

# Revisiting the finite temperature string method for the calculation of reaction tubes and free energies

Eric Vanden-Eijnden, and Maddalena Venturoli

Citation: *The Journal of Chemical Physics* **130**, 194103 (2009);

View online: <https://doi.org/10.1063/1.3130083>

View Table of Contents: <http://aip.scitation.org/toc/jcp/130/19>

Published by the *American Institute of Physics*

---

## Articles you may be interested in

[Simplified and improved string method for computing the minimum energy paths in barrier-crossing events](#)

*The Journal of Chemical Physics* **126**, 164103 (2007); 10.1063/1.2720838

[String method in collective variables: Minimum free energy paths and isocommittor surfaces](#)

*The Journal of Chemical Physics* **125**, 024106 (2006); 10.1063/1.2212942

[Free energy of conformational transition paths in biomolecules: The string method and its application to myosin VI](#)

*The Journal of Chemical Physics* **134**, 085103 (2011); 10.1063/1.3544209

[Markovian milestoning with Voronoi tessellations](#)

*The Journal of Chemical Physics* **130**, 194101 (2009); 10.1063/1.3129843

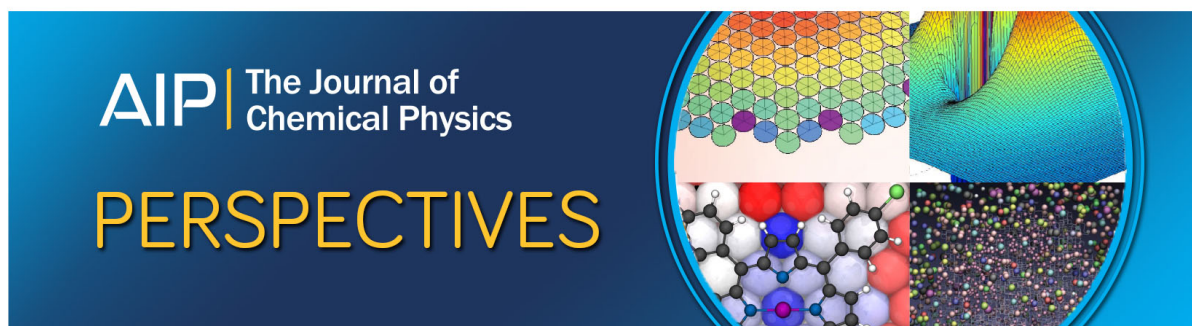
[Transition pathways in complex systems: Application of the finite-temperature string method to the alanine dipeptide](#)

*The Journal of Chemical Physics* **123**, 134109 (2005); 10.1063/1.2013256

[A climbing image nudged elastic band method for finding saddle points and minimum energy paths](#)

*The Journal of Chemical Physics* **113**, 9901 (2000); 10.1063/1.1329672

---



# Revisiting the finite temperature string method for the calculation of reaction tubes and free energies

Eric Vanden-Eijnden<sup>a)</sup> and Maddalena Venturoli<sup>b)</sup>

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

(Received 7 April 2009; accepted 15 April 2009; published online 15 May 2009)

An improved and simplified version of the finite temperature string (FTS) method [W. E, W. Ren, and E. Vanden-Eijnden, *J. Phys. Chem. B* **109**, 6688 (2005)] is proposed. Like the original approach, the new method is a scheme to calculate the principal curves associated with the Boltzmann–Gibbs probability distribution of the system, i.e., the curves which are such that their intersection with the hyperplanes perpendicular to themselves coincides with the expected position of the system in these planes (where perpendicular is understood with respect to the appropriate metric). Unlike more standard paths such as the minimum energy path or the minimum free energy path, the location of the principal curve depends on global features of the energy or the free energy landscapes and thereby may remain appropriate in situations where the landscape is rough on the thermal energy scale and/or entropic effects related to the width of the reaction channels matter. Instead of using constrained sampling in hyperplanes as in the original FTS, the new method calculates the principal curve via sampling in the Voronoi tessellation whose generating points are the discretization points along this curve. As shown here, this modification results in greater algorithmic simplicity. As a by-product, it also gives the free energy associated with the Voronoi tessellation. The new method can be applied both in the original Cartesian space of the system or in a set of collective variables. We illustrate FTS on test-case examples and apply it to the study of conformational transitions of the nitrogen regulatory protein C receiver domain using an elastic network model and to the isomerization of solvated alanine dipeptide. © 2009 American Institute of Physics. [DOI: [10.1063/1.3130083](https://doi.org/10.1063/1.3130083)]

## I. INTRODUCTION

In the study of rare reactive events such as conformation changes in molecules, chemical reactions, kinetic phase transitions, etc., it is customary to try to explain the mechanism of the reaction by identifying one (or more) specific path(s). How to define these paths is a nontrivial question, however. It is reasonable to require that each of the paths be made of the typical configurations of the system during the reaction or, more generally, that it lies at the center of a channel through which the system is likely to pass as the reaction occurs. This requirement, however, falls short of being a working definition, which probably explains why various candidates for transition path have been proposed. Probably the most widely known and used among those is the minimum energy path (MEP) (also known as intrinsic reaction coordinate).<sup>1</sup> The MEP can be justified as being the most likely path by which a reaction occurs under specific assumptions.<sup>2,3</sup> Without going into the specifics of what these assumptions are, it is however quite clear that they often fail. The reason why is quite simple: The MEP is an object that depends on *local* characteristics of the energy landscape (recall that the MEP is everywhere parallel to the gradient of the energy, i.e., to the force). Yet, reaction channels, assuming they exist, may be influenced by more global features of the underlying landscape. For instance, the land-

scape can be rough on a scale smaller than the thermal energy, in which case many of the detailed features contained in the MEP are irrelevant for the mechanism of the reaction and should be averaged over. In addition, nonlocal features of the energy away from the MEP in the direction perpendicular to it are not accounted for by the MEP but may matter for the reaction. For instance, the energy may climb very steeply away from the MEP, i.e., it may lie in a narrow channel which therefore turns out to be less important for the reaction than another channel which is less favorable energetically but much broader. One way to go beyond the concept of MEP, smooth out some of the irrelevant features of the energy landscape, and account for some nonlocal features of the energy landscape is to work with some appropriately chosen collective variables, define the free energy associated with these variables, and look for a minimum free energy path (MFEP) on the free energy.<sup>4,5</sup> After the MEP, the MFEP probably is the second most widely used candidate for a reaction path and indeed it has recently been shown to be able to explain the mechanism of reactions in various nontrivial setups.<sup>4–8</sup> Yet the MFEP may still miss important nonlocal features of the free energy landscape in the directions perpendicular to itself since it also depends only on local characteristics of this landscape (the MFEP is everywhere parallel to the gradient of the free energy calculated in some appropriate metric). As a result, the MFEP too may wrongly predict that the reaction occurs via a channel with low free energy barrier whereas the actual channel is another one with

<sup>a)</sup>Electronic mail: eve2@cims.nyu.edu.

<sup>b)</sup>Electronic mail: mventuro@cims.nyu.edu.

a higher free energy barrier but broader. Other paths, e.g., such as the maximum-flux path,<sup>9,10</sup> which have been invoked to explain the mechanism of a reaction are also local in nature and thereby suffer from similar limitations.

The above considerations indicate that if using paths to explain the mechanism of a reaction is at all possible, the definition of these paths has to be nonlocal in general, i.e., their location must depend on global features of the underlying energy or free energy landscapes. This led to the proposal made in Refs. 11–13 to use principal curve(s) to explain the mechanism of a reaction. Roughly, a principal curve is such that its intersection with each of the hyperplanes perpendicular to itself in some appropriate metric coincides with the averaged position of the system in these planes.<sup>14</sup> As a result, the principal curves tend to follow low lying channels on the underlying energy or free energy landscapes and their location depends more globally on the properties of these landscapes. In particular, they permit to smooth out the features of the landscape below the thermal energy scale and account for nonlocal contributions related to the width of the channels. These features make the principal curves natural candidates to explain the mechanism of a reaction, and indeed they can be justified within the framework of transition path theory<sup>15–18</sup> (TPT) under appropriate assumptions, mainly that the reaction channels are not too wide nor too complicated (e.g., they do not form an intricate network of interlaced tubes). Since these assumptions will not always be valid, the principal curves cannot be the ultimate objects to explain all reactions. On the other hand, these curves do represent a significant advancement when compared to existing paths such as the MEP or the MFEP, and this justifies developing efficient algorithms to calculate them.

In Refs. 11–13 one such algorithm, termed the finite temperature string (FST) method, was proposed. The objective of the present paper is to present an improved and simplified version of the FTS algorithm. The main idea behind the new FTS method is to eliminate the hyperplanes used in the original formulation of FTS and replace these planes by the Voronoi cells whose respective generating points are the discretization points (or images) along the string. Each Voronoi cell is made of the region in configuration space which is closer to its generating image than to any other image along the string. The Voronoi cells converge toward the original planes in the limit of infinitely many images along the string. When the number of images along the string is finite, however, using Voronoi cells rather than planes permits to simplify the FTS algorithm. Essentially, the algorithm amounts to sampling the system in the Voronoi cells using standard molecular dynamics (MD) plus a collision rule at the boundaries of the cells to keep the trajectory inside these cells, while at the same time moving these cells consistently with the sampling output so as to drag their generating points toward the average position in the cells. As shown below, this guarantees that, at convergence, the images along the string approximate the discretization points along a principal curve. At convergence, the Voronoi cells also have the remarkable property that their generating points (i.e., the images along the string) are, up to reparametrization along the

string, the centroids of their associated cells, that is, the expected position of the system in these cells. Voronoi tessellations which have such property are called *centroidal Voronoi tessellation*<sup>19</sup> and have been shown to be useful in applications as diverse as data compression, clustering analysis, meshless computing, etc. The new FTS method is a way to generate approximate centroidal Voronoi tessellations under the additional constraints that the generating points of the Voronoi cells also are discretization points along a curve with specified parametrization.

Besides greater algorithmic simplicity, the new FTS also has other advantages. One is that it allows to simplify the calculation of the free energy of the reaction. The method presented here approximates this free energy by the free energy of the Voronoi cells (that is, minus  $k_B T$  times the logarithm of the equilibrium probability to find the system in each cell). The free energy of the Voronoi cells is calculated in a straightforward way by matching the (virtual) probability fluxes through the boundaries of the cells. These fluxes are obtained as by-products of the sampling used in the new FTS method by monitoring the number of collisions of the MD trajectory with the boundary of the cells. This technique to compute the free energy associated with a Voronoi tessellation is general and can be used in other contexts as well (that is, in situations when the tessellation is not necessarily associated with a string). Another advantage of the new FTS is that it can be generalized to work in collective variable space. That is to say, the algorithm can be straightforwardly modified to analyze a reaction in (possibly many) prescribed collective variables rather than in the original Cartesian space. This introduces additional assumptions (in order to justify which collective variables to use) but may be necessary if there are no well-defined reactions channels in the original Cartesian space (e.g., due to symmetry by translation or rotation) and these channels can only be defined in the collective variable space. Yet another advantage of the new FTS is that the Voronoi tessellation it computes can be used within other recent techniques to analyze rare events, such as e.g., milestoneing,<sup>20,21</sup> transition interface sampling (TIS),<sup>22</sup> or forward flux sampling (FFS).<sup>23</sup>

The remainder of this paper is organized as follows. In Sec. II we describe the type of situations we are interested in, define what a principal curve is, and explain why it is relevant to explain the mechanism of reactions in complicated setups. In Sec. III we present the new FTS algorithm and test it on a few simple examples. In Sec. IV we show how the new FTS can be used to compute the free energy of the reaction. In Sec. V, we then generalize FTS to systems at finite friction and/or for which we want to describe the reaction in some prescribed collective variables. As we will show, this generalization is straightforward and amounts to changing the metric to make it consistent with the collective variables which are curvilinear coordinates in general. In Sec. VI we apply FTS to compute the reaction pathway and the free energy in two examples: the conformational transition of the nitrogen regulatory protein C receiver domain using a two-state elastic network model<sup>7</sup> and the isomerization of the alanine dipeptide (AD) in explicit solvent. A few concluding remarks are given in Sec. VII. Finally, Appendix



A summarizes the link between FTS and TPT, whereas Appendix B is devoted to a technical justification related to the use of Voronoi cells.

## II. DEFINITION AND SIGNIFICANCE OF THE PRINCIPAL CURVE

To introduce the main ideas behind FTS, we will first consider a system whose state space is  $\Omega \subset \mathbb{R}^d$  and whose dynamics is governed by the following overdamped Langevin equation:

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

Here  $V(\mathbf{x})$  is the potential,  $\beta = 1/(k_B T)$  is the inverse temperature,  $\boldsymbol{\eta}(t)$  is a Gaussian white noise with mean zero and covariance  $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$ , and we have used units of time such that the friction coefficient is 1 in these units. Throughout this paper we use boldface characters to denote vectors and normal characters to denote their components, e.g.,  $\mathbf{x} = (x_1, x_2, \dots, x_d)$ . The equilibrium probability distribution for Eq. (1) is the Boltzmann–Gibbs distribution with density

$$\rho(\mathbf{x}) = Z^{-1} e^{-\beta V(\mathbf{x})}, \quad (2)$$

where  $Z = \int_{\Omega} e^{-\beta V(\mathbf{x})} d\mathbf{x}$  is a normalization constant. The generalization of the concepts exposed below to other dynamics and to collective variable space will be the topic of Sec. V.

We are interested in studying Eq. (1) in the following situations. We assume that there exist two metastable sets  $A$  and  $B$  in  $\Omega$ , i.e., regions of space that the system quickly reaches from nearby but between which it seldom makes transitions. This is the situation of interest, for example, in reactive systems, where sets  $A$  and  $B$  are associated to the reactant and product states, respectively. We stress that we do not restrict ourselves to situations where the sets  $A$  and  $B$  are small neighborhoods around specific local minima of the potential  $V(\mathbf{x})$ . Rather, we are mainly interested in situations where  $V(\mathbf{x})$  is rugged, i.e., it possesses very many critical points such as minima, saddle points, etc., most of which are irrelevant for the definition of  $A$  and  $B$  because the location of these sets depends on more global (e.g., large scale) features of the potential. A simple example of such potential in two dimensions is the rugged version of the Müller potential<sup>24</sup> which is depicted in Fig. 1 and which we will use as a test case throughout this section. This potential was constructed by adding to the standard Müller potential the term,

$$a \sin(2\pi kx) \sin(2\pi ky), \quad (3)$$

with  $a=9$  and  $k=5$ , and analyzing the dynamics at the temperature  $k_B T = a=9$ . This example potential has many critical points but because the thermal energy available is of the same order of the small potential barriers between nearby minima, these minima are not the primary cause of metastability. Rather, metastability arises because of larger scale energetic features, i.e., the large barrier between the dark regions in the upper left and lower right corners in Fig. 1. These dark regions can be taken as the sets  $A$  and  $B$  (as we will see below, a precise identification of these sets is not

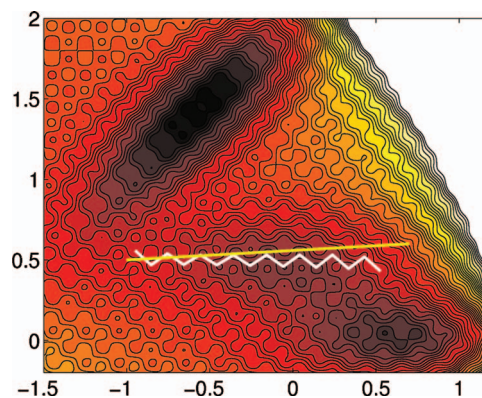


FIG. 1. (Color) Contour plot of the rugged Müller potential obtained by adding the term (3) to the standard Müller potential. Because of this extra term, the potential contains a large number of critical points. Also shown in white is one among many MEPs connecting two local minima via a succession of saddle points and other minima. As can be seen, this MEP is totally uninformative about the mechanism of reaction between the metastable sets in this example, which are the two dark regions around  $(x,y)=(-0.5,1.5)$  and  $(x,y)=(0.5,0)$ . The MEP was calculated by the zero-temperature string method (Refs. 25 and 26) using as initial string a piece of straight line between  $(x,y)=(-1,0.5)$  and  $(x,y)=(0.7,0.6)$  (shown in yellow).

necessary). To describe the reaction between these sets, the MEPs are inadequate: indeed there are a very large number of them on a potential such as this one and most of them are irrelevant (see Fig. 1). What needs to be used instead to describe the reaction is a tube in  $\Omega$  which connects the metastable regions where  $A$  and  $B$  lie and through which most reactive trajectories (i.e., these pieces of trajectories by which the reaction occurs) go (see Fig. 2: the objects shown in this figure will be explained in more details below). The reaction tubes allow one to average irrelevant energy features of the potential on the scale of  $k_B T$  or smaller (like the barriers between nearby minima in the example of the rugged Müller potential), and focus on the large scale features of the

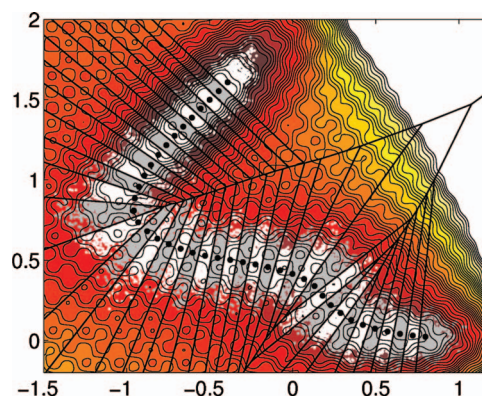


FIG. 2. (Color) Results of the improved FTS on the rugged Müller potential. The black dots are the images  $\varphi_\alpha^n$  along the string; the black lines show the boundary of the Voronoi cells  $B_\alpha^n$  associated with each of these images; the white and gray dots show the last  $10^3$  positions of the replicas  $\mathbf{x}_\alpha^n$  and determine the transition tube (we use two colors to distinguish points in neighboring cells). The boundaries of the Voronoi cells are approximations of isosurfaces of the function  $s_\alpha(\mathbf{x})$ , and they are also approximations of the level sets of the committor function  $q(\mathbf{x})$  (see Appendix A) for the transition between the first and the last cells along the string, i.e.,  $B_0^n$  and  $B_N^n$ , which can be chosen to identify the metastable sets,  $A$  and  $B$ . This calculation was initiated using the same initial string as the one used in Fig. 1, i.e., a piece of straight line between  $(x,y)=(-1,0.5)$  and  $(x,y)=(0.7,0.6)$ .

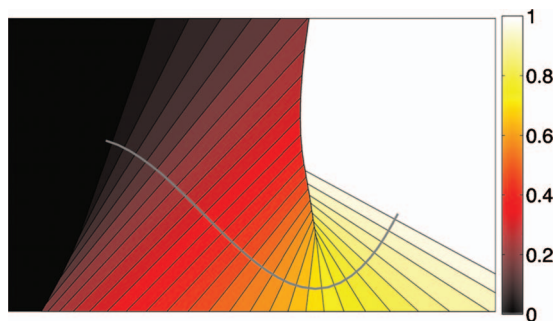


FIG. 3. (Color) Contour plot of the function  $s_\gamma(\mathbf{x})$  associated with a curve  $\gamma = \{\varphi(s) : s \in [0, 1]\}$ . The curve is shown as the thick gray line. The straight lines perpendicular to this curve are isosurfaces where  $s_\gamma(\mathbf{x}) = s$  for several values of  $s$  in  $(0, 1)$ . The regions where  $s_\gamma(\mathbf{x}) = 0$  and  $s_\gamma(\mathbf{x}) = 1$  are the regions at the left and right of the curve which are uniformly black and white, respectively.

potential which control the mechanism of the transition. The width of the tubes also accounts for the entropy of the reaction, i.e., it is what controls its free energy and rate. FTS is based on the idea that principal curves lie, under certain assumption, at the center of these tubes. Next in Sec. II A we will define precisely what these principal curves are. Then in Sec. II B, we will justify why they are relevant in the context of reactive events.

### A. Definition of the principal curve

We have said earlier that the FTS path is a principal curve associated with the density  $\rho(\mathbf{x})$  in Eq. (2) that is (using the imagery of the statistics literature) a curve which is its own expectation.<sup>14</sup> We will now make this statement precise. Given a curve  $\gamma$  in  $\Omega$ , let us parametrize this curve as  $\gamma = \{\varphi(s) : s \in [0, 1]\}$  for some  $\varphi : [0, 1] \rightarrow \Omega$  (unlike in the other string papers we will use  $s$  as continuous parameter along the string and we will reserve  $\alpha$  to denote the index of the discretization points along the string, see Sec. III; these notations are consistent with the ones recently used in Ref. 21). Associated with the curve  $\gamma$ , we then define a function  $s_\gamma(\mathbf{x})$ ,<sup>8,15</sup> which takes at point  $\mathbf{x}$  the value of the parameter  $s$  identifying the point  $\varphi(s)$  along the curve which is the closest to  $\mathbf{x}$ . In other words,  $s_\gamma(\mathbf{x})$  is such that for every  $\mathbf{x} \in \Omega$  it satisfies

$$|\mathbf{x} - \varphi(s_\gamma(\mathbf{x}))| = \min_{s \in [0, 1]} |\mathbf{x} - \varphi(s)|, \quad (4)$$

where  $|\cdot|$  denotes the Euclidean norm. The function  $s_\gamma(\mathbf{x})$  associated with a given curve is shown in Fig. 3. If the curve  $\gamma$  is smooth (i.e.,  $\varphi$  is differentiable), then Eq. (4) implies that the function  $s_\gamma(\mathbf{x})$  must satisfy

$$0 = \varphi'(s_\gamma(\mathbf{x})) \cdot (\mathbf{x} - \varphi(s_\gamma(\mathbf{x}))), \quad (5)$$

where the prime denotes derivative with respect to  $s$ . It also means that  $s_\gamma(\mathbf{x})$  is uniquely defined for every  $\mathbf{x} \in \Omega$  except for a set of Lebesgue measure zero in  $\Omega$  corresponding to those points for which the minimum at the right-hand side of Eq. (4) is achieved by more than one  $s$ . At those points,  $s_\gamma(\mathbf{x})$  is discontinuous and for definiteness we conventionally assign to  $s_\gamma(\mathbf{x})$  the value of the smallest  $s$  achieving the minimum (this choice will make no difference). Each of the level

sets (or isosurfaces) where  $s_\gamma(\mathbf{x}) = s \in (0, 1)$  is the subset of an hyperplane defined by the following three conditions (see Fig. 3 for illustration):

- (i) this subset contains the point  $\varphi(s)$ ,
- (ii) it is perpendicular to  $\gamma$  at  $\varphi(s)$ , and
- (iii) it contains only the points on the hyperplane which are closer to  $\varphi(s)$  than to any other  $\varphi(s')$  with  $s' \neq s$ .

The regions where  $s_\gamma(\mathbf{x}) = 0$  and  $s_\gamma(\mathbf{x}) = 1$ , i.e., the regions associated with the end points of the curve, are different and consist of the two sets on the sides of  $s_\gamma(\mathbf{x}) = 0^+$  and  $s_\gamma(\mathbf{x}) = 1^-$  opposite to the curve (they are the regions in black in white in Fig. 3). We will see later what role these sets play.

The curve  $\gamma$  is said to be a principal curve with respect to  $\rho(\mathbf{x})$  if the following condition holds:<sup>14</sup>

$$\forall s \in (0, 1) : \varphi(s) = \langle \mathbf{x} | s_\gamma(\mathbf{x}) = s \rangle_\rho, \quad (6)$$

where  $\langle \cdot | s_\gamma(\mathbf{x}) = s \rangle_\rho$  denotes the expectation with respect to  $\rho(\mathbf{x})$  conditional on  $s_\gamma(\mathbf{x}) = s$ ; explicitly, this expectation is given by

$$\langle \mathbf{x} | s_\gamma(\mathbf{x}) = s \rangle_\rho = \frac{\int_\Omega \mathbf{x} e^{-\beta V(\mathbf{x})} \delta(s_\gamma(\mathbf{x}) - s) d\mathbf{x}}{\int_\Omega e^{-\beta V(\mathbf{x})} \delta(s_\gamma(\mathbf{x}) - s) d\mathbf{x}}. \quad (7)$$

Condition (6) is what justifies the terminology that the principal curve is its own expectation. If the density  $\rho(\mathbf{x})$  has finite variance and one imposes that the length of the curve solution of Eq. (6) be finite, then it can be shown that a solution of this equation exists although it may not be unique.<sup>27</sup> It is also easy to show that the graph (i.e., location) of the principal curve  $\gamma$  is independent of its parametrization by  $\varphi(s)$ . Here, in the spirit of the string method, we will parametrize this curve by normalized arc length, i.e., we will impose the condition that  $|\varphi'(s)| = \text{const} = L_\gamma$ , where  $L_\gamma$  is the length of the curve  $\gamma$  (other parametrization can be used as well, but this one is convenient for numerical purposes).

Principal curves were originally introduced in the context of unsupervised (machine) learning.<sup>14</sup> Algorithms have been designed in this field to identify the principal curve but these algorithms are tailored to situations where the data set used to approximate the expectation in Eq. (7) is given and fixed, whereas we can (and have to) generate this data set as we go along. These computational aspects are discussed in Sec. III. Before going there, however, we recall a few elementary properties of the principal curve which help justify its relevance in the context of reactive events (see also Appendix A).

### B. Elementary properties and significance of the principal curve

Unlike the MEP, the principal curve associated with  $\rho(\mathbf{x})$  is an object which depends on *global* characteristics of the potential  $V(\mathbf{x})$ . Indeed its location is determined by an average in each isosurface  $s_\gamma(\mathbf{x}) = s$  rather than by the local value of the potential below the curve as with the MEP. Nevertheless, the principal curve reduces to a MEP in one specific limit, namely when  $\beta \rightarrow \infty$  (zero-temperature limit). Let us make this point first and then consider what happens at finite  $\beta$ . When  $\beta \rightarrow \infty$  the conditional expectation in Eq. (7) is

dominated by the point on  $s_\gamma(\mathbf{x})=s$  where the potential is minimum, i.e.,

$$\lim_{\beta \rightarrow \infty} \langle \mathbf{x} | s_\gamma(\mathbf{x}) = s \rangle = \mathbf{x}^*, \quad (8)$$

where  $\mathbf{x}^*$  is defined via

$$V(\mathbf{x}^*) = \min_{\mathbf{x}} V(\mathbf{x}) \text{ with } \mathbf{x} \text{ on } s_\gamma(\mathbf{x}) = s. \quad (9)$$

Since Eq. (9) requires that  $\mathbf{x}^*$  satisfies

$$0 = [\nabla V(\mathbf{x}^*)]^\perp, \quad (10)$$

where  $[\nabla V(\mathbf{x}^*)]^\perp$  denotes the component of  $\nabla V(\mathbf{x}^*)$  perpendicular to the curve  $\gamma$ , Eq. (6) for the principal curve then implies that this curve reduces to a MEP in the limit as  $\beta \rightarrow \infty$ .

When the temperature is nonzero ( $\beta < \infty$ ), the principal curve is, in general, different from a MEP and, as mentioned above, its location depends on global features of the potential. Some intuition on the principal curve can be gained from an alternative definition for it which states that the principal curve minimizes the following objective function:

$$D^2(\gamma) = Z^{-1} \int_{\Omega} |\mathbf{x} - \boldsymbol{\varphi}(s_\gamma(\mathbf{x}))|^2 e^{-\beta V(\mathbf{x})} d\mathbf{x}. \quad (11)$$

The equivalence between the definition involving minimizing Eq. (11) and the one involving Eq. (6) is proven in Ref. 14. The objective function (11) is a measure of the average square distance from the curve of the points in configuration space, in which points where the probability density is high count more than points where it is low. Thus, minimizing this objective function amounts to finding a curve to which the points in configuration space with significant probability weight are as close as possible. This implies that the metastable states will be located along the curve since by definition, these are states of high probability the curve wants to be close to. In between these states, the curve will tend to go through channels whose probability is globally high, where globally is understood in the sense implied by Eq. (11). For instance, since the probability weight is high where  $V(\mathbf{x})$  is low, the curve will tend to stay low on the potential, joining regions where  $V(\mathbf{x})$  is low by going through channels where it will stay low as well. However, let us stress that this last description does not exhaust the possibilities: To minimize Eq. (11) it may be preferable to place the curve in a wide channel with a higher barrier rather than a narrow one with a lower barrier. In other words, it is not how low  $V(\mathbf{x})$  is directly below the curve that matters, but how low it is more globally around the curve, with all the irrelevant features of  $V(\mathbf{x})$  on the thermal scale being smoothed out.

These properties of the principal curve are illustrated in Fig. 2 on the example of the rugged Müller potential. In the figure, the black dots are discretization points along the principal curve, whereas the black lines are isosurfaces of the function  $s_\gamma(\mathbf{x})$  associated with this curve. As it can be seen, the location of the principal curve is much less affected than the MEP shown in Fig. 1 by the small energy bumps in the potential  $V(\mathbf{x})$  and it roughly lies at the center of the tube carrying most of the flux of the reactive trajectories. As ex-

plained before, the reaction tube can be identified with the union for  $s \in (0,1)$  of the regions in the isosurfaces  $s_\gamma(\mathbf{x}) = s$  which concentrate most of the mass of  $Z^{-1}e^{-\beta V(\mathbf{x})}$  in these surfaces. The metastable states can be defined, as explained before, as two regions where  $s_\gamma(\mathbf{x})=0$  and  $s_\gamma(\mathbf{x})=1$  (i.e., the two cells made of the points in the configuration space which are closest to the two end points of the curve). The location of the reaction tube is indicated by the white and gray dots in Fig. 2, which are sample points of  $Z^{-1}e^{-\beta V(\mathbf{x})}$  used in the FTS method (see Sec. III).

The properties of the principal curve discussed above are the main intuitive reasons why the principal curve may be relevant to identify reaction channels: A more technical justification based directly on TPT was also proposed in Ref. 15 and for the reader's convenience, some of these calculations are repeated in Appendix A. In particular we recall there how the function  $s_\gamma(\mathbf{x})$  associated with the principal curve is related to the committor function of the reaction.

We conclude this section by pointing out that the discussion above focuses on situations where Eq. (6) has a unique solution, i.e., the principal curve is unique. Roughly, this corresponds to situations where there are only two metastable sets and a single channel connecting them. If there are more metastable sets and/or more channels between them, things are more complicated: There may be more than one principal curve or, even worse, there may be a single one but it becomes less informative (e.g., it may be influenced by several channels at once and lie in between them). In such situations, the definition of the principal curve has to be amended. One possibility is to localize the definition, i.e., when computing the conditional expectation in Eq. (7) only account for points within a certain distance from the curve. Another option is to use more than one curve simultaneously. In this case one function  $s_\gamma(\mathbf{x})$  must be associated with each curve, and each is only defined in the region the closest to its curve so that for each curve the conditional expectation in Eq. (7) is again restricted in a region close to this curve. Such amendments and modifications will permit to extend the applicability of principal curves to situations where the reaction occurs by several channels. It is clear, however, that these modifications will not be sufficient in all situations, and we may have to resort to objects other than principal curves to explain certain reactions.

### III. COMPUTATIONAL ASPECTS: THE IMPROVED FTS

The general idea behind the string method is to evolve a curve (or string) in some suitable way while controlling its parametrization and let this evolution go until some criterion is satisfied: In the context of the present paper, this criterion is Eq. (6) defining a principal curve. In practice this is done by discretizing the string into  $N+1$  points, referred to as images and denoted as  $\boldsymbol{\varphi}_\alpha$  with  $\alpha=0, \dots, N$ , and evolving these images. The string itself can be reconstructed from the images by interpolation.

In the original version of FTS, the positions of the images along the string are updated by performing a sampling in the planes perpendicular to the string to estimate the mean position of the system in these planes, and by dragging the



images toward these means until their positions coincide. This gives a curve which satisfies Eq. (6) up to numerical errors related, in particular, to the discrete nature of the string. This also means that the sampling in the planes performed in the original method can be replaced by any other sampling strategy which is consistent with the first within discretization errors. The new FTS method uses one such alternative sampling strategies which can be motivated as follows.

With each image  $\varphi_\alpha$  along the string we can associate its Voronoi cell  $B_\alpha$ , i.e., the region in  $\Omega$  which contains all the points  $\mathbf{x}$  which are closer to  $\varphi_\alpha$  than to any other image (see Fig. 2 for an illustration),

$$B_\alpha = \{\mathbf{x} \text{ such that } |\mathbf{x} - \varphi_\alpha| < |\mathbf{x} - \varphi_{\alpha'}| \text{ for all } \alpha' \neq \alpha\}. \quad (12)$$

Introducing these Voronoi cells is useful because they provide a way to approximate the conditional expectation in Eq. (7). More precisely, if we denote by  $\gamma = \{\varphi(s) : s \in [0, 1]\}$  a continuous curve interpolated through the images  $\varphi_\alpha$ ,  $\alpha = 0, \dots, N$  and by  $s_\gamma(\mathbf{x})$  the function defined in Eq. (4) associated with the curve, then we have that for any  $\alpha = 0, \dots, N$ ,

$$\langle \mathbf{x} | s_\gamma(\mathbf{x}) = \alpha/N \rangle = \frac{\int_{B_\alpha} \mathbf{x} e^{-\beta V(\mathbf{x})} d\mathbf{x}}{\int_{B_\alpha} e^{-\beta V(\mathbf{x})} d\mathbf{x}} + \text{discretization errors}. \quad (13)$$

The size of the discretization errors depends on the type of interpolation used to reconstruct  $\varphi(s)$  from the images  $\varphi_\alpha$ , but these errors always vanish as a power (which depends on the order of the interpolation) of  $1/N$  as  $N \rightarrow \infty$ . Qualitatively, Eq. (13) can be understood because the Voronoi cells are thickened versions of the isosurfaces of  $s_\gamma(\mathbf{x})$  which close up on these surfaces as  $N \rightarrow \infty$ . A more precise justification of Eq. (13) is provided in Appendix B.

Identity (13) suggests an algorithm to evolve the images along the string in which one updates the positions of these images as follows:

- (i) *Simulate Eq. (1) inside each  $B_\alpha$  with appropriate reflecting boundary conditions at the boundary of  $B_\alpha$ . This should be done in such a way that the generated trajectory permits to compute the mean position of the system inside each  $B_\alpha$  via time average, and thereby estimate the expectation in Eq. (13).*
- (ii) *Update the images  $\varphi_\alpha$  along the string (and hence the cells  $B_\alpha$ ) to make their positions closer to the mean position of the system inside each cell  $B_\alpha$ . This operation is performed to satisfy condition (6), i.e., in order that the string becomes an approximation of the principal curve up to discretization errors.*

Since the cells  $B_\alpha$  change each time the images  $\varphi_\alpha$  are updated, and in turn this influences the location of the mean position inside the cells, it is necessary to iterate upon steps (i) and (ii) until convergence. Step (ii) must also be performed in such a way that some constraint on the parametrization of the string be imposed as the string evolves e.g., to

keep the images equidistant, and using a numerical trick to temporarily smooth out possible kinks along the string which may arise at the initial stage of the calculation due to statistical errors in the computation of the mean position in step (i). Putting all these requirements together, we arrive at the new FTS algorithm presented next.

### A. The new FTS algorithm

Start with an initial set of images,  $\varphi_\alpha^0$  with  $\alpha = 0, \dots, N$  which is such that  $|\varphi_{\alpha+1}^0 - \varphi_\alpha^0| = |\varphi_\alpha^0 - \varphi_{\alpha-1}^0|$  for all  $\alpha = 1, \dots, N-1$ . With every image  $\varphi_\alpha^0$ , associate a replica of the original system,  $\mathbf{x}_\alpha^0$ , and set  $\mathbf{x}_\alpha^0 = \varphi_\alpha^0$  initially. Then, for  $n \geq 0$ , update the position of these images by iterating upon the following steps:

- (1) Update  $\mathbf{x}_\alpha^n$  via a time-discretized version of Eq. (1) with a reflecting boundary condition at the boundary of the Voronoi cell associated with the image  $\varphi_\alpha^n$ . For example, calculate

$$\mathbf{x}_\alpha^* = \mathbf{x}_\alpha^n - \Delta t \nabla V(\mathbf{x}_\alpha^n) + \sqrt{2\beta^{-1}\Delta t} \xi_\alpha^n \quad (14)$$

and set

$$\mathbf{x}_\alpha^{n+1} = \begin{cases} \mathbf{x}_\alpha^*, & \text{if } \mathbf{x}_\alpha^* \in B_\alpha^n \\ \mathbf{x}_\alpha^n, & \text{otherwise,} \end{cases} \quad (15)$$

where

$$B_\alpha^n = \{\mathbf{x} \text{ such that } |\mathbf{x} - \varphi_\alpha^n| < |\mathbf{x} - \varphi_{\alpha'}^n| \text{ for all } \alpha' \neq \alpha\}. \quad (16)$$

In Eq. (14)  $\Delta t$  denotes the time step used for numerical integration of Eq. (1) and  $\xi_\alpha^n$  are independent Gaussian variables with mean 0 and variance 1.

- (2) Compute the running average of each  $\mathbf{x}_\alpha^n$ ,

$$\bar{\mathbf{x}}_\alpha^n = \frac{1}{n} \sum_{n'=0}^n \mathbf{x}_\alpha^{n'}. \quad (17)$$

- (3) Update each image along the string toward the running average  $\bar{\mathbf{x}}_\alpha^n$  while keeping the string smooth. To do so use

$$\varphi_\alpha^* = \varphi_\alpha^n - \Delta \tau (\varphi_\alpha^n - \bar{\mathbf{x}}_\alpha^n) + \mathbf{r}_\alpha^*, \quad (18)$$

where  $\Delta \tau > 0$ ,  $\mathbf{r}_0^* = \mathbf{r}_N^* = 0$ , and for  $\alpha = 1, \dots, N-1$ ,

$$\mathbf{r}_\alpha^* = \kappa^n (\varphi_{\alpha+1}^* + \varphi_{\alpha-1}^* - 2\varphi_\alpha^*). \quad (19)$$

Here  $\kappa^n > 0$  is an adjustable parameter, whose value is commented upon below and which controls how aggressive the smoothing is.

- (4) Enforce the equal arc-length parametrization by interpolating a piecewise linear curve through  $\{\varphi_\alpha^*\}_{\alpha=0, \dots, N}$  and redistributing points at equal distance along this curve to obtain  $\{\varphi_\alpha^{n+1}\}_{\alpha=0, \dots, N}$ .
- (5) If  $\mathbf{x}_\alpha^{n+1} \in B_\alpha^{n+1}$  go to step 1, otherwise set  $\mathbf{x}_\alpha^{n+1} = \varphi_\alpha^{n+1}$  and go to step 1. Repeat until convergence of  $\{\varphi_\alpha^{n+1}\}_{\alpha=0, \dots, N}$ .

The various steps of this algorithm are commented in Sec. III B. Before going in the details, however, let us stress that this algorithm is simple. Indeed the modification it

amounts to introducing on an existing integrator for Eq. (1) is the acceptance/rejection step in Eq. (15), which consists of a mere distance check (“is  $\mathbf{x}_\alpha^*$  closest to  $\boldsymbol{\varphi}_\alpha^n$ ?”). All the other steps in the algorithm [including the solution of the tridiagonal system to obtain  $\boldsymbol{\varphi}_\alpha^*$  from Eq. (18) and the reparametrization step 4] are simple as well. It is also worth noting that the FTS algorithm is trivially parallelizable since the simulations of the replicas associated with each image along the string can be performed on independent nodes.

Compared to the original FTS method, the new method does not require computing the tangent vector along the string at every image, which was needed to define the hyperplanes perpendicular to the string at these images, and it avoids the constrained sampling in these hyperplanes. It also avoids having to restrict the extent of the sampling in these planes to avoid overlaps since the Voronoi cells automatically prevent such overlaps (we may, however, have to restrict the extent of sampling in the Voronoi cells if there are more than one reaction channel, as explained at the end of Sec. II B).

## B. Analysis of the FTS algorithm

Let us now comment on the various steps of the new FTS algorithm described in Sec. III A. Step 1 is a Metropolized version of the forward Euler scheme for Eq. (1) which is used to sample the probability density (2) conditional in  $B_\alpha^n$  (assuming this cell is fixed, which is eventually the case when the algorithm above has converged). The probability density (2) conditional in  $B_\alpha^n$  is explicitly

$$\rho_{B_\alpha^n}(\mathbf{x}) = Z_{B_\alpha^n}^{-1} e^{-\beta V(\mathbf{x})} \mathbf{1}_{B_\alpha^n}(\mathbf{x}), \quad (20)$$

where  $Z_{B_\alpha^n} = \int_{B_\alpha^n} e^{-\beta V(\mathbf{x})} d\mathbf{x}$  and  $\mathbf{1}_{B_\alpha^n}(\mathbf{x}) = 1$  if  $\mathbf{x} \in B_\alpha^n$  and  $\mathbf{1}_{B_\alpha^n}(\mathbf{x}) = 0$  otherwise. Using the updating step (14) together with the acceptance/rejection step (15) is a way to sample Eq. (20) approximately, i.e., up to time-discretization errors of order  $\Delta t$ . These time-discretization errors could be eliminated using a full Metropolis–Hastings sampling algorithm.<sup>28</sup> We also note that sampling procedures based on dynamics other than Eqs. (14) and (15) could be used, as long as they allow to compute expectations with respect to  $\rho_{B_\alpha^n}(\mathbf{x})$ . This means that we could replace the updating step in Eq. (14) e.g., by a standard MD update using Nosé–Hoover, or isokinetic, or Langevin as thermostat, or even by a Monte Carlo sampling procedure. This flexibility will be used Sec. V B. In addition it is not even necessary to use for the replica a dynamics which is constrained to stay in  $B_\alpha^n$  as long as the running average in Eq. (17) is computed consistently (i.e., by assigning each replica to the cell it is currently in if they are allowed to go from cell to cell). Here we stuck to the procedure in step 1 because it is simple to implement and proved to be accurate and efficient enough for our purposes. Finally, we stress again that the calculations in step 1 can be trivially parallelized since Eq. (14) can be run on different nodes [communication is only required for the acceptance/rejection step (15)].

Equation (18) in step 3 is the update of the string. In this equation the time step  $\Delta\tau$  can be taken different from  $\Delta t$  in Eq. (14). In fact, the first term at the right-hand side of Eq.

(18) can be written as  $(1 - \Delta\tau)\boldsymbol{\varphi}_\alpha^n + \Delta\tau\bar{\mathbf{x}}_\alpha^n$ , i.e., for  $0 \leq \Delta\tau \leq 1$ , it is a weighted average between  $\boldsymbol{\varphi}_\alpha^n$  and  $\bar{\mathbf{x}}_\alpha^n$  where the relative weight of both terms is controlled by  $\Delta\tau$ . In practice, we observed that setting  $\Delta\tau=0.1$  offers a good balance between speed of convergence (which favors  $\Delta\tau$  close to 1) and stability (which favors  $\Delta\tau$  close to 0). The extra term  $\mathbf{r}_\alpha^*$  in Eq. (18) is added to keep the string relatively smooth despite possible statistical fluctuations in  $\bar{\mathbf{x}}_\alpha^n$ . The updating step (18) was made implicit in  $\boldsymbol{\varphi}_\alpha^*$  to guarantee that  $\Delta\tau$  can be chosen independently of  $N$  without affecting the stability of the scheme. This leads to no significant loss in efficiency since the linear equation for  $\boldsymbol{\varphi}_\alpha^*$  in Eq. (18) can be solved in  $2N$  operations by Thomas algorithm which is a simplified form of a Gaussian elimination for tridiagonal systems (see e.g., Ref. 29). The strength of the term  $\mathbf{r}_\alpha^*$  is controlled by the value of  $\kappa^n$  in Eq. (19) and this value can be decreased as  $n$  increases since the size of the fluctuations in  $\bar{\mathbf{x}}_\alpha^n$  decreases like  $O(1/\sqrt{n})$  as  $n \rightarrow \infty$ . Here we simply kept  $\kappa^n$  fixed at the value  $\kappa^n = \kappa N \Delta\tau$  with  $\kappa=0.1$ . Note that by making  $\kappa^n$  proportional to  $N$ , we guarantee that the effect of the smoothing term also disappears in the continuous limit ( $N \rightarrow \infty$ ) since one would have to use  $\kappa^n = \kappa N^2 \Delta\tau$  instead of  $\kappa^n = \kappa N \Delta\tau$  in order that  $\mathbf{r}_\alpha^*$  converges toward the diffusion term  $\kappa\boldsymbol{\varphi}''$  in the limit as  $N \rightarrow \infty$  (here the double prime denotes the second derivative with respect to  $s$ ). In other words, working at a finite  $N$ , the extra term  $\mathbf{r}_\alpha^*$  in Eq. (18) with  $\kappa^n = \kappa N \Delta\tau$  affects the string by no more than the discretization errors inherent to using only a finite number of images.

Step 4 involves the standard reparametrization step of the string method which enforces the constraint that the images remain equidistant, i.e.,

$$|\boldsymbol{\varphi}_{\alpha+1}^n - \boldsymbol{\varphi}_\alpha^n| = |\boldsymbol{\varphi}_\alpha^n - \boldsymbol{\varphi}_{\alpha-1}^n| \quad (21)$$

for all  $n \geq 0$  and  $\alpha=1, \dots, N-1$ . For completeness, let us recall the details of this step when one uses linear interpolation. Denote by  $L(\alpha)$  the length up to  $\boldsymbol{\varphi}_\alpha^*$  of the piecewise linear curve interpolated across these points, i.e.,  $L(0)=0$  and for  $\alpha=1, \dots, N$ ,

$$L(\alpha) = \sum_{\alpha'=1}^{\alpha} |\boldsymbol{\varphi}_{\alpha'}^* - \boldsymbol{\varphi}_{\alpha'-1}^*|. \quad (22)$$

Then  $L(N)$  is the total length of the piecewise linear curve and  $\ell(\alpha) = L(N)\alpha/N$ ,  $\alpha=1, \dots, N$  is the distance along this curve from the end point  $\boldsymbol{\varphi}_0^*$  up to the point at which we need to set the new  $\alpha^{\text{th}}$  point  $\boldsymbol{\varphi}_\alpha^{n+1}$  in order to make these points equidistant along the curve. This is done by setting  $\boldsymbol{\varphi}_0^{n+1} = \boldsymbol{\varphi}_0^*$  and  $\boldsymbol{\varphi}_N^{n+1} = \boldsymbol{\varphi}_N^*$  (by construction the end points are unaffected by the reparametrization step), and for  $\alpha=1, \dots, N-1$ , taking

$$\boldsymbol{\varphi}_\alpha^{n+1} = \boldsymbol{\varphi}_{a(\alpha)-1}^* + (\ell(\alpha) - L(a(\alpha)-1)) \frac{\boldsymbol{\varphi}_{a(\alpha)}^* - \boldsymbol{\varphi}_{a(\alpha)-1}^*}{|\boldsymbol{\varphi}_{a(\alpha)}^* - \boldsymbol{\varphi}_{a(\alpha)-1}^*|}, \quad (23)$$

where  $a(\alpha)=1, \dots, N$  is such that  $L(a(\alpha)-1) < \ell(\alpha) \leq L(a(\alpha))$ . We then have  $|\boldsymbol{\varphi}_{\alpha+1}^{n+1} - \boldsymbol{\varphi}_\alpha^{n+1}| = |\boldsymbol{\varphi}_\alpha^{n+1} - \boldsymbol{\varphi}_{\alpha-1}^{n+1}|$  up to  $O(1/N^2)$  for all  $\alpha=1, \dots, N-1$ . Note that the reparametrization step can be done in  $O(N)$  operations.



It should also be stressed that the motion of the end points of the string,  $\varphi_0^n$  and  $\varphi_N^n$ , is unaffected by the term  $r_\alpha^*$  in Eq. (18) (since  $r_0^* = r_N^* = 0$  by definition) and by the reparametrization step. In other words, these images simply follow the mean positions  $\bar{x}_0^n$  and  $\bar{x}_N^n$  in the cells  $B_0^n$  and  $B_N^n$ . Because the replicas in each of the cells are free to move except for the constraint that they must remain in their cell, the end points of the string move toward the metastable sets even if initially they are not exactly in these sets (this was the case for the FTS simulation on the rugged Müller potential reported in Fig. 2). Thus, the improved FTS algorithm allows one to identify the metastable sets, or at least refine upon the initial guess we made for these sets.

Finally, a remark about step 5. After steps 3 and 4, since  $B_\alpha^{n+1} \neq B_\alpha^n$  in general, we may have that  $x_\alpha^{n+1} \notin B_\alpha^{n+1}$  even if we used Eq. (15). To avoid problems, we therefore test again whether  $x_\alpha^{n+1} \in B_\alpha^{n+1}$  and if this is not the case, we simply reset  $x_\alpha^{n+1} = \varphi_\alpha^{n+1}$ . Note that this resetting becomes more and more unlikely as  $n$  increases since the images  $\varphi_\alpha^n$  and the cells  $B_\alpha^n$  converge and change less and less at every update.

### C. Illustrative examples

The improved FTS algorithm was first tested on the rugged Müller potential, and the results are shown in Fig. 2. This figure should be compared with Fig. 1 which, we recall, shows a MEP in the rugged Müller potential. For this computation, we used  $N+1=40$  images along the string,  $\Delta t = 10^{-4}$ ,  $\Delta \tau = 0.1$ ,  $\kappa = 0.1$  and we ran the algorithm for  $n=10^4$  steps. The figure shows the images  $\varphi_\alpha^n$  along the string at convergence, the Voronoi cells  $B_\alpha^n$  of these images, and the location of the last  $10^3$  updates of the replicas  $x_\alpha^n$ . The locations of these replicas give an idea of the width of the transition tube, and we checked that the reactive trajectories (which are not shown in Fig. 2) remain mostly inside the tube as they should. Notice that the validity of the assumption on which FTS relies [see Eq. (A2) in Appendix A] requires that the probability to find the replica  $x_\alpha^n$  near a corner of the cell  $B_\alpha^n$  is small. This is approximately true here even though the transition tube is fairly wide. The only dangerous region is around the elbow of the string at  $(x,y)=(-1,0.75)$  where the replicas in these cells sometimes visit the regions inside the elbow where several Voronoi cells come in contact. Note however that this also shows that the FTS algorithm is robust even when several cells come in contact.

To show that the FTS algorithm can also be used in high dimension, we applied the method to the potential

$$V(\mathbf{x}) = \frac{1}{4}(1 - x_1^2)^2 + \frac{1}{2} \sum_{i=2}^d g_i(x_i + a_i x_1^2 + b_i x_1)^2. \quad (24)$$

Here  $d$  is the dimensionality of the system, and  $a_i$ ,  $b_i$ , and  $g_i$  are parameters which were chosen randomly in the intervals  $[-1,1]$ ,  $[-1,1]$ , and  $[0.3,1]$ , respectively, and which control the shape and the curvature of the potential. The potential in Eq. (24) has two minima located at  $x_1 = \pm 1$  and  $x_i = -a_i \mp b_i$ ,  $i=2, \dots, d$  and a saddle point at zero. We tested the FTS method on potential (24) with  $d=100$  and with a temperature  $\beta^{-1}=0.025$  [for comparison the energy barrier between the two minima of Eq. (24) is 0.25]. We used various

values of the number  $N$  of images between  $N+1=25$  and  $N+1=60$ , with  $\Delta t=10^{-3}$  and the same values  $\Delta \tau=0.1$ ,  $\kappa=0.1$  as before. Because the string cannot be visualized directly and its low dimensional projections are rather uninformative, we do not show them. However, we checked that the results of FTS were robust against the choice of  $N$  by comparing the free energy associated to the transition tube calculated at different values of  $N$  (see Fig. 6 in Sec. IV). Note that the computational cost of the algorithm scales linearly with  $d$ .

## IV. FREE ENERGY CALCULATION

The free energy  $G(s)$  associated with  $s_\gamma(\mathbf{x})$  is defined as

$$G(s) = -k_B T \log Z^{-1} \int_{\Omega} e^{-\beta V(\mathbf{x})} \delta(s - s_\gamma(\mathbf{x})) d\mathbf{x}. \quad (25)$$

In this section, we show how to relate this free energy to the free energy of the Voronoi tessellation and how to compute it. Here we present the calculation for the Cartesian case: The generalization to collective variables is straightforward and will be presented in Sec. V C. Note that the free energy  $G(s)$  can be related to the committor of the reaction, see Appendix A for details.

### A. Computing the free energy via flux matching

Define

$$G_\alpha = -k_B T \log Z^{-1} \int_{B_\alpha} e^{-\beta V(\mathbf{x})} d\mathbf{x}, \quad (26)$$

where  $B_\alpha$  is the Voronoi cell associated with the image  $\varphi_\alpha$  along the string and, for economy, we have dropped the superscript  $n$  since we assume that we work with the converged string (i.e., no more updates of it are needed). If we denote by  $\varphi(s)$  a continuous string interpolated through the images  $\{\varphi_\alpha\}_{\alpha=0, \dots, N}$ , then Eq. (26) is related to Eq. (25) as

$$e^{-\beta G_\alpha} = \int_{s_{\alpha-(1/2)}}^{s_{\alpha+(1/2)}} e^{-\beta G(s)} ds, \quad (27)$$

where  $s_{\alpha-(1/2)}$  and  $s_{\alpha+(1/2)}$  are the values of  $s$  along  $\varphi(s)$  which delimit the boundary of the cell  $B_\alpha$  (i.e.,  $B_\alpha = \{\mathbf{x} : s_{\alpha-(1/2)} \leq s_\gamma(\mathbf{x}) \leq s_{\alpha+(1/2)}\}$ ). Notice that since the images  $\varphi_\alpha$  are equidistant from one another and  $|\varphi'(s)| = c s t$ ,  $s_{\alpha+(1/2)} - s_{\alpha-(1/2)} = 1/N$ .

The quantity  $e^{-\beta G_\alpha}$  is the equilibrium probability to find the system inside the cell  $B_\alpha$ , and it can be computed as a by-product of the updating Eqs. (14) and (15) for  $x_\alpha^n$  in step 2 of the improved FTS algorithm. For convenience, let us recall these steps assuming that the Voronoi cell in the acceptance/rejection step is taken to be the fixed cell  $B_\alpha$ : given  $x_\alpha^n$ , compute

$$x_\alpha^* = x_\alpha^n - \Delta t \nabla V(x_\alpha^n) + \sqrt{2\beta^{-1}\Delta t} \xi_\alpha^n \quad (28)$$

and set

$$\mathbf{x}_{\alpha}^{n+1} = \begin{cases} \mathbf{x}_{\alpha}^*, & \text{if } \mathbf{x}_{\alpha}^* \in B_{\alpha} \\ \mathbf{x}_{\alpha}^n, & \text{otherwise.} \end{cases} \quad (29)$$

Suppose that as we generate sequences  $\{\mathbf{x}_{\alpha}^0, \mathbf{x}_{\alpha}^1, \mathbf{x}_{\alpha}^2, \dots\}$  for  $\alpha=0, \dots, N$  (starting from  $n=0$  by resetting the clock once the string has converged), we compute the number of rejections, i.e.,

$$N_{\alpha\alpha'}^n = \left( \begin{array}{l} \text{number of rejections up to time } n \\ \text{caused by } \mathbf{x}_{\alpha}^* \in B_{\alpha'} \text{ with } \alpha' \neq \alpha \end{array} \right), \quad (30)$$

where  $\alpha, \alpha'=0, \dots, N$ . If we now set

$$k_{\alpha\alpha'} = \frac{N_{\alpha\alpha'}^n}{n\Delta t}, \quad (31)$$

this quantity gives an estimate of the instantaneous rate of escape (or flux) from cell  $B_{\alpha}$  to cell  $B_{\alpha'}$  given that the system is in cell  $B_{\alpha}$  (this estimate would become exact, up to discretization errors, as  $n \rightarrow \infty$ ). Therefore, if we denote by  $(\pi_0, \pi_1, \dots, \pi_N)$  the equilibrium probability to find the system in cell  $B_{\alpha}$  and account for the fact that at statistical steady state, the fluxes in and out each cell  $B_{\alpha}$  are zero, we arrive at the following relation fixing  $\pi_{\alpha}$  in terms of  $k_{\alpha\alpha'}$ :

$$\sum_{\substack{\alpha'=0 \\ \alpha' \neq \alpha}}^N \pi_{\alpha'} k_{\alpha'\alpha} = \sum_{\substack{\alpha'=0 \\ \alpha' \neq \alpha}}^N \pi_{\alpha} k_{\alpha\alpha'}, \quad \sum_{\alpha=0}^N \pi_{\alpha} = 1. \quad (32)$$

Since  $(\pi_0, \pi_1, \dots, \pi_N) = (e^{-\beta G_0}, e^{-\beta G_1}, \dots, e^{-\beta G_N})$  by definition of  $G_{\alpha}$  [see Eq. (26)], Eq. (32) also is a way to compute the free energy of each cell  $B_{\alpha}$ . Note that the value of  $k_{\alpha\alpha'}$  may depend on the specific dynamics used to sample in the cells  $B_{\alpha}$ , but the solution of Eq. (32) is independent of this choice (provided, of course, that the same dynamics is used in each cell and that it is consistent with the canonical density  $Z^{-1}e^{-\beta V(\mathbf{x})}$ ). In particular, while the expressions above share some resemblance to those used in Markov state models,<sup>30,31</sup> we note that these results do not rely on any Markov assumption. This is because we are only interested in free energies, rather than actual rates of transition between the cells (the calculation of the reaction rate will be commented upon below in Sec. VII). More refined approximations of  $G(s)$  can be obtained by binning the trajectories  $s_{\gamma}(\mathbf{x}_{\alpha}^n)$  inside each cell  $B_{\alpha}$ , then using  $e^{-\beta G_{\alpha}}$  to weight these histograms against each other.

### B. Illustrative examples

To illustrate the procedure, we first used a two dimensional example of Eq. (1) with

$$V(x, y) = \frac{1}{4}(1 - x^2)^2 + \frac{1}{2}y^2(1.1 + a \tanh(4x)), \quad (33)$$

where  $a$  is a parameter which controls the shape of the potential (see Fig. 4). We analyzed Eq. (33) for two values of  $a$ :  $a=0$ , for which the width of the potential in the  $y$ -direction is independent of  $x$  (see Fig. 4, top panel), and  $a=1$ , for which the width of the potential in the  $y$ -direction is a function of  $x$  (see Fig. 4, bottom panel). By symmetry, the principal curve must lie along the  $x$ -axis for both values of  $a$ , but the asso-

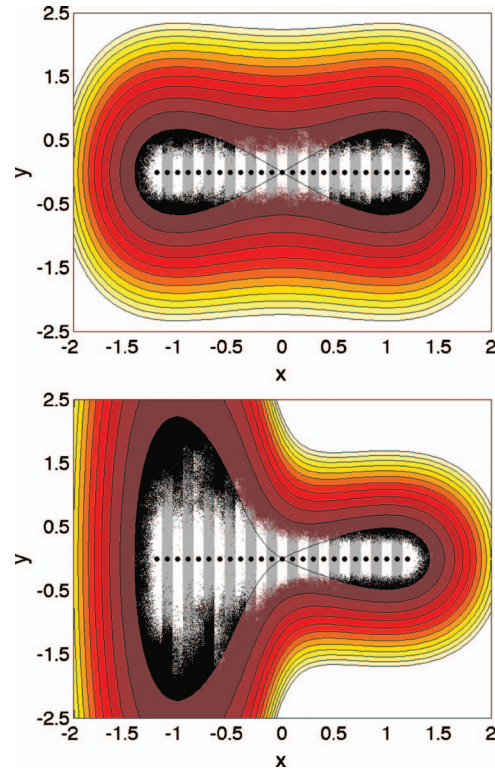


FIG. 4. (Color) Principal curve (black dots) and transition tube in the double-well potential in Eq. (33) with uniform ( $a=0$ , top panel) and non-uniform ( $a=1$ , bottom panel) width along the  $x$ -axis.

ciated free energy is different in the two cases since it is given by

$$G(x) = -\beta^{-1} \log \left( Z^{-1} \int_{\mathbb{R}} e^{-\beta V(x, y)} dy \right) \\ = \frac{1}{4}(1 - x^2)^2 + \frac{1}{2}\beta^{-1} \ln(1.1 + a \tanh(4x)) + C, \quad (34)$$

where  $C$  is an irrelevant constant. When  $a=0$ , the term  $\frac{1}{2}\beta^{-1} \ln(1.1 + a \tanh(4x))$  is constant, so it does not matter, but it gives a nontrivial contribution when  $a=1$ , as illustrated in Fig. 5. First, we checked that FTS does indeed identify the  $x$ -axis as principal curve. Then we used the procedure proposed in Sec. IV A to compute the associated free energy. This was done by using the following Voronoi cells in  $\mathbb{R}^2$ :  $B_0 = \{(x, y) : x \leq x_{\min} + 0.5\Delta x\}$ ,  $B_N = \{(x, y) : x \geq x_{\min} + (N - 0.5)\Delta x\}$ , and  $B_{\alpha} = \{(x, y) : x_{\min} + (\alpha - 0.5)\Delta x \leq x \leq x_{\min} + (\alpha + 0.5)\Delta x\}$  for  $\alpha = 1, \dots, N - 1$  with  $N + 1 = 25$ ,  $x_{\min} = -1.2$ , and  $\Delta x = L_{\gamma}/N$ , where  $L_{\gamma} = 2|x_{\min}|$  is the length of the string. We used  $\beta^{-1} = 0.025$ ,  $\Delta t = 10^{-3}$ , and run the simulation for  $n = 10^7$  steps in each cell [which corresponds to an interval of time for which the chance to observe any transition between the metastable sets around  $(x, y) = (-1, 0)$  and  $(x, y) = (1, 0)$  by the original dynamics with Eq. (33) is extremely small, and so direct sampling with this equation is very inaccurate]. The results obtained by the procedure described in Sec. IV A are shown in Fig. 5 and agree well with the analytical expression (34).

As a second example, we calculated the free energy for the case of the potential in Eq. (24) with  $d=100$  using dif-

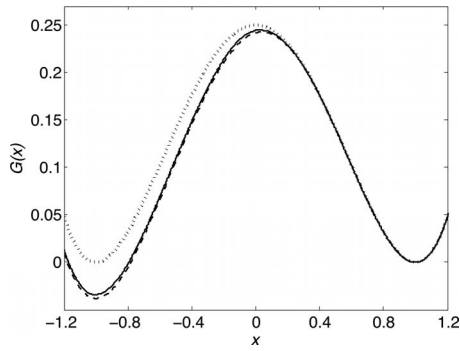


FIG. 5. Free energy  $G(x)$  for the double-well potential in Eq. (33) with uniform width ( $a=0$ , dotted line) and nonuniform width ( $a=1$ , solid line) reconstructed using the procedure described in Sec. IV A. The free energy for  $a=1$  is compared with the analytical expression (34) (dashed line). For  $a=0$  the analytical expression coincides exactly with the FTS calculation and we did not plot it. When the tube has nonuniform width, a larger entropic contribution at the metastable state located at  $(x,y)=(-1,0)$  lowers the free energy of this state.

ferent numbers of images  $N$  along the string. The results are displayed in Fig. 6 and show very little dependency in  $N$  as they should.

## V. GENERALIZATIONS

The results so far are specific to systems governed by the overdamped Eq. (1). Typically, we are interested in analyzing reactions in systems governed by other types of dynamics, e.g.,

$$m\ddot{\mathbf{x}} = -\nabla V(\mathbf{x}(t)) + \text{thermostat}, \quad (35)$$

where we can use any thermostat such that the configurational part of the equilibrium probability density of Eq. (35) is Eq. (2), i.e., it could be Langevin, Nosé–Hoover, isokinetic, etc. We may also want (or need) to introduce collective variables to describe the reaction. One reason to introduce these collective variables is to remove some degeneracies (i.e., invariance by rotation, translation, etc.) which make that the assumption that the reaction occurs by localized channels breaks down. Another reason is that it permits to analyze the reaction in a smaller set of variables such as

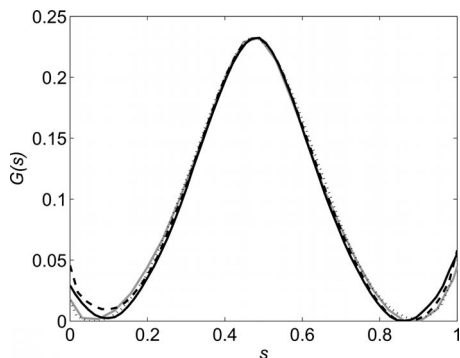


FIG. 6. Free energy for the potential in Eq. (24) with  $d=100$  where a different number of images was used in FTS to discretize the principal curve:  $N+1=25$  (gray solid line),  $N+1=35$  (dotted line),  $N+1=50$  (dashed line), and  $N+1=60$  (black solid line). In each corresponding Voronoi cell a replica of the system was simulated for  $10^7$  steps, with  $\beta^{-1}=0.025$  and  $\Delta t=10^{-3}$ .

dihedral angles, interatomic distances, local density fields, etc. Here we consider how to generalize the results in Secs. II–IV when the dynamics is as in Eq. (35) and when we use collective variables.

## A. Introducing collective variables

In Ref. 4 a framework was introduced to handle systems governed by Eq. (35) and we will use those results as background. Suppose that one starts with a system described by Eq. (35) and assume that the mechanism of the reaction can be described adequately using the collective variables  $\boldsymbol{\theta}(\mathbf{x}) = (\theta_1(\mathbf{x}), \dots, \theta_D(\mathbf{x}))$ , where  $D \in \mathbb{N}$  (note that we assume that these variables depend only on position  $\mathbf{x}$  and not on velocities  $\dot{\mathbf{x}}$ ). As shown in Ref. 4 (see also Appendix A) this is equivalent to assuming that this mechanism (if not the actual dynamics) can be captured by the following pseudodynamics in collective variable space

$$\begin{aligned} \dot{z}_k(t) = & - \sum_{k'=1}^D \left( M_{kk'}(\mathbf{z}(t)) \frac{\partial A(\mathbf{z}(t))}{\partial z_{k'}} - \beta^{-1} \frac{\partial M_{kk'}(\mathbf{z}(t))}{\partial z_{k'}} \right) \\ & + \sqrt{2\beta^{-1}} \sum_{k'=1}^D \sigma_{kk'}(\mathbf{z}(t)) \eta_{k'}(t). \end{aligned} \quad (36)$$

Here  $A(\mathbf{z})$  is the free energy associated with the collective variables [not to be confused with the free energy of the reaction  $G(s)$  defined in Eq. (25)] defined as

$$A(\mathbf{z}) = -k_B T \log Z^{-1} \int_{\Omega} e^{-\beta V(\mathbf{x})} \delta(\boldsymbol{\theta}(\mathbf{x}) - \mathbf{z}) d\mathbf{x}, \quad (37)$$

where  $\delta(\boldsymbol{\theta}(\mathbf{x}) - \mathbf{z}) = \prod_{k=1}^D \delta(\theta_k(\mathbf{x}) - z_k)$ . In Eq. (36) we have also introduced the tensor  $M(\mathbf{z}) = (M_{kk'}(\mathbf{z}))$  with entries

$$M_{kk'}(\mathbf{z}) = \langle \nabla \theta_k(\mathbf{x})^T m^{-1} \nabla \theta_{k'}(\mathbf{x}) | \boldsymbol{\theta}(\mathbf{x}) = \mathbf{z} \rangle, \quad (38)$$

where  $\langle \cdot | \boldsymbol{\theta}(\mathbf{x}) = \mathbf{z} \rangle$  denotes the conditional average

$$\langle g(\mathbf{x}) | \boldsymbol{\theta}(\mathbf{x}) = \mathbf{z} \rangle = \frac{\int_{\Omega} g(\mathbf{x}) e^{-\beta V(\mathbf{x})} \delta(\boldsymbol{\theta}(\mathbf{x}) - \mathbf{z}) d\mathbf{x}}{\int_{\Omega} e^{-\beta V(\mathbf{x})} \delta(\boldsymbol{\theta}(\mathbf{x}) - \mathbf{z}) d\mathbf{x}}. \quad (39)$$

The tensor  $M(\mathbf{z})$  accounts for the curvilinear nature of the collective variables. Finally, the tensor  $\sigma(\mathbf{z}) = (\sigma_{kk'}(\mathbf{z}))$  is such that  $\sum_{k''} \sigma_{kk''}(\mathbf{z}) \sigma_{k''k'}(\mathbf{z}) = M_{kk'}(\mathbf{z})$  and  $\eta_k(t)$  is a white noise.

The presence of the tensor  $M(\mathbf{z})$  affects in a nontrivial way the mechanism of the reaction. To proceed and use the results of Secs. II–IV in the context of Eq. (36), we now need to distinguish two cases, depending on whether the tensor  $M_{kk'}(\mathbf{z})$  in Eq. (38) is independent of  $\mathbf{z}$  or not (both situations arise in applications, see the examples in Secs. VI A and VI B).

### 1. The case of a constant $M(\mathbf{z}) \equiv \bar{M}$

If  $M(\mathbf{z})$  is independent of  $\mathbf{z}$ , i.e., if  $M(\mathbf{z}) = \bar{M}$  for some constant tensor  $\bar{M}$ , then Eq. (36) reduces to



$$\dot{z}_k(t) = - \sum_{k'=1}^D \bar{M}_{kk'} \frac{\partial A(z(t))}{\partial z_{k'}} + \sqrt{2\beta^{-1}} \sum_{k'=1}^D \bar{\sigma}_{kk'} \eta_{k'}(t), \quad (40)$$

where  $\bar{\sigma} = (\bar{\sigma}_{kk'})$  satisfies  $\sum_{k''=1}^D \bar{\sigma}_{kk''} \bar{\sigma}_{k''k'} = \bar{M}_{kk'}$ . Let us introduce the variables

$$u_k = \sum_{k'=1}^D \bar{\sigma}_{kk'}^{-1} z_{k'}, \quad k = 1, \dots, D, \quad (41)$$

where  $\bar{\sigma}_{kk'}^{-1}$  denotes the entries of the tensor  $\bar{\sigma}^{-1}$ , and the rescaled free energy

$$\mathcal{A}(\mathbf{u}) = A(\bar{\sigma}\mathbf{u}), \quad (42)$$

where  $\bar{\sigma}\mathbf{u}$  denotes the vector with components  $\sum_{k'=1}^D \sigma_{kk'} u_{k'}$ . Then it can be readily checked that in terms of the new variables  $\mathbf{u}$ , Eq. (40) reads

$$\dot{\mathbf{u}} = -\nabla_{\mathbf{u}} \mathcal{A}(\mathbf{u}) + \sqrt{2\beta^{-1}} \boldsymbol{\eta}(t). \quad (43)$$

Substituting  $\mathbf{x}$  for  $\mathbf{u}$  and  $V(\mathbf{x})$  for  $\mathcal{A}(\mathbf{u})$ , Eq. (43) becomes identical to Eq. (1), i.e., we are back to the situation considered in Sec. II.

In practice, rather than changing variables, it is simpler to change the metric and define the distance between any two points  $\mathbf{z}$  and  $\mathbf{z}'$  in collective variable space as

$$d(\mathbf{z}, \mathbf{z}') = \left( \sum_{k,k'=1}^D (z_k - z'_k) \bar{M}_{kk'}^{-1} (z_{k'} - z'_{k'}) \right)^{1/2}. \quad (44)$$

Given a curve  $\gamma_z$  in collective variable space parametrized as  $\gamma_z = \{\boldsymbol{\varphi}(s) : s \in [0, 1]\}$ , let us replace condition (4) defining the function  $s_\gamma(\mathbf{x})$  by the following condition defining a function  $s_\gamma(\mathbf{z})$ :

$$d(\mathbf{z}, \boldsymbol{\varphi}(s_\gamma(\mathbf{z}))) = \min_{s \in [0, 1]} d(\mathbf{z}, \boldsymbol{\varphi}(s)). \quad (45)$$

Finally, let us define the principal curve by the equivalent of Eq. (6),

$$\boldsymbol{\varphi}(s) = \langle \mathbf{z} | s_\gamma(\mathbf{z}) = s \rangle \text{ for every } s \in (0, 1), \quad (46)$$

where  $\langle \cdot | s_\gamma(\mathbf{z}) = s \rangle$  denotes the expectation with respect to  $\varrho(\mathbf{z}) = e^{-\beta A(\mathbf{z})}$  conditional on  $s_\gamma(\mathbf{z}) = s$ ,

$$\langle \mathbf{z} | s_\gamma(\mathbf{z}) = s \rangle = \frac{\int_{\Omega_z} \mathbf{z} e^{-\beta A(\mathbf{z})} \delta(s_\gamma(\mathbf{z}) - s) d\mathbf{z}}{\int_{\Omega_z} e^{-\beta A(\mathbf{z})} \delta(s_\gamma(\mathbf{z}) - s) d\mathbf{z}}, \quad (47)$$

where  $\Omega_z \subset \mathbb{R}^D$  is the domain of definition of the collective variables  $\boldsymbol{\theta}(\mathbf{x})$ . Then it is easy to see that the principal curve associated with  $\varrho(\mathbf{z}) = e^{-\beta A(\mathbf{z})}$  in the new metric (44), i.e., the curve satisfying Eq. (46), is, upon changing variables as in Eq. (41), the same as the principal curve associated with  $\exp(-\beta \mathcal{A}(\mathbf{u}))$  using the Euclidean metric in the variables  $\mathbf{u}$ . Note also that, using definition (37) of  $A(\mathbf{z})$ , expectation (47) can be written as

$$\langle \mathbf{z} | s_\gamma(\mathbf{z}) = s \rangle = \frac{\int_{\Omega} \boldsymbol{\theta}(\mathbf{x}) e^{-\beta V(\mathbf{x})} \delta(s_\gamma(\boldsymbol{\theta}(\mathbf{x})) - s) d\mathbf{x}}{\int_{\Omega} e^{-\beta V(\mathbf{x})} \delta(s_\gamma(\boldsymbol{\theta}(\mathbf{x})) - s) d\mathbf{x}}, \quad (48)$$

i.e., it can be computed by working in the original Cartesian space. The formulas above offer the possibility to generalize the improved FTS algorithm presented in Sec. III by defining

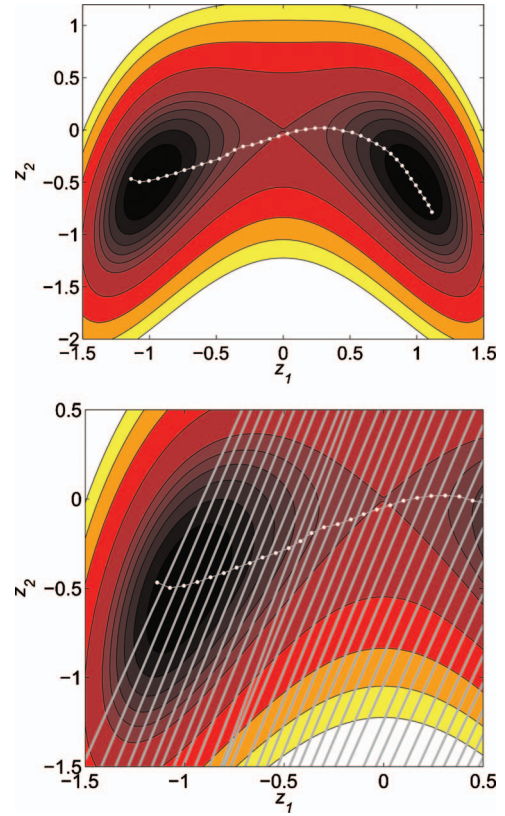


FIG. 7. (Color) The images (white dots) along the string (white line) when the dynamics is governed by Eq. (40) with the double-well potential (49) and the tensor  $\bar{M}$  defined in Eq. (50). The bottom panel shows the region around the minimum located at  $(-1, -0.5)$  with the associated Voronoi cells (gray lines). The presence of the tensor  $\bar{M}$  forces us to change the metric and use Eq. (44). This is why the cells do not look perpendicular (by eyeball, i.e., in Euclidean metric) to the string in this example, and we have to transform the coordinates to make  $\bar{M}$  reduce to the identity in order to recover the familiar picture (see Fig. 8). Note that the FTS calculation was performed in the original space where  $\bar{M}$  is not the identity.

the Voronoi cells according to metric (44). We will show how to do that in Sec. V B.

To illustrate the effect of the tensor  $\bar{M}$  on the principal curve, we applied it to an example of Eq. (40) in two dimensions with the following hypothetical free energy

$$A(\mathbf{z}) = \frac{1}{4}(1 - z_1^2)^2 + \frac{1}{2}\left(z_2 + \frac{1}{2}z_1^2\right)^2 \quad (49)$$

and the following matrix  $\bar{M}$ :

$$\bar{M} = \begin{pmatrix} 0.25 & 0.5 \\ 0.5 & 1.3 \end{pmatrix}. \quad (50)$$

As can be seen in Fig. 7, the presence of  $\bar{M}$  changes the location of both the principal curve and the Voronoi cells. In particular, one can see that the edges of the Voronoi cells are not perpendicular to the principal curve in the original Euclidean metric, which is to be expected since perpendicularity has to be defined according to the new metric (44). If however, we redisplay the result after making the change in coordinates defined in Eq. (41), we see that perpendicularity in the Euclidean metric is restored (although this makes the tube and the potential look different), as shown in Fig. 8.

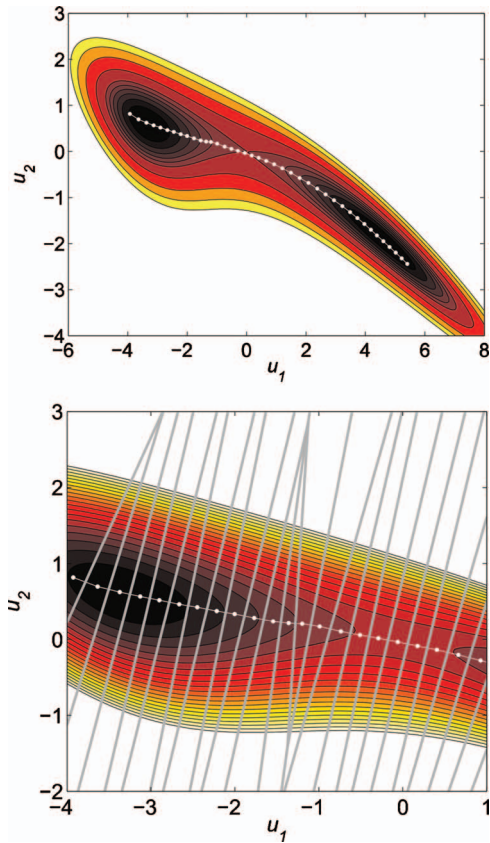


FIG. 8. (Color) Same as in Fig. 7 but after the coordinate transformation (41) which makes  $\bar{M}$  become the identity. Now the edges of the Voronoi cells look perpendicular (again by eyeball, i.e., in Euclidean metric) to the string (bottom panel).

## 2. The case of a nonconstant $M(z)$

If the tensor  $M(z)$  depends on  $z$  the situation is more complex. Indeed, the change in variables that should now be employed to turn Eq. (36) into Eq. (43) is not as simple as Eq. (41). Rather it involves solving the system of differential equations,

$$\frac{\partial u_k}{\partial z_{k'}} = \sigma_{kk'}^{-1}(z). \quad (51)$$

Solving such equations is not easily doable in practice, especially since the computation of  $\sigma$  requires sampling. To avoid this difficulty we will therefore recur to an additional approximation in situations where  $M(z)$  depends on  $z$ , namely, we will assume that we can replace  $M(z)$  by some constant matrix  $\bar{M}$  which can be obtained, e.g., by averaging  $M(z)$  (see Sec. V B). This approximation should retain the main geometric features of the problem encompassed in  $M(z)$  while at the same time making practical calculations possible along the lines discussed in Sec. V A 1. This assumption is, however, not easily justifiable *a priori*.

## B. FTS in collective variable space

Assuming that  $M(z) = \bar{M}$  as in Sec. V A 1, we can make the following modifications in the new FTS algorithm of Sec. III to identify the principal curve in collective variables satisfying Eq. (47):

- Given the images  $\varphi_\alpha^n$  of the string in collective variable space (i.e.,  $\varphi_\alpha^n \in \Omega_z$ ), we define their Voronoi cells  $B_\alpha^n$  in the original Cartesian space of the system, i.e., we replace Eq. (12) by

$$B_\alpha^n = \{x \text{ such that } d(\theta(x), \varphi_\alpha^n) < d(\theta(x), \varphi_{\alpha'}^n) \text{ for all } \alpha' \neq \alpha\}, \quad (52)$$

where the distance  $d(z, z')$  is computed as in Eq. (44);

- in step 1 we use the Voronoi cells (52) in Eq. (15);
- In step 2, we compute the running average as

$$\bar{z}_\alpha^n = \frac{1}{n} \sum_{n'=0}^n \theta(x_{\alpha'}^{n'}) \quad (53)$$

and in step 3, we update the string as

$$\varphi_\alpha^* = \varphi_\alpha^n - \Delta\tau(\varphi_\alpha^n - \bar{z}_\alpha^n) + r_\alpha^*, \quad (54)$$

where  $\Delta\tau > 0$  and  $r_\alpha^*$  is as in Eq. (19);

- we obtain  $\{\varphi_\alpha^{n+1}\}_{\alpha=0,\dots,N}$  from  $\{\varphi_\alpha^*\}_{\alpha=0,\dots,N}$  by reparametrization as in step 4. Note that in this step one can use the Euclidean distance.

We stress that this algorithm can again be implemented via a simple modification of an existing MD code since the main modification consists in verifying that  $x \in B_\alpha^n$ , and this is again a simple distance check. We also stress that the way the variables  $x$  are updated in Eq. (14) is not the only possible one since any sampling procedure consistent with the NVT ensemble will do. It will often be more practical to replace Eq. (1) by Eq. (35), and hence Eq. (14) by a time-discretized version of Eq. (35). In this case, besides the positions  $x_\alpha^n$ , there are also velocities  $v_\alpha^n$ , and after having updated  $(x_\alpha^n, v_\alpha^n)$  into  $(x_\alpha^*, v_\alpha^*)$  by standard MD, the acceptance/rejection step (15) should also include a rule for the velocities. A simple choice is to take

$$v_\alpha^{n+1} = \begin{cases} v_\alpha^*, & \text{if } x_\alpha^* \in B_\alpha^n \\ -v_\alpha^*, & \text{otherwise.} \end{cases} \quad (55)$$

If we are in a situation where  $M(z)$  depends on  $z$  but decide to approximate it by a constant,  $M(z) \approx \bar{M}$ , as explained in Sec. V A 2, then we can supplement the algorithm above with some steps to estimate the approximate  $\bar{M}$  as we go along. One possibility is to define the current value of  $\bar{M}^n$  at iteration  $n$  via the double average

$$\bar{M}_{kk'}^n = \frac{1}{n(N+1)} \sum_{\alpha=0}^N \sum_{n'=0}^n \nabla \theta_k(x_\alpha^{n'})^T m^{-1} \nabla \theta_{k'}(x_\alpha^{n'}). \quad (56)$$

This updating rule for  $\bar{M}$  is consistent with our assumption that this tensor is approximately constant and it then uses all the information available to compute it (hence the average over the images).

## C. Free energy calculation in collective variables

The calculation of free energy presented in Sec. IV can be straightforwardly generalized to the collective variable

case. In this case, the cells  $B_\alpha$  are defined according to Eq. (52) with  $\varphi_\alpha^n$  replaced by the converged string, i.e.,

$$B_\alpha = \{\mathbf{x} \text{ such that } d(\boldsymbol{\theta}(\mathbf{x}), \varphi_{\alpha'}) < d(\boldsymbol{\theta}(\mathbf{x}), \varphi_\alpha) \text{ for all } \alpha' \neq \alpha\}. \quad (57)$$

Thus, to compute  $G_\alpha$ , we can proceed as in Sec. IV provided that in Eqs. (29) and (30) we use Eq. (57) for  $B_\alpha$ . We can also replace Eq. (28) by the relevant updating formula [e.g., a time-discretized version of Eq. (35)].

## VI. APPLICATIONS OF FTS IN COLLECTIVE VARIABLES

### A. Elastic network model of the NtrC<sup>r</sup> protein domain

As a first illustration of the FTS algorithm in collective variable space we analyze the conformational transition of a two-state elastic network model of the nitrogen regulatory protein C receiver domain (NtrC<sup>r</sup>). This allosteric protein is involved in bacterial signal transduction and has two metastable conformations referred to as the inactive and active state. In Ref. 7, the elastic network model for this protein was studied using the swarm string method to calculate the pathway of transition between the two metastable states and the associated free energy. Elastic network models are simplified representations of a protein where only the  $\alpha$ -carbons ( $C_\alpha$ ) of the protein backbone are explicitly represented and the interactions between the atoms are reduced to harmonic potentials with respect to equilibrium conformations of the protein. While on the one hand these models are too simplistic to give a quantitative (and, often, qualitative) description of the dynamics of the true protein, on the other hand they can result quite useful and efficient when one wants to generate initial paths for the string method (or other path methods).

#### 1. Elastic network model

We denote by  $\mathbf{x} \in \mathbb{R}^d$  the coordinates of the full system, where  $d$  is three times the number  $M$  of residues in the protein (which is equal to the number of  $C_\alpha$ 's) and by  $\mathbf{x}_i \in \mathbb{R}^3$  the coordinates of the  $i^{\text{th}}$  residue  $C_\alpha$ . Following Ref. 7, the total energy of the system is given by

$$U(\mathbf{x}) = -\frac{1}{\beta_m} \ln(e^{-\beta_m U^A(\mathbf{x})} + e^{-\beta_m U^B(\mathbf{x})}) + U^R(\mathbf{x}), \quad (58)$$

where the superscripts  $A$  and  $B$  relate to the two reference conformations of the protein. For the NtrC<sup>r</sup> domain  $A$  and  $B$  were taken as the positions of the 124  $C_\alpha$  atoms from the NMR structures<sup>32</sup> 1DC7.pdb ( $A$ ) and 1DC8.pdb ( $B$ ) shown in Fig. 9. The parameter  $\beta_m$  in Eq. (58) controls the overlap of the two potentials  $U^A(\mathbf{x})$  and  $U^B(\mathbf{x})$  [when  $\beta_m \rightarrow \infty$ , the first function at the right-hand side of Eq. (58) reduces to  $\min(U^A(\mathbf{x}), U^B(\mathbf{x}))$ ]. These two potentials are given by

$$U^A(\mathbf{x}) = \frac{1}{2} \sum_{ij}^M k_{ij} D_{ij}^A (\Delta x_{ij} - \Delta x_{ij}^A)^2 \quad (59)$$

and similarly for  $U^B(\mathbf{x})$ . Here  $\Delta x_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ ;  $\Delta x_{ij}^A = |\mathbf{x}_i^A - \mathbf{x}_j^A|$  and  $\Delta x_{ij}^B = |\mathbf{x}_i^B - \mathbf{x}_j^B|$  are the distances between corresponding

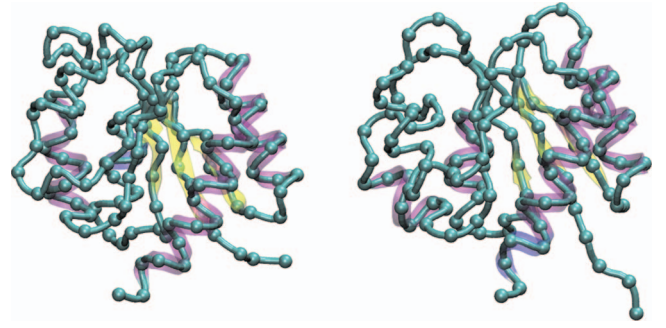


FIG. 9. (Color) Conformations from the NMR structures (Ref. 32) of the NtrC<sup>r</sup> protein domain in the inactive ( $A$ , left) and active ( $B$ , right) states. The  $C_\alpha$  atoms used as interaction sites in the elastic network model are shown as spheres.

residues in states  $A$  and  $B$ , respectively;  $D_{ij}^A$  and  $D_{ij}^B$  are the contact matrices given by

$$D_{ij}^A = \begin{cases} 1, & \Delta x_{ij}^A < d^A \\ 0, & \text{otherwise,} \end{cases} \quad (60)$$

where  $d^A$  is some cutoff distance, and similarly for  $D_{ij}^B$ ;  $k_{ij}$  are the force constants, which are the same for  $A$  and  $B$ , but depend on the interacting residues through

$$k_{ij} = \min\left(\frac{\varepsilon_k}{(\Delta x_{ij}^A - \Delta x_{ij}^B)^2}, k_{\max}\right). \quad (61)$$

The term  $U^R(\mathbf{x})$  in Eq. (58) takes into account repulsive contributions due to steric hindrance and avoids unphysical conformations. Explicitly it is given by

$$U^R(\mathbf{x}) = \varepsilon \sum_{\substack{i,j=1 \\ i \neq j}}^M \left(\frac{\sigma}{\Delta x_{ij}}\right)^{12}. \quad (62)$$

The parameters in the potentials above were taken with the values:  $d^{A,B} = 11.5 \text{ \AA}$ ,  $\varepsilon_k = 0.5 \text{ kcal mol}^{-1}$ ,  $k_{\max} = 0.2 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ,  $\varepsilon = 1.0 \text{ kcal mol}^{-1}$ ,  $\sigma = 2.5 \text{ \AA}$ ,  $\beta_m = 0.02 \text{ kcal}^{-1} \text{ mol}$ , and the masses of the particles were all taken equal to  $m = 100 \text{ amu}$ . Note that the values of these parameters are the same as in Ref. 7, with the exception of  $\beta_m$ , which in Ref. 7 is  $0.005 \text{ kcal}^{-1} \text{ mol}$ . By performing energy minimization of the two reference structures using the smaller value of the coupling parameter we have found that it is too small to be able to maintain the reference conformations and to capture the full transition, and results in equilibrium conformations which are a mix of the two reference states  $A$  and  $B$ . The value  $\beta_m = 0.02 \text{ kcal}^{-1} \text{ mol}$ , on the other hand, stabilizes the two reference structures.

#### 2. FTS calculation

To apply FTS, we introduce collective variables to remove degeneracy by rotations and translations. This was done by defining the collective variables  $\boldsymbol{\theta}(\mathbf{x})$  from the Cartesian variables  $\mathbf{x}$  via a root mean square minimization procedure using  $\mathbf{x}^A$ , the coordinates of the system in conformation  $A$ , as reference structure:  $\boldsymbol{\theta}(\mathbf{x}) = U\mathbf{x} + \mathbf{X}$  where  $U$  is a rotation matrix and  $\mathbf{X} \in \mathbb{R}^d$  is translation vector adjusted so that  $|\boldsymbol{\theta}(\mathbf{x}) - \mathbf{x}^A|$  is minimum (thus, in the present case  $\boldsymbol{\theta}(\mathbf{x})$  has



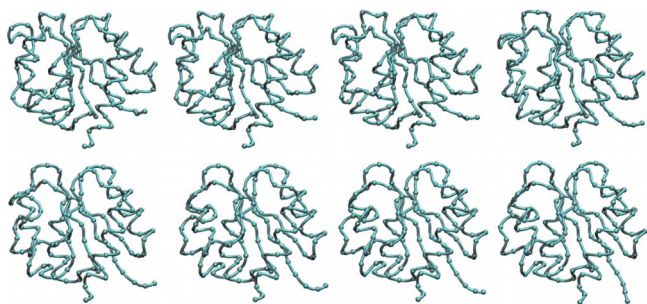


FIG. 10. (Color online) Snapshots along the principal curve for the conformational transition of the NtrC<sup>r</sup> protein domain. From left to right and top to bottom:  $\alpha=0, 5, 6, 11, 14, 18, 20, 23$ .

the same dimensionality as  $\mathbf{x}$ , i.e., it involves the positions of the 124 residues in the elastic network model). Therefore, instead of using Eq. (44) to compute the distance, in the present situation it is simpler to use

$$d(\theta(\mathbf{x}), \varphi_{\alpha}^n) = \min_{\substack{\mathbf{y} \text{ such that} \\ \theta(\mathbf{y})=\mathbf{x}}} |\mathbf{y} - \varphi_{\alpha}^n|. \quad (63)$$

The minimization on  $\mathbf{y}$  then simply amounts to finding the structure which is the closest to  $\varphi_{\alpha}^n$  after rotation and translation of  $\mathbf{x}$ . Note that with the distance in Eq. (63)  $M$  is constant and reduces to the identity.

The FTS calculation was started from the MEP (Refs. 25 and 26) with the string in the full coordinate space and represented using  $N+1=24$  images. The replicas of the system in the Voronoi cells associated with the string were simulated using Langevin dynamics with a friction coefficient  $\gamma = 30 \text{ ps}^{-1}$  and temperature  $T=300 \text{ K}$ . The second-order integrator of Ref. 33 was used to integrate the resulting equation of motion, with a time step  $\Delta t=2.5 \text{ fs}$ . The time step for the string evolution was  $\Delta \tau=0.1$  and the smoothing coefficient  $\kappa=0.1$ . Upon rejection of a dynamic steps due to crossing of a cell boundary, the replica was put back to its previous position and its velocity was updated according to Eq. (55). The initial string converged to the principal curve in  $10^6$  iterations of the FTS algorithm. Snapshots of the protein conformation along the principal curve are shown in Fig. 10.

### 3. Free energy calculation

The free energy associated to the principal curve was computed using the sampling procedure described in Sec. V C, where in each Voronoi cell a replica of the system was simulated for  $10^7$  steps with Langevin dynamics. The resulting free energy is plotted in Fig. 11.

### B. Isomerization of the solvated alanine dipeptide

As a second test of the FTS algorithm in collective variables, we used the example of AD in solution. This small peptide ( $\text{CH}_3\text{--CO--NH--C}_{\alpha}\text{HCH}_3\text{--CO--NH--CH}_3$ ) has multiple metastable states and it is often used as a test case for numerical algorithms and methods in MD since it is sufficiently small to allow for efficient atomistic MD simulations, while still possessing a nontrivial dynamics. We refer the interested reader to Refs. 4, 13, and 34–38 for more background material on AD. Conformational transitions of AD

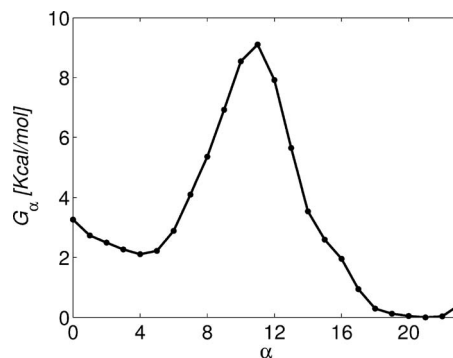


FIG. 11. Free energy  $G_{\alpha}$  associated to the Voronoi tessellation of the images along the principal curve for the elastic network model of the conformational transformation of the NtrC<sup>r</sup> protein domain.

are usually described in terms of its torsion angles along the backbone. Here we use as collective variables two of these angles,  $\varphi$  and  $\psi$ , defined by the quadruplets of atoms (C,N,C<sub>α</sub>,C) and (N,C<sub>α</sub>,C,N), respectively. This is a situation in which  $M(\mathbf{z})$  depends on  $\mathbf{z}$ , and we approximated it by a constant using the updating procedure explained in Sec. V A 2. The final value of the approximation  $\bar{M}$  for  $M(\mathbf{z})$  was

$$\bar{M} = \begin{pmatrix} 0.26 & -0.05 \\ -0.05 & 0.30 \end{pmatrix}. \quad (64)$$

Note that  $\bar{M}$  is approximately diagonal in this case. The transition tube resulting from FTS calculation is shown in Fig. 12 and the free energy of the corresponding tessellation is shown in Fig. 13. The free energy associated to the Voronoi tessellation calculated using the procedure described in Sec. V C is compared to the free energy calculated by brute force simulation, i.e., by binning a very long free trajectory, and the agreement is excellent.

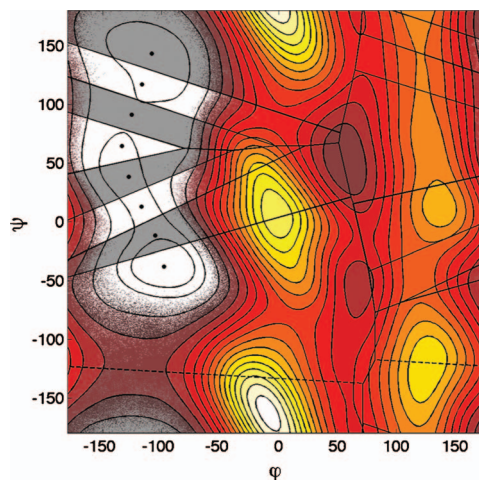


FIG. 12. (Color) Transition tube in the space of the torsion angles  $\varphi$  and  $\psi$  for the AD. The black dots are the images along the principal curve and the black lines show the boundaries of the Voronoi cells. Note that this calculation identified the transition channel with  $\varphi < 0^\circ$ ; another channel exists with  $\varphi > 0^\circ$  which should be calculated separately. The underlying free energy landscape [which is the free energy  $A(\mathbf{z})$  in Eq. (37) with  $\mathbf{z}$  made of the two dihedral angles] was computed using the single sweep method of Ref. 39.

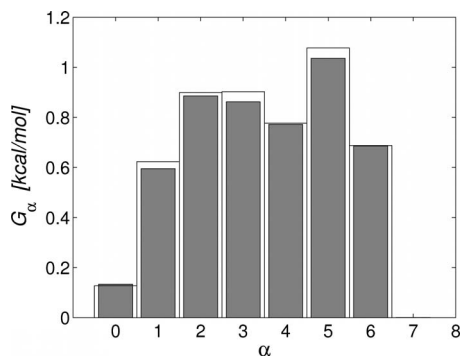


FIG. 13. Free energy of the Voronoi tessellation shown in Fig. 12 for the AD. The free energy calculated using the procedure described in Sec. V C (gray histogram) is in excellent agreement with the free energy calculated by binning a very long free trajectory (white histogram).

### 1. Simulation details

All MD simulations were performed with a version of the MOIL code<sup>40</sup> that was parallelized to handle the different replicas of the system. The AMBER/OPLS (Ref. 41) force field as implemented in the code was used. A starting structure for the AD molecule was solvated in a box of volume  $(20 \text{ \AA})^3$  with 252 water molecules. Periodic boundary conditions were used. Van der Waals interactions were truncated at  $9 \text{ \AA}$ . Electrostatic interactions were treated with the particle mesh Ewald method<sup>42</sup> with real space cutoff  $9 \text{ \AA}$ , a grid of 323 points, and fourth order  $B$ -splines for the interpolation of the structure factor (in order to be in the high accuracy range). The TIP3 model<sup>43</sup> was used for the water molecules, with the bonds kept fixed with the SHAKE algorithm,<sup>44,45</sup> while all the bonds in AD were modeled as harmonic springs. Amide planes were restrained to be always in *trans* configuration. The velocity Verlet algorithm was used for the dynamics of the Cartesian variables with time step 1 fs, and all velocities were scaled at every step to keep the temperature at 300 K. As initial string for the FTS calculation the MFEP obtained by the on-the-fly string method<sup>5</sup> was used.

## VII. OUTLOOK AND CONCLUSIONS

To summarize, we have proposed an improved and simplified version of the FTS method. The new method can be used either in the original Cartesian space or in some collective variable space and, essentially, it amounts to an adaptive sampling in Voronoi cells whose generating points are the images along the string. The sampling can be done by standard MD by simple addition of a velocity reversal step at the boundary of the cells to keep the MD replicas in their respective Voronoi cells. The output of the method is a Voronoi tessellation which is close to being centroidal. The method also permits as a by-product to obtain the free energy associated with this tessellation, which is also the free energy of the reaction (i.e., it properly accounts for averaging over the degrees of freedom perpendicular to the string identified by FTS).

As mentioned before, the method proposed to calculate the free energy associated with a Voronoi tessellation is general and could be used with tessellations generated by other methods than FTS. The tessellation associated with the string

of FTS is quite special, however, since it approximates the committor function of the reaction. As such, this tessellation can be used in other techniques such as milestoning, TIS or FFS to compute the rate of the reaction. Specifically, because the boundaries of the cells approximate locally the isocommittor surfaces, these interfaces should be the optimal ones to use in these techniques (in the case of milestoning, this statement can actually be proved, as was shown in Ref. 46). In the context of milestoning this idea has been exploited in Refs. 21 and 46 and we refer the reader to these papers for details.

## ACKNOWLEDGMENTS

We thank Luca Maragliano for providing us with the data used in the AD example. This work was partially supported by the NSF under Grant Nos. DMS02-09959 and DMS02-39625, and by the ONR under Grant No. N00014-04-1-0565.

## APPENDIX A: LINK WITH TPT

In this Appendix, we briefly review some important concepts of TPT and the role they play in the context of the FTS method. TPT characterizes the statistical properties of the reactive trajectories i.e., those pieces of trajectories by which the transition from the reactant set  $A$  to a product set  $B$  actually occurs. The theory shows that a key object in the statistical description of the reactive trajectories is the committor function, which is the function whose value at any given point is the probability that a trajectory initiated from this point reaches first the product state  $B$  rather than the reactant state  $A$ . In the context of the overdamped Eq. (1), the committor function is the minimizer of the following Dirichlet form:<sup>17,47</sup>

$$E(q) = \int_{\Omega} |\nabla q(\mathbf{x})|^2 Z^{-1} e^{-\beta V(\mathbf{x})} d\mathbf{x}, \quad (\text{A1})$$

where the minimization must be performed over all function  $q(\mathbf{x})$  such that  $q(\mathbf{x})=0$  if  $\mathbf{x} \in A$  and  $q(\mathbf{x})=1$  if  $\mathbf{x} \in B$ . The committor function enters the expressions of the probability density of reactive trajectories, given by  $\rho_R(\mathbf{x}) = Z^{-1} e^{-\beta V(\mathbf{x})} q(\mathbf{x})(1-q(\mathbf{x}))$ , their probability current, given by  $\mathbf{j}_R(\mathbf{x}) = Z^{-1} e^{-\beta V(\mathbf{x})} \nabla q(\mathbf{x})$ , and their rate, given by  $k_{AB} = Z^{-1} \int_{\Omega} e^{-\beta V(\mathbf{x})} |\nabla q(\mathbf{x})|^2 d\mathbf{x}$ . The key approximation used in the string method<sup>11-13</sup> (see also Ref. 15) is that locally around the principal curve, the level sets of  $q(\mathbf{x})$  can be approximated by the level sets of the function  $s_{\gamma}(\mathbf{x})$  defined in Eq. (4). In other words,

$$q(\mathbf{x}) \approx f(s_{\gamma}(\mathbf{x})) \quad (\text{A2})$$

for some function  $f: [0, 1] \rightarrow [0, 1]$  to be determined [see Eq. (A10) below]. In Sec. II, we have already justified at least qualitatively why the choice of the principal curve permits to focus attention on the regions that are of importance for the reaction. Next we show how to compute the function  $f$  in Eq. (A2).

Using ansatz (A2) in Eq. (A1) turns this objective function into an objective function for  $f$ . Our goal is to show that this minimization can be done explicitly and show that the

minimizing  $f$  is the one in Eq. (A10). Before going there, however, we need to take care of a technical difficulty that arises with ansatz (A2). Since  $s_\gamma(\mathbf{x})$  is not differentiable everywhere (indeed it is discontinuous at places, see Fig. 3), Eq. (A2) is not differentiable either and this leads to difficulties in Eq. (A1). To avoid these difficulties, we must make an approximation which embeds the idea that Eq. (A2) is an ansatz that is valid only locally around the principal curve. This approximation can be introduced as follows. Taking the gradient of Eq. (5) with respect to  $\mathbf{x}$  gives

$$\nabla s_\gamma(\mathbf{x}) (|\boldsymbol{\varphi}'(s_\gamma(\mathbf{x}))|^2 + \boldsymbol{\varphi}''(s_\gamma(\mathbf{x})) \cdot (\boldsymbol{\varphi}(s_\gamma(\mathbf{x})) - \mathbf{x})) = \boldsymbol{\varphi}'(s_\gamma(\mathbf{x})). \quad (\text{A3})$$

This relation implies that

$$|\nabla s_\gamma(\mathbf{x})|^2 = \frac{|\boldsymbol{\varphi}'(s_\gamma(\mathbf{x}))|^2}{||\boldsymbol{\varphi}'(s_\gamma(\mathbf{x}))|^2 + \boldsymbol{\varphi}''(s_\gamma(\mathbf{x})) \cdot (\boldsymbol{\varphi}(s_\gamma(\mathbf{x})) - \mathbf{x})|^2} \quad (\text{A4})$$

and this gradient square can indeed blow up when  $|\boldsymbol{\varphi}'(s_\gamma(\mathbf{x}))|^2 + \boldsymbol{\varphi}''(s_\gamma(\mathbf{x})) \cdot (\boldsymbol{\varphi}(s_\gamma(\mathbf{x})) - \mathbf{x}) = 0$ . Since  $\boldsymbol{\varphi}(s_\gamma(\mathbf{x})) = \mathbf{x}$  if  $\mathbf{x}$  is on the curve, this always happens away from the curve. The additional approximation that we make is that this occurs sufficiently far away from the principal curve that it does not matter. More precisely, we assume that, under any average  $\langle \cdot | s_\gamma(\mathbf{x}) = s \rangle$ , the term involving  $\boldsymbol{\varphi}''(s_\gamma(\mathbf{x}))$  in Eq. (A4) is negligible, i.e., we can set

$$\langle |\nabla s_\gamma(\mathbf{x})|^2 | s_\gamma(\mathbf{x}) = s \rangle \approx |\boldsymbol{\varphi}'(s)|^{-2}. \quad (\text{A5})$$

This approximation should in principle be checked *a posteriori*. Using Eq. (A5), we see that the square norm of the gradient of the function  $q(\mathbf{x}) = f(s_\gamma(\mathbf{x}))$  in Eq. (A1) can be approximated as

$$\langle |\nabla q(\mathbf{x})|^2 | s_\gamma(\mathbf{x}) = s \rangle = \langle |\nabla s_\gamma(\mathbf{x})|^2 | s_\gamma(\mathbf{x}) = s \rangle (f'(s))^2 \approx |\boldsymbol{\varphi}'(s)|^{-2} (f'(s))^2. \quad (\text{A6})$$

If we now insert ansatz (A2) into Eq. (A1), and use approximation (A6), this objective function becomes the following objective function for  $f$ :

$$\begin{aligned} E(f(s_\gamma)) &= \int_{\Omega} |\nabla s_\gamma(\mathbf{x})|^2 (f'(s_\gamma(\mathbf{x})))^2 Z^{-1} e^{-\beta V(\mathbf{x})} d\mathbf{x} \\ &= \int_0^1 \langle |\nabla s_\gamma(\mathbf{x})|^2 | s_\gamma(\mathbf{x}) = s \rangle (f'(s))^2 e^{-\beta G(s)} ds \\ &\approx \int_0^1 |\boldsymbol{\varphi}'(s)|^{-2} (f'(s))^2 e^{-\beta G(s)} ds, \end{aligned} \quad (\text{A7})$$

where  $G(s)$  is the free energy defined in Eq. (25). It is now straightforward to minimize Eq. (A7) over  $f$ . The Euler–Lagrange equation for the function  $f$  is the second-order differential equation:

$$(f'(s) |\boldsymbol{\varphi}'(s)|^{-2} e^{-\beta G(s)})' = 0, \quad (\text{A8})$$

$$f(0) = 0, \quad f(1) = 1,$$

and the solution of this equation is

$$f(s) = \frac{\int_0^s |\boldsymbol{\varphi}'(s')|^2 e^{G(s')} ds'}{\int_0^1 |\boldsymbol{\varphi}'(s')|^2 e^{\beta G(s')} ds'}. \quad (\text{A9})$$

If we impose that  $|\boldsymbol{\varphi}'(s')| = cst$ , then this expression can be further simplified into

$$f(s) = \frac{\int_0^s e^{G(s')} ds'}{\int_0^1 e^{\beta G(s')} ds'}, \quad (\text{A10})$$

which relates the free energy of  $s_\gamma(\mathbf{x})$  with the committor function along the curve  $\gamma$  [and also away from it through Eq. (A2)].

Finally we note that if the dynamics is given by Eq. (35) and includes velocities,  $\mathbf{v} = \dot{\mathbf{x}}$ , then the committor is a function of both  $\mathbf{x}$  and  $\mathbf{v}$  which gives the probability to reach first  $B$  rather than  $A$  starting from position  $\mathbf{x}$  with velocity  $\mathbf{v}$ . Denoting this function by  $q(\mathbf{x}, \mathbf{v})$ , the main approximation made in Ref. 4 which we used in Sec. V is that  $q(\mathbf{x}, \mathbf{v}) \approx Q(\boldsymbol{\theta}(\mathbf{x}))$  for some function  $Q(\mathbf{z})$ . This is the approximation that leads to the pseudodynamics in Eq. (36): Indeed, it can be shown that if we make the approximation  $q(\mathbf{x}, \mathbf{v}) \approx Q(\boldsymbol{\theta}(\mathbf{x}))$ , then  $Q(\mathbf{z})$  must be the committor function for Eq. (36).

## APPENDIX B: JUSTIFICATION OF EQUATION (13)

Consider a curve  $\gamma = \{\boldsymbol{\varphi}(s) : s \in [0, 1]\}$ . Pick a point along this curve,  $\boldsymbol{\varphi}^* = \boldsymbol{\varphi}(s^*)$  for some  $s^* \in (0, 1)$ . Pick also  $\delta > 0$  sufficiently small so that  $0 \leq s^* - \delta < s^* + \delta \leq 1$ , and consider the region

$$B_\delta^* = \{\mathbf{x} : s^* - \delta \leq s_\gamma(\mathbf{x}) \leq s^* + \delta\}. \quad (\text{B1})$$

By definition of the Dirac delta function, for any function  $g(\mathbf{x})$  such that these integrals exist we have that

$$\lim_{\delta \rightarrow 0} \frac{1}{\delta} \int_{B_\delta^*} g(\mathbf{x}) e^{-\beta V(\mathbf{x})} d\mathbf{x} = \int_{\Omega} g(\mathbf{x}) e^{-\beta V(\mathbf{x})} \delta(s_\gamma(\mathbf{x}) - s^*) d\mathbf{x}. \quad (\text{B2})$$

As a result

$$\begin{aligned} \lim_{\delta \rightarrow 0} \frac{\int_{B_\delta^*} \mathbf{x} e^{-\beta V(\mathbf{x})} d\mathbf{x}}{\int_{B_\delta^*} e^{-\beta V(\mathbf{x})} d\mathbf{x}} &= \lim_{\delta \rightarrow 0} \frac{\delta^{-1} \int_{B_\delta^*} \mathbf{x} e^{-\beta V(\mathbf{x})} d\mathbf{x}}{\delta^{-1} \int_{B_\delta^*} e^{-\beta V(\mathbf{x})} d\mathbf{x}} \\ &= \frac{\lim_{\delta \rightarrow 0} \delta^{-1} \int_{B_\delta^*} \mathbf{x} e^{-\beta V(\mathbf{x})} d\mathbf{x}}{\lim_{\delta \rightarrow 0} \delta^{-1} \int_{B_\delta^*} e^{-\beta V(\mathbf{x})} d\mathbf{x}} \\ &= \frac{\int_{\Omega} \mathbf{x} e^{-\beta V(\mathbf{x})} \delta(s_\gamma(\mathbf{x}) - s^*) d\mathbf{x}}{\int_{\Omega} e^{-\beta V(\mathbf{x})} \delta(s_\gamma(\mathbf{x}) - s^*) d\mathbf{x}} \\ &= \langle \mathbf{x} | s_\gamma(\mathbf{x}) = s^* \rangle. \end{aligned} \quad (\text{B3})$$

This shows that the conditional expectation of  $\mathbf{x}$  in  $B_\delta^*$  is an approximation of  $\langle \mathbf{x} | s_\gamma(\mathbf{x}) = s^* \rangle$  whose quality improves as  $\delta \rightarrow 0$ . If  $\boldsymbol{\varphi}(s)$  is discretized into  $N+1$  images in such a way that  $\boldsymbol{\varphi}(s^*)$  is one such image, and  $\delta = 1/2N$ , then the region  $B_\delta^*$  is an approximation of the Voronoi cell associated with the image  $\boldsymbol{\varphi}(s^*)$ , the quality of which improves as  $\delta = 1/2N \rightarrow 0$ . Since  $B_\delta^*$  itself is an approximation of the Voronoi cell associated with the image  $\boldsymbol{\varphi}(s^*)$  whose quality improves as  $\delta \rightarrow 0$ , this justifies Eq. (13).



- <sup>1</sup>K. Fukui, *J. Phys. Chem.* **74**, 4161 (1970).
- <sup>2</sup>R. Olender and R. Elber, *J. Mol. Struct.: THEOCHEM* **398–399**, 63 (1997).
- <sup>3</sup>E. Vanden-Eijnden and M. Heymann, *J. Chem. Phys.* **128**, 061103 (2008).
- <sup>4</sup>L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, *J. Chem. Phys.* **125**, 024106 (2006).
- <sup>5</sup>L. Maragliano and E. Vanden-Eijnden, *Chem. Phys. Lett.* **446**, 182 (2007).
- <sup>6</sup>M. Venturoli, E. Vanden-Eijnden, and G. Ciccotti, *J. Math. Chem.* **45**, 188 (2009).
- <sup>7</sup>A. C. Pan, D. Sezer, and B. Roux, *J. Phys. Chem. B* **112**, 3432 (2008).
- <sup>8</sup>D. Branduardi, F. L. Gervasio, and M. Parrinello, *J. Chem. Phys.* **126**, 054103 (2007).
- <sup>9</sup>M. Berkowitz, J. D. Morgan, J. A. McCammon, and S. H. Northrup, *J. Chem. Phys.* **79**, 5563 (1983).
- <sup>10</sup>S. Huo and J. E. Straub, *J. Chem. Phys.* **107**, 5000 (1997).
- <sup>11</sup>W. E, W. Ren, and E. Vanden-Eijnden, *J. Phys. Chem. B* **109**, 6688 (2005).
- <sup>12</sup>W. E, W. Ren and E. Vanden-Eijnden, *Chem. Phys. Lett.* **413**, 242 (2005).
- <sup>13</sup>W. Ren, E. Vanden-Eijnden, P. Maragakis, and W. E, *J. Chem. Phys.* **123**, 134109 (2005).
- <sup>14</sup>T. Hastie and W. Stuetzle, *J. Am. Stat. Assoc.* **84**, 502 (1989).
- <sup>15</sup>W. E and E. Vanden-Eijnden, in *Multiscale Modelling and Simulation*, Lecture Notes in Computational Science and Engineering, edited by S. Attinger and P. Koumoutsakos (Springer, Berlin, 2004), Vol. 39.
- <sup>16</sup>W. E and E. Vanden-Eijnden, *J. Stat. Phys.* **123**, 503 (2006).
- <sup>17</sup>E. Vanden-Eijnden, in *Computer Simulations in Condensed Matter: From Materials to Chemical Biology*, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer, Berlin, 2006), Vol. 1, pp. 439–478.
- <sup>18</sup>P. Metzner, C. Schütte, and E. Vanden-Eijnden, *J. Chem. Phys.* **125**, 084110 (2006).
- <sup>19</sup>Q. Du, V. Faber, and M. Gunzburger, *SIAM Rev.* **41**, 637 (1999).
- <sup>20</sup>A. K. Faradjian and R. Elber, *J. Chem. Phys.* **120**, 10880 (2004).
- <sup>21</sup>E. Vanden-Eijnden and M. Venturoli, *J. Chem. Phys.* **130**, 194101 (2009).
- <sup>22</sup>D. Moroni, T. S. van Erp, and P. G. Bolhuis, *Physica A* **340**, 395 (2004).
- <sup>23</sup>R. J. Allen, D. Frenkel, and P. R. ten Wolde, *J. Chem. Phys.* **124**, 024102 (2006).
- <sup>24</sup>K. Müller, *Angew. Chem.* **19**, 1 (1980).
- <sup>25</sup>W. E, W. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).
- <sup>26</sup>W. E, W. Ren, and E. Vanden-Eijnden, *J. Chem. Phys.* **126**, 164103 (2007).
- <sup>27</sup>B. Kégl, A. Krzyzak, T. Linder, and K. Zeger, *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 281 (2000).
- <sup>28</sup>J. S. Liu, *Monte Carlo Strategies in Scientific Computing* (Springer, New York, 2001).
- <sup>29</sup>K. W. Morton and D. F. Mayers, *Numerical Solution of Partial Differential Equations*, 2nd ed. (Cambridge University Press, Cambridge, 2005).
- <sup>30</sup>W. C. Swope, J. W. Pitera, and F. Suits, *J. Phys. Chem. B* **108**, 6571 (2004).
- <sup>31</sup>W. C. Swope, J. W. Pitera, F. Suits, M. Pitman, M. Eleftheriou, B. G. Fitch, R. S. Germain, A. Rayshubski, T. J. C. Ward, Y. Zhestkov, and R. Zhou, *J. Phys. Chem. B* **108**, 6582 (2004).
- <sup>32</sup>D. Kern, B. F. Volkman, P. Luginbuhl, M. J. Nohaile, S. Kustu, and D. Wemmer, *Nature (London)* **402**, 894 (1999).
- <sup>33</sup>E. Vanden-Eijnden and G. Ciccotti, *Chem. Phys. Lett.* **429**, 310 (2006).
- <sup>34</sup>B. M. Montgomery Pettitt and M. Karplus, *Chem. Phys. Lett.* **121**, 194 (1985).
- <sup>35</sup>R. Czerminski and R. Elber, *J. Chem. Phys.* **92**, 5580 (1990).
- <sup>36</sup>D. J. Tobias and C. L. Brooks, *J. Chem. Phys.* **96**, 3864 (1992).
- <sup>37</sup>P. G. Bolhuis, C. Dellago, and D. Chandler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5877 (2000).
- <sup>38</sup>A. Ma and A. R. Dinner, *J. Phys. Chem. B* **109**, 6769 (2005).
- <sup>39</sup>L. Maragliano and E. Vanden-Eijnden, *J. Chem. Phys.* **128**, 184110 (2008).
- <sup>40</sup>R. Elber, A. Roitberg, C. Simmerling, R. Goldstein, H. Li, C. Verkhivker, C. Keasar, J. Zhang, and A. Ulitsky, *Comput. Phys. Commun.* **91**, 159 (1995).
- <sup>41</sup>W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.* **110**, 1657 (1988).
- <sup>42</sup>U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *J. Chem. Phys.* **103**, 8577 (1995).
- <sup>43</sup>W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- <sup>44</sup>J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- <sup>45</sup>Y. Weinbach and R. Elber, *J. Comput. Phys.* **209**, 193 (2005).
- <sup>46</sup>E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, *J. Chem. Phys.* **129**, 174102 (2008).
- <sup>47</sup>R. Durrett, *Stochastic Calculus* (CRC, Washington, D.C., 1996).