MODELING RARE TRANSITION EVENTS

WEINAN E AND ERIC VANDEN-EIJNDEN

ABSTRACT. Dynamics in nature often proceed in the form of rare transition events: The system under study spends very long periods of time at various metastable states; only very rarely it hops from one metastable state to another. Understanding the dynamics of such systems requires us to study the ensemble of transition paths between the different metastable states. Transition path theory is a general mathematical framework developed for this purpose. It is also the foundation for developing modern numerical algorithms such as the string method for finding the transition pathways. We review the basic ingredients of the transition path theory and discuss connections with the more classical transition state theory. We also discuss how the string method arises in order to find approximate solutions in the framework of the transition path theory.

CONTENTS

1. INTRODUCTION

A fundamental fact is that dynamics in nature works on very disparate time scales. Atomic vibration occurs on femto- to pico-second time scales (10^{-15} to 10^{-12} sec.), whereas our daily lives are organized on the time scale of hours or days. Although many physical or biological processes occur on intermediate time scales, there are still huge gaps between the different time scales. Consequently, most dynamic processes proceed in the form of rare events. The system under consideration spends most of its time in localized regions in configuration space. Only very rarely, it hops from one localized region to another. These localized regions are called metastable states. Chemical reactions, conformation changes of molecules, nucleation events in a first order phase transition are all examples of rare events, and so are many other dynamic processes in material science, biology and chemistry.

Intuitively, one can think of the dynamics of a system as the process of navigation over its potential or free energy landscape, under the action of small amplitude noise. The metastable states are the local minima of the energy, or the basins of attraction of the local minima, all called potential wells. Without the noise, the system will simply be stuck at the local minima. Noise makes it possible to move between different local minima. However, transition events are rare because the system has to overcome some barriers. These barriers can either be of an energetic nature or an entropic nature. An entropic barrier will arise for example when a diffusing particle tries to find a narrow exit over a flat but vast landscape.

Broadly speaking, there are three classes of problems we are interested in:

(1) **The $A \to B$ problem.** Here A and B refer to the initial and final states (also called the reactant and product states, respectively) during a transition. Knowing them, we would like to find out the mechanism of the transition and the transition rates. For this purpose, we need to find the most probable transition paths from A to B or the bottleneck of the transition. From an application viewpoint, this means understanding the mechanism of chemical reactions, nucleations, subcritical instabilities and conformation changes, etc.
(2) **The exploration problem.** This does not require knowing the final states. Instead, we are interested in finding out the possible outcomes of a transition and the relative likelihood of each outcome.
(3) **Accelerating dynamic simulations.** The effectiveness of molecular dynamics or kinetic Monte Carlo simulation methods is often hindered by the occurence of rare events. It is of great practical interest to be able to develop tools that can accelerate these dynamic simulation algorithms without compromising their statistical accuracy.

Note that our objective is not to keep track of the detailed dynamics of the system, but rather to capture statistically the sequence of hops or transitions between different local minima or metastable states. This means that effectively, the dynamics of the system is modeled by a Markov chain: the metastable states are the states of the chain and the hopping rates are the transition rates between different states. When designing algorithms for accelerating molecular dynamics, our purpose is not to reproduce the detailed dynamics, but rather to capture the effective dynamics of the Markov chain.

In this review, we will focus on the first class of problems.

From a theoretical viewpoint, the well-known transition state theory (TST) [26, 75, 34], a cornerstone of chemical physics, has been very successful in providing the language, the intuition as well as the foundation for developing computational tools for studying barrier

crossing events. TST states basically that to go from the reactant state to the product state, the system has to navigate itself to the transition state which is a saddle point on the potential energy surface. In addition, the transition rate can be obtained by computing the probability flux of particles that cross the dividing surface which contains the transition state and separates the basins of attraction of the reactant and product states.

The same problem has been studied in the mathematics literature within the framework of the powerful large deviation theory (LDT) [**?**, **?**]. LDT allows us to compute the probability that a diffusion process stays in a small neighborhood of a particular path. This allows to define the most probable transition path from one local minimum to another. For overdamped systems, the most probable path is simply the minimum energy path (MEP), to be defined below.

Numerical algorithms have also been developed to look for saddle points or to find the MEP. Saddle point search algorithms include the eigenvector following method [**?**, **?**, **?**], the dimer method [**?**], the climbing image nudged elastic band method [**?**] and the gentlest ascent dynamics (GAD) [**?**]. Algorithms for finding the MEP include the string method [**?**] and the nudged elastic band method [**?**]. These algorithms have been extended and improved in many different ways, and they are generally quite effective for situations when the energy landscape is relatively smooth.

It has also been realized for a long time that TST is limited to situations when the crossing through the transition state region is ballistic, since it assumes that every crossing gives rise to a successful reaction. If the crossing is diffusive, then TST overestimates the reaction rate. The situation is actually a bit worse. For a system with a rugged energy landscape with many small barriers, or when entropic (e.g. volume) effects matter as they typically do in high dimensions, the saddle points do not necessarily play the role of transition states. In fact, in complex situations, the very notion of the transition state becomes questionable. Essentially the same kind of difficulties arise in LDT. The notion of most probable paths becomes irrelevant. Instead, an ensemble of paths contribute to the transition process. Intuitively, for systems with rugged energy landscapes, what one would like to do is to replace the notion of transition state by the notion of transition state ensemble, and to replace the notion of most probable transition paths by that of the transition tubes (inside which most of the flux of the transition paths are concentrated).

A popular approach in chemistry is to reduce the problem to a free energy landscape corresponding to some "reaction coordinates". Intuitively, reaction coordinates should be something that can be used to parametrize the transition paths. This is certainly a very useful notion (and one that we will discuss in detail below), but is also one that has been greatly abused: Oftentimes one has to guess what the reaction coordinates are based on intuition, and there are plenty of examples showing that our intuition does not always serve us in the right way. In particular, as has been pointed out in [**?**], the slow variables may not have anything to do with the reaction coordinates.

Therefore to study he transition processes in complex systems with rugged energy landscape, it is necessary to examine the ensemble of all the transition paths as a probability space. This viewpoint has been developed in two directions. As a numerical algorithm, the "transition path sampling" (TST) technique has been developed by Bolhuis, Chandler, Dellago and Geissler [6, 5] as a way of Monte Carlo sampling of the ensemble of transition paths. As a theoretical framework, the "transition path theory" (TPT) has been developed by E, Vanden-Eijnden and co-workers to characterize the ensemble of transition paths [23] (see also [65, **?**, 44]. TPT addresses questions like:

(1) What is the probability distribution of the particles in the transition path ensemble?

(2) What is the transition rate?
(3) What is the probability current associated with the transition paths?

TPT is exact no matter how complicated the transition is. Being exact, TPT provides the foundation for making approximations and developing efficient numerical algorithms. For example, as we will see later, the finite temperature string method [19, 20, 20, 54, 21, 41, 43, 70], which is a rather powerful tool for analyzing transition in systems with rough energy landscapes, can be derived from TPT under the approximation that with high probability, the flux of the transition paths is concentrated inside one or a few localized thin tubes. One can envision other approximations, which will lead to other numerical algorithms.

In this paper we begin with a short review of the classical transition state theory and large deviation theory. We then focus on the recently developed transition path theory and the numerical algorithms for finding the most probable paths.

1.1. **The setup.** Let us first introduce the kind of dynamics that we will consider. For the most part we will focus on systems governed by the Langevin equation

$$\begin{cases} \dot{\boldsymbol{x}} = m^{-1}\boldsymbol{p} \\ \dot{\boldsymbol{p}} = -\nabla U - \gamma \boldsymbol{p} + \sqrt{2\gamma k_B T} m^{1/2} \dot{\boldsymbol{w}} \end{cases} \tag{1}$$

Here m is the mass matrix and U denotes the potential energy of the system. The last two terms in the second equation are thermostat terms which represents the effects of the heat bath: γ is the friction coefficient, which can also be a tensor although we will only consider the case when it is a scalar, T is the temperature, k_B is Boltzmann's constant and $\dot{\boldsymbol{w}}$ is the standard Gaussian white noise, i.e. the Gaussian process with mean 0 and covariance

$$\langle \dot{w}_i(t) \dot{w}_j(s) \rangle = \delta_{ij} \delta(t-s) \tag{2}$$

Formally, one may also consider the special case when $\gamma = 0$, corresponding to the NVE ensemble, when (1) reduces to the Hamiltonian dynamics, with no explicit noise

$$\begin{cases} \dot{\boldsymbol{x}} = m^{-1}\boldsymbol{p} \\ \dot{\boldsymbol{p}} = -\nabla U \end{cases} \tag{3}$$

In this case, the chaotic nature of the dynamics plays the role of the noise, and indeed one typically uses the Langevin equation (1) with a friction coefficient small enough in order that the solutions of (1) do not look very different from the ones of (3). Yet the presence of the stochastic noise in (1) is important to justify the analysis below – in contrast making rigorous statements with (3) is much more difficult.

Another limit of (1) which is probably less relevant in applications but is useful for discussing concepts is the overdamped dynamics obtained when $\gamma \gg 1$:

$$\gamma m \dot{\boldsymbol{x}} = -\nabla U + \sqrt{2\gamma k_B T} m^{1/2} \dot{\boldsymbol{w}} \tag{4}$$

Note that both (1) and (4) can be cast into the same form:

$$\dot{\boldsymbol{z}} = -K \nabla V(\boldsymbol{z}) + \sqrt{2k_B T} \sigma \dot{\boldsymbol{w}}(t) \tag{5}$$

For the Langevin dynamics, we have $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{p})^T$, $V(\boldsymbol{z}) = H(\boldsymbol{x}, \boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T m^{-1} \boldsymbol{p} + U(\boldsymbol{x})$ and

$$K = \begin{pmatrix} 0 & -I \\ I & \gamma m \end{pmatrix} \qquad \sigma = \begin{pmatrix} 0 & 0 \\ 0 & (\gamma m)^{1/2} \end{pmatrix}, \tag{6}$$

For the overdamped dynamics, we have $\boldsymbol{z} = \boldsymbol{x}$, $V(\boldsymbol{z}) = U(\boldsymbol{x})$, $K = \gamma^{-1}m^{-1}$, and $\sigma = \gamma^{-1/2}m^{-1/2}$. Note that we also have in both cases

$$\tfrac{1}{2}(K + K^T) = \sigma\sigma^T \tag{7}$$

More generally, we will consider the stochastic process $\boldsymbol{z} \in \Omega \subset \mathbb{R}^d$ described by the Itō stochastic differential equation (SDE):

$$\dot{\boldsymbol{z}} = \boldsymbol{b}(\boldsymbol{z}) + \sqrt{2}\,\sigma(\boldsymbol{z})\dot{\boldsymbol{w}} \tag{8}$$

The generator of this process is the operator given by [18]

$$L = \boldsymbol{b}(\boldsymbol{z}) \cdot \nabla + a(\boldsymbol{z}) : \nabla\nabla = \sum_{i=1}^{d} b_i(\boldsymbol{z})\frac{\partial}{\partial z_i} + \sum_{i,j=1}^{d} a_{i,j}(\boldsymbol{z})\frac{\partial^2}{\partial z_i \partial z_j} \tag{9}$$

where

$$a(\boldsymbol{z}) = \sigma\sigma^T(\boldsymbol{z}) \qquad \text{i.e.} \qquad a_{i,j}(\boldsymbol{z}) = \sum_{k=1}^{d} \sigma_{i,k}(\boldsymbol{z})\sigma_{j,k}(\boldsymbol{z}) \tag{10}$$

The operator L is such that

$$\lim_{t\to 0+} \frac{1}{t}\left(\mathbb{E}^{\boldsymbol{z}}\Phi(\boldsymbol{z}(t)) - \Phi(\boldsymbol{z})\right) = (L\Phi)(\boldsymbol{z}) \tag{11}$$

where $\mathbb{E}^{\boldsymbol{z}}$ denotes expectation conditional on $\boldsymbol{z}(0) = \boldsymbol{z}$ and $\Phi(\boldsymbol{z})$ is any suitable test function. Associated with L is its adjoint operator L^* whose action on the test function $\Phi(\boldsymbol{z})$ is defined as:

$$\begin{aligned}(L^*\Phi)(\boldsymbol{z}) &= -\nabla \cdot (\boldsymbol{b}(\boldsymbol{z})\Phi(\boldsymbol{z})) + \nabla\nabla : (a(\boldsymbol{z})\Phi(\boldsymbol{z})) \\ &= -\sum_{i=1}^{d} \frac{\partial}{\partial z_i}(b_i(\boldsymbol{z})\Phi(\boldsymbol{z})) + \sum_{i,j=1}^{d} \frac{\partial^2}{\partial z_i \partial z_j}(a_{i,j}(\boldsymbol{z})\Phi(\boldsymbol{z}))\end{aligned} \tag{12}$$

The operator L and its adjoint L^* enter respectively in the backward and forward Kolmogorov equations associated with the process (8), and these equations will play important roles in the sequel. For instance, we will assume that the system has a unique invariant distribution (albeit not necessarily an equilibrium one) with density $\rho(\boldsymbol{z})$, which satisfies the forward equation:

$$(L^*\rho)(\boldsymbol{z}) = 0 \tag{13}$$

For example in the case when the dynamics of the system is governed by (1), then

$$\rho(\boldsymbol{x}, \boldsymbol{p}) = Z_H^{-1} e^{-\beta H(\boldsymbol{x},\boldsymbol{p})} \tag{14}$$

where $\beta = 1/(k_B T)$ and $Z_H = \int e^{-\beta H(\boldsymbol{x},\boldsymbol{p})} d\boldsymbol{x} d\boldsymbol{p}$. When the dynamics is governed by (4), then

$$\rho(\boldsymbol{x}) = Z^{-1} e^{-\beta U(\boldsymbol{x})} \tag{15}$$

where $Z = \int e^{-\beta U(\boldsymbol{x})} d\boldsymbol{x}$.

Before ending this introduction, we make a comment about terminology. One class of the most commonly studied transition events are chemical reactions. For this reason, we also call transition events, "reactions", and transition paths, "reaction paths".

2. METASTABLE STATES

Metastability is characterized by two disparate time scales, a short relaxation time scale and long transition time scale. A metastable state is stable over the relaxation time scale, but unstable over the transition time scale. The key parameter that enters in the discussion is the ratio between noise amplitude and typical barrier height for the potential.

In a simple system, each local minimum of the potential is a metastable set. For complex systems, metastable sets are less well-defined since we must deal with a set of wells, but their intuitive meaning is quite clear. In practice, they can be defined with the help of some coarse-grained variables or order parameters. For example, for a molecule, we can define the metastable sets by restricting the values of a set of torsion angles.

Precise definition of the metastable sets can be found using the spectral theory of the transfer operators or the generators associated with the underlying stochastic dynamics. See for example [**?**, **?**, **?**]. Roughly speaking, metastability can be related to the gaps in the spectrum of the relevant generators of the Markov process and metastable sets can be defined through the eigenfunctions associated with the leading eigenvalues. Indeed, it can be shown that these eigenfunctions are approximately piecewise constant, and the subsets on which the eigenfunctions are approximately constant are the metastable sets.

3. TRANSITION STATE THEORY

The classical theory that describes barrier-crossing events over an energy landscape is the transition state theory (TST), developed by Eyring, Polanyi, Wigner, Kramers, et al [**?**]. There are two main components in the transition state theory, the notion of transition states and an expression for the transition rate. Transition states are the dynamic bottlenecks for the transition. They are important for the following reasons:

(1) With very high probability, transition paths have to pass through a very small neighborbood of the transition states. Exceptional paths have exponentially small relative probability.
(2) Once the transition state is passed, the system can relax to the new stable state via its own intrinsic dynamics, on the relaxation time scale.

For simple systems, given two neighboring local minima, the transition state is nothing but the saddle point that separates the two minima.

The transition rate, i.e. the average number of transitions per unit time interval, can be computed from the probability flux of particles that pass through the neighborhood of the transition state. As an example, let us consider the canonical ensemble of a Hamiltonian system. For simplicity, we will focus on an example of a one-dimensional system [**?**]. Let V be a one-dimensional potential with one local minimum at $x = -1$ and another at $x = 1$ separated by a saddle point at $x = 0$. We are interested in computing the transition rate from the local minimum at $x = -1$ to the local minimum at $x = 1$ in a canonical ensemble with temperature T. According to the transition state theory, the transition rate is given by the ratio of two quantities: The first is the flux of particles that tranverse the transition state region from the side of the negative real axis to the side of the positive real axis. The second is the equilibrium distribution of particles on the side of the negative real axis, which is the basin of attraction of the initial state at $x = -1$. This second quantity is given by:

$$Z_A = \frac{1}{Z} \int_{x<0} e^{-\beta H(x,p)} dx dp \tag{16}$$

where $H(x,p) = \frac{1}{2m}p^2 + U(x)$ is the Hamiltonian of the system, $Z = \int_{\mathbb{R}^2} e^{-\beta H(x,p)} dxdp$ is the normalization constant, $\beta = (k_B T)^{-1}$. The first quantity is given by

$$Z_{AB} = \frac{1}{Zm} \int_{x=0} p\theta(p) e^{-\beta H(x,p)} dp \tag{17}$$

where θ is the Heaviside function: $\theta(z) = 1$ if $z > 0$ and $\theta(z) = 0$ if $z < 0$. To leading order, we have

$$Z_{AB} \sim \frac{1}{Z\beta} e^{-\beta U(0)} \tag{18}$$

If we approximate U further by a quadratic function centered at $x = -1$, $U(x) \sim U(-1) + 1/2U''(-1)(x+1)^2$, we obtain:

$$Z_A \sim \frac{1}{Z} \frac{2\pi}{\beta} \sqrt{\frac{m}{U''(-1)}} e^{-\beta U(-1)} \tag{19}$$

The ratio of the two quantities gives the rate we are interested in:

$$\nu_R = \frac{1}{2\pi} \sqrt{\frac{U''(-1)}{m}} e^{-\beta \delta E} \tag{20}$$

where $\delta E = U(0) - U(-1)$ is the energy barrier.

Note that this rate depends on the mass of the particle and the second derivative of the potential at the initial state, but it does not depend on the second derivative of the potential at the saddle point.

More general situations will be discussed later after introducing the transition path theory.

4. Large deviation theory

Large deviation theory is a rigorous mathematical theory for characterizing rare events in general Markov processes [**?**, **?**]. The special case for random perturbations of dynamical systems was considered by Freidlin and Wentzell [**?**]. This theory is often referred to as the Wentzell-Freidlin theory.

Consider the stochastically perturbed dynamical system:

$$\dot{\boldsymbol{x}} = \boldsymbol{b}(\boldsymbol{x}) + \sqrt{\varepsilon}\dot{\boldsymbol{w}}, \quad \boldsymbol{x} \in \mathbb{R}^d \tag{21}$$

Fix two points $\boldsymbol{A}$ and $\boldsymbol{B}$ in $\mathbb{R}^d$ and a time parameter $T > 0$. Let $\phi : [-T, T] \to \mathbb{R}^d$ be a continuous and differentiable path such that $\phi(-T) = \boldsymbol{A}, \phi(T) = \boldsymbol{B}$. The Wentzell-Freidlin theory asserts roughly that

$$\text{Prob.}\{\boldsymbol{x} \text{ stays in a neighborhood of } \phi \text{ over the interval } [-T, T]\} \sim e^{-S(\phi)/\varepsilon} \tag{22}$$

where the *action functional* S is defined by:

$$S_T(\phi) = \frac{1}{2} \int_{-T}^{T} |\dot{\phi}(t) - \boldsymbol{b}(\phi(t))|^2 dt \tag{23}$$

With (22), finding the path with maximum probability is turned into the problem of finding the path with minimum action:

$$\inf_T \inf_\phi S_T(\phi) \tag{24}$$

subject to the constraint that $\phi(-T) = \boldsymbol{A}, \phi(T) = \boldsymbol{B}$. Consider the case when $\boldsymbol{b}(\boldsymbol{x}) = -\nabla U(\boldsymbol{x})$. Assume that $\boldsymbol{A}$ and $\boldsymbol{B}$ are two neighboring local minima of U separated by the saddle point $\boldsymbol{C}$. Then we have:

Lemma (Wentzell-Freidlin [**?**]):

(1) The minimum of (24) is given by:

$$S^* = 2(U(\boldsymbol{C}) - U(\boldsymbol{A})) \tag{25}$$

(2) Consider the paths

$$\dot{\phi_1}(s) = \nabla U(\phi_1(s)), \quad \phi_1(-\infty) = \boldsymbol{A}, \phi_1(\infty) = \boldsymbol{C} \tag{26}$$

$$\dot{\phi_2}(s) = -\nabla U(\phi_2(s)), \quad \phi_2(-\infty) = \boldsymbol{C}, \phi_2(\infty) = \boldsymbol{B} \tag{27}$$

then

$$S^* = S_\infty(\phi_1) + S_\infty(\phi_2) = S_\infty(\phi_1) \tag{28}$$

The most probable transition path is then the combination of ϕ_1 and ϕ_2: ϕ_1 goes against the original dynamics and therefore requires the action of the noise. ϕ_2 simply follows the original dynamics and therefore does not require the help from the noise.

It is not difficult to convince oneself that the minimum in T in (24) is attained when $T = \infty$. To see why the minimization problem in (24) is solved by the path defined above, note that

$$S_\infty[\phi_1] = 2(U(\boldsymbol{C}) - U(\boldsymbol{A})), S_\infty[\phi_2] = 0. \tag{29}$$

In addition, for any path ϕ connecting $\boldsymbol{A}$ and a point on the separatrix that separates the basins of attraction of $\boldsymbol{A}$ and $\boldsymbol{B}$, we have

$$\begin{aligned} S_\infty[\phi] &= \frac{1}{2}\int_{\mathbb{R}} \langle \dot{\phi} - \nabla U, (\dot{\phi} - \nabla U)\rangle dt + 2\int_{\mathbb{R}} \dot{\phi}\cdot\nabla U dt \\ &\geq 2\int_{\mathbb{R}} \dot{\phi}\cdot\nabla U dt \\ &= 2\int_{\mathbb{R}} \dot{U} dt \\ &\geq 2(U(\boldsymbol{C}) - U(\boldsymbol{A})) \end{aligned}$$

since $\boldsymbol{C}$ is the minimum of U on the separatrix.

This result can also be generalized to the case when there are intermediate stable states between $\boldsymbol{A}$ and $\boldsymbol{B}$. In that case the most probable transition path is a combination of paths that satisfy:

$$\dot{\phi}(s) = \pm\nabla U(\phi(s)) \tag{30}$$

Paths that satisfy this equation are called the *minimum energy path* (MEP). One can write (30) as:

$$\nabla U(\phi(s))^\perp = 0 \tag{31}$$

where $\nabla U(\phi(s))^\perp$ denotes the component of $\nabla U(\phi(s))$ normal to the curve described by ϕ.

An alternative characterization of MEP is as follows. Given a path γ, let $\boldsymbol{z}$ be an arbitrary point on γ, and $P_{\boldsymbol{z}}$ be the hyperplane that contains $\boldsymbol{z}$ and is normal to γ. γ is a MEP if for any point $\boldsymbol{z}$ on γ, $\boldsymbol{z}$ is a local minimum of V restricted to $P_{\boldsymbol{z}}$. As we will see later in this section, this characterization is more directly linked with the transition tubes or principal curves obtained from the transition path theory.

Example: A frequently used example is the *Mueller potential*:

$$U(x,y) = \sum_{i=1}^{4} A_i \exp(a_i(x-x_i)^2 + b_i(x-x_i)(y-y_i) + c_i(y-y_i)^2), \tag{32}$$

where the parameters are

$$\begin{aligned} A &= (-200, -100, -170, 15), & a &= (-1, -1, -6.5, 0.7), \\ b &= (0, 0, 11, 0.6), & c &= (-10, -10, -6.5, 0.7), \\ x &= (1, 0, -0.5, -1), & y &= (0, 0.5, 1.5, 1). \end{aligned}$$

There are three local minima in this example. Shown in Figure 1 is the minimum energy path that connects two of the local minima. This minimum energy path passes through the third local minimum. In other words, the third local minimum is an intermediate state in the transition between the two end points.

FIGURE 1. Mueller potential and a minimum energy path that connects two local minima.

5. TRANSITION PATH THEORY (TPT)

5.1. **The main objects in transition path theory.** Given two sets A and B in phase-space, viewed respectively as reactant and product states, TPT analyzes the reaction from A to B by characterizing the statistical properties of the transition paths by which this reaction occurs. More precisely, if we consider an infinitely long trajectory $\boldsymbol{z}(t)$ with $t \geq 0$, we can cut out from this trajectory the pieces during which the trajectory is on its way from A to B without coming back to A (see figure 2 for a schematic illustration). We call each such piece a *reactive trajectory from A to B*, an *A-B reactive trajectory* or a *transition path from A to B*, and we refer to the set of transition paths as the *transition path ensemble from A to B*. This ensemble is closely related to the one considered in TPS, except that we do not put any restriction on the length of the reactive trajectories. Given the trajectory $\boldsymbol{z}(t)$ with $t \geq 0$, we will denote by R the subset of all the times at which $\boldsymbol{z}(t)$ belongs to an A-B reactive path (i.e. R is the set of times during which the trajectory is red in figure 2).

For the transition path ensemble, we can ask several questions:

Where do these paths spend their time? We are looking for a density $\rho_R(\boldsymbol{z})$ that gives the probability density to find a trajectory at $\boldsymbol{z}$ *and* that it be reactive. In the setup described above, $\rho_R(\boldsymbol{z})$ can be defined via

$$\rho_R(\boldsymbol{z}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \delta(\boldsymbol{z} - \boldsymbol{z}(t)) \mathbf{1}_R(t) dt \tag{33}$$

where $\mathbf{1}_R(t) = 1$ if $t \in R$ (i.e. at that time $\boldsymbol{z}(t)$ belongs to an A-B reactive path) and $\mathbf{1}_R(t) = 0$ otherwise. Note that if we consider the whole trajectory (not just the reactive pieces) and ask the same question, then the probability density at the right hand side should be given by the density $\rho(\boldsymbol{z})$ of the invariant distribution:

$$\rho(\boldsymbol{z}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \delta(\boldsymbol{z} - \boldsymbol{z}(t)) dt \tag{34}$$

Note also that $\rho_R(\boldsymbol{z})$ is not normalized: the total integral of $\rho_R(\boldsymbol{z})$ gives the probability of that the trajectory is A-B reactive.

The density $\rho_R(\boldsymbol{z})$ gives information about how likely it is for the A-B reactive paths to go through a given region, but not about how they go from regions to regions. To

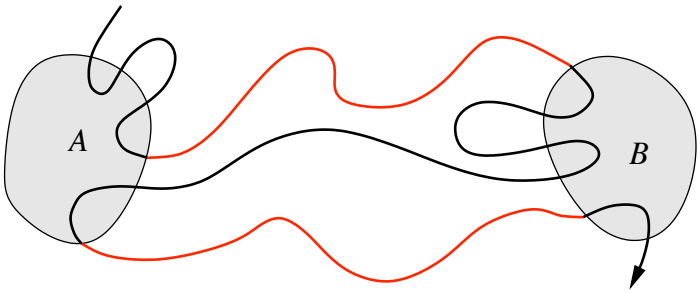

FIGURE 2. Schematic illustration of a long trajectory oscillating between the reactant state A and the product state B. The reactive pieces of this trajectory during which it travels from A to B are shown in red.

characterize how these paths flow from A to B on the average, we need another quantity, namely their probability current. To understand how the probability current comes about, let us consider first the instantaneous probability density $\rho(\boldsymbol{z},t)$ of the (full) process $\boldsymbol{z}(t)$. Formally, this probability density can be defined as

$$\rho(\boldsymbol{z},t) = \langle \delta(\boldsymbol{z} - \boldsymbol{z}(t)) \rangle \tag{35}$$

where the bracket denotes expectation with respect to the realizations of the noise in (8). Differentiating (35) with respect to time we can write[1]

$$\frac{\partial}{\partial t}\rho(\boldsymbol{z},t) = -\langle \dot{\boldsymbol{z}}(t) \cdot \nabla \delta(\boldsymbol{z} - \boldsymbol{z}(t)) \rangle \equiv -\nabla \cdot \boldsymbol{J}(\boldsymbol{z},t) \tag{36}$$

where we defined the instantaneous probability current

$$\boldsymbol{J}(\boldsymbol{z},t) = \langle \dot{\boldsymbol{z}}(t) \delta(\boldsymbol{z} - \boldsymbol{z}(t)) \rangle \tag{37}$$

(36) has the form of a conservation law (the total probability being the quantity that is conserved). The ergodicity assumption implies that $\rho(\boldsymbol{z},t) \to \rho(\boldsymbol{z})$ as $t \to \infty$, i.e. the time derivative at the left hand side of (36) tends to 0 in the long time limit. In turns this implies that $\boldsymbol{J}(\boldsymbol{z},t) \to \boldsymbol{J}(\boldsymbol{z})$ for some divergence free $\boldsymbol{J}(\boldsymbol{z})$, i.e. such that $\nabla \cdot \boldsymbol{J}(\boldsymbol{z}) = 0$. By ergodicity, this steady state current can also be defined via the time average

$$\boldsymbol{J}(\boldsymbol{z}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \dot{\boldsymbol{z}}(t) \delta(\boldsymbol{z} - \boldsymbol{z}(t)) dt \tag{38}$$

Time averages of this type can be evaluated explicitly. In the present case the components of the current $\boldsymbol{J}(\boldsymbol{z})$ can also be read directly from (13) (using the explicit form of L^* given in (12)):

$$J_i(\boldsymbol{z}) = b_i(\boldsymbol{z})\rho(\boldsymbol{z}) - \sum_{j=1}^{d} \frac{\partial}{\partial z_j}(a_{i,j}(\boldsymbol{z})\rho(\boldsymbol{z})) \tag{39}$$

If the statistical steady state is an equilibrium one, then by definition $\boldsymbol{J}(\boldsymbol{z}) = 0$. For nonequilibrium steady states, however, $\boldsymbol{J}(\boldsymbol{z})$ does not vanish, and the definition (38) shows that this quantity is like an average velocity field of the process $\boldsymbol{z}(t)$ at point $\boldsymbol{z}$ which thereby permits to characterize how this process flows on the average. This suggests asking the same question for the A-B reactive paths, i.e. ask:

[1]Note that these manipulations require interpreting the various products in Stratonovich sense in order that standard differentiation rules apply. This is just a technical point, however.

What is the current associated with the reactive trajectories? The current associated with the reactive paths is the vector field $\boldsymbol{J}_R : \Omega \mapsto \mathbb{R}^d$ defined as

$$\boldsymbol{J}_R(\boldsymbol{z}) = \lim_{T\to\infty} \frac{1}{T} \int_0^T \dot{\boldsymbol{z}}(t)\delta(\boldsymbol{z} - \boldsymbol{z}(t))\mathbf{1}_R(t)dt \tag{40}$$

This current must be divergence free outside of A and B, but it does not vanish. Indeed, by construction A is a source and B is a sink of reactive trajectories (but there are no other sources nor sinks of reactive trajectories outside of A and B). The flowlines of $\boldsymbol{J}_R(\boldsymbol{z})$ indicate how the reactive trajectories go from A to B on the average. The current $\boldsymbol{J}_R(\boldsymbol{z})$ is also useful to answer the following question:

What is the reaction rate? If N_T is the number of A-B reactive trajectories (of the given trajectory $\boldsymbol{z}(t)$) in $[0, T]$, what is the limit of N_T/T as $T \to \infty$? This is the reaction rate

$$\nu_R = \lim_{T\to\infty} \frac{N_T}{T} \tag{41}$$

The rate ν_R is the total probability flux of reactive trajectories out of A and into B, which (since $\boldsymbol{J}_R(\boldsymbol{z})$ is divergence free) is also the probability flux through any dividing surface S leaving A on one side and B on the other:

$$\nu_R = \int_S \hat{\boldsymbol{n}}_S(\boldsymbol{z}) \cdot \boldsymbol{J}_R(\boldsymbol{z}) d\sigma_S(\boldsymbol{z}) \tag{42}$$

where $\hat{\boldsymbol{n}}_S(\boldsymbol{z})$ is the unit normal to S pointing towards B and $d\sigma_S(\boldsymbol{z})$ is the surface element in S. Another expression for ν_R will be given below, see (66). Notice that ν_R is also the rate of the back reaction from B to A. Indeed, each time the process makes a transition from A to B, it needs to make a transition from B to A before making another transition from A to B, i.e. the number of A-B trajectories in $[0, T]$ is the same as the number of B-A trajectories plus or minus one. Of course, this does not mean that the average time it takes the process to go to B starting from A is the same as the one it takes to go to A starting from B. To compute these average times, let T_A be the total time in $[0, T]$ during which the last set hit by the trajectory was A, and T_B the total time during which the last set was B. Then $T_A + T_B = T$, and we can define the rates

$$k_{A,B} = \lim_{T\to\infty} \frac{N_T}{T_A}, \qquad k_{B,A} = \lim_{T\to\infty} \frac{N_T}{T_B} \tag{43}$$

These two rates are different (and different from ν_R) and their inverses give an estimate of the average transition times from A to B and B to A. How to compute the rates in (43) and relate them to ν_R is explained below.

5.2. **The committor functions.** To answer the questions listed in section 5.1, we need the notion of *committor functions* [23, 65]. These are also referred as the *capacitance functions* in the mathematics literature [17, 61].

Let $q_+ : \Omega \mapsto [0, 1]$ be the solution of

$$\begin{cases} Lq_+ = \boldsymbol{b} \cdot \nabla q_+ + a : \nabla\nabla q_+ = 0, & \boldsymbol{z} \notin A \cup B, \\ q_+ = 0, & \boldsymbol{z} \in A, \\ q_+ = 1, & \boldsymbol{z} \in B, \end{cases} \tag{44}$$

The function q_+ is the called *the forward committor function* for the transition process from A to B. In a similar fashion, we can define the *backward committor function* q_- :

$\Omega \mapsto [0,1]$ as the solution of

$$\begin{cases} L^\dagger q_- := -\boldsymbol{b}\cdot\nabla q_- + \dfrac{2}{\rho}\mathrm{div}(a\rho)\cdot\nabla q_- + a:\nabla\nabla q_- = 0, & \boldsymbol{z}\notin A\cup B,\\ q_- = 1, & \boldsymbol{z}\in A,\\ q_- = 0, & \boldsymbol{z}\in B, \end{cases} \tag{45}$$

The operator $L^\dagger$ (not to be confused with L^*) is the generator of the time-reversed process. The committor functions have a very simple interpretation: $q_+(\boldsymbol{z})$ is the probability that a trajectory initiated at $\boldsymbol{z}$ will reach B before it reaches A, and $q_-(\boldsymbol{z})$ is the probability that a trajectory arriving at $\boldsymbol{z}$ came from A, i.e. before arriving at $\boldsymbol{z}$, it left from A after it left from B. The isosurfaces of these functions are called *isocommittor surfaces*. The isosurface $\{\boldsymbol{z} : q_+(\boldsymbol{z}) = \frac{1}{2}\}$ is of particular interest: points on this surface have an equal probability of first reaching A or B.

For non-equilibrium processes, $q_- \neq 1 - q_+$ because $L^\dagger \neq L$ (the process is not statistically invariant under time reversal). In particular, for the Langevin equation (1), $q_-(\boldsymbol{x},\boldsymbol{p}) \neq 1-q_+(\boldsymbol{x},\boldsymbol{p})$ since time reversal involves a momentum flip. In fact, $q_-(\boldsymbol{x},\boldsymbol{p}) = 1 - q_+(\boldsymbol{x},-\boldsymbol{p})$. For equilibrium processes, however, $q_- = 1 - q_+$ since $L^\dagger = L$. This is the case, in particular, of the overdamped equation (4). The symmetry under time-reversal implies that the committor function can be obtained from a minimization principle. In the case of the overdamped equation, this involves minimizing the objective function

$$I(q) = k_B T\gamma^{-1}Z^{-1}\int \sum_{i=1}^{d} m_i^{-1}\left(\frac{\partial q(\boldsymbol{x})}{\partial x_i}\right)^2 e^{-\beta U(\boldsymbol{x})}d\boldsymbol{x} \tag{46}$$

over all q such that $q(\boldsymbol{x}) = 0$ if $\boldsymbol{x}\in A$ and $q(\boldsymbol{x}) = 1$ if $\boldsymbol{x}\in B$.

The committor equations (44) and (45) cannot be solved in closed form except in the simplest cases. One such case is that of the overdamped equation in one dimension, when (44) reduces to

$$k_B T q_+'' - U' q_+' = 0 \tag{47}$$

where the prime denotes derivative with respect to x. The solution of (47) with the boundary condition $q_+(a) = 0$, $q_+(b) = 1$ with $a \le b$ (i.e. $A = \{x \le a\}$ and $B = \{x \ge b\}$) is

$$q(x) = \frac{\int_a^x e^{\beta U(x')}dx'}{\int_a^b e^{\beta U(x')}dx'} \tag{48}$$

In particular, if $U(x)$ is a double-well potential with a barrier much higher than $k_B T = 1/\beta$, and a and b are two points at the bottom of the wells, e.g. $U(x) = (1-x^2)^2$, $a = -1$ and $b = 1$, we see that the function $q(x)$ is close to zero on the left of the barrier, close to 1 on the right, and it goes rapidly from 0 to 1 in the region near the top of the barrier. In particular, $q(x) = \frac{1}{2}$ is attained close to the barrier top.

For more complicated situations, the committor equations (44) and (45) have to be solved numerically. How to do this in high dimension will be the object of the later sections of this paper since in such situations standard numerical techniques based e.g. on finite element discretization are ineffective. One example where such discretization can be used, however, is the overdamped equation on the rugged Mueller potential in two dimensions shown in figure 3. Below, we will use this example for illustration. The isosurfaces of the committor function in this example are shown in figure 4. Other low dimensional examples, involving multiple pathways, entropic effects, etc. are discussed in [**?**].

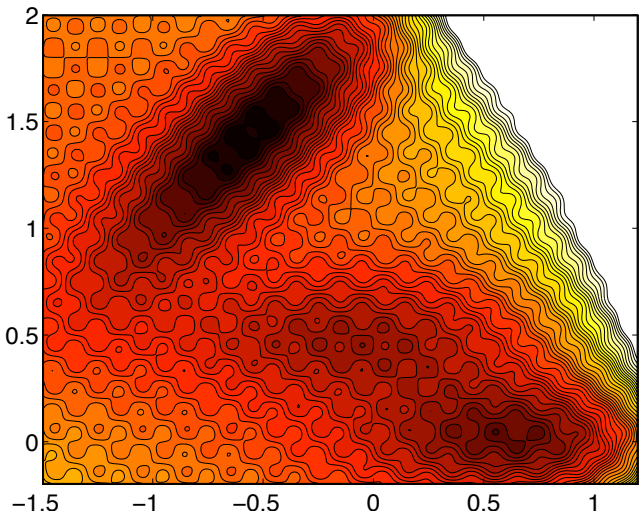

FIGURE 3. Isosurfaces of the rugged Mueller potential. The darker the region, the lower the potential.

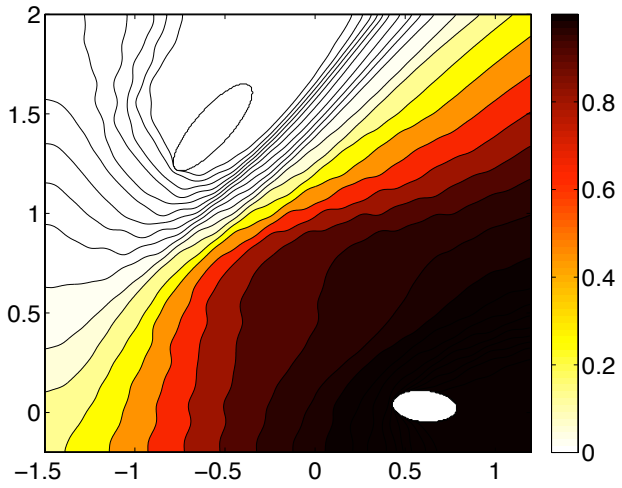

FIGURE 4. Isosurfaces the committor function in the rugged Mueller potential. The reactant state A is the ellipse in the upper left corner; the product state is the ellipse in the lower right corner. Both sets are regions where the potential shown in figure 3 is relatively low.

Remark: committor vs. averaged committor functions. q_+ and q_- are function of the full phase space $\boldsymbol{z}$ of the process. It is sometime useful to work with averaged versions of q_+ and q_-. For instance, in the case of the Langevin equation, the committor functions depend on both the positions $\boldsymbol{x}$ and the momenta $\boldsymbol{p}$, but it is useful to consider the averaged function

$$\bar{q}(\boldsymbol{x}) = \langle q_+(\boldsymbol{x},\boldsymbol{p})\rangle_{\boldsymbol{p}} \tag{49}$$

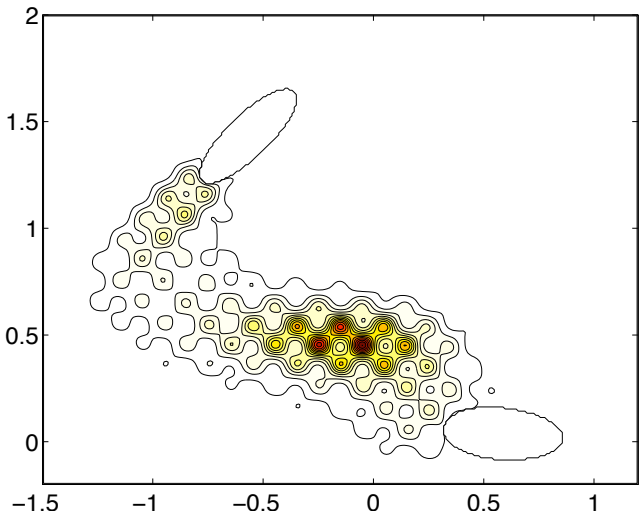

FIGURE 5. Isosurfaces of the probability density of reactive trajectories $\hat{\rho}_R$ in the example of the rugged Mueller potential. The density is as rugged as the underlying potential and it is mostly peaked in the shallow region in between A and B since this region plays the role of dynamical trap for the transition. This figure should be compared with figure 6 which shows the flowlines of the probability current of reactive trajectories. Comparison between these figures indicates that the information contained in probability density of reactive trajectories is somewhat limited and the object which is most useful to characterize the mechanism of the reaction is the probability current of these trajectories.

where $\langle\cdot\rangle_{\boldsymbol{p}}$ denotes canonical expectation with respect to the momenta. In fact, $\bar{q}(\boldsymbol{x})$ is what some authors define as the committor function [6]. It is important to stress, however, that, unlike $q_+(\boldsymbol{x},\boldsymbol{p})$, $\bar{q}(\boldsymbol{x})$ does not satisfy a closed form equation such as (44).

5.3. **Probability density of reactive trajectories.** Using the Markov nature of the dynamics, it is easy to see that the probability density to find a reactive trajectory at $\boldsymbol{z}$ can be expressed as the product between $\rho(\boldsymbol{z})$ (which give the probability density to find a trajectory at $\boldsymbol{z}$) times $q_+(\boldsymbol{z})q_-(\boldsymbol{z})$ (which gives the probability that the trajectory came last from A and will go next to B, i.e. that it is A-B reactive). As a result

$$\rho_R(\boldsymbol{z}) = q_+(\boldsymbol{z})q_-(\boldsymbol{z})\rho(\boldsymbol{z}) \tag{50}$$

If we limit ourselves to the ensemble of reactive trajectories and ask what is the probability density of finding them at $\boldsymbol{z}$ (i.e. we look at the probability density to find a trajectory at $\boldsymbol{z}$ conditional on it being reactive), we need to normalize ρ_R i.e.

$$\hat{\rho}_R(\boldsymbol{z}) = Z_R^{-1} q_+(\boldsymbol{z})q_-(\boldsymbol{z})\rho(\boldsymbol{z}) \tag{51}$$

where the normalization factor

$$Z_R = \int_\Omega q_+(\boldsymbol{z})q_-(\boldsymbol{z})\rho(\boldsymbol{z})d\boldsymbol{z} \tag{52}$$

is the probability to be A-B reactive. The formula above were first derived in [35] as a mean to analyze the transition path ensemble sampled by TPS.

The isosurfaces of $\hat{\rho}$ in the example of the rugged Mueller potential are shown in figure 5. Not surprisingly, this density is peaked in the shallow region in between A and B since this region is a dynamical trap for the transition. This example shows that the information given by ρ_R about the mechanism of the reaction is somewhat limited. Note in particular that ρ_R is not peaked at the transition state region.

5.4. **The current of A-B reactive trajectories.** The current of reactive trajectories can be computed directly from (40). The only difficulty in the calculation is that it requires to give a precise meaning to this integral since the process $\boldsymbol{z}(t)$, being solution of (8), is not differentiable in time. For the convenience of the reader, this derivation is given at the end of this section. The result is

$$\boldsymbol{J}_R = q_+ q_- \boldsymbol{J} + \rho q_- a \nabla q_+ - \rho q_+ a \nabla q_- \tag{53}$$

where $\boldsymbol{J}$ was given in (39). Componentwise, (53) is

$$J_{R,i} = q_+ q_- J_i + \rho q_- \sum_{j=1}^{d} a_{i,j} \frac{\partial q_+}{\partial z_j} - \rho q_+ \sum_{j=1}^{d} a_{i,j} \frac{\partial q_-}{\partial z_j} \tag{54}$$

Note that $\nabla \cdot \boldsymbol{J}_R = 0$ outside of A and B, consistent with the fact that there are no other source or sink of reactive trajectories besides A and B. Note also that

$$\begin{aligned} \hat{\boldsymbol{n}}_{\partial A}(\boldsymbol{z}) \cdot \boldsymbol{J}_R(\boldsymbol{z}) &= \rho(\boldsymbol{z}) \sum_{i,j=1}^{d} \hat{\boldsymbol{n}}_{\partial A,i}(\boldsymbol{z}) a_{i,j}(\boldsymbol{z}) \frac{\partial q_+(\boldsymbol{z})}{\partial z_j} \\ &= \rho(\boldsymbol{z}) |\nabla q_+(\boldsymbol{z})|^{-1} \sum_{i,j=1}^{d} a_{ij}(\boldsymbol{z}) \frac{\partial q_+(\boldsymbol{z})}{\partial z_i} \frac{\partial q_+(\boldsymbol{z})}{\partial z_j} \geq 0. \end{aligned} \tag{55}$$

where $\hat{\boldsymbol{n}}_{\partial A}(\boldsymbol{x})$ is the unit outward normal of ∂A which can also be defined as $|\nabla q_+|^{-1} \nabla q_+$. (55) means that all reactive trajectories must go out of A, consistent with their definition. Similarly, one can also check that the current is consistent with all the reactive trajectories going into B since

$$\hat{\boldsymbol{n}}_{\partial B}(\boldsymbol{z}) \cdot \boldsymbol{J}_R(\boldsymbol{z}) \leq 0 \tag{56}$$

where $\hat{\boldsymbol{n}}_{\partial B}$ is the unit outward normal of ∂B.

For the Langevin equation (1), (53) reduces to

$$\boldsymbol{J}_R = Z_H^{-1} e^{-\beta H} q_+ q_- \begin{pmatrix} \boldsymbol{p} \\ -\nabla U \end{pmatrix} + k_B T \gamma Z_H^{-1} e^{-\beta H} \begin{pmatrix} 0 \\ q_- \dfrac{\partial q_+}{\partial \boldsymbol{p}} - q_+ \dfrac{\partial q_-}{\partial \boldsymbol{p}} \end{pmatrix} \tag{57}$$

where $q_-(\boldsymbol{x}, \boldsymbol{p}) = 1 - q_+(\boldsymbol{x}, -\boldsymbol{p})$. For the overdamped equation (4), we have simply

$$\boldsymbol{J}_R = k_B T \gamma^{-1} Z^{-1} e^{-\beta U} m \nabla q \tag{58}$$

The probability current $\boldsymbol{J}_R$ can be analyzed in various ways. Of special interest are the flowlines of this current, i.e. the solutions of the artificial dynamics

$$\frac{d\boldsymbol{z}(\tau)}{d\tau} = \boldsymbol{J}_R(\boldsymbol{z}(\tau)) \tag{59}$$

If we solve this equation with initial conditions $\boldsymbol{z}(0)$ on ∂A, then each solution travels toward B and eventually reaches B. The set of all these solutions are the flowlines of $\boldsymbol{J}_R$ and they indicate how on average the reactive trajectories proceed to go from A to B. Note also that each flowline can be weighted according to the probability flux that it carries. This weighting can be done in various ways, each of which gives some type of information. For

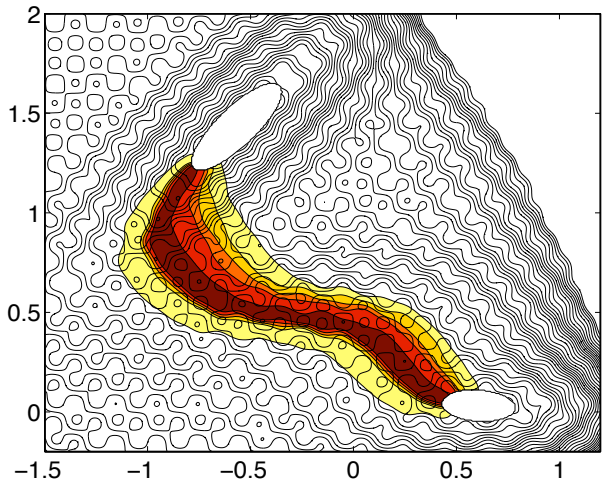

FIGURE 6. The flowlines of the probability current of reactive trajectories in the example of the rugged Mueller potential. Each flowline is weighted by the amplitude of the current through the isocommittor $\frac{1}{2}$ surface: the darker the line, the higher this amplitude. Thus, the darker the region, the more flux of reactive trajectories going through this region. Note that this permits to define transition tubes carrying a certain percentage of the flux of reactive trajectories. Note also that looking at the reaction this way, the effect of the dynamical trap that affected the probability density of reactive trajectories (see figure 5) has been filtered out and the flowlines go smoothly through this trap.

instance, we can weight each flowline by the minimum of the amplitude of the current along it, i.e. using

$$\min_{0\leq\tau\leq\tau_B} |\boldsymbol{J}_R(z(\tau))| \tag{60}$$

where $\boldsymbol{z}(0) \in \partial A$ and $\boldsymbol{z}(\tau_B) \in \partial B$. Another possibility is to assign as weight the amplitude of the current normal to a given dividing surface S, i.e. a hypersurface in Ω that separates A and B. This amounts to using

$$\hat{\boldsymbol{n}}_S(z(\tau^*)) \cdot \boldsymbol{J}_R(z(\tau^*)) \tag{61}$$

where $\hat{\boldsymbol{n}}_S(\boldsymbol{z})$ is the unit normal to S pointing toward B and $\boldsymbol{z}(\tau^*) \in S$ is the first location where the flowline $\boldsymbol{z}(\tau)$ crosses S. We applied this second procedure in the example of the rugged Mueller potential, using the isocommittor surface $\{\boldsymbol{x} : q_+(\boldsymbol{x}) = \frac{1}{2}\}$ as dividing surface, see figure 6. This procedure can be used to define transition tubes, an issue on which we will come back in section 6. The integral of the current through a dividing surface gives the total probability flux through this surface, i.e. the reaction rate ν_R. This is discussed next in section 5.5.

Remark: reactive trajectories vs flowlines. The flowlines of the probability current of reactive trajectories should not be confused with the reactive trajectories themselves. While the latter are random and typically quite complicated, the former are averaged objects in which all unnecessary details about the reaction have been filtered out. For instance, if

there are dead ends along the way between the sets A and B or dynamical traps, this affects the reactive trajectories that will wander in these dead ends and traps for what can be long periods of time. But typically these wanders amount to the trajectories going back and forth on themselves, and so they do not contribute to the current nor its flowlines. We have already illustrated this point on the maze example shown in figure **??** and it is why looking at the flowlines of the current of reactive trajectories rather than the reactive trajectories themselves is a much better way to characterize the mechanism of the reaction.

Derivation of (53). The integral in (40) can be interpreted as

$$\boldsymbol{J}_R(\boldsymbol{z}) = \lim_{h\to 0}\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\delta(\boldsymbol{z}-\boldsymbol{z}_n)a_{n+1}^{+}a_{n-1}^{-}\frac{\boldsymbol{z}_{n+1}-\boldsymbol{z}_{n-1}}{2h} \tag{62}$$

where $\boldsymbol{z}_n = \boldsymbol{z}(nh)$, $a_n^+ = 1$ if the trajectory starting from $\boldsymbol{z}_n$ will reach B before A and $a_n^+ = 0$ otherwise, and $a_n^- = 1$ if the trajectory arriving at $\boldsymbol{z}_n$ left A rather than B last and $a_n^- = 0$ otherwise. By the ergodicity assumption, the limit as $N\to\infty$ of (62) is easy to compute:

$$\begin{aligned}\boldsymbol{J}_R(\boldsymbol{z}) = \lim_{h\to 0}\frac{1}{2h}\int \delta(\boldsymbol{z}-\boldsymbol{z}_0)q^+(\boldsymbol{z}_1)q^-(\boldsymbol{z}_{-1})\rho(\boldsymbol{z}_{-1})p_h^{\boldsymbol{z}_{-1}}(\boldsymbol{z}_0)p_h^{\boldsymbol{z}_0}(\boldsymbol{z}_1)\\ \times(\boldsymbol{z}_1-\boldsymbol{z}_{-1})d\boldsymbol{z}_{-1}d\boldsymbol{z}_0d\boldsymbol{z}_1\end{aligned} \tag{63}$$

where $p_h^{\boldsymbol{z}}(\boldsymbol{z}')$ is the probability density of $\boldsymbol{z}(h)$ conditional on $\boldsymbol{z}(0)=\boldsymbol{z}$. To compute the limit as $h\to 0$, we can now use

$$p_h^{\boldsymbol{z}}(\boldsymbol{z}') = \delta(\boldsymbol{z}'-\boldsymbol{z}) + hL^*_{\boldsymbol{z}'}\delta(\boldsymbol{z}'-\boldsymbol{z}) + o(h) \tag{64}$$

where $L^*_{\boldsymbol{z}'}$ is the forward operator defined in (12), with the subscript added to stress that it acts on the variable $\boldsymbol{z}'$. Inserting this expression in (63) permits to take the limit as $h\to 0$ and gives (53) after some straightforward algebra.

5.5. **The transition rates.** Let S be any dividing surface. Then by definition

$$\nu_R = \int_S \hat{\boldsymbol{n}}_S(\boldsymbol{z})\cdot\boldsymbol{J}_R(\boldsymbol{z})d\sigma_S(\boldsymbol{z}) \tag{65}$$

Since $\boldsymbol{J}_R$ is divergence-free, it is easy to see that the integral over S defined above is actually independent of S. In fact, one can show that [65]

$$\nu_R = \int_\Omega \sum_{i,j=1}^{d} a_{i,j}(\boldsymbol{z})\frac{\partial q_+(\boldsymbol{z})}{\partial z_i}\frac{\partial q_+(\boldsymbol{z})}{\partial z_j}\rho(\boldsymbol{z})d\boldsymbol{z}. \tag{66}$$

For the Langevin equation (1) this is

$$\nu_R = k_BT\gamma Z_H^{-1}\int\sum_{i=1}^{d} m_i\left(\frac{\partial q_+}{\partial p_i}\right)^2 e^{-\beta H(\boldsymbol{x},\boldsymbol{p})}d\boldsymbol{x}d\boldsymbol{p}. \tag{67}$$

and for the overdamped equation

$$\nu_R = k_BT\gamma^{-1}Z^{-1}\int\sum_{i=1}^{d} m_i\left(\frac{\partial q_+}{\partial x_i}\right)^2 e^{-\beta U(\boldsymbol{x})}d\boldsymbol{x}. \tag{68}$$

How (67) reduces to (68) in the limit when $\gamma\gg 1$ is explained in [65].

Let us now consider the rates $k_{A,B}$ and $k_{B,A}$ defined in (43). Clearly, these rates can be expressed as

$$k_{A,B} = \frac{\nu_R}{\rho_A}, \qquad k_{B,A} = \frac{\nu_R}{\rho_B} \tag{69}$$

where ρ_A and ρ_B are defined as

$$\rho_A = \lim_{T\to\infty} \frac{T_A}{T}, \qquad \rho_B = \lim_{T\to\infty} \frac{T_B}{T} = 1 - \rho_A \tag{70}$$

Since the time t contributes to T_A if at that time the trajectory visited A rather than B last, is easy to see that

$$\rho_A = \int_\Omega \rho(\boldsymbol{z}) q_-(\boldsymbol{z}) d\boldsymbol{z}, \qquad \rho_B = \int_\Omega \rho(\boldsymbol{z})(1 - q_-(\boldsymbol{z})) d\boldsymbol{z}. \tag{71}$$

5.6. **Definition of the sets and reduced dynamics.** The framework of TPT does not require any specific property about the sets A and B and it can be used to characterize the reaction between any pair of such sets. In most situations, however, we are not interested in understanding the reaction between arbitrary sets, but rather between sets that are metastable. How should such metastable sets be defined? Intuitively, one would like that when the trajectory reaches one of these metastable sets, it spends a long time around this set and lose memory of where it came from before visiting another metastable set – the standard jargon is then to say that the trajectory commits to the metastable set before visiting another. Only if that is the case can we expect that the dynamics of the system will be reducible to a continuous-time Markov chain with the sets playing the role of the states in this chain and with some appropriate rates for the transition between the sets. The questions then become how to characterize concretely metastable sets with the property above and how to compute the rates.

Definition of the sets. To answer the first question, a suitable set of metastable sets should be such that (i) any trajectory launched from outside the sets reaches one of these sets quickly and (ii) any trajectory launched from inside one of these sets takes a very long time to reach another set[2]. Given a collection of sets, testing for these two properties is in principle doable. How to obtain such a collection of sets in practice is a much more complicated question, and one we shall not dwell upon further here since, typically, answering this question will require some *a priori* knowledge of the system, i.e. it will be system specific (see however section **??**). Let us simply note that the metastable sets and their number can vary and depend on the time scale at which one looks at the system. Indeed this specifies what one means by quick and slow in the properties (i) and (ii) above, and there can be different timescales for which these two requirements are satisfied with different sets.

Reduced dynamics. Regarding the rates to be used to characterize the transition between metastable sets, the main additional difficulty with respect to what we discussed in section 5.5 is that there may be more than two metastable sets. However handling such situations is straightforward and amounts to repeating the calculations made before for different partitions of the sets. Suppose indeed that the metastable sets are B_1, B_2, ..., B_N. For every $i = 1, \ldots, N$, take $A = B_i$ and $B = \cup_{j\neq i} B_j$. Denoting by $k_{i,j}$ the rate at which the transition from set B_i to set B_j occurs, it is then easy to see that the total rate of escape from $B_i = A$ is

$$\sum_{\substack{j=1 \\ j\neq i}}^{N} k_{i,j} = k_{A,B}. \tag{72}$$

[2]To avoid confusion, note that these sets do not need to partition state-space. To the contrary it will typically be easier to satisfy requirements (i) and (ii) with rather small sets with lot of space in between.

To get the individual rates $k_{i,j}$ we then simply need to estimate the probability that the system reaches a specific B_j starting from B_i. This probability is simply the portion of the flux of reactive trajectories going into B_j divided by their total flux going into $\cup_{j\neq i}B_j$:

$$p_{i,j} = \frac{\int_{\partial B_i} \hat{\boldsymbol{n}}_{\partial B_i}(\boldsymbol{z}) \cdot \boldsymbol{J}_R(\boldsymbol{z}) d\sigma_{\partial B_i}(\boldsymbol{z})}{\sum_{j\neq i} \int_{\partial B_j} \hat{\boldsymbol{n}}_{\partial B_j}(\boldsymbol{z}) \cdot \boldsymbol{J}_R(\boldsymbol{z}) d\sigma_{\partial B_j}(\boldsymbol{z})}. \tag{73}$$

The denominator in this equation is just $\nu_R = k_{A,B}\rho_A$. Combining (72) and (73) we see that

$$k_{i,j} = \rho_A^{-1} \int_{\partial B_i} \hat{\boldsymbol{n}}_{\partial B_i}(\boldsymbol{z}) \cdot \boldsymbol{J}_R(\boldsymbol{z}) d\sigma_{\partial B_i}(\boldsymbol{z}). \tag{74}$$

Thus, by successively treating every set B_i as reactant set and its complement $\cup_{j\neq i}B_j$ as product set, we can obtain all the rates $k_{i,j}$ using TPT. Assuming that requirements (i) and (ii) above are satisfied, the reduced dynamics where one maps the trajectory onto the index of the last metastable set B_j it visited can then be approximated by a continuous-time Markov chain with the rates given by (74).

5.7. **Generalization: discrete TPT.** It is straightforward to generalize TPT to discrete systems whose dynamics is described by a discrete- or a continuous-time Markov chain [44] (see also [72, 3]). Discrete TPT has become the tool of choice for analyzing the Markov State Models (MSMs) that have recently become popular to postprocess long MD simulation data data [55, 36, 40, 60, 59, 58, 13, 47, 50, 10]. In particular, like its continuous counterpart, discrete TPT gives expressions for the probability density and current of the reactive path, and thereby permits to calculate the equivalent of the flow lines of the current of reactive trajectories as well as the rate of the reaction. Often in these examples, the size of the chain, while big, remains small enough that the objects of TPT (in particular the committor functions) can be calculated directly using standard tools from numerical linear algebra and without having to make additional approximations. We refer the readers to [44] for details.

5.8. **Connection with transition state theory.** First, let us formulate TST in more general terms. TST computes the average frequency through a dividing surface S between A and B rather than the reaction rate between these sets themselves as in TPT. More precisely, let N_T^S be the number of times the trajectory crosses S from the side where A is to the one where B is in the interval $[0, T]$ (note that this number is, up to plus or minus 1, the same as the number of crossings of S in the other direction, and half the total number of crossings of S). Then the TST gives an exact expression for

$$\nu_{\text{TST}} = \lim_{T\to\infty} \frac{N_T^S}{T}. \tag{75}$$

To see how this limit can be computed, let $\theta(\boldsymbol{z})$ be such that $\theta(\boldsymbol{z}) < 0$ if $\boldsymbol{z}$ is on the side of S where A is, $\theta(\boldsymbol{z}) > 0$ if $\boldsymbol{z}$ is on the side where B is, and $\theta(\boldsymbol{z}) = 0$ if $\boldsymbol{z} \in S$. Then we can write the following exact counter for N_T^S:

$$N_T^S = \int_0^T H(\dot{\boldsymbol{z}}(t) \cdot \nabla\theta(\boldsymbol{z}(t))) \frac{d}{dt} H(\theta(\boldsymbol{z}(t)))dt \tag{76}$$

where $H(\theta)$ is the Heaviside theta function. Indeed the integral of $\frac{d}{dt}H(\theta(\boldsymbol{z}(t)))$ on any interval such that the trajectory crosses S once from the side where A is to the one where B is counts for 1, and only these crossings are accounted for because of the factor $H(\dot{\boldsymbol{z}}(t) \cdot$

$\nabla\theta(\boldsymbol{z}(t)))$. The limit in (75) can be computed exactly using the ergodicity assumption and the equality

$$\frac{d}{dt}H(\theta(\boldsymbol{z}(t))) = \dot{\boldsymbol{z}}(t)\cdot\nabla\theta(\boldsymbol{z}(t))\delta(\theta(\boldsymbol{z}(t)))$$

The calculation gives

$$\nu_{\text{TST}} = \int_\Omega (\tilde{\boldsymbol{b}}(\boldsymbol{z})\cdot\nabla\theta(\boldsymbol{z}))_+\delta(\theta(\boldsymbol{z}))\rho(\boldsymbol{z})d\boldsymbol{z} \tag{77}$$

where $(\tilde{\boldsymbol{b}}\cdot\nabla\theta)_+ = \max(\tilde{\boldsymbol{b}}\cdot\nabla\theta, 0)$ and $\tilde{\boldsymbol{b}}(\boldsymbol{z})$ is the modified drift with components[3]

$$\tilde{b}_i(\boldsymbol{z}) = b_i(\boldsymbol{z}) + \sum_{j,k=1}^{d}\frac{\partial\sigma_{i,k}(\boldsymbol{z})}{\partial z_j}\sigma_{j,k}(\boldsymbol{z}) \tag{78}$$

To derive (77) we had to assume that the unit normal to the surface does not span the direction in which the noise term in (8) acts, i.e. $\hat{\boldsymbol{n}}_S(\boldsymbol{z})\cdot\sigma(\boldsymbol{z})\dot{\boldsymbol{w}}(t) = 0$ or, equivalently, $\sigma^T(\boldsymbol{z})\hat{\boldsymbol{n}}_S(\boldsymbol{z}) = 0$. If that is not the case, then we cannot count the number of crossings of S by $\boldsymbol{z}(t)$ and $\nu_{\text{TST}} = \infty$. Dividing surfaces such that $\sigma^T\hat{\boldsymbol{n}}_S = 0$ exist for the Langevin equation (1): it suffices to take $S = \{(\boldsymbol{x},\boldsymbol{p}) : \theta(\boldsymbol{x}) = 0\}$ where $\theta(\boldsymbol{x})$ is a function of the positions but not the momenta, in which case (77) reduces to

$$\begin{aligned}\nu_{\text{TST}} &= Z_H^{-1}\int(\boldsymbol{p}^T m^{-1}\nabla\theta(\boldsymbol{x}))_+\delta(\theta(\boldsymbol{x}))e^{-\beta H(\boldsymbol{x},\boldsymbol{p})}d\boldsymbol{x}d\boldsymbol{p}\\ &= \sqrt{\frac{k_BT}{2\pi}}Z^{-1}\int_\Omega \left|m^{-1/2}\nabla\theta(\boldsymbol{x})\right|\delta(\theta(\boldsymbol{x}))e^{-\beta U(\boldsymbol{x})}d\boldsymbol{x}\end{aligned} \tag{79}$$

where $Z = \int e^{-\beta U(\boldsymbol{x})}d\boldsymbol{x}$ and we obtained the second equality by performing explicitly the integration over the momenta. However, surfaces such that $\sigma^T\hat{\boldsymbol{n}}_S = 0$ do not exist for the overdamped equation (4), i.e $\nu_{\text{TST}} = \infty$ in that case.

(77) is the standard expression for the TST rate. It can be re-expressed as a surface integral to make explicit that it does not depend on the way we parametrized S by $\theta(\boldsymbol{z})$:

$$\nu_{\text{TST}} = \int_S (\tilde{\boldsymbol{b}}(\boldsymbol{z})\cdot\hat{\boldsymbol{n}}_S(\boldsymbol{z}))_+\rho(\boldsymbol{z})d\sigma_S(\boldsymbol{z}). \tag{80}$$

Clearly $\nu_{\text{TST}} \geq \nu_R$ since every transition from A to B is associated to at least one one-sided crossing of S, but not every one-sided crossing of S leads to a transition from A to B. This means that ν_{TST} can be a poor approximation of ν_R if the reactive trajectories tend to cross several times the dividing surface. When that is the case, the standard jargon is to say that there are many recrossing events, and the ratio ν_R/ν_{TST} is what is called the transmission coefficient of the dividing surface. We will come back to this issue in section 6.2.1 where we will analyze when (and for which dividing surface) ν_{TST} is an accurate approximation of ν_R. This will also allow us to discuss the concept of transition state which is not readily apparent in the formulas above.

Within the TST context, it is also possible to account for the recrossing events and correct ν_{TST} to get ν_R. To do so, it suffices replace the counter in (76) by the following counter for N_T:

$$N_T = \int_0^T \mathbf{1}_R(t)\frac{d}{dt}H(\theta(\boldsymbol{z}(t)))dt \tag{81}$$

[3]The modified drift $\tilde{\boldsymbol{b}}(\boldsymbol{z})$ rather than $\boldsymbol{b}(\boldsymbol{z})$ enters (77) because the calculation requires to reinterpret (8) in Stratonovich sense, in which case the drift term becomes $\tilde{\boldsymbol{b}}(\boldsymbol{z})$. Notice however that $\tilde{\boldsymbol{b}}(\boldsymbol{z}) = \boldsymbol{b}(\boldsymbol{z})$ if σ is independent of $\boldsymbol{z}$, e.g. like for the Langevin equation.

Indeed the factor $\mathbf{1}_R(t)$ permits to focus on the reactive trajectories, and every such trajectory crosses S one time more going from the side where A is to the one where B is than the other way around. As a result, the integral of $\frac{d}{dt}H(\theta(\boldsymbol{z}(t)))$ along each reactive trajectory gives exactly 1. Dividing N_T by T and letting $T \to \infty$, (81) gives the following expression for ν_R:

$$\begin{aligned}\nu_R &= \int_\Omega \tilde{\boldsymbol{b}}(\boldsymbol{z}) \cdot \nabla\theta(\boldsymbol{z})\delta(\theta(\boldsymbol{z}))q_+(\boldsymbol{z})q_-(\boldsymbol{z})\rho(\boldsymbol{z})d\boldsymbol{z} \\ &= \int_S \tilde{\boldsymbol{b}}(\boldsymbol{z}) \cdot \hat{\boldsymbol{n}}_S(\boldsymbol{z})q_+(\boldsymbol{z})q_-(\boldsymbol{z})\rho(\boldsymbol{z})d\sigma_S(\boldsymbol{z}).\end{aligned} \tag{82}$$

This expression is less general than the TPT expressions (65) and (66) (recall that the normal to S cannot spend the direction in which the noise act for (82) to be valid), but it can be seen that (82) is equivalent to (65) in that case. Indeed, using (53) and $\sigma^T\hat{\boldsymbol{n}}_S = 0$, we see that

$$\hat{\boldsymbol{n}}_S \cdot \boldsymbol{J}_R = q_+q_-\hat{\boldsymbol{n}}_S \cdot \boldsymbol{J} = q_+q_-\hat{\boldsymbol{n}}_S \cdot \tilde{\boldsymbol{b}}\rho \tag{83}$$

which is precisely the integrand in (82).

The fact that the TST expression for ν_R is less general than that of TPT is not the most important limitation of TST, however. Its most important limitation is that, unlike TPT, TST gives no information about the reaction other than its rate.

6. Asymptotic solutions based on TPT

6.1. **Transition tubes.** TPT indicates that the committor functions are key objects that characterize the mechanism of a reaction since they determine both the probability density and the current of reactive trajectories. A main issue then is how to estimate these functions and their gradients in systems of actual interest. As already mentioned in section 5.2 the main difficulty stems from the fact that *such systems are typically high dimensional*, and therefore standard numerical methods such as finite difference or finite element are ineffective. Here we discuss one possible way of approximating the committor function which is the basis of the string method that we shall discuss in sections 7, 8 and 9. It also allows to establish a rationale behind the objects calculated by other chain of state methods.

For simplicity, we will discuss first the case of the overdamped dynamics:

$$\dot{\boldsymbol{x}} = -\nabla U(\boldsymbol{x}) + \sqrt{2k_BT}\dot{\boldsymbol{w}}(t) \tag{84}$$

This is (4) in which we set $m_i = \gamma = 1$; this simply amounts to using mass-weighted coordinates, i.e. use $m^{1/2}\boldsymbol{x}$ as new positions and $\gamma^{-1}t$ as new time. Later we will translate our results to the original variables and explain how to generalize our results to the Langevin equation (1). The committor functions associated with (84) are such that $q_+(\boldsymbol{x}) = q(\boldsymbol{x})$ and $q_-(\boldsymbol{x}) = 1 - q(\boldsymbol{x})$ where $q(\boldsymbol{x})$ satisfies

$$\begin{cases} -\nabla U \cdot \nabla q + k_BT\Delta q = 0, & \boldsymbol{x} \notin A \cup B, \\ q = 0, & \boldsymbol{x} \in A, \\ q = 1, & \boldsymbol{x} \in B, \end{cases} \tag{85}$$

To proceed we shall make two main assumptions: (i) most of the flux of reactive trajectories goes through one or a few isolated and localized tubes and (ii) in these tubes, the isosurfaces of $q(\boldsymbol{x})$ are locally planar. To avoid confusions, we stress that the localized tube assumption is about the flux of reactive trajectories and not the reactive trajectories themselves: *the flux may very well be localized in a few tubes even in situations where the reactive trajectories are not*. Under these two assumptions, the function $q(\boldsymbol{x})$ can locally

be approximated using its values along a suitably chosen curve – the value of $q(\boldsymbol{x})$ at any point $\boldsymbol{x}$ outside the curve is the same as its value at the point closest to $\boldsymbol{x}$ on the curve. The main question then becomes where to place the curve(s) in such a way that it lies at the center of the reaction tube(s) carrying most of the flux of reactive trajectories. Before addressing this question, let us make precise the concept of reaction tube as well as the locally planar isocommittor surface ansatz for $q(\boldsymbol{x})$.

Defining transition tubes via the flux density across the isocommittor surfaces. The current of reactive trajectories associated with (84) is

$$\boldsymbol{J}_R(\boldsymbol{x}) = Z^{-1} e^{-\beta U(\boldsymbol{x})} \nabla q(\boldsymbol{x}) \tag{86}$$

Suppose that we look at the flux induced by this current through a given set of surfaces that foliate space between A and B, i.e. such that they each are a dividing surface, they do not intersect, and their union is the region between A and B. Suppose in addition that there exists regions where this flux is localized in each surface (e.g carry a certain percentage of the total flux through each surface). Then the ensemble of these regions will form one or more of the transition tubes carrying this percentage of the flux of reaction trajectories.

Which surfaces should we use in this construction? The correct ones are the isocommittor surfaces $\{q(\boldsymbol{x}) = q^*\}$ with $q^* \in [0,1]$. These surfaces form a foliation. In addition, they have the remarkable property that the flux intensity through these surfaces, i.e. $j_R(\boldsymbol{x}) = \hat{\boldsymbol{n}} \cdot \boldsymbol{J}_R = \nabla q \cdot \boldsymbol{J}_R / |\nabla q|$ or explicitly from (86)

$$j_R(\boldsymbol{x}) = Z^{-1} e^{-\beta U(\boldsymbol{x})} |\nabla q(\boldsymbol{x})| \tag{87}$$

is, up to a normalization constant, proportional to the probability density of the last hitting point of the reactive trajectories in the surface. This property was established in [71] and its proof is quite technical, but it can be understood as follows. Consider the reactions from the extended reactant sets $\mathcal{A}(q^*) = \{\boldsymbol{x} : q(\boldsymbol{x}) \leq q^*\}$, where $q^* \in [0,1]$, to the product set B. Because the isocommittor surfaces foliate space between A and B, the extended reactant sets are nested into one another, i.e. $\mathcal{A}(q) \subset \mathcal{A}(q')$ if $q < q'$, and it is easy to see that the committor functions for the reaction from $\mathcal{A}(q^*)$ to B is simply

$$\tilde{q}_{q^*}(\boldsymbol{x}) = \frac{q(\boldsymbol{x}) - q^*}{1 - q^*} \tag{88}$$

since this function solves (85) in the region between $\mathcal{A}(q^*)$ and B with the right boundary conditions. In particular, the flowlines of the current of the reaction from the extended set $\mathcal{A}(q^*)$ to B are the same as those of the reaction from A to B past the isosurface $q(\boldsymbol{x}) = q^*$, and the flux they carry is simply proportional in both cases. This is to say, the reaction from A to B past the isosurface $q(\boldsymbol{x}) = q^*$ happens exactly as the reaction from $\mathcal{A}(q^*)$ to B and what happened before $q(\boldsymbol{x}) = q^*$ has become irrelevant. This is why it is natural to look at the flux of reactive trajectories out of these extended reactant sets, i.e. across the isocommittor surfaces, to define the transition tubes.

Approximations of the committor function. Suppose that we connect the center of the regions where a transition tube intersect the isocommittor surfaces. This defines a curve γ. Let us parametrize the points along this curve by $\boldsymbol{\varphi}(s)$ with $s \in [0, L]$: here s is the arclength along γ and L is the length of this curve so that $|\boldsymbol{\varphi}'(s)| = 1$ (later it will be more convenient to use other parametrizations, e.g. based on normalized arclength, but this one is simpler for the calculations in this section). Define the function $s : \Omega \to [0, L]$ via

$$s(\boldsymbol{x}) = \arg\min_s |\boldsymbol{\varphi}(s) - \boldsymbol{x}| \tag{89}$$

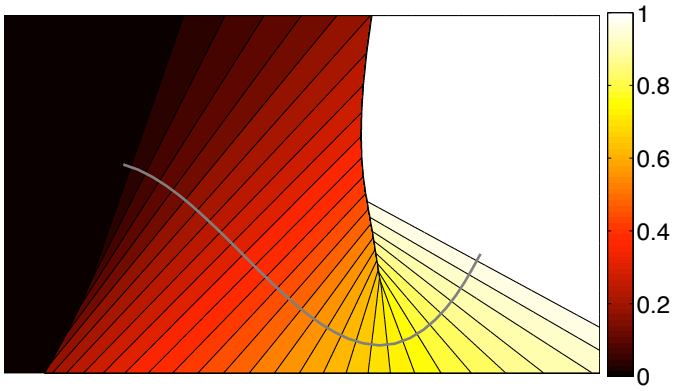

FIGURE 7. Isosurfaces of the function $s(\boldsymbol{x})$ defined in (89) and associated with the unit length ($L = 1$) curve shown in gray. The ansatz (91) amounts to assuming that locally around such a curve the isosurfaces of $s(\boldsymbol{x})$ and those of $q(\boldsymbol{x})$ coincide.

The value of $s(\boldsymbol{x})$ at point $\boldsymbol{x}$ is the arclength of the point $\boldsymbol{\varphi}(s(\boldsymbol{x}))$ along the curve which is the closest to $\boldsymbol{x}$. The isosurfaces of $s(\boldsymbol{x})$ are subsets of planes which are perpendicular to the curve at point $\boldsymbol{\varphi}(s(\boldsymbol{x}))$, see figure 7 for an illustration. The function $s(\boldsymbol{x})$ can also be characterized by the equation

$$\boldsymbol{\varphi}'(s(\boldsymbol{x})) \cdot (\boldsymbol{\varphi}(s(\boldsymbol{x})) - \boldsymbol{x}) = 0. \tag{90}$$

Assuming that the isocommittor surfaces are locally planar around the curve γ at the center of a transition tube is then equivalent to assuming that locally in this tube $q(\boldsymbol{x})$ can be represented as

$$q(\boldsymbol{x}) = f(s(\boldsymbol{x})) \tag{91}$$

In words, the function $f : [0, L] \to [0, 1]$ specifies the value of $q(\boldsymbol{x})$ along the curve, and the value of $q(\boldsymbol{x})$ at any point $\boldsymbol{x}$ not on the curve is the same as the one at the point on the curve closest to $\boldsymbol{x}$, $\boldsymbol{\varphi}(s(\boldsymbol{x}))$. (91) implies that the isosurfaces of $q(\boldsymbol{x})$ are the same as the ones of $s(\boldsymbol{x})$, up to relabeling. Note that (91) is supposed to hold only in the transition tube, i.e. only locally in the vicinity of the curve γ.

The catch of course is that we do not know the isocommittor surfaces beforehand, so to proceed we will have to turn the construction above upside down and identify the curve γ first, then deduce the structure of the committor function locally around this curve via the ansatz (91). This is what we do next in sections 6.2, 6.3 and 6.6 under various assumptions. As we will see this will also allow us to deduce the value of the function $f(s)$ in (91).

6.2. **The minimum energy path.** In this section we will assume that there exists one (or more generally a few) line of current γ that carries most of the flux of reactive trajectories (i.e. the transition tubes are very thin). Then, following the construction outlined in 6.1, this line of current must be such that it maximizes $j_R(\boldsymbol{x})$ in each isosurface of $q(\boldsymbol{x})$. Using the ansatz (91), which implies that $\nabla q(\boldsymbol{x}) = f'(s(\boldsymbol{x}))\nabla s(\boldsymbol{x})$, in (87) this is equivalent to the requirement that

$$\text{In } s(\boldsymbol{x}) = s, \;\; Z^{-1} e^{-\beta U(\boldsymbol{x})} f'(s) |\nabla s(\boldsymbol{x})| \;\; \text{must be maximum at } \boldsymbol{x} = \boldsymbol{\varphi}(s) \tag{92}$$

This also means that the gradient of $j_R(\boldsymbol{x})$ must be parallel to $\nabla s(\boldsymbol{x})$ along γ, i.e. (neglecting factors that are constants in $s(\boldsymbol{x}) = s$)

$$\nabla \left(e^{-\beta U(\boldsymbol{\varphi})} |\nabla s(\boldsymbol{\varphi})| \right) \text{ must be parallel to } \nabla s(\boldsymbol{\varphi}) \text{ along } \gamma \tag{93}$$

To make this equation explicit, take the gradient of (90) with respect to $\boldsymbol{x}$ to obtain

$$\nabla s(\boldsymbol{x}) \left(1 + \boldsymbol{\varphi}''(s(\boldsymbol{x})) \cdot (\boldsymbol{\varphi}(s(\boldsymbol{x})) - \boldsymbol{x})\right) = \boldsymbol{\varphi}'(s(\boldsymbol{x})) \tag{94}$$

where we used $|\boldsymbol{\varphi}'| = 1$. Note that if $\boldsymbol{x}$ is a point along the curve, $\boldsymbol{x} = \boldsymbol{\varphi}(s)$ and (94) reduces to

$$\nabla s(\boldsymbol{\varphi}) = \boldsymbol{\varphi}'(s) \tag{95}$$

(94) implies that

$$\nabla \Big(e^{-\beta U(\boldsymbol{x})} |\nabla s(\boldsymbol{x})| \Big) = e^{-\beta U(\boldsymbol{x})} |\nabla s(\boldsymbol{x})| \left(-\beta \nabla U(\boldsymbol{x}) + |\nabla s(\boldsymbol{x})| \boldsymbol{\varphi}''(s(\boldsymbol{x}))\right) \tag{96}$$

Along the curve we have $|\nabla s(\boldsymbol{\varphi})| = 1$ from (95) and (96) reduces to

$$\nabla \Big(e^{-\beta U(\boldsymbol{\varphi})} |\nabla s(\boldsymbol{\varphi})| \Big) = e^{-\beta U(\boldsymbol{\varphi})} \left(-\beta \nabla U(\boldsymbol{\varphi}) + \boldsymbol{\varphi}''\right) \tag{97}$$

Inserting (95) and (97) in (93) gives the condition

$$-\beta \nabla U(\boldsymbol{\varphi}) + \boldsymbol{\varphi}'' \text{ must be parallel to } \boldsymbol{\varphi}' \text{ along } \gamma \tag{98}$$

Equivalently this can be written as

$$-\beta [\nabla U]^{\perp} + \boldsymbol{\varphi}'' = 0 \tag{99}$$

where $[\nabla U]^{\perp} = \nabla U - (\nabla U \cdot \boldsymbol{\varphi}')\boldsymbol{\varphi}'$ is the component of ∇U perpendicular to γ and we used $[\boldsymbol{\varphi}'']^{\perp} = \boldsymbol{\varphi}''$. Condition (99) is well-known: it is the equation for the path of max-flux originally derived in [4] using a different approach. As we will see next in section **??**, when the most of the flux of reactive trajectory is indeed carried by the flowline satisfying (99), the term $\boldsymbol{\varphi}''$ is a small correction compared to $-\beta[\nabla U]^{\perp}$ and therefore (99) can be approximated by

$$[\nabla U]^{\perp} = 0 \tag{100}$$

We recover the equation for the minimum energy path (MEP) defined earlier. Algorithms to solve (99) and (100) will be discussed in section 7.

What about the function f? Let us start from expression (65) for the rate ν_R using the isocommittor surface $q(\boldsymbol{x}) = f(s)$ as particular dividing surface:

$$\nu_R = \int_{q(\boldsymbol{x})=f(s)} \hat{\boldsymbol{n}}(\boldsymbol{x}) \cdot \boldsymbol{J}_R(\boldsymbol{x}) d\sigma(\boldsymbol{x}) \tag{101}$$

where $\hat{\boldsymbol{n}}(\boldsymbol{x})$ denotes the unit normal to $q(\boldsymbol{x}) = f(s)$ and $d\sigma(\boldsymbol{x})$ the surface element in this surface. Using (86) as well as the ansatz (91) and rewriting the surface integral as a volume integral using the Dirac delta function gives

$$\nu_R = f'(s) \int |\nabla s(\boldsymbol{x})|^2 Z^{-1} e^{-\beta U(\boldsymbol{x})} \delta(s(\boldsymbol{x}) - s) d\boldsymbol{x} \tag{102}$$

By our assumption that most of the flux of reactive trajectories goes through a thin tube around γ the integral in (102) is dominated by a small region around γ. This means that we can approximate $|\nabla s(\boldsymbol{x})|^2$ by its value along γ, i.e. by $|\boldsymbol{\varphi}'|^2 = 1$ according to (95), and get

$$\nu_R = f'(s) e^{-\beta F(s)} \tag{103}$$

where we defined the free energy of $s(\boldsymbol{x})$:

$$F(s) = -k_B T \ln \int Z^{-1} e^{-\beta U(\boldsymbol{x})} \delta(s(\boldsymbol{x}) - s) d\boldsymbol{x} \tag{104}$$

Since ν_R is independent of s, differentiating (103) with respect to s give the following second order ordinary differential equation for f:

$$\left(f' e^{-\beta F}\right)' = 0, \tag{105}$$

to be solved with the boundary conditions $f(0) = 0$, $f(L) = 1$. This gives

$$f(s) = \frac{\int_0^s e^{\beta F(s')} ds'}{\int_0^L e^{\beta F(s')} ds'}. \tag{106}$$

Note also that inserting back this expression in (103) implies that

$$\nu_R = \Big(\int_0^L e^{\beta F(s')} ds' \Big)^{-1} \tag{107}$$

In terms of the non-mass-weighted coordinates and the original units of time, (104) remain valid but (106) must be replaced by

$$f(s) = \frac{\int_0^s |m^{-1/2} \boldsymbol{\varphi}'(s')|^2 e^{\beta F(s')} ds'}{\int_0^L |m^{-1/2} \boldsymbol{\varphi}'(s')|^2 e^{\beta F(s')} ds'}. \tag{108}$$

and we need to multiply (107) by γ^{-1}. Algorithms to calculate the free energy $F(s)$ will be discussed in section 9.1 in the context of the finite temperature string method.

6.2.1. *Connection with variational transition state theory.* Recall that for Langevin dynamics, the TST rate is given by (79)[4]. In that formula, the dividing surface is arbitrary. Since, as was explained in section 5.8, ν_{TST} always overestimates the actual reaction rate ν_R, a natural thing to do is to optimize the dividing surface so as to minimize the TST rate. This is what is referred to as variational transition state theory (VTST) and it amounts to viewing (79) as an objective function to be minimized over all possible dividing surfaces. The Euler-Lagrange equation for this minimization problem can be written down explicitly [34, 67] (for simplicity we go back to mass-weighted coordinates)

$$-\beta \nabla U \cdot \hat{\boldsymbol{n}}_S + \kappa_S = 0 \tag{109}$$

where $\hat{\boldsymbol{n}}_S$ is the unit normal to the surface and $\kappa_S = \nabla \cdot \hat{\boldsymbol{n}}_S$ its local curvature, and we must look for a solution of (109) that leaves the reactant set on one side and the product set on the other. (109) is a complicated equation. It becomes much simpler, however, if we assume that the curvature term in this equation is small and can be neglected. In this case, (109) reduces to

$$\nabla U \cdot \hat{\boldsymbol{n}}_S = 0 \tag{110}$$

The solutions are the stable manifold of any saddle point, i.e. the set of all initial conditions such that the solutions of $\dot{\boldsymbol{x}} = -\nabla U(\boldsymbol{x})$ converge to the saddle point as $t \to \infty$. This set forms a dividing surface.

The above considerations show why the MEP is relevant also in the context of VTST. Indeed, having identified a MEP between two minima, we can localize the saddle point along it, and thereby calculate at least locally (e.g. by using a planar approximation) the

[4]Note that this formula is independent of the friction coefficient γ, so we can in principle take the overdamped limit as $\gamma \gg 1$ without changing the result. There is a catch, of course: while ν_{TST} is independent of γ, ν_R is not, which also means that how much recrossings events there are depends on γ in a way that we cannot predict *a priori*. Since we are primarily interested in ν_R, not ν_{TST}, this is a problem.

optimal dividing surface according to (110). The question, however, is when is this a good approximation? Clearly, we again need the potential to be smooth rather than rugged since otherwise there would be too many critical points in the system. In fact, to justify neglecting the curvature term in (109) and to derive (110), we need to be in a situation where LDT applies: the potential $U(\boldsymbol{x})$ must have only a few critical points, among which a few local minima that are all separated by energy barriers that are much larger than $k_B T$. As we have already mentioned, such situations are quite restrictive.

Finally, notice that even the TST rate associated with the optimal dividing surface may be a poor approximation of ν_R if there are substantial recrossings. These problems are overcome in TPT since TPT considers a family of isocommittor surfaces that foliates the configuration (or phase) space between the reactant and product sets, not just a single dividing surface.

6.3. **Working with collective variables: the minimum free energy path.** As we have explained in section 6.2, a main difficulty in realistic examples is that their potential energy surface is typically rugged, and when that is the case, it becomes the wrong question to look for most probable transition paths. One way to go around this problem is to introduce some appropriate collective variables, i.e. a set of functions $\theta_i : \Omega \to \mathbb{R}$ with $i = 1, \ldots, D$, and try to reformulate the problem in terms of these variables rather than the original ones. As we will see in a moment, this amounts to analyzing the reaction using the free energy landscape associated with the variables $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D)$ rather than the original energy landscape $U(\boldsymbol{x})$. If the collective variables are well-chosen (which is a nontrivial issue that we will come back to later) and $D \ll d$, we may expect that the free energy landscape will be smooth even if $U(\boldsymbol{x})$ is rugged, and entropic effects will be less of an issue as well. Note that there is an obvious similarity to the notion of reaction coordinates that are commonly used in the literature. Here we use "collective variables" for two reasons: The first is that "reaction coordinates" is by now a very much abused terminology. The second is that reaction coordinates usually involve one or very few variables, whereas collective variables can live in rather high dimensions, i.e. D can still be very large even if it is smaller than d.

How can we formalize the idea of analyzing the reaction by means of the collective variables? Let us answer this question first in the context of the overdamped equation (1) (in this section we will work with the original non-mass-weighed coordinates) and consider the Langevin equation later in section 6.5. A set of collective variables $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_D)$ is adequate to describe the reaction if the committor function of the reaction can be parametrized using these variables, i.e.

$$q(\boldsymbol{x}) \approx Q(\boldsymbol{\theta}(\boldsymbol{x})) \tag{111}$$

for some suitable $Q(\boldsymbol{\theta})$. Inserting this ansatz into the objective function (46) for $q(\boldsymbol{x})$, it is easy to see that this objective function can be re-expressed as

$$I(Q) = \int_{\mathbb{R}^D} \sum_{i,j=1}^{D} M_{i,j}(\boldsymbol{\theta}) \frac{\partial Q}{\partial \theta_i} \frac{\partial Q}{\partial \theta_j} e^{-\beta G(\boldsymbol{\theta})} d\boldsymbol{\theta} \tag{112}$$

where we defined the free energy associated with the variables $\boldsymbol{\theta}$,

$$G(\boldsymbol{\theta}^*) = -k_B T \ln \left(Z^{-1} \int_{\mathbb{R}^d} e^{-\beta U(\boldsymbol{x})} \prod_{i=1}^{D} \delta(\theta_i(\boldsymbol{x}) - \theta_i^*) d\boldsymbol{x} \right) \tag{113}$$

and the metric tensor $M_{i,j}(\boldsymbol{\theta})$

$$M_{i,j}(\boldsymbol{\theta}^*) = k_B T \gamma^{-1} \left\langle \sum_{k=1}^{d} m_k^{-1} \frac{\partial \theta_i}{\partial x_k} \frac{\partial \theta_j}{\partial x_k} \right\rangle_{\boldsymbol{\theta}(\boldsymbol{x})=\boldsymbol{\theta}^*} \tag{114}$$

where $\langle \cdot \rangle_{\boldsymbol{\theta}(\boldsymbol{x})=\boldsymbol{\theta}^*}$ denotes canonical expectation conditional on $\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{\theta}^*$. The Euler Lagrange equation associated with the minimization of (112) is

$$\begin{cases} -\beta M \nabla_{\boldsymbol{\theta}} G \cdot \nabla_{\boldsymbol{\theta}} Q + \mathrm{div}_{\boldsymbol{\theta}} M \cdot \nabla_{\boldsymbol{\theta}} Q + M : \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} Q = 0, & \boldsymbol{\theta} \notin \mathcal{A} \cup \mathcal{B} \\ Q = 0, & \boldsymbol{\theta} \in \mathcal{A} \\ Q = 1, & \boldsymbol{\theta} \in \mathcal{B} \end{cases} \tag{115}$$

where $\mathcal{A}$ and $\mathcal{B}$ are the projections to the collective variable space of the reactant set A and the product set B, respectively.

Remarkably, (115) is the equation for the (forward) committor of a reaction associated with the following evolution equation for the collective variables $\boldsymbol{\theta}$:

$$\dot{\boldsymbol{\theta}} = -\beta M \nabla_{\boldsymbol{\theta}} G + \mathrm{div}_{\boldsymbol{\theta}} M + \sqrt{2}\, M^{1/2}\, \dot{\boldsymbol{w}} \tag{116}$$

which is similar to the original overdamped equation (4), except that the dimensionality of space is reduced (D instead of d) and the friction has become space-dependent – the metric tensor $M(\boldsymbol{\theta})$ plays the role of $k_B T \gamma^{-1} m^{-1}$ in (4) and it depends on $\boldsymbol{\theta}$ in general because of the curvilinear nature of the coordinates $\boldsymbol{\theta}$. Therefore, using TPT, we can define the probability density of reactive trajectories associated with (116), their probability current, etc. and get information about the original reaction under the assumption that (111) holds.

This also means that we can generalize the construction made in sections 6.1 and 6.2 to (116). The equivalent of the ansatz (91) is obtained by picking a curve γ in collective variable space, parametrizing it by $\boldsymbol{\varphi}(s)$ as before (but with $\boldsymbol{\varphi}(s) \in \mathbb{R}^D$), defining the equivalent of the function $s(\boldsymbol{x})$, i.e. the function $s(\boldsymbol{\theta})$ specified by

$$M^{-1}(\boldsymbol{\varphi}(s(\boldsymbol{\theta})))\boldsymbol{\varphi}'(s(\boldsymbol{\theta})) \cdot (\boldsymbol{\varphi}'(s(\boldsymbol{\theta})) - \boldsymbol{\theta}) = 0, \tag{117}$$

and assuming that

$$Q(\boldsymbol{\theta}) = f(s(\boldsymbol{\theta})) \tag{118}$$

As in section 6.2, under the assumption that most of the flux of reactive trajectories associated with (116) are carried by the thin tubular neighborhood of one or a few isolated curves, we can identify these curves using the look-forward property. The only additional difficulty is the presence of the tensor $M(\boldsymbol{\theta})$, but this difficulty can be handled easily by an appropriate change of metric. Skipping the details, the equivalent of (**??**) is (assuming for simplicity that $M(\boldsymbol{\theta})$ is a slowly varying function of $\boldsymbol{\theta}$)

$$-\beta [M \nabla_\theta G]^\perp + \frac{\boldsymbol{\varphi}''}{|M^{-1/2}\boldsymbol{\varphi}'|} = 0, \tag{119}$$

the equivalent of (**??**) is

$$[M \nabla_{\boldsymbol{\theta}} G]^\perp = 0 \tag{120}$$

and the equivalent of (108) is

$$f(s) = \frac{\int_0^s |M^{1/2}(\boldsymbol{\varphi}(s))\boldsymbol{\varphi}'(s')|^2 e^{\beta F(s')} ds'}{\int_0^L |M^{1/2}(\boldsymbol{\varphi}(s))\boldsymbol{\varphi}'(s')|^2 e^{\beta F(s')} ds'}. \tag{121}$$

where $F(s)$ is the free energy of $s(\boldsymbol{\theta})$

$$\begin{aligned} F(s^*) &= -k_B T \ln \int e^{-\beta G(\boldsymbol{\theta})} \delta(s(\boldsymbol{\theta}) - s^*) d\boldsymbol{\theta} \\ &= -k_B T \ln \int Z^{-1} e^{-\beta U(\boldsymbol{x})} \delta(s(\boldsymbol{\theta}(\boldsymbol{x})) - s^*) d\boldsymbol{x} \end{aligned} \tag{122}$$

Equation (119) defines the path of maxflux in collective variable space whereas equation (120) defines the minimum free energy path (MFEP). As before, when the assumptions leading to (119) are satisfied, the second term in this equation is a small correction, i.e. its solution can be approximated by the MFEP. Algorithms for identifying MFEPs will be presented in section 8.

6.4. **What is a good set of collective variables?** For the MFEP solution of (120) to be relevant, clearly we need to be in a situation where LDT applies to (116). This requires that the free energy $G(\boldsymbol{\theta})$ has only a few critical point, among which a few local minima that are all separated by free energy barriers that are much larger than $k_B T$. In other words, $G(\boldsymbol{\theta})$ must be smooth which, as explained before, is a less stringent requirement than asking that $U(\boldsymbol{x})$ is smooth.

Unfortunately, the smoothness of $G(\boldsymbol{\theta})$ is necessary but not sufficient to validate the MFEP. Indeed we also need that the manipulations that led to (116) be justified which, ultimately, amounts to justifying the approximation (111). This is a difficult question to which a definite answer is not yet available. We should, however, dispel a possible source of confusion: good collective variables (in the sense that (111) holds) are *not* necessarily slow variables. At first sight, we may think that they should be slow, but the following example shows that this is not necessarily the case and that slow variables can even be the wrong ones to describe a reaction. Consider the two-dimensional overdamped system governed by

$$\begin{cases} \gamma_x \dot{x} = -\partial_x U(x,y) + \sqrt{2k_B T \gamma_x}\, \dot{w}_x(t) \\ \gamma_y \dot{y} = -\partial_y U(x,y) + \sqrt{2k_B T \gamma_y}\, \dot{w}_y(t) \end{cases} \tag{123}$$

where $U(x,y) = (1-x^2)^2 + y^2$. Assume that $\gamma_y \ll \gamma_x$. This means that x evolves much faster than y. At the same time, there is metastability in the x-direction and this is also the direction in which the reaction occurs. This is not a paradox: while x is faster than y, there is a hidden slow time scale in the x variable related to the hopping over the barrier, and if the barrier ΔU is high enough that $\gamma_x^{-1} e^{-\beta \Delta U} \gg \gamma_y^{-1} \gg \gamma_x^{-1}$, the slow time scale associated with the transition will be the slowest in the system. Therefore $\theta(x,y) = x$ is a good collective variable for describing the reaction (indeed $q(x,y) \approx q(x)$ for this reaction), but it is not a slow variable in the usual sense. This example illustrates reactive events are rare because they involve many failed attempts, each of which can be pretty fast, as the ones that do succeed.

The example above is specific, but the message it carries is quite general. To see why, just think about picking $q(\boldsymbol{x})$ itself as single collective variable, i.e. use $\theta(\boldsymbol{x}) = q(\boldsymbol{x})$. Then, obviously, (111) is exact with Q being the identity function. But $q(\boldsymbol{x})$ is *not* a slow variable in general, in the sense that it does not satisfy a closed equation. In fact, $q(\boldsymbol{x})$ satisfies

$$\dot{q} = \sqrt{2k_B T \gamma^{-1}} m^{-1/2} \nabla q \cdot \dot{\boldsymbol{w}}(t), \tag{124}$$

which is not closed (and cannot be closed in general because ∇q is not a function of q). Note that the equivalent of (116) in this case is the averaged version of (124)

$$\dot{q} = \sqrt{2k_B T \gamma^{-1}} \sigma(q) \dot{w}(t) \tag{125}$$

where $\sigma(q^*) = \sqrt{\langle m^{-1}\nabla q \cdot \nabla q\rangle_{q(\boldsymbol{x})=q^*}}$. This raises the question: what is the range of validity of the dynamical equation (116)? The above considerations shows that it can be used to analyze the mechanism of a reaction, but it does not necessarily capture the true dynamics of the collective variables. Again, this is not a paradox: the reactive trajectories from different dynamics can have the same statistical properties.

6.5. **The case of the Langevin equation.** Generalizing the ideas of the previous sections to the Langevin equation (1) poses an additional difficulty, namely that the equivalent of the objective function (46) does not exist for $q_+(\boldsymbol{x},\boldsymbol{p})$. This is related to the fact that, unlike the overdamped dynamics, the Langevin dynamics is not statistically invariant under time reversal (invariance requires to also flip the momenta). One way to go around this difficulty is to replace (46) by the following least square principle in which we minimizes

$$\tilde{I}(q_+) = Z_H^{-1}\int |Lq_+|^2 e^{-\beta H(\boldsymbol{x},\boldsymbol{p})} d\boldsymbol{x}d\boldsymbol{p} \tag{126}$$

where L is the generator of the Langevin equation:

$$Lq_+ = m^{-1}\boldsymbol{p}\cdot\nabla_{\boldsymbol{x}}q_+ - \nabla U\cdot\nabla_{\boldsymbol{p}}q_+ - \gamma\boldsymbol{p}\cdot\nabla_{\boldsymbol{p}}q_+ + \gamma k_B T m : \nabla_{\boldsymbol{p}}\nabla_{\boldsymbol{p}}q_+ \tag{127}$$

Clearly, the solution of $Lq_+ = 0$ is a minimizer of (126). (126) can be used as the starting point for further approximations. One such approximation is based on the assumption that the committor function $q_+(\boldsymbol{x},\boldsymbol{p})$ can be represented by a function of the positions only, i.e.

$$q_+(\boldsymbol{x},\boldsymbol{p}) \approx q(\boldsymbol{x}) = \langle q_+(\boldsymbol{x},\boldsymbol{p})\rangle_{\boldsymbol{p}} \tag{128}$$

If one inserts this ansatz into (126) and performs explicitly the integration over the momenta, it is easy to see that $\tilde{I}(q)$ reduces *exactly* to the objective function (46) that arises in the context of the overdamped equation (4). As explained in section 6.3, this does not mean that the overdamped equation necessarily captures the actual dynamics of the system. It simply means that it can be used to explain the mechanism of the reaction if one assumes that (128) holds. Under this assumption, we can use all the technology developed in sections 6.2 and 6.3 to analyze the reaction either via identifying the MEPs (if appropriate), or MFEPs (if appropriate). Under this assumption, we can also use the procedure explained next in section 6.6 when neither the MEP nor the MFEP are the relevant object.

6.6. **Going beyond the thin tube assumption via principal curves.** One main difficulty with the MEP is that it gives a viewpoint on the reaction that is too local. Indeed by trying to identify a single line of current that carries most of the flux of reactive trajectories, one ignores reaction channels that are less favorable energetically, but are broader and may indeed carry more flux. The same criticism may apply to the MFEP, though to a less extend.

In order to go beyond the concepts of MEP or MFEP, we must look at the problem more globally. One possibility, advocated in [20, 54, 21, 70], is as follows. Instead of trying to identify the point that maximizes the current intensity on each isocommittor surfaces, as we did in section 6.2 to get the MEP, we can look for the position in these surfaces which is centroidal with respect to the current intensity $j_R(\boldsymbol{x})$ defined in (87). In other words, assuming that (91) holds, we can look for $\boldsymbol{\varphi}(s)$ that solves for every $s\in[0,L]$

$$\begin{aligned} 0 &= \int_{s(\boldsymbol{x})=s} (\boldsymbol{x}-\boldsymbol{\varphi}(s)) j_R(\boldsymbol{x}) d\sigma(\boldsymbol{x}) \\ &\equiv Z^{-1} f'(s) \int_{s(\boldsymbol{x})=s} e^{-\beta U(\boldsymbol{x})} (\boldsymbol{x}-\boldsymbol{\varphi}(s)) |\nabla s(\boldsymbol{x})|^2 \delta(s(\boldsymbol{x})-s) d\boldsymbol{x} \end{aligned} \tag{129}$$

where $d\sigma(\boldsymbol{x})$ denotes the surface element in $s(\boldsymbol{x}) = s$ and we used $|\nabla q| = f'(s)|\nabla s|$. If we neglect curvature effects and approximate $|\nabla s|$ by its value $|\boldsymbol{\varphi}'| = 1$ along the curve (see (95)), up to an irrelevant proportionality factor (129) reduces to the following equation for $\boldsymbol{\varphi}(s)$

$$0 = \int (\boldsymbol{x} - \boldsymbol{\varphi}(s)) e^{-\beta U(\boldsymbol{x})} \delta(s(\boldsymbol{x}) - s) d\boldsymbol{x} \tag{130}$$

(130) means that the conditional canonical expectation of $\boldsymbol{x}$ in $s(\boldsymbol{x}) = s$ must be $\boldsymbol{\varphi}(s)$ for every $s \in [0, L]$, i.e.

$$\boldsymbol{\varphi}(s) = \langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=s} \tag{131}$$

The curve satisfying this requirement is an object familiar in the statistics literature [29, 30] called the *principal curve* associated with the density $Z^{-1} e^{-\beta U(\boldsymbol{x})}$. The main advantage of the principal curve over the MEP is that it permits to look more globally at the flux going through each isocommittor surface. Indeed, the principal curve γ identified via (131) is not anymore meant to be a single line of current that carries most of the flux, but rather the center of a tube carrying most of the flux, and its location is influenced by the width of the tube. Since (131) also involves an average in each isocommittor surface, this smoothes out irrelevant details on the thermal scale or below in these surfaces and makes that the curve γ is less sensitive to these details. This is illustrated in the context of the rugged Mueller potential example in figure 8.

Note that we can include the next order term in (131) by using (94) to approximate $|\nabla s|^2$ as

$$|\nabla s|^2 \approx 1 - 2\boldsymbol{\varphi}'' \cdot (\boldsymbol{x} - \boldsymbol{\varphi}) \tag{132}$$

It is easy to see that using this approximation in (129) gives the equation

$$0 = \langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=s} - \boldsymbol{\varphi}(s) + C(s)\boldsymbol{\varphi}''(s) \tag{133}$$

where we defined the tensor

$$C(s) = 2\langle (\boldsymbol{x} - \boldsymbol{\varphi})(\boldsymbol{x} - \boldsymbol{\varphi})^T \rangle_{s(\boldsymbol{x})=s} \tag{134}$$

(133) is to (131) what the equation (99) for the line of maxflux is to (100) for the MEP.

Numerical algorithms to solve (131) and (133) will be presented in section 9. These equations can also be generalized in collective variable space by an appropriate change of metric. Since it is easier to explain what this amounts to doing algorithmically, we postpone the discussion of this point till section 9.

Finally, note that, under the assumptions above, formula (106) for the function f remains valid.

7. Numerical algorithms for computing the minimum energy path

For systems with smooth energy landscapes for which the original TST or Kramers' rate theory gives a sufficiently accurate description of the transition process, the main object of interest are the transition states which are saddle points on the potential energy landscape. From a numerical viewpoint, we are naturally interested in algorithms for finding such saddle points. Ideas such as Newton's method, the dimer method, conjugate peak refinement, etc. are developed for this purpose. These algorithms are intended for searching directly the saddle points. However, in cases when the initial and final states are separated by intermediate stable states, one is interested in a sequence of transition states. In that case, algorithms for finding saddle points are no longer sufficient for determining the relevant sequence of transition states. Instead, one must look for the MEP.

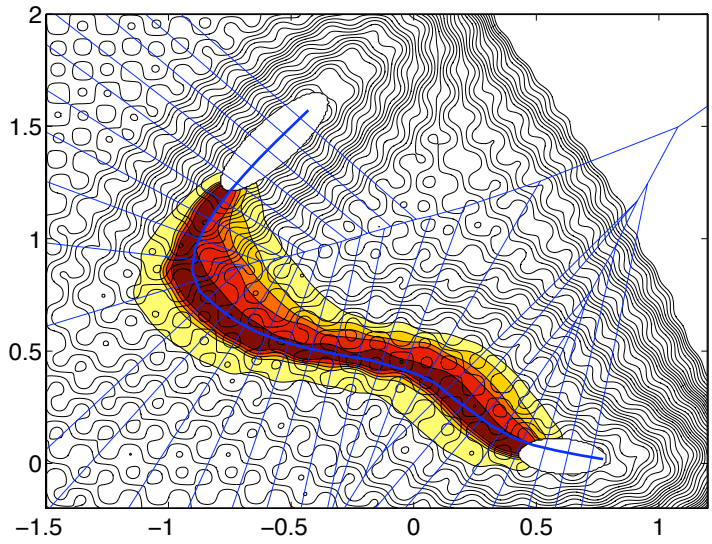

FIGURE 8. The principal curve solution of (130) in the example of the rugged Mueller potential. Also shown are the isosurfaces of the function $s(\boldsymbol{x})$ associated with the curve, and, in the background, the actual current of reactive trajectories already shown in figure 6. As can be seen, the principal curve is a the center of the tube carrying most of the flux of reactive trajectories. Inside this tube, the isosurfaces of the function $s(\boldsymbol{x})$ are good approximation of the isocommittor surfaces, i.e. $q(\boldsymbol{x}) \approx f(s(\boldsymbol{x}))$ with f given in (106). Note also that the principal curve is rather smooth, i.e. it is not affected much by the ruggedness of the potential.

In this section, we will mostly focus our discussion on the (zero-temperature) string method. Even though it shares a lot of similarities with several other strategies that were developed earlier, particularly for the discretized version, the string method distinguishes itself by two important features that are worthy to be pointed out from the onset.

First, as we have seen in section 6.2, the MEP is defined as a curve in configuration space. Therefore to look for the MEP, it is most natural to use an algorithm that moves curves in the configuration space. The string method does exactly that: It is an intrinsic formulation of the dynamics of curves in the configuration space. This is to be contrasted with algorithms that are formulated as chain-of-states algorithms in the first place. It is true that after discretizing the curves, the string method also becomes a chain-of-states method. Nevertheless, it is much better to start with an intrinsic formulation and then discretize. After all, this is why calculus is so important and useful. In addition, since the states are continuously been reinterpolated, the states in the string method do not have the kind of fixed identity as other chain-of-states methods.

Secondly, there are practical advantages of the intrinsic formulation. One is that it is much easier to be extended to the case with rough energy landscapes (the finite temperature string method). The other is that it is much easier to improve the accuracy of the algorithms that are based an intrinsic formulation.

For simplicity, throughout this section we use mass-weighted coordinates. Most of the formulas below can be translated back in the original non-mass weighted coordinates by the substitution $\nabla \to m^{-1/2}\nabla$ and $\boldsymbol{\varphi} \to m^{1/2}\boldsymbol{\varphi}$.

7.1. **The zero temperature string method.** One main motivation for the string method is to formulate a strategy based on the intrinsic dynamics of curves connecting two local minima of the potential $U(\boldsymbol{x})$ located, say, at $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$. Dynamics of curves are determined by the normal velocity at each point along the curve. The tangential component does not affect the graph of the curve. Since the MEP satisfies (**??**), the simplest dynamics for the evolution of the curve toward a MEP is given abstractly by

$$\boldsymbol{v}_n = -[\nabla U]^\perp \tag{135}$$

where $\boldsymbol{v}_n$ denotes the normal velocity of the curve. This formulation guarantees that it is gauge invariant, i.e. it is invariant under the change of the parametrization of the curve.

To translate (135) to a form that can be readily used in numerical computations, we assume that we have picked a particular parametrization of the evolving curve, $\gamma(t)$. For numerical purposes, the simplest choice is to use equal-arclength parametrization. In this case the curve $\gamma(t)$ is represented as $\gamma(t) = \{\boldsymbol{\varphi}(\alpha, t) : \alpha \in [0, 1]\}$, where α is a constant multiple of the arclength from $\boldsymbol{\varphi}(0, t)$ to the point $\boldsymbol{\varphi}(\alpha, t)$[5]. Denote by prime the derivative we respect to α, we then have $|\boldsymbol{\varphi}'|$ = constant in α (this constant is the length of the curve $\gamma(t)$). Other parametrizations are possible as well: for instance, we can add an energy-dependent weight function along the curve in order to enhance the accuracy near the saddle points [19]. In this case $w(\alpha)|\boldsymbol{\varphi}'(\alpha, t)|$ = constant in α, where $w(\alpha)$ is the weight function. This kind of ideas has been used quite extensively since the 1980's, for example, in the work of Brower *et al.* on geometric models of interface evolution [8, 9].

Two slightly different forms of the string method have been suggested [19, 22]. The original form was based on the evolution equation

$$\dot{\boldsymbol{\varphi}} = -[\nabla U(\boldsymbol{\varphi})]^\perp + \lambda\hat{\boldsymbol{\tau}} \tag{136}$$

(136) is just the parametric version of (135): $\dot{\boldsymbol{\varphi}}$ denotes the time derivative of $\boldsymbol{\varphi}$, $\hat{\boldsymbol{\tau}} = \boldsymbol{\varphi}'/|\boldsymbol{\varphi}'|$ is the unit tangent vector along the string and $\lambda\hat{\boldsymbol{\tau}}$ is a Lagrange multiplier term for the purpose of enforcing the particular parametrization of the string, such as the equal-arclength parametrization. The action of this term is very easy to compute and amounts to a reparametrization step, as explained below. There is, however, a subtle point associated with the discretization of (136) related to the fact that the component of $-[\nabla U(\boldsymbol{\varphi})]^\perp$ parallel to $\hat{\boldsymbol{\tau}}$ is a convection term that needs to be discretized carefully in order to avoid numerical instabilities [19, 53]. This point was noted already in [32]. On the other hand, since the component of $-[\nabla U(\boldsymbol{\varphi})]^\perp$ parallel to $\hat{\boldsymbol{\tau}}$ can be absorbed into the Lagrange multiplier term, (136) can be recast into

$$\dot{\boldsymbol{\varphi}} = -\nabla U(\varphi) + \bar{\lambda}\hat{\boldsymbol{\tau}}. \tag{137}$$

where the particular parametrization of the string is now enforced by the action of $\bar{\lambda}\hat{\boldsymbol{\tau}}$ [22]. This new form permits us to eliminate the numerical instability issue discussed above.

In actual computations, the string is discretized into a number of images $\{\boldsymbol{\varphi}_i(t), i = 0, 1, \ldots, N\}$ where $\boldsymbol{\varphi}_i(t) = \boldsymbol{\varphi}(\alpha = i/N, t)$. The images along the string are evolved by iterating upon the following two-step procedure:

[5]We use a constant multiple of the arclength, instead of the arclenght itself as in section 6.1, so that the range of the parameter α is fixed, $\alpha \in [0, 1]$, even if the length of the curve $\gamma(t)$ varies during its evolution. This is more convenient in the numerics.

1. Evolution step: The images on the string are evolved over some time-interval Δt according to the potential force at that point:

$$\dot{\boldsymbol{\varphi}}_i = -\nabla U(\boldsymbol{\varphi}_i) \tag{138}$$

This equation can be integrated in time by any stable ODE solver, e.g. the forward Euler method or Runge-Kutta methods.

2. Reparametrization step: The points (images) are redistributed along the string using a simple interpolation procedure (see [22]). Take the example of equal-arclength parametrization. Then the reparametrization step again involves two steps. The first step is to compute the arclength function along the string, i.e. the piecewise linear function $\ell(\alpha)$ whose values at $\alpha = i/N$ is given by

$$\ell(0) = 0, \qquad \ell(i/N) = \sum_{j=1}^{i} |\boldsymbol{\varphi}_j - \boldsymbol{\varphi}_{j-1}|, \quad i = 1, \ldots, N \tag{139}$$

This allows us to compute the new parametrization, the values of $\{\alpha_i, i = 0, 1, \ldots, N\}$ which make up an equal-distance parametrization, i.e.

$$\ell(\alpha_i) = \frac{i}{N}\ell(1) \tag{140}$$

The second step is to compute the images at these new parameter values using standard piecewise polynomial interpolation. For instance, if ones uses linear interpolation, this amounts to constructing the piecewise linear function $\boldsymbol{\varphi}(\alpha)$ such that its values at $\alpha = i/N$ are the images before reparametrization, then set $\boldsymbol{\varphi}_i = \boldsymbol{\varphi}(\alpha_i)$.

Note that the time interval Δt between the reparametrization steps can be bigger than the actual time step used to evolve (138). A practical way to determine when to invoke the reparametrization procedure is to monitor the distances between neighboring points along the string by (138) and go to the reparametrization step if the ratio of the largest to the smallest distances goes above a given prescribed tolerance. Note also that the position of the end points along the curve, $\boldsymbol{\varphi}_0(t)$ and $\boldsymbol{\varphi}_N(t)$, are not affected by the reparametrization step since $\alpha_0 = 0$ and $\alpha_N = 1$ from (140). In other words, they find on their own the location of the nearest minima of $U(\boldsymbol{x})$ even if the end points of the curve where not initially there.

One issue with the approach based on (137) is that even after the string has converged to a steady state, points on the string will still move back and forth along the string: they move down in energy by the evolution step (138), then back up by the reparametrization step. To avoid this, we may modify the evolution step in (138) and project out the tangential component of the potential force. By using the projected force instead of the bare one, one eliminates the tendency for the images to move down in energy at the evolution step. But, as mentioned before, some care must then be taken on how to discretize this term, see[32, 53, 22] for details. For this reason, it is harder to design uniformly high order accurate string method in this form, though it is possible [53].

The version of the string method described above is a very simple but very effective technique for finding the minimum energy paths. Its implementation requires only a simple ODE solver for the evolution step, and a simple interpolation routine for the parametrization step. It is therefore very easy to use and it can be readily incorporated in any existing code as long as the code provides force evaluation.

7.2. **Broyden accelerated string method.** Instead of using the evolution equations (136) or (137), another way to proceed is to solve directly (**??**). This procedure, too, requires the curve to be parametrized and discretized, but it permits to use accelerated methods for solving non-linear equations, such as the quasi-Newton method due to Broyden. Denote again by $\{\boldsymbol{\varphi}_i, i = 0, 1, \ldots, N\}$ the images along the string and assuming that the end points of the string are located at two minima of the potential, i.e. $\boldsymbol{\varphi}_0 = \boldsymbol{x}_a$ and $\boldsymbol{\varphi}_N = \boldsymbol{x}_b$, the discretized version of (**??**) for the remaining images reads

$$0 = \nabla U(\boldsymbol{\varphi}_i) + (\nabla U(\boldsymbol{\varphi}_i) \cdot \hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i, \qquad i = 1, \ldots, N-1 \tag{141}$$

where $\hat{\boldsymbol{\tau}}_i$ is a suitable discrete approximation of the unit tangent vector along the string, for instance

$$\hat{\boldsymbol{\tau}}_i = \frac{\boldsymbol{\varphi}_{i+1} - \boldsymbol{\varphi}_{i-1}}{|\boldsymbol{\varphi}_{i+1} - \boldsymbol{\varphi}_{i-1}|} \tag{142}$$

Solving the nonlinear equation (141) for $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{N-1}$ by Newton's method requires to calculate the inverse of the Jacobian matrix: if $\boldsymbol{G}_i(\varphi_1, \ldots, \varphi_{N-1})$ denotes the right hand side of (141), this Jacobian is the $(N-1)d \times (N-1)d$ matrix with $d \times d$ blocks

$$\frac{\partial \boldsymbol{G}_i(\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_{N-1})}{\partial \boldsymbol{\varphi}_j}, \qquad i, j = 1, \ldots N-1 \tag{143}$$

Computing and inverting this matrix can be impractical since it involves evaluating the Hessian of the potential, $\nabla\nabla U$. The idea behind Broyden's method is to approximate the inverse of (143) during the calculation and update its value at every iterative step used to solve (141). In the limited memory version of the method, this is done via a low-rank approximation of the inverse Jacobian which makes the method practical even for systems too large to store a $(N-1)d \times (N-1)d$ tensor. The details of how this is done in practice can be found in standard textbooks. The only point worth mentioning is that the method needs to be interfaced with the reparametrization step discussed in section 7.1 to enforce the parametrization of the string. The limited memory Broyden's string method was originally proposed in [19], and shown to converge faster than the simpler string method discussed in section 7.1. Broyden's method was also recently used in the context of NEB in [49].

7.3. **Action based methods to calculate the MEP.** As was originally noted by Elber and Olender [48] and rederived in [66], it is possible to write down a variational formulation for the MEP. Specifically, the MEP is the minimizer of

$$E(\boldsymbol{\varphi}) = \int_0^1 |\nabla U(\boldsymbol{\varphi})||\boldsymbol{\varphi}'| d\alpha \tag{144}$$

Indeed, the functional derivative of (144) with respect to $\boldsymbol{\varphi}$ is (using $|\boldsymbol{\varphi}'|$ = constant)

$$\frac{\delta E}{\delta \boldsymbol{\varphi}} = \frac{|\boldsymbol{\varphi}'|}{|\nabla U|}[\nabla\nabla U \nabla U]^\perp - \frac{|\nabla U|}{|\boldsymbol{\varphi}'|}\boldsymbol{\varphi}'' \tag{145}$$

From (100), the MEP is such that $\boldsymbol{\varphi}'$ must be parallel to $\nabla U(\boldsymbol{\varphi})$, i.e. we must have

$$|\boldsymbol{\varphi}'|\nabla U - |\nabla U|\boldsymbol{\varphi}' = 0 \tag{146}$$

Differentiating this equation with respect to α using $|\boldsymbol{\varphi}'|' = 0$, one obtains the equation $\delta E/\delta\boldsymbol{\varphi} = 0$, which means that the solution of (146) is a critical point of (144). It can be checked that it is also a minimum. This suggests the possibility of identifying MEPs by solving

$$[\nabla\nabla U \nabla U]^\perp - \frac{|\nabla U|^2}{|\boldsymbol{\varphi}'|^2}\boldsymbol{\varphi}'' = 0. \tag{147}$$

where we multiplied $\delta E/\delta\boldsymbol{\varphi} = 0$ by $|\nabla U|/|\boldsymbol{\varphi}'|$ to avoid singularities at the critical points of the potential where $\nabla U = 0$. (148) can be discretized as

$$\begin{aligned}\nabla\nabla U(\boldsymbol{\varphi}_i)\nabla U(\boldsymbol{\varphi}_i) - (\nabla\nabla U(\boldsymbol{\varphi}_i)\nabla U(\boldsymbol{\varphi}_i)\cdot\hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i \\ - \frac{|\nabla U(\boldsymbol{\varphi}_i)|^2}{|\boldsymbol{\varphi}_{i+1}-\boldsymbol{\varphi}_{i-1}|^2}(\boldsymbol{\varphi}_{i+1}+\boldsymbol{\varphi}_{i-1}-2\boldsymbol{\varphi}_i) = 0, \qquad i = 1,\dots,N-1\end{aligned} \tag{148}$$

where $\boldsymbol{\varphi}_0 = \boldsymbol{x}_a$ and $\boldsymbol{\varphi}_N = \boldsymbol{x}_b$. This equation can then be solved e.g. by the Broyden's method discussed in section 7.2 in the context of (141). A disadvantage of (148) is that it requires computing $\nabla\nabla U\nabla U$. In practice, this can be done via finite difference using two force evaluation per image, i.e.

$$\nabla\nabla U(\boldsymbol{\varphi}_i)\nabla U(\boldsymbol{\varphi}_i) \approx \delta^{-1}[\nabla U(\boldsymbol{\varphi}_i + \delta\nabla U(\boldsymbol{\varphi}_i)) - \nabla U(\boldsymbol{\varphi}_i)] \tag{149}$$

where δ is a small parameter. An advantage of (148) is that it automatically enforces the equal-arclenght parametrization of the string. Indeed, multiplying this equation by $\hat{\boldsymbol{\tau}}_i$ implies that

$$\hat{\boldsymbol{\tau}}_i\cdot(\boldsymbol{\varphi}_{i+1}+\boldsymbol{\varphi}_{i-1}-2\boldsymbol{\varphi}_i) = 0, \qquad i = 1,\dots,N-1 \tag{150}$$

Using (142), it is easy to see that (150) can be cast into

$$\frac{|\boldsymbol{\varphi}_{i+1}-\boldsymbol{\varphi}_i|^2 - |\boldsymbol{\varphi}_i-\boldsymbol{\varphi}_{i-1}|^2}{|\boldsymbol{\varphi}_{i+1}-\boldsymbol{\varphi}_{i-1}|} = 0, \qquad i = 1,\dots,N-1 \tag{151}$$

i.e. when (148) is satisfied, the images must be equidistant.

Remark: calculating the line of max-flux. The idea discussed above can also be used to calculate the line of max-flux solution of (99). This has been recently discussed in [11]. (99) can be discretized as

$$-\beta\nabla U(\boldsymbol{\varphi}_i) + \beta(\nabla U(\boldsymbol{\varphi}_i)\cdot\hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i - \frac{\boldsymbol{\varphi}_{i+1}+\boldsymbol{\varphi}_{i-1}-2\boldsymbol{\varphi}_i}{|\boldsymbol{\varphi}_{i+1}-\boldsymbol{\varphi}_{i-1}|^2} = 0, \tag{152}$$

where $i = 1,\dots,N-1$ and $\boldsymbol{\varphi}_0 = \boldsymbol{x}_a$ and $\boldsymbol{\varphi}_N = \boldsymbol{x}_b$. Like (148), this equation automatically enforces the equal-arclenght parametrization of the string. Solving (152) by Broyden's method is simpler than solving (148), but it should be stressed that the solution of (152) is a line of max-flux, not a MEP.

7.4. **Chain-of-states methods.**

7.4.1. *Elastic band method.* The basic idea is to connect the two stable states $\boldsymbol{x}_a$ and $\boldsymbol{x}_b$ by a chain of states (replicas or images), and evolve this chain of states. In an early attempt [52], Pratt proposed to use Monte Carlo methods to sample chains of states between the initial and final states in order to find the transition state region. Pratt's idea has been developed in two directions. One is Monte Carlo algorithms for sampling true dynamical trajectories between the initial and final states. This is the well-known transition path sampling (TPS) algorithm, developed by Bolhuis, Chandler, Dellago and Geissler [15, 6]. The second is the class of optimization algorithms for finding the MEP using a chain of states. The elastic band method [28] is an example of this type of algorithms.

Given a chain of states $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$, where $\boldsymbol{x}_0 = \boldsymbol{x}_a, \boldsymbol{x}_N = \boldsymbol{x}_b$, let us define an energy for the chain:

$$E(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{N-1}) = \sum_{i=1}^{N-1} U(\boldsymbol{x}_i) + \frac{k\Delta\alpha}{2}\sum_{i=1}^{N}\frac{|\boldsymbol{x}_i-\boldsymbol{x}_{i-1}|^2}{\Delta\alpha^2} \tag{153}$$

where $\Delta\alpha = 1/N$ and $k > 0$ is a parameter. Alternative energy functions have been proposed by Elber and Karplus [25, 62]q.

In the elastic band method (also called plain elastic band method, to be contrasted with the nudged elastic band method discussed below), the idea is to move the chain of states according to the gradient flow of the energy E [28]:

$$\dot{\boldsymbol{x}}_i = -\frac{\partial E}{\partial \boldsymbol{x}_i} = -\nabla U(\boldsymbol{x}_i) + k\frac{\boldsymbol{x}_{i+1} + \boldsymbol{x}_{i-1} - 2\boldsymbol{x}_i}{\Delta\alpha}, \quad i = 1, \ldots, N-1. \tag{154}$$

The first term at the right hand side is the potential force, the second term is the spring force. Note how the scaling in the coefficient of the second term is chosen: Because of that, if we use an explicit ODE solver to evolve (154), the time step size allowed is $\Delta t \sim \Delta\alpha$ in order to guarantee long time numerical stability. However, in this scaling, the second term drops out in the continuum limit as $\Delta\alpha \to 0$.

The elastic band method is extremely simple and intuitive. However, this method fails to converge to the MEP because the spring force tends to make the chain straight which leads to corner-cutting [28].

7.4.2. *The nudged elastic band method.* To overcome the corner-cutting problem, Jónsson *et al.* introduced the *nudged elastic band method* (NEB) [37, 31, 57]. This is a very simple modification of the elastic band method, but one that made the method truly useful. Instead of using the total potential force and spring force to move the chain, one uses only the normal component of the potential force and the tangential component of the spring force:

$$\dot{\boldsymbol{x}}_i = -[\nabla U(\boldsymbol{x}_i)]^\perp + (\boldsymbol{F}_i^s \cdot \hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i, \qquad i = 1, \cdots, N-1. \tag{155}$$

where $\boldsymbol{F}_i^s = k(\boldsymbol{x}_{i+1} + \boldsymbol{x}_{i-1} - 2\boldsymbol{x}_i)/\Delta\alpha$ and $\hat{\boldsymbol{\tau}}_i$ denotes the tangent vector along the elastic band at $\boldsymbol{x}_i$.

It is easy to see that if the chain converges to a steady state, it should be (a discretized approximation of) a MEP. In fact, from (155), we see that if the left hand side vanishes, then

$$-[\nabla U(\boldsymbol{x}_i)]^\perp + (\boldsymbol{F}_i^s \cdot \hat{\boldsymbol{\tau}}_i)\hat{\boldsymbol{\tau}}_i = 0, \qquad i = 1, \cdots, N-1. \tag{156}$$

Since the two terms in this equation are normal to each other, each has to vanish. In particular, we have

$$[\nabla U(\boldsymbol{x}_i)]^\perp = 0, \qquad i = 1, \cdots, N-1. \tag{157}$$

which is the discretized version of (100). As already mentioned in section 7.2, we can also solve (156) directly using accelerated methods such as Broyden's [49].

The choice of the elastic constant k is a crucial issue for the performance of NEB. If k is too large, then the elastic band is too stiff and one has to use very small time steps to solve the set of ODEs in (155). If k is too small, then there is not enough force to prevent the states on the chain from moving away from the saddle point, hence the accuracy of the saddle point will be reduced. This difficulty can be alleviated by using the climbing-image version of NEB [33], a strategy which can also be adapted to the string method [22].

It is natural to ask whether the elastic band method can also be formulated in terms of evolution of continuous curves. As we remarked earlier, if we use the original scaling of the spring constant, then the term that represents the elastic force disappears in the continuum limit (which suggests that the elastic term in (154) is not enough to prevent the clustering of the states near the local minima as $\Delta\alpha \to 0$). In order to retain the spring term, one has to replace the spring constant k by $k/\Delta\alpha$. In that case, we can take the continuum limit of the (nudged) elastic band method and obtain:

$$\dot{\boldsymbol{\varphi}} = -[\nabla U(\boldsymbol{\varphi})]^\perp + k(\boldsymbol{\varphi}'' \cdot \hat{\boldsymbol{\tau}})\hat{\boldsymbol{\tau}} \tag{158}$$

However, to evolve this dynamics, one has to use $\Delta t \sim (\Delta\alpha)^2$ and it will take many more iterations in order for the dynamics to converge to MEP.

8. FINDING MINIMUM FREE ENERGY PATHS

8.1. **The zero temperature string method in collective variable space.** The zero temperature string method discussed in section 7.1 can be easily modified to calculate MFEPs which are solutions of (120). The equivalent of (137) simply becomes

$$\dot{\boldsymbol{\varphi}} = -[M(\boldsymbol{\varphi})\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\varphi})]^{\perp} + \bar{\lambda}\hat{\boldsymbol{\tau}} \tag{159}$$

where $\boldsymbol{\varphi}(\alpha, t)$ is now a parametrized curve in collective variable space. If we discretize this curve into $N+1$ images $\{\boldsymbol{\varphi}_i, i = 0, \ldots, N\}$, these images can be evolved by the two step procedure discussed in section 7.1. The reparametrization step is identical. The evolution step amounts to integrating the following equation over some time-interval Δt

$$\dot{\boldsymbol{\varphi}}_i = -M(\boldsymbol{\varphi}_i)\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\varphi}_i) \tag{160}$$

This requires calculating the gradient of the free energy (i.e. the mean force) $\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\varphi}_i)$ and the tensor $M(\boldsymbol{\varphi}_i)$ at every time step. Since both quantities can be expressed in terms of canonical averages conditional on $\boldsymbol{\theta}(\boldsymbol{x}) = \boldsymbol{\varphi}_i$ (see (114) for the entries of M), they can both be evaluated using standard constrained or restrained MD simulation, such as the blue moon sampling [12, 14]. The details can be found in [41]. We will not repeat these here since the method proposed in [41] can be simplified to avoid any explicit calculation of conditional averages, as discussed next.

8.2. **On-the-fly string method.** The basic idea of the on-the-fly variant of the string method proposed in [43] is to approximate (160) by the following coupled system of equations in which two independent copies, or replica, $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ of the actual system are associated with every image $\boldsymbol{\varphi}_i$ along the string:

$$\begin{cases} \dot{\boldsymbol{\varphi}}_i = -\kappa\tilde{M}(\boldsymbol{x}_i)(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\bar{\boldsymbol{x}}_i)) \\ m\ddot{\boldsymbol{x}}_i = -\nabla U(\boldsymbol{x}_i) + \kappa\nabla\boldsymbol{\theta}(\boldsymbol{x}_i)(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\boldsymbol{x}_i)) \\ m\ddot{\bar{\boldsymbol{x}}}_i = -\nabla U(\bar{\boldsymbol{x}}_i) + \kappa\nabla\boldsymbol{\theta}(\bar{\boldsymbol{x}}_i)(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\bar{\boldsymbol{x}}_i)) \end{cases} \qquad i = 0, 1, \ldots, N \tag{161}$$

where $\kappa > 0$ is a parameter and we defined the $D \times D$ tensor $\tilde{M}(\boldsymbol{x})$ with entries

$$\tilde{M}_{i,j}(\boldsymbol{x}) = k_B T \gamma^{-1} \sum_{k=1}^{d} m_k^{-1} \frac{\partial\theta_i}{\partial x_k}\frac{\partial\theta_j}{\partial x_k}, \qquad i, j = 1, \ldots, D \tag{162}$$

From (161), the replica $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ evolve by a modified force field arising from the potential in which the original potential $U(\boldsymbol{x})$ has been adjoined with

$$\frac{\kappa}{2}|\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\boldsymbol{x})|^2 \tag{163}$$

This is a restraining potential which, if κ is adjusted appropriately, guarantees that the replica remain slaved to the images during the evolution in the sense that $\boldsymbol{\theta}(\boldsymbol{x}_i) \approx \boldsymbol{\varphi}_i$ and $\boldsymbol{\theta}(\bar{\boldsymbol{x}}_i) \approx \boldsymbol{\varphi}_i$. The key idea of the method is to artificially adjust the friction coefficient γ entering $\tilde{M}_{i,j}(\boldsymbol{x})$ to make the evolution of $\boldsymbol{\varphi}_i$ slower than that of $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$. Then at any given moment the replica $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ are at thermal equilibrium conditional on $\boldsymbol{\theta}(\boldsymbol{x}_i) \approx \boldsymbol{\varphi}_i$ and $\boldsymbol{\theta}(\bar{\boldsymbol{x}}_i) \approx \boldsymbol{\varphi}_i$, and the images $\boldsymbol{\varphi}_i$ feel only their average effect. In other words, the evolution of $\boldsymbol{\varphi}_i$ is effectively captured by

$$\dot{\boldsymbol{\varphi}}_i = -\langle\tilde{M}(\boldsymbol{x}_i)\rangle_\kappa\langle\kappa(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\bar{\boldsymbol{x}}_i))\rangle_\kappa \tag{164}$$

where the brackets denote canonical averages with respect to Boltzmann distribution involving the extended potential entering the equation for $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ in (161) (assuming that a proper thermostat has been added in these equations). Explicitly

$$\langle \tilde{M}(\boldsymbol{x}_i)\rangle_\kappa = \frac{\int \tilde{M}(\boldsymbol{x}) e^{-\beta U(\boldsymbol{x}) - \frac{1}{2}\beta\kappa|\boldsymbol{\varphi}_i - \theta(\boldsymbol{x})|^2} d\boldsymbol{x}}{\int e^{-\beta U(\boldsymbol{x}) - \frac{1}{2}\beta\kappa|\boldsymbol{\varphi}_i - \theta(\boldsymbol{x})|^2} d\boldsymbol{x}} \tag{165}$$

and similarly for $\langle \kappa(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\bar{\boldsymbol{x}}_i))\rangle_\kappa$. A simple calculation (see []) indicates that, for sufficiently large κ,

$$\langle \tilde{M}(\boldsymbol{x}_i)\rangle_\kappa \approx M(\boldsymbol{\varphi}_i) \quad \text{and} \quad \langle \kappa(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\bar{\boldsymbol{x}}_i))\rangle_\kappa \approx \nabla_{\boldsymbol{\theta}} G(\boldsymbol{\varphi}_i) \tag{166}$$

meaning that the evolution of the images $\boldsymbol{\varphi}_i$ by (161) approximates the evolution of these images by (160). Note that we need two replicas per image to guarantee that the two terms at the right hand side of the equation for $\boldsymbol{\varphi}_i$ in (161) self-average independently to give (164).

The on-the-fly string method is very simple to implement since it amounts to evolving concurrently the images and the replicas according to (161) and does not require any explicit averaging nor any restarting of the MD calculations. Beside the string discretization errors, there are two sources of errors in the method. The first is due to the finiteness of κ and it can be eliminated by replacing the retraining terms in (161) by Lagrange multiplier terms to enforce the constraint that $\boldsymbol{\theta}(\boldsymbol{x}_i) = \boldsymbol{\varphi}_i$ and $\boldsymbol{\theta}(\bar{\boldsymbol{x}}_i) = \boldsymbol{\varphi}_i$ exactly. The second source of errors are statistical errors which arise to the fact that the time scale separation between the images and the replica is not infinite in practice. These errors can be eliminated if, after a certain time t^*, we replace the terms $\tilde{M}(\boldsymbol{x}_i)$ and $\kappa(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\bar{\boldsymbol{x}}_i))$ by their running time-averages,

$$\frac{1}{t-t^*}\int_{t^*}^t \tilde{M}(\boldsymbol{x}_i(t'))dt' \quad \text{and} \quad \frac{\kappa}{t-t^*}\int_{t^*}^t (\boldsymbol{\varphi}_i(t) - \boldsymbol{\theta}(\boldsymbol{x}_i(t')))dt' \tag{167}$$

This also guarantees convergence of the method. At that stage of the calculation, a single replica per image becomes sufficient since we time-average $\tilde{M}(\boldsymbol{x}_i)$ and $\kappa(\boldsymbol{\varphi}_i - \boldsymbol{\theta}(\boldsymbol{x}_i))$ separately.

8.3. **String method with swarms.** In [51] another variant of the string method has been proposed in which the reparametrization step is as before, but the evolution step is done as follows. Given the current location of the images, $\boldsymbol{\varphi}_i(t)$, for each image a swarm of initial conditions $\{\boldsymbol{x}_i^r, r = 1, \ldots, R\}$ are prepared which sample the canonical distribution conditional on $\boldsymbol{\theta}(\boldsymbol{x}_i^r) = \boldsymbol{\varphi}_i$. These initial conditions are then evolved independently under the original force field (i. e. without any restraint or constraint) over a time interval Δt, and their average positions at that time is used to update the image positions:

$$\boldsymbol{\varphi}_i(t+\Delta t) = \frac{1}{R}\sum_{r=1}^R \boldsymbol{\theta}(\boldsymbol{x}_i^r(\Delta t)), \qquad i = 0, 1, \ldots, N \tag{168}$$

where $\boldsymbol{x}_i^r(\Delta t)$ denotes the position of the system evolved from the initial condition $\boldsymbol{x}_i^r$. A reparametrization step is then applied to the images $\{\boldsymbol{\varphi}_i(t+\Delta t), i = 0, \ldots, N\}$, and the procedure is repeated.

If Δt is small enough and the number M of replicas in the swarm is large enough, a direct calculation [42] shows that the updating rule (168) is consistent with

$$\boldsymbol{\varphi}_i(t+\Delta t) = \boldsymbol{\varphi}_i(t) - \Delta t\, M(\boldsymbol{\varphi}_i(t))\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\varphi}_i(t)) + \Delta t\, k_B T\, \mathrm{div}_{\boldsymbol{\theta}} M(\boldsymbol{\varphi}_i(t)) \tag{169}$$

up to corrections in Δt^2 and statistical errors due the finiteness of R. (169) is a time-discretized version of the evolution equation

$$\dot{\boldsymbol{\varphi}}_i = -M(\boldsymbol{\varphi}_i)\nabla_{\boldsymbol{\theta}} G(\boldsymbol{\varphi}_i) + k_B T \, \mathrm{div}_{\boldsymbol{\theta}} M(\boldsymbol{\varphi}_i) \tag{170}$$

Thus, if one is willing to accept the error introduced by the second term at the right hand side of (169), the string method with swarms can indeed be used as another way to calculate the MFEP approximately. Since the procedure requires to recreate R independent initial conditions canonically distributed conditional on $\theta(\boldsymbol{x}_i^r) = \boldsymbol{\varphi}_i$ per images at every evolution step, it is more expensive than the on-the-fly version of the method. The swarm version of the string method has however the advantage that it can be interfaced easily with existing MD packages using simple scripts.

Other variants of the string method to integrate (159) have been developed in [7, 38, 39].

9. Finding the transition tubes

9.1. **The finite temperature string method.** Methods to solve (131) and find principal curves have been developed in the statistics literature, see [30]. These methods are adaptations of the expectation-maximization (EM) algorithm in which the maximization step is replaced by a relaxation step. Given the current configuration of the string, the new configuration is found through the following steps:

1. Expectation step: Sample on the isosurfaces of the function $s(\boldsymbol{x})$ defined in (89), i.e. the pieces of hyperplanes normal to the current configuration of the string. How to do so in practice is explained next.

2. Relaxation step: Compute the empirical center of mass $\langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=s}$ on each hyperplane and move the string to a new configuration according to:

$$\boldsymbol{\varphi}^{n+1} = \boldsymbol{\varphi}^n + \Delta t \, \left(\langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=s} - \boldsymbol{\varphi}^n \right) \tag{171}$$

where $\boldsymbol{\varphi}^n$ denotes the current configuration of the string and Δt is a step size.

In practice, the string is discretized into $N+1$ images $\{\boldsymbol{\varphi}_i, i = 0, 1, \ldots, N\}$, and (171) is used for each images along the string. It becomes the counterpart of the evolution step performed in the zero temperature string method, whereas the reparametrization step is done as before (see section 7.1). In the original version of the finite temperature string method [20, 54, 21], $\langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=s}$ was estimated via constrained or restrained simulations in the hyperplanes perpendicular to the string. More recently in [70], it was realized that there is a simpler way to estimate $\langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=s}$. The idea is to replace the surface $s(\boldsymbol{x}) = s$ associated with $\boldsymbol{\varphi}(s)$ in the continuous setting by the Voronoi cell associated with image $\boldsymbol{\varphi}_i$, i.e. the region of space that contains all the points that are closer to $\boldsymbol{\varphi}_i$ than to any other image:

$$B_i = \{\boldsymbol{x} : |\boldsymbol{x} - \boldsymbol{\varphi}_i| < |\boldsymbol{x} - \boldsymbol{\varphi}_j| \text{ for all } j \neq i\} \tag{172}$$

The cells B_0 and B_N play a special role: they are, respectively, the reactant and the product states A and B. For $i = 1, \ldots, N-1$, however, the cell B_i is just a mollified version of the surface $s(\boldsymbol{x}) = i/N$, and we then have

$$\langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=i/N} \approx \langle \boldsymbol{x} \rangle_{B_i}, \qquad i = 1, \ldots, N-1 \tag{173}$$

where $\langle \boldsymbol{x} \rangle_{B_i}$ denotes canonical average conditional on $\boldsymbol{x} \in B_i$. This conditional average can be computed by a simple modification of the MD integrator in which we add a momentum reversal rule at collision with the boundaries of cell B_i to confine the simulation

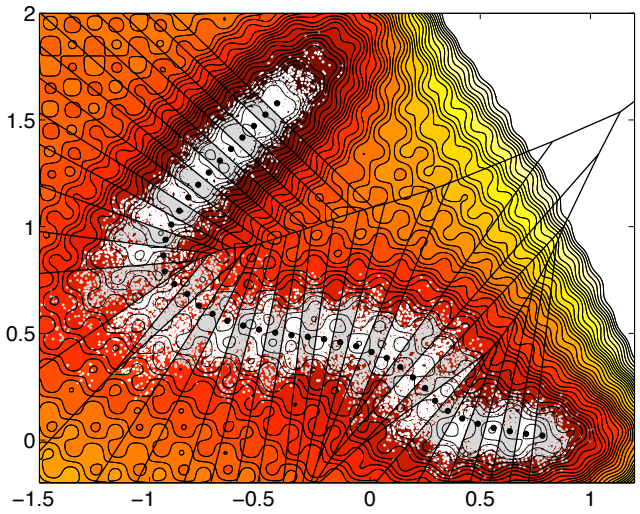

FIGURE 9. Result of a finite temperature string method calculation made in the context of the rugged Mueller potential. The images along the string are shown as black dots. The straight black lines show the boundaries of the Voronoi cells associated with each image, and the white and gray dots are the last 1000 positions of the MD replicas used to perform the conditional sampling in these cells. This figure should be compared to figure 8. Notice in particular that the replica sample the transition tube concentrating the flux of reactive trajectories shown in figure 8.

inside this cell. Thus, if $(\boldsymbol{x}_i(t), \boldsymbol{p}_i(t))$ denotes the position and momentum at time t of the independent copy (or replica) of the MD system used to do the sampling in cell B_i, we set

$$(\boldsymbol{x}_i(t+\delta t), \boldsymbol{p}_i(t+\delta t)) = \begin{cases} (\boldsymbol{x}_i^*(t+\delta t), \boldsymbol{p}_i^*(t+\delta t)) & \text{if } \boldsymbol{x}_i^*(t+\delta t) \in B_i \\ (\boldsymbol{x}_i(t), -\boldsymbol{p}_i(t)) & \text{if } \boldsymbol{x}_i^*(t+\delta t) \notin B_i \end{cases} \tag{174}$$

where $(\boldsymbol{x}_i^*(t+\delta t), \boldsymbol{p}_i^*(t+\delta t))$ denotes the time-evolved value of $(\boldsymbol{x}_i(t), \boldsymbol{p}_i(t))$ after one standard MD step of size δt. We then have

$$\langle \boldsymbol{x} \rangle_{B_i} = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \boldsymbol{x}_i^*(k\delta t) \tag{175}$$

up to small errors due to time-discretization. Note that the test made in (174) is a simple distance check: is $\boldsymbol{x}_i^*(t+\delta t)$ still closest to $\boldsymbol{\varphi}_i$? Note also that these MD calculations can be performed in parallel in each cell B_i. Finally, note that it is not necessary to compute with high precision the average (175) at every update of the string. Rather, it is more efficient to evolve concurrently the images $\boldsymbol{\varphi}_i$ and the MD replicas and replace (175) by a running time average. Details can be found in [70].

The scheme above can be straightforwardly generalized to solve (133) rather than (131). This simply amounts to accounting for the action of the term $C(s)\boldsymbol{\varphi}''$ in the updating rule (171). Since the diffusion tensor $C(s)$ is given by a conditional average (see (134)), it

can be estimated by the procedure above: Denoting $C_i = C(i/N)$, we have

$$C_i = \lim_{N\to\infty} \frac{2}{N} \sum_{k=0}^{N-1} (\boldsymbol{x}_i^*(k\delta t) - \boldsymbol{\varphi}_i)(\boldsymbol{x}_i^*(k\delta t) - \boldsymbol{\varphi}_i)^T \tag{176}$$

In practice, it is again more efficient to evolve the images $\boldsymbol{\varphi}_i$ and the MD replica concurrently, and replace the average (176) by a running average. Note also that, even when we solve (131), one typically needs to smooth the string to remove possible kinks arising from statistical errors. In effect, this amounts to adding a diffusion term to the updating rule (171), not unlike the one provided by the additional term $C(s)\boldsymbol{\varphi}''$ in (133), except that we then need to pick a value for $C(s)$ beforehand and keep it fixed once and for all.

The result of a finite temperature string method calculation made in the context of the rugged Mueller potential is shown in figure 9.

9.2. **Generalization to collective variable space.** The finite temperature string method can be easily generalized in collective variable space. As in section 8, the string now lives in collective variable space. The updating rule (171) is left unchanged, except that $\langle \boldsymbol{x} \rangle_{s(\boldsymbol{x})=s}$ must be replaced by $\langle \boldsymbol{\theta}(\boldsymbol{x}) \rangle_{s(\boldsymbol{\theta}(\boldsymbol{x}))=s}$ where $s(\boldsymbol{\theta})$ is the function defined in (117). In practice, this conditional expectation can be approximated as

$$\langle \boldsymbol{\theta}(\boldsymbol{x}) \rangle_{s(\boldsymbol{\theta}(\boldsymbol{x}))=i/N} \approx \langle \boldsymbol{\theta}(\boldsymbol{x}) \rangle_{B_i}, \qquad i = 1, \ldots, N-1 \tag{177}$$

Here $\langle \boldsymbol{\theta}(\boldsymbol{x}) \rangle_{B_i}$ denotes canonical average conditional on $\boldsymbol{x} \in B_i$ where B_i is the Voronoi cell

$$B_i = \{\boldsymbol{x} : \|\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\varphi}_i\|_{i,j} < \|\boldsymbol{\theta}(\boldsymbol{x}) - \boldsymbol{\varphi}_j\|_{i,j} \text{ for all } j \neq i\} \tag{178}$$

in which the Euclidean norm $|\cdot|$ used in (172) has been replaced by the norm

$$\|\boldsymbol{\theta}\|_{i,j}^2 = \sum_{k,l=1}^{D} \theta_k \frac{M_{k,l}^{-1}(\boldsymbol{\varphi}_i) + M_{k,l}^{-1}(\boldsymbol{\varphi}_j)}{2} \theta_l \tag{179}$$

In practice, $\langle \boldsymbol{\theta}(\boldsymbol{x}) \rangle_{B_i}$ can therefore be evaluated using MD simulations confined in cell B_i as before, except that the updating rule in (174) now involves a distance check using the norm (179).

9.3. **Free energy calculations.** The free energy of the reaction defined in (104) (or (122) if one uses collective variables) can be estimated from the equilibrium probability to find the system in each Voronoi cell B_i, which we shall denote by π_i. Indeed, if $F(s)$ is given by (104), then to leading in $1/N$ we have

$$\int_{(i-\frac{1}{2})/N}^{(i+\frac{1}{2})/N} e^{-\beta F(s)} ds = Z^{-1} \int_{B_i} e^{-\beta U(\boldsymbol{x})} \equiv \pi_i. \tag{180}$$

for $i = 1, \ldots, N-1$, whereas (since $B_0 = A$ and $B_N = B$):

$$\pi_0 = Z^{-1} \int_A e^{-\beta U(\boldsymbol{x})}, \qquad \pi_N = Z^{-1} \int_B e^{-\beta U(\boldsymbol{x})}, . \tag{181}$$

We can also define the free energy of the Voronoi tessellation as the vector $(F_0, F_1, \ldots, F_N)$ where $F_i = -k_B T \ln \pi_i$.

The MD simulations used in the finite temperature string method offer a simple way to calculate π_i. Suppose that as we run these MD simulations in each cell, we store the total number of times the MD replica in cell B_i made a collision with the boundary of cell B_j. Let us denote by $N_{i,j}$ this number of collisions observed during the interval T_i during

which we ran the simulation in cell B_i (this time could be different in each cell, hence the subscript i), and define

$$\nu_{i,j} = \frac{N_{i,j}}{T_i} \tag{182}$$

For large T_i, $\nu_{i,j}$ gives an estimate of the rate of escape from cell B_i to cell B_j. At statistical steady state, by conservation of probability, we must then have

$$\sum_{\substack{j=0 \\ j\neq i}}^{N} \pi_j \nu_{j,i} = \sum_{\substack{j=0 \\ j\neq i}}^{N} \pi_i \nu_{i,j}, \qquad \sum_{i=1}^{N} \pi_i = 1 \tag{183}$$

Thus, by monitoring collisions with the boundaries of the cells, estimating $\nu_{i,j}$, and finally solving (183), we have a simple way to calculate π_i and thereby estimate the free energy $F(s)$ from (180).

Note that the method we just described can be used to calculate the free energy associated with any (discretized) path: zero temperature string, finite temperature string, in Cartesian or collective variables, etc.

9.4. **Rate calculations: milestoning, etc.** The setup with Voronoi cells used in the finite temperature string method can also be conveniently combined with several techniques that have been introduced recently to calculate reaction rates: milestoning [27, 56, 24, 74, 71, 69], transition interface sampling (TIS) [64, 45, 46], forward flux sampling (FFS) [1, 2, 63], etc. All these methods analyze a reaction by decomposing it into a series of steps, each of which corresponds to the system going from one intermediate state to another. These intermediates are called milestones in milestoning and interfaces in TIS and FFS, and a basic issue is where to place them since this choice will affect the accuracy and/or the efficiency of the procedures.

Recently in [69] it has been proposed to use as intermediates the boundaries of the Voronoi cells along the string. This has two advantages. First it places the intermediate optimally with respect to the reaction since the boundaries of the Voronoi cells are approximations of the isocommittor surfaces which, as we know, have special properties as a reaction coordinate. Second it offers computational advantages. Let us briefly explain these advantages in the context of Markovian milestoning.

The basic idea of Markovian milestoning is to reduce the dynamics of the original MD system to that of the index of the last milestone it crossed, and to approximate the latter by a continuous-time Markov chain. To estimate the rate matrix of this chain, in the original procedure, initial conditions are generated on each milestones, then released and followed in time till they hit another milestone. The rate matrix of the chain can then be estimated easily from data collected from these short pieces of trajectories. The caveat with this approach is that it requires reinitialization on the milestones, and the probability density to perform this reinitialization is not explicitly known (in particular, as shown in [71] it is not the equilibrium density). In the new procedure proposed in [69], the reinitialization is avoided by extracting all the information to build the rate matrix from the MD simulations confined in the cells. If we denote by a, b, etc. the index of the milestones (i.e. the index of the boundaries of the cells B_i, not to be confused with the index i of these cells), the entry $k_{a,b}$ of the rate matrix can then be estimated as

$$k_{a,b} = \frac{\sum_{i=0}^{N} \pi_i N_{a,b}^i / T_i}{\sum_{i=0}^{N} \pi_i T_a^i / T_i}, \qquad a \neq b \tag{184}$$

Here π_i is the probability to be in cell B_i (see (183)), $N^i_{a,b}$ is the total number of transitions from edge a to edge b observed in the simulation confined to cell B_i, T^i_a is the total time during this simulation during which the last edge hit was a, and T_i is the total duration of the simulation confined to cell B_i (i.e. $\sum_a T^i_a = T_i$). Details can be found in [69].

Simulations confined to Voronoi cells can also be used to perform rate calculations *à la* TIS and FFS using again the boundaries of the cells as interfaces, see [68]. In this case, the procedure becomes a generalization of the nonequilibrium sampling method introduced in [73, 16].

10. CONCLUSIONS

To conlcude this review, we emphasize again the generality of the problems discussed here – such rare transition events are not rare, they are quite obiquitous. They are rare on the microscopic time scale of the system, such as the atomic vibration time scale, but they are not rare on the macroscopic time scales of interest.

In this review, we have focused on the $A \to B$ problem. Much progress has also been made on the other two classes of problems. In particular, the gentlest ascent dynamics (GAD) and its variants have emerged as an elegant tool for exploring the configuration space [**?**, **?**]. Voter's work on hyperdynamics and temperature accelerated dynamics has opened the door for accelerating molecular dynamics [**?**, **?**]. These topics are very interesting but beyond the scope of the present review.

REFERENCES

[1] 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.

[2] 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.

[3] A. Berezhkovskii, G. Hummer, and A. Szabo. Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J. Chem. Phys.*, 130(20):205102, 2009.

[4] M. Berkowitz, J. D. Morgan, J. A. McCammon, and S. H. Northrup. Diffusion-controlled reactions: A variational formula for the optimum reaction coordinate. *J. Chem. Phys.*, 79(11):5563–5565, 1983.

[5] P. Bolhuis. Transition-path sampling of β-hairpin folding. *Proc. Natl. Acad. Sci. USA*, 100(21):12129–12134, 2003.

[6] P. Bolhuis, D. Chandler, C. Dellago, and P. Geissler. T ransition p ath s ampling: Throwing ropes over rough mountain passes, in the dark. *Ann. Rev. Phys. Chem.*, 53(1):291–318, 2002.

[7] D. Branduardi, F. L. Gervasio, and M. Parrinello. From a to b in free energy space. *J. Chem. Phys.*, 126(5):054103, 2007.

[8] R. C. Brower, D. A. Kessler, J. Koplik, and H. Levine. Geometrical approach to moving-interface dynamics. *Phys. Rev. Lett.*, 51(13):1111–1114, 1983.

[9] R. C. Brower, D. A. Kessler, J. Koplik, and H. Levine. Geometrical models of interface evolution. *Phys. Rev. A*, 29(3):1335–1342, 1984.

[10] N. Buchete and G. Hummer. Coarse master equations for peptide folding dynamics. *J. Phys. Chem. B*, 112(19):6057–6069, 2008.

[11] M. Cameron and E. Vanden-Eijnden. Revisting the maxflux method. In preparation.

[12] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral. Constrained reaction coordinate dynamics for the simulation of rare events. *Chem. Phys. Lett.*, 156:472–477, 1989.

[13] 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(15):155101, 2007.

[14] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden. Blue moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics. *ChemPhysChem*, 6(9):1809–1814, 2005.

[15] C. Dellago, P. G. Bolhuis, and P. L. Geissler. Transition path sampling. *Adv. Chem. Phys.*, 123(1):1–78, 2002.

[16] A. Dickson, A. Warmflash, and A. R. Dinner. Nonequilibrium umbrella sampling in spaces of many order parameters. *J. Chem. Phys.*, 130(7):074104, 2009.
[17] J. L. Doob. *Classical potential theory and its probabilistic counterpart*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1984 edition.
[18] R. Durrett. *Stochastic Calculus*. CRC Press, Washington DC, 1996.
[19] W. E, W. Ren, and E. Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B.*, 66:052301, 2002.
[20] W. E, W. Ren, and E. Vanden-Eijnden. Finite temperature string method for the study of rare events. *J. Phys. Chem. B*, 109:6688–6693, 2005.
[21] W. E, W. Ren, and E. Vanden-Eijnden. Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.*, 413:242–247, 2005.
[22] W. E, W. Ren, and E. Vanden-Eijnden. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.*, 126:164103, 2007.
[23] W. E and E. Vanden-Eijnden. Towards a theory of transition paths. *J. Stat. Phys.*, 123:503–523, 2006.
[24] R. Elber. A milestoning study of the kinetics of an allosteric transition: Atomically detailed simulations of deoxy Scapharca hemoglobin. *Biophys. J.*, 92:L85–L87, 2007.
[25] R. Elber and M. Karplus. A method for determining reaction paths in large molecules: application to myoglobin. *Chem. Phys. Lett.*, 139(5):375–380, 1987.
[26] H. Eyring. The activated complex and the absolute rate of chemical reactions. *Chemical Reviews*, 17(1):65–77, 1935.
[27] A. K. Faradjian and R. Elber. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.*, 120:10880–10889, 2004.
[28] R. E. Gillilan and R. H. Lilien. Optimization and dynamics of protein-protein complexes using b-splines. *J. Comput. Chem.*, 25(13):1630–1646, 2004.
[29] T. Hastie. *Principal curves and surfaces*. PhD thesis, Stanford University, 1984.
[30] T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. Springer New York, 2001.
[31] G. Henkelman, G. Jóhannesson, and H. Jónsson. Methods for finding saddle points and minimum energy paths. *Progress on Theoretical Chemistry and Physics*, 111:269–300, 2000.
[32] G. Henkelman and H. Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *J. Chem. Phys.*, 113:9978, 2000.
[33] G. Henkelman, B. P. Uberuaga, and H. Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Chem. Phys.*, 113:9901, 2000.
[34] J. Horiuti. On the statistical mechanical treatment of the absolute rate of chemical reaction. *Bull. Chem. Soc. Japan*, 13(1):210–216, 1938.
[35] G. Hummer. From transition paths to transition states and rate coefficients. *J. Chem. Phys.*, 120(2):516–23, 2004.
[36] G. Hummer and I. G. Kevrekidis. Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computation. *J. Chem. Phys.*, 118(23):10762–10773, 2003.
[37] H. Jónsson, G. Mills, and K. W. Jacóbsen. Nudged Elastic Band Method for finding minimum energy paths of transitions. In B. J. Berne, G. Ciccotti, and D. F. Coker, editors, *Classical and Quantum Dynamics in Condensed Phase Simulations*, page 385. World Scientific, 1998.
[38] I. V. Khavrutskii, K. Arora, and C. L. Brooks, 3rd. Harmonic fourier beads method for studying rare events on rugged energy surfaces. *J. Chem. Phys.*, 125(17):174108, 2006.
[39] I. V. Khavrutskii and J. A. McCammon. Generalized gradient-augmented harmonic fourier beads method with multiple atomic and/or center-of-mass positional restraints. *J. Chem. Phys.*, 127(12):124901, 2007.
[40] S. Krivov and M. Karplus. Free energy disconnectivity graphs: Application to peptide models. *J. Chem. Phys.*, 117:10894–10903, 2004.
[41] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.*, 125(2):024106, 2006.
[42] L. Maragliano, B. Roux, and E. Vanden-Eijnden. Comparison of the on-the-fly string method and the string method with swarms for the calculation of minium free energy paths. In preparation.
[43] L. Maragliano and E. Vanden-Eijnden. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.*, 446(1–3):182–190, 2007.
[44] P. Metzner, C. Schütte, and E. Vanden-Eijnden. Transition path theory for markov jump processes. *Multiscale Model. Sim.*, 7(3):1192–1219, 2009.
[45] D. Moroni, P. G. Bolhuis, and T. S. van Erp. Rate constants for diffusive processes by partial path sampling. *J. Chem. Phys.*, 120:4055–4065, 2004.

[46] D. Moroni, T. S. van Erp, and P. G. Bolhuis. Investigating rare events by transition interface sampling. *Physica A*, 340:395–401, 2004.
[47] F. Noe, I. Horenko, C. Schütte, and J. Smith. Hierarchical analysis of conformational dynamics in biomolecules: Transition networks of metastable states. *J. Chem. Phys.*, 126(15):155102, 2007.
[48] R. Olender and R. Elber. Yet another look at the steepest descent path. *Journal of Molecular Structure: THEOCHEM*, 398:63–71, 1997.
[49] R. A. Olsen, G. J. Kroes, G. Henkelman, A. Arnaldsson, and H. Jónsson. Comparison of methods for finding saddle points without knowledge of the final states. *J. Chem. Phys.*, 121:9776, 2004.
[50] A. C. Pan and B. Roux. Building Markov state models along pathways to determine free energies and rates of transitions. *J. Chem. Phys.*, 129(6):064107, 2008.
[51] A. C. Pan, D. Sezer, and B. Roux. Finding transition pathways using the string method with swarms of trajectories. *J. Phys. Chem. B*, 112(11):3432–3440, 2008.
[52] L. Pratt. A statistical method for identifying transition states in high dimensional problems. *J. Chem. Phys.*, 85:5045–5048, 1986.
[53] W. Ren. Higher order string method for finding minimum energy paths. *Comm. Math. Sci.*, 1(2):377–384, 2003.
[54] W. Ren, E. Vanden-Eijnden, P. Maragakis, and W. E. Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.*, 123:134109, 2005.
[55] D. Shalloway. Macrostates of classical stochasic systems. *J. Chem. Phys.*, 105(22):9986–10007, 1996.
[56] D. Shalloway and A. K. Faradijan. Efficient computation of the first passage time distribution of the generalized master equation by steady-state relaxation. *J. Chem. Phys.*, 124:054112, 2006.
[57] D. Sheppard, R. Terrell, and G. Henkelman. Optimization methods for finding minimum energy paths. *J. Chem. Phys.*, 128(13):134106, 2008.
[58] S. Sriraman, I. G. Kevrekidis, and G. Hummer. Coarse master equation from bayesian analysis of replica molecular dynamics simulations. *J. Phys. Chem. B*, 109(14):6479–84, 2005.
[59] W. Swope, J. Pitera, and F. Suits. Describing protein folding kinetics by molecular dynamics simulations. 1. theory. *J. Phys. Chem. B*, 108(21):6571–6581, 2004.
[60] W. Swope, J. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. Fitch, R. Germain, A. Rayshubski, T. Ward, Y. Zhestkov, and R. Zhou. Describing protein folding kinetics by molecular dynamics simulations. 2. example applications to alanine dipeptide and beta-hairpin peptide. *J. Phys. Chem. B*, 108(21):6582–6594, 2004.
[61] A.-S. Sznitman. *Brownian motion, obstacles and random media*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998.
[62] A. Ulitsky and R. Elber. A new technique to calculate steepest descent paths in flexible polyatomic systems. *J. Chem. Phys.*, 92:1510, 1990.
[63] C. Valeriani, R. J. Allen, M. J. Morelli, D. Frenkel, and P. R. ten Wolde. Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling. *J. Chem. Phys.*, 127(11):114109, 2007.
[64] 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.
[65] E. Vanden-Eijnden. Transition path theory. In M. Ferrario, G. Ciccotti, and K. Binder, editors, *Computer Simulations in Condensed Matter: From Materials to Chemical Biology - Vol. 1*, pages 439–478, Berlin, 2006. Springer.
[66] E. Vanden-Eijnden and M. Heymann. The geometric minimum action method for computing minimum energy paths. *J. Chem. Phys.*, 128(6):061103, 2008.
[67] E. Vanden-Eijnden and F. Tal. Transition state theory: Variational formulation, dynamical corrections, and error estimates. *J. Chem. Phys.*, 123:184103, 2005.
[68] E. Vanden-Eijnden and M. Venturoli. Exact rate calculations by trajectory parallelization and tilting. *J. Chem. Phys.*, 131(4):044120, 2009.
[69] E. Vanden-Eijnden and M. Venturoli. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.*, 130:194101, 2009.
[70] E. Vanden-Eijnden and M. Venturoli. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.*, 130:194103, 2009.
[71] E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber. On the assumptions underlying milestoning. *J. Chem. Phys.*, 129:174102, 2008.
[72] M. Venturoli, E. Vanden-Eijnden, and G. Ciccotti. Kinetics of phase transitions in two dimensional ising models studied with the string method. *J. Math. Chem.*, 45(1):188–222, 2009.

[73] A. Warmflash, P. Bhimalapuram, and A. R. Dinner. Umbrella sampling for nonequilibrium processes. *J. Chem. Phys.*, 127(15):154112, 2007.
[74] A. M. A. West, R. Elber, and D. Shalloway. Extending molecular dynamics timescales with milestoning: Example of complex kinetics in a solvated peptide. *J. Chem. Phys.*, 126:145104, 2007.
[75] E. Wigner. The transition state method. *Transactions of the Faraday Society*, 34:29–41, 1938.

DEPARTMENT OF MATHEMATICS AND PROGRAM IN APPLIED AND COMPUTATIONAL MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NJ 08544
E-mail address: weinan@princeton.edu

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, NEW YORK, NY 10012
E-mail address: eve2@cims.nyu.edu