

The effect of time resolution on the observed first passage times in diffusive dynamics

Kevin Song¹, Dmitrii E. Makarov^{2,3,*} and Etienne Vouga¹

¹Department of Computer Science, University of Texas at Austin, Austin, Texas 78712, USA

² Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, USA

³ Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, Texas 78712, USA

Abstract.

Single-molecule and single-particle tracking experiments are typically unable to resolve fine details of thermal motion at short timescales where trajectories are continuous. We show that, when a diffusive trajectory $x(t)$ is sampled at finite time intervals δt , the resulting error in measuring the first passage time to a given domain can exceed the time resolution of the measurement by more than an order of magnitude. Such surprisingly large errors originate from the fact that the trajectory may enter and exit the domain while being unobserved, thereby lengthening the apparent first passage time by an amount that is larger than δt . Such systematic errors are particularly important in single-molecule studies of barrier crossing dynamics. We show that the correct first passage times, as well as other properties of the trajectories such as splitting probabilities, can be recovered via a stochastic algorithm that reintroduces unobserved first passage events probabilistically.

* makarov@cm.utexas.edu

Consider a particle undergoing stochastic dynamics, with its location $x(t)$, as a function of time t , sampled at finite time intervals δt . We are interested in measuring the average time before the particle, starting at $x(0) = x_0$, crosses some specified boundary b for the first time (i.e. the first passage time). At first glance, it seems the error in such a measurement should be comparable to δt , but in fact it can be much greater! The reason is illustrated in Figure 1, showing a trajectory $x(t)$ that crosses b while the sampled discrete version of the trajectory does not – as a result of such missed crossings the trajectory arrives at b at a later time, introducing an error much greater than δt .

Our interest in this problem is motivated by recent single-molecule/single-particle-tracking experiments, whose aim is to detect transient molecular phenomena such as conformational changes underlying the action of molecular machines¹ or molecular intermediates and barrier-crossing dynamics in protein folding and binding²⁻⁷. Of particular interest is the temporal span of short pieces of molecular trajectories spent in thermodynamically unfavorable regions of configuration space such as transition paths quickly traversing activation barriers⁸⁻¹⁰ or exit paths escaping the barrier region¹¹: the distributions of transition path times and other (conditional) first passage times can, for example, inform one about molecular intermediates encountered along the way^{4, 12, 13} and about the topology of the underlying energy landscape¹⁴⁻¹⁶. These times are often relatively short (e.g., microseconds or milliseconds), only 1-2 orders of magnitude longer than the experimental time resolution. For example, a typical transition path of a micron-sized bead hopping between two optical traps, which takes a few milliseconds¹⁷ to cross the activation barrier, will consist of only $\sim 10 - 10^2$ data points when observed with a fast camera with a typical sampling rate of $1/\delta t = 10^4$

frames per second. As will be shown below, the relative error in measuring transition path times caused by missed crossings can be quite large in such systems.

The purpose of this note is to explain and to quantify the errors in measuring conditional first passage times caused by a finite sampling rate, and to introduce a simple method that corrects for such errors by reintroducing first passage events probabilistically between times when the system's location is unobserved.

Figure 2 illustrates the effect of finite sampling rate on the conditional exit times from an interval $(0, L)$. The mean exit time^{11, 18, 19} $\langle \tau(x_0 \rightarrow L) \rangle_E$ is the average time it takes to escape the interval, starting from $x_0 \in (0, L)$, through its boundary at $x = L$ conditional upon reaching this boundary before crossing the other boundary at $x = 0$. Exit times have attracted recent attention as measures of barrier crossing dynamics¹¹ and as “reaction coordinates” that include temporal information¹⁹. They are related to the much studied transition path times^{5, 8, 20, 21}: the mean transition path time from 0 to L , $\langle t(0 \rightarrow L) \rangle_{TP}$, is the average time it takes to traverse the interval $(0, L)$ conditional upon entering the interval through the left boundary and exiting through the right boundary while staying continuously within the interval in between these two events. The exit times are more general than transition path times, and they can also be used to describe failed attempts to cross the interval¹¹; by definition, we have

$$\langle t(0 \rightarrow L) \rangle_{TP} = \lim_{x_0 \rightarrow 0} \langle \tau(x_0 \rightarrow L) \rangle_E. \quad (1)$$

Exit times shown in Fig. 2 were computed using the overdamped Langevin equation

$$\frac{k_B T}{D} \frac{dx}{dt} = -U'(x) + \zeta(t). \quad (2)$$

Here the potential $U(x) = Fx$ describes the motion of a particle in the presence of a constant force F acting from right to left, D is the particle's diffusivity, and $\zeta(t)$ is a Gaussian-distributed

random force, which has zero mean and which satisfies the fluctuation-dissipation theorem

$\langle \zeta(t)\zeta(t') \rangle = \frac{2(k_B T)^2}{D} \delta(t - t')$, where T is the temperature. Eq. 2 was integrated using the

Euler-Maruyama scheme with a timestep of $\delta t_0 = 5 \times 10^{-6} \frac{L^2}{D}$ for 1.5×10^{10} steps. To mimic single-particle tracking data, which yield subsampled versions of continuous-time trajectories, we then studied coarser versions of the same trajectory sampled at multiples of the integration time step (i.e. defined as $x(i\delta t)$, where i is an integer and $\delta t = \delta t_0, 10\delta t_0, 100\delta t_0, 1000\delta t_0, 2000\delta t_0$).

The results are shown in Fig. 2 for three values for the force: zero force, “moderate” force $F = k_B T/L$, and “high” force $F = 5k_B T/L$, where L is the length of the interval. As can be seen from Fig. 2, top, the exit times estimated from such subsampled trajectories are significantly longer than the corresponding theoretical values¹⁸; moreover, the error introduced by subsampling is much greater than the timestep δt .

Finite sampling time introduces significant errors in computing another important characteristic of barrier crossing dynamics, the splitting probability (Figure 3, top). Splitting probability $\phi(x_0 \rightarrow L) = 1 - \phi(x_0 \rightarrow 0)$, also known as committor or “pfold”²²⁻²⁷, is the probability that a trajectory starting from $x_0 \in (0, L)$, will reach the boundary L before reaching 0. Splitting probabilities can be measured experimentally^{28, 29} and are useful, for example, for reconstructing effective potentials governing the dynamics^{28, 30, 31} as well as coordinate-dependent diffusivities³². It is instructive to note that, while the magnitude of the force has only a modest effect on the relative error in measuring the exit times (Fig. 2), its effect on the splitting probability is more dramatic. In particular, for the largest value of the force F used here (note that it is directed from right to left), the splitting probabilities $\phi(x_0 \rightarrow L)$ measured

near the right interval boundary ($x_0 \approx L$) can be significantly lower than the theoretical values (Fig 3, top/right), because many trajectories that start near L and exit while unobserved during the time interval δt drift back left and into the interval before finally crossing at $x = 0$; these paths will be misclassified as exiting left when the time step δt is large.

These results show that a finite sampling time may introduce significant errors into typical observables characterizing diffusive dynamics within a finite interval (such as a transition region in protein folding). The same problem also arises in the analysis of molecular dynamics trajectories: because of data storage limitations molecular configurations are usually not saved at every simulation timestep, with a timestep between stored configurations typically being orders of magnitude longer than the simulation timestep.

While the errors introduced by finite time resolution could be surprisingly large, it is possible to correct for them. Specifically, suppose we suspect that a trajectory $x(t)$ has crossed a boundary located at $x = 0$ between two observations, where the particle was observed at $x(t_1) = x_1 > 0$ and at $x(t_2 = t_1 + \delta t) = x_2 > 0$, such that the discretized trajectory does not cross 0. The probability p_c that an unobserved crossing has taken place can be determined by dividing the conditional probability $\tilde{G}(x_2, t_2 | x_1, t_1)$ that a trajectory arrives at x_2 at time t_2 having crossed $x = 0$ at some time $t' \in (t_1, t_2)$ given that it originated at x_1 at time t_1 , by the conditional probability $G(x_2, t_2 | x_1, t_1)$ that the particle is found at $x(t_2) = x_2$ given that it has started at $x(t_1) = x_1$. For overdamped Langevin dynamics, Eq. 2, the latter probability is just the Green's function of the diffusion equation, $G(x_2, t_2 | x_1, t_1) = \frac{1}{\sqrt{4\pi D \delta t}} \exp\left(-\frac{(x_2 - x_1)^2}{4D \delta t}\right)$. We note that this result is valid for sufficiently small value of the time step δt even in the presence of a nonzero potential $U(x)$, as diffusion dominates over the deterministic force $-U'(x)$ at

short times²¹; we expect, however, that this approximation to become less accurate (given the same δt) in the presence of larger forces. The former conditional probability is also easily estimated by using the reflection principle¹⁸: each trajectory $x(t)$ that has crossed $x = 0$ at time t' can be replaced by a statistically equivalent trajectory $\tilde{x}(t)$ that is reflected with respect to the boundary $x = 0$ after the first crossing event: $\tilde{x}(t) = x(t), t < t', \tilde{x}(t) = -x(t), t \geq t'$.

Then we have $\tilde{G}(x_2, t_2 | x_1, t_1) = G(-x_2, t_2 | x_1, t_1) = \frac{1}{\sqrt{4\pi D \delta t}} \exp\left(-\frac{(x_2 + x_1)^2}{4D \delta t}\right)$, and the crossing probability is

$$p_c = \frac{\tilde{G}(x_2, t + \delta t | x_1, t)}{G(x_2, t + \delta t | x_1, t)} = \exp\left(-\frac{x_1 x_2}{D \delta t}\right) \quad (3)$$

As seen from this equation, the crossing probability is typically non-negligible when the trajectory points are located within a distance of $\sim \sqrt{D \delta t}$ from the boundary. Eq. 3 can be used to correct for sampling errors. For example, if we want to compute the first passage time to the boundary $x = 0$, we generate a successful first passage event according to Eq. 3 every time a trajectory is close enough to $x = 0$, even if the sampled trajectory has not crossed the boundary. Of course, the exact boundary crossing time is only known to within the sampling time δt , which determines the resulting error. More generally, in application to a trajectory sampled at time intervals δt , this method reintroduces the crossing of a boundary b (e.g. $b = 0$ or $b = L$ in exit time calculations) during time step i with a probability of

$$p_c = \exp - \frac{[x(i\delta t) - b][x((i+1)\delta t) - b]}{D \delta t} \text{ whenever } x(i\delta t) - b \text{ and } x((i+1)\delta t) - b \text{ have the same}$$

sign and thus the discretized trajectory has not crossed this boundary. Figures 2-3 (bottom) show that this correction procedure does indeed significantly reduce the errors in estimating both average exit times and splitting probabilities.

An obvious limitation of the method is the requirement that the diffusivity D is known, as it is needed to estimate p_c using Eq. 3. At first glance, this appears to limit the applicability of the method only to cases where the model of Eq. 2 is known to be true and where, moreover, the value of D is known. The value of the diffusion coefficient, however, can be estimated using a variety of methods; for example, one can use the first and second moments of the particle's displacement to estimate the diffusion coefficient^{30, 33, 34}:

$$\langle \Delta x^2 \rangle - \langle \Delta x \rangle^2 \approx 2D\delta t, \quad (4)$$

where $\Delta x = x(t + \delta t) - x(t)$. This idea can be extended to certain classes of systems that are not described by Eq. 2; for example, systems with position-dependent diffusion coefficients $D(x)$ can be treated in a similar way, and the function $D(x)$ can again be estimated using Eq. 4. In using Eq. 3, then, one should use the value of the diffusivity that corresponds to the boundary whose crossing is being considered (i.e. $D(0)$ or $D(L)$ in the above case).

For systems with memory, such as those described by generalized Langevin equations (see, e.g., refs.^{21, 35}), the diffusivity is effectively frequency-dependent³⁶⁻³⁸, with short-time/high frequency dynamics being described by the high-frequency diffusivity limit³⁹. Assuming the timestep δt is short enough that the high-frequency limit is reached, the coefficient D estimated from Eq. 4 should be treated as such high frequency limit, and Eq. 3 should still apply despite the fact that the dynamics, at longer timescales, is not truly diffusive. Indeed, the high-frequency limit of the diffusion coefficient can be estimated both from simulated and experimental trajectories^{17, 40-42}.

To test the applicability of the method to trajectories $x(t)$ whose dynamics is non-Markovian, we have considered a system of two particles, with their respective coordinates obeying the two-dimensional Langevin equation of the form

$$\begin{aligned}\frac{k_B T}{D_x} \frac{dx}{dt} &= -\partial U(x, y)/\partial x + \zeta_x(t) \\ \frac{k_B T}{D_y} \frac{dy}{dt} &= -\partial U(x, y)/\partial y + \zeta_y(t)\end{aligned}\quad (5)$$

with

$$U(x, y) = Fx + \frac{k(y-x)^2}{4} \quad (6)$$

and with $\zeta_{x,y}$ being the Gaussian noise satisfying the fluctuation-dissipation theorem

$$\langle \zeta_\alpha(t) \zeta_\beta(t') \rangle = \frac{2(k_B T)^2}{D_\alpha} \delta(t - t') \delta_{\alpha\beta}, \quad \alpha, \beta \in \{x, y\}. \quad (7)$$

The particle whose coordinate is x is subjected to a force F , as in the previous examples. The trajectory $x(t)$ is viewed as the “experimental” data, while the effect of the “unobserved” particle (with coordinate y) that is coupled harmonically to x is to introduce memory effects in the dynamics of x . Indeed, it is known that the resulting trajectory $x(t)$ in this case obeys a generalized Langevin equation with an exponential memory kernel^{21, 43}.

In our simulations, we use the parameters $D_y = D_x/5$ and $k = 1000 \frac{k_B T}{L^2}$, which result in highly non-Markovian dynamics of the observed trajectory. Indeed, the long-time dynamics is governed by a diffusion coefficient $D_{long} = (D_x^{-1} + D_y^{-1})^{-1} = \frac{D_x}{6}$, (which describes the collective motion of the two particles) while its short-time properties are governed by the high-frequency diffusion coefficient $D_{short} = D_x$, where the observed degree of freedom x is effectively decoupled from y (see, e.g., ref.³⁹). Using the same sampling timesteps δt as in the

previous example, the diffusion coefficient estimated from the coarse trajectories $x(i\delta t)$ using Eq. 4 is virtually identical to the correct short-time limit $D_{short} = D_x$.

Figure 4 shows analysis of exit times for the non-Markov trajectory $x(t)$ sampled at different time intervals. Although an exact analytical solution is not known for this system, we observe the same trends as in the case of Langevin dynamics: while the exit times measured directly from the sampled trajectories strongly depend on the sampling time interval δt , they become virtually independent of δt after the correction of Eq. 3. Similar behavior was observed for the splitting probabilities (data not shown.)

In summary, inference of underlying dynamics, and particularly of first passage times and transition path times, from imperfect observations has attracted considerable recent interest—the present work is complementary to studies that treat imperfect observations as “gating”⁴⁴ and to studies of the effect of smoothing that is sometimes used to reduce experimental noise⁴⁵. Here we have shown that finite (and often disappointingly large) sampling time intervals, typical of both experimental trajectories and of molecular simulations, introduce errors to various (conditional) first passage times that far exceed the time resolution. Such errors are caused by missed first passage events. Under certain assumptions about the underlying dynamics, it is possible to correct for such errors using a simple stochastic algorithm that reintroduces missed crossing events probabilistically. Here, we have explored the simplest possible version of such algorithm, in which the dynamics of the particle in the vicinity of the target is approximated as free diffusion, leading to Eq. 3. It is further possible, if needed, to improve the crossing probability estimate using a linear approximation, $U(x) \approx U(b) +$

$U'(b)(x - b)$ (where b is the location of the “target”) and using Green’s function for diffusion with drift instead of that of free diffusion.

Acknowledgments

This work was supported by the Robert A. Welch Foundation (Grant No. F- 1514 to DEM), the National Science Foundation (Grant Nos. CHE 1955552 to DEM and IIS 2212048 to EV), and by the Adobe Inc.

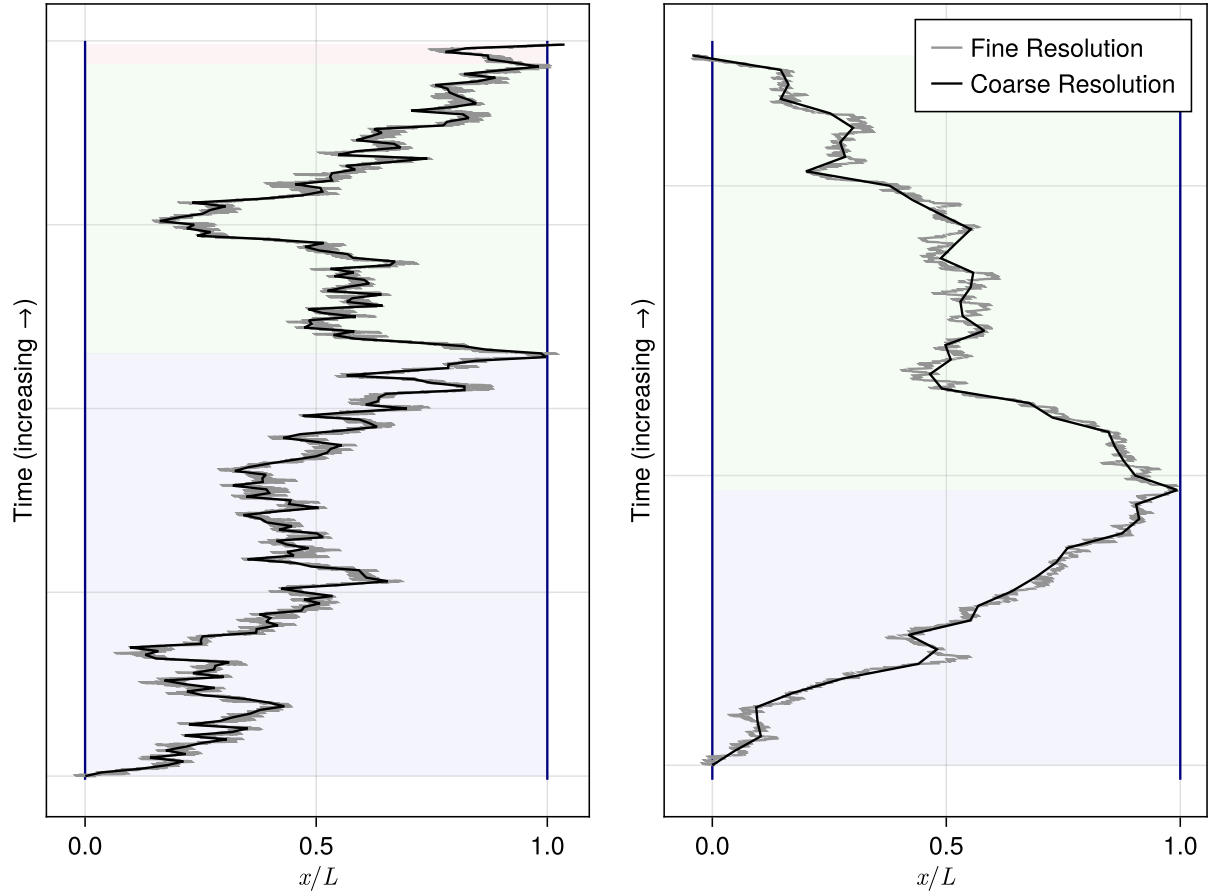


Figure 1: Examples of crossings missed when a trajectory $x(t)$ is sampled at finite time intervals. The particle's position as a function of time is shown by the gray line. A coarser-resolution trajectory, obtained by sampling the particle's location every 1000 timesteps, is shown in black. The background color changes whenever the particle leaves the interval $(0, L)$; the coarse-resolution trajectory, however, remains within the interval. *Left:* the particle exits the right side of the interval three times, but only the final exit event is detected for the trajectory that is sampled with a coarser resolution. This leads to an overestimate of how long it takes the particle to transition across the interval (i.e. of the transition path time). *Right:* the trajectory exits the interval $(0, L)$ to the right; thus two transition paths are observed, one from left to right and the other from right to left. The coarse-resolution trajectory, however, never sees the particle cross the right endpoint of the interval, and thus no transition paths are observed at all.

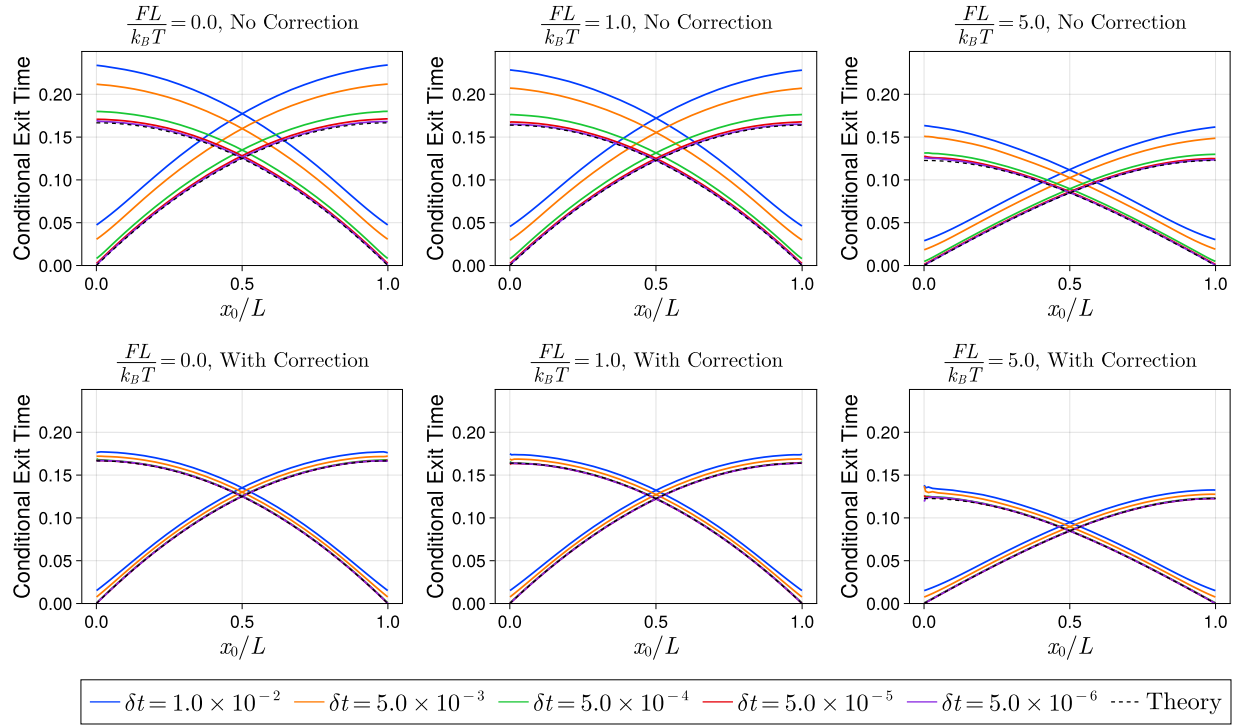


Figure 2: Conditional exit times to the right and left boundaries of the interval $(0, L)$ shown as a function of the starting position x_0 for unbiased diffusion (left), moderately biased diffusion (middle), and strongly biased diffusion (right), obtained from trajectories sampled at various timesteps δt . The upper row shows the values obtained directly from sampled trajectories, along with the theory. Due to the effects illustrated in Figure 1, the typical error in these curves is much larger than the sampling timestep, ranging from $3\delta t$ to $60\delta t$. The lower row shows the values obtained by applying our correction procedure based on Eq. 3 to the same trajectories. All timescales are reported in reduced units, where L^2/D sets the timescale.

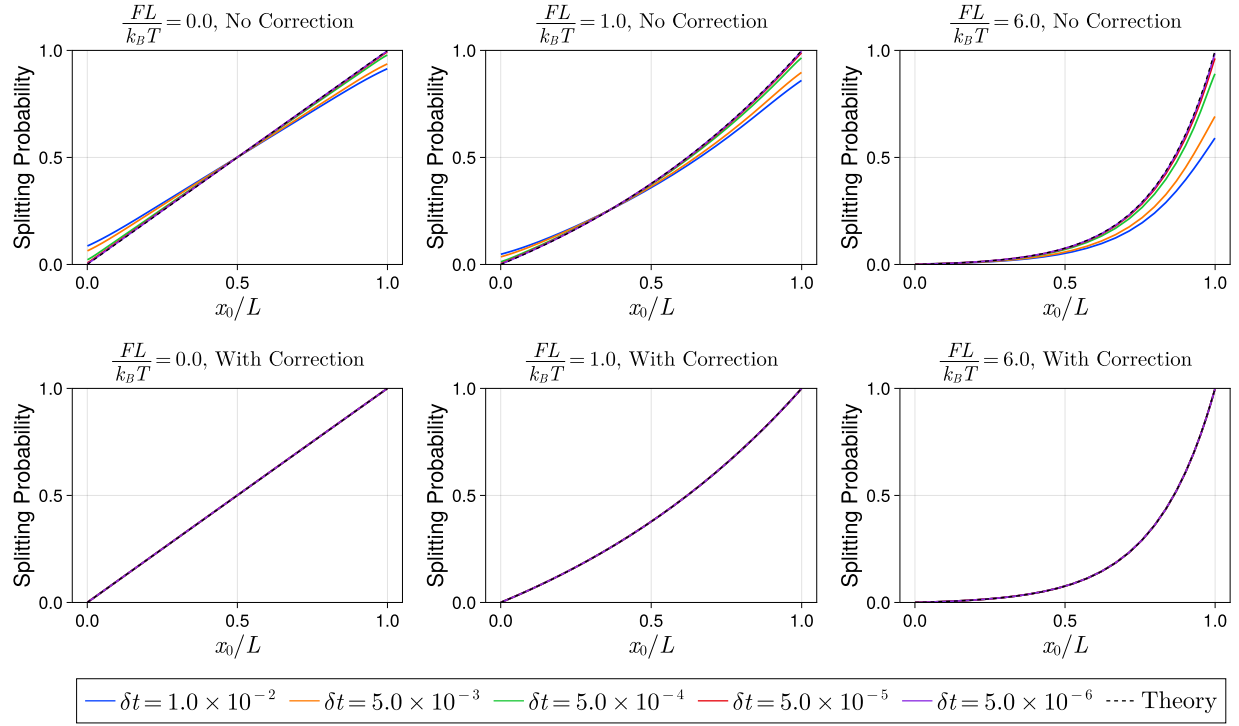


Figure 3: Splitting probability for leaving an interval $(0, L)$ plotted as a function of the initial position x_0 for unbiased diffusion (left), moderately biased diffusion (middle) and strongly biased diffusion (right) and compared with their respective theoretical values¹⁸. The upper row shows the values obtained from sampled trajectories using different sampling timesteps δt . As in Figure 2, the splitting probability estimates become worse with larger time steps. The lower row shows the splitting probabilities obtained after the correction procedure based on Eq. 3 was applied to the same trajectories. All timescales are reported in reduced units, where L^2/D sets the timescale.

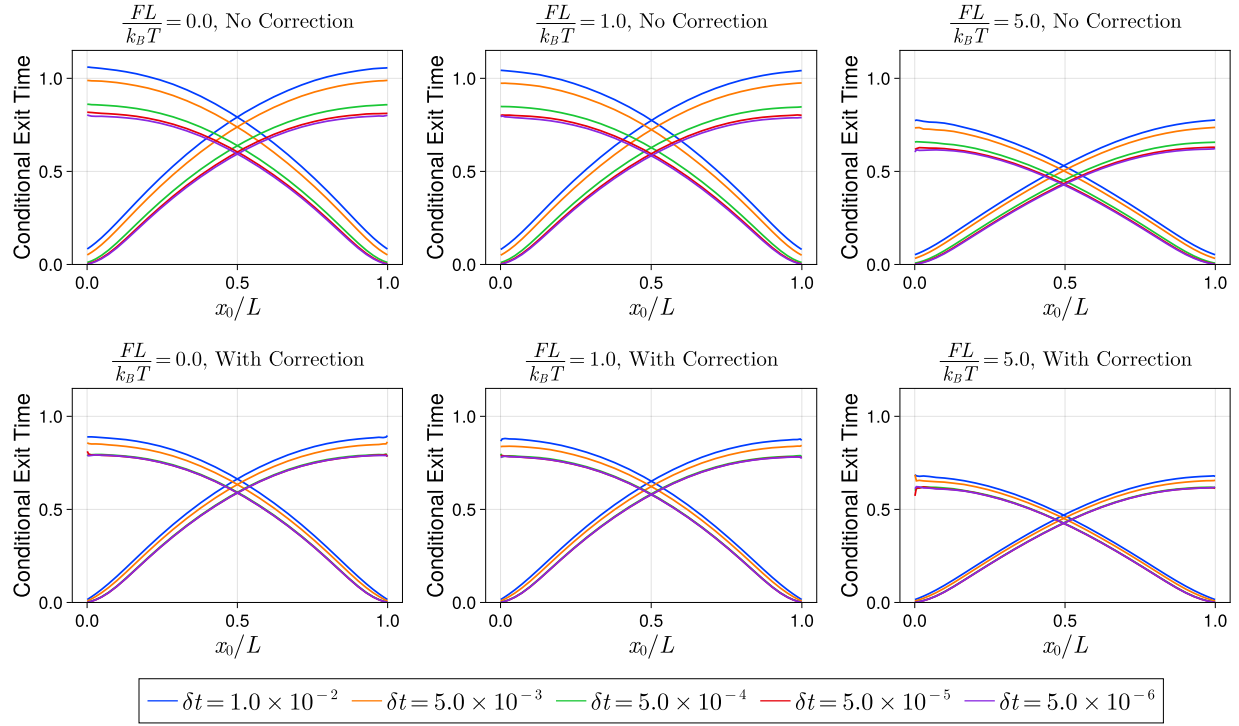


Figure 4: Conditional exit times to the right and left boundaries of the interval $(0, L)$ shown as a function of the starting position x_0 for the non-Markovian trajectory $x(t)$ simulated using Eqs. 5-7 sampled at different time intervals δt , as indicated. Upper row shows raw data and lower row shows the data after the correction of Eq. 3 was implemented. Other parameters are the same as in Figure 2. All timescales are reported in reduced units, where L^2/D_x sets the timescale.

References

1. J. Andrecka, Y. Takagi, K. J. Mickolajczyk, L. G. Lippert, J. R. Sellers, W. O. Hancock, Y. E. Goldman and P. Kukura, *Methods Enzymol* **581**, 517-539 (2016).
2. K. Neupane, D. A. Foster, D. R. Dee, H. Yu, F. Wang and M. T. Woodside, *Science* **352** (6282), 239-242 (2016).
3. K. Neupane, D. B. Ritchie, H. Yu, D. A. Foster, F. Wang and M. T. Woodside, *Phys Rev Lett* **109** (6), 068102 (2012).
4. F. Sturzenegger, F. Zosel, E. D. Holmstrom, K. J. Buholzer, D. E. Makarov, D. Nettels and B. Schuler, *Nat Commun* **9** (1), 4708 (2018).
5. H. S. Chung and W. A. Eaton, *Curr Opin Struct Biol* **48**, 30-39 (2018).
6. H. S. Chung, J. M. Louis and W. A. Eaton, *Science* **335**, 981-984 (2012).
7. J. Y. Kim and H. S. Chung, *Science* **368** (6496), 1253-1257 (2020).
8. D. E. Makarov, *Single Molecule Science: Physical Principles and Models*. (CRC Press, Taylor & Francis Group, Boca Raton, 2015).
9. H. S. Chung, J. M. Louis and W. A. Eaton, *Proc Natl Acad Sci U S A* **106** (29), 11837-11844 (2009).
10. G. Hummer, *J Chem Phys* **120**, 516-523 (2004).
11. A. M. Berezhkovskii and D. E. Makarov, *Biophysical Reports* **1**, 100029 (2021).
12. X. Li and A. B. Kolomeisky, *J Chem Phys* **139** (14), 144106 (2013).
13. A. L. Thorneywork, J. Gladrow, Y. Qing, M. Rico-Pasto, F. Ritort, H. Bayley, A. B. Kolomeisky and U. F. Keyser, *Sci Adv* **6** (18), eaaz4642 (2020).
14. R. Satija, A. M. Berezhkovskii and D. E. Makarov, *Proc Natl Acad Sci U S A* **117** (44), 27116-27123 (2020).
15. A. M. Berezhkovskii, S. M. Bezrukov and D. E. Makarov, *J Chem Phys* **154** (11), 111101 (2021).
16. D. Hartich and A. Godec, *Phys. Rev. X* **11**, 041047 (2021).
17. N. Zijlstra, D. Nettels, R. Satija, D. E. Makarov and B. Schuler, *Phys. Rev. Letters* **125**, 146001 (2020).
18. S. Redner, *A Guide to First Passage Times*. (Cambridge University Press, 2001).
19. P. Ma, R. Elber and D. E. Makarov, *J Chem Theory Comput* **16** (10), 6077-6090 (2020).
20. N. Q. Hoffer and M. T. Woodside, *Curr Opin Chem Biol* **53**, 68-74 (2019).
21. R. Elber, D. E. Makarov and H. Orland, *Molecular Kinetics in Condense Phases: Theory, Simulation, and Analysis*. (Wiley and Sons, 2020).
22. C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. (Springer-Verlag, Berlin, 1983).
23. B. Peters, *Annu Rev Phys Chem* **67**, 669-690 (2016).
24. W. E and E. Vanden-Eijnden, *J. Stat. Phys.* **123** (3), 503-523 (2006).
25. R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka and E. S. Shakhnovich, *J. Chem. Phys* **108**, 334-350 (1998).
26. A. M. Berezhkovskii and A. Szabo, *J Chem Phys* **150** (5), 054106 (2019).
27. A. M. Berezhkovskii and A. Szabo, *J Phys Chem B* **117** (42), 13115-13119 (2013).

28. A. P. Manuel, J. Lambert and M. T. Woodside, *Proc Natl Acad Sci U S A* **112** (23), 7183-7188 (2015).
29. K. Neupane, A. P. Manuel, J. Lambert and M. T. Woodside, *The Journal of Physical Chemistry Letters* **6**, 1005-1010 (2015).
30. A. M. Berezhkovskii and D. E. Makarov, *J Phys Chem Lett* **11** (5), 1682-1688 (2020).
31. R. Covino, M. T. Woodside, G. Hummer, A. Szabo and P. Cossio, *J Chem Phys* **151** (15), 154115 (2019).
32. A. M. Berezhkovskii and D. E. Makarov, *J Chem Phys* **147** (20), 201102 (2017).
33. R. J. de Oliveira, *J Chem Phys* **149** (23), 234107 (2018).
34. F. C. Freitas, A. N. Lima, V. G. Contessoto, P. C. Whitford and R. J. Oliveira, *J Chem Phys* **151** (11), 114106 (2019).
35. R. Zwanzig, *Nonequilibrium Statistical Mechanics*. (Oxford University Press, 2001).
36. E. Pollak, *J Chem Phys* **85** (2), 865-867 (1986).
37. R. F. Grote and J. T. Hynes, *The Journal of Chemical Physics* **73** (6), 2715-2732 (1980).
38. J. E. Straub, M. Borkovec and B. J. Berne, *J Chem Phys* **84**, 1788-1794 (1986).
39. D. E. Makarov, *J. Chem. Phys.* **138**, 014102 (2013).
40. C. Ayaz, L. Tepper, F. N. Brunig, J. Kappler, J. O. Daldrop and R. R. Netz, *Proc Natl Acad Sci U S A* **118** (31) (2021).
41. J. O. Daldrop, J. Kappler, F. N. Brunig and R. R. Netz, *Proc Natl Acad Sci U S A* **115** (20), 5169-5174 (2018).
42. R. Satija and D. E. Makarov, *J Phys Chem B* **123** (4), 802-810 (2019).
43. E. Medina, R. Satija and D. E. Makarov, *J Phys Chem B* **122** (49), 11400-11413 (2018).
44. A. Kumar, Y. Scher, S. Reuveni and M. S. , 2022).
45. D. E. Makarov, A. M. Berezhkovskii, G. Haran and E. Pollak, *J. Phys. Chem. B* **126** (40), 7966–7974 (2022).