis, *A replica exchange transition interface sampling method with multiple interface sets for investigating networks of rare events*, J. Chem. Phys., 141 (2014), 044101.
[51] A. Ulitsky and R. Elber, *A new technique to calculate steepest descent paths in flexible polyatomic systems*, J. Chem. Phys., 92 (1990), 1510.
[52] T. S. van Erp, D. Moroni, and P. G. Bolhuis, *A novel path sampling method for the calculation of rate constants*, J. Chem. Phys., 118 (2003), pp. 7762–7774.
[53] E. Vanden-Eijnden and M. Venturoli, *Exact rate calculations by trajectory parallelization and tilting*, J. Chem. Phys., 131 (2009), pp. 1–7 (see also arXiv:0904.3763).
[54] E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, *On the assumptions underlying milestoning*, J. Chem. Phys., 129 (2008), 174102.
[55] A. Warmflash, P. Bhimalapuram, and A. R. Dinner, *Umbrella sampling for nonequilibrium processes*, J. Chem. Phys., 127 (2007), 154112.
[56] A. M. A. West, R. Elber, and D. Shalloway, *Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide*, J. Chem. Phys., 126 (2007), 145104.
[57] B. W. Zhang, D. Jasnow, and D. M. Zuckerman, *The "weighted ensemble" path sampling method is statistically exact for a broad class of stochastic processes and binning procedures*, J. Chem. Phys., 132 (2010), 054107.
[58] R. Zwanzig, *Diffusion in a rough potential*, Proc. Natl. Acad. Sci. USA, 85 (1988), pp. 2029–2030.