A fuzzy-set theoretical framework for computing exit rates of rare events in potential-driven diffusion processes

Marcus Weber[1,⋆] and *Natalia* Ernst[1,2]

[1] *Zuse Institute Berlin (ZIB), Takustr. 7, 14195 Berlin, Germany*
[2] *Technical University Berlin (TUB), Straße des 17. Juni 135, 10623 Berlin, Germany*

Abstract. This article is about molecular simulation. However, the theoretical results apply for general overdamped Langevin dynamics simulations. Molecular simulation is often used for determining the stability of a complex (e.g., ligand-receptor). The stability can be measured by computing the expected holding time of the complex before its dissociation. This dissociation can be seen as an exit event from a certain part S of the conformational state space Γ. Determining exit rates (i.e, for SDE-based simulations exiting from a given starting set S) for a stochastic process in which the exit event occures very rarely is obviously hard to solve by straight forward simulation methods. Finding a low variance procedure for computing rare event statistics is still an open problem. Imagine now, e.g., a simulation of a diffusion process. As long as the time-dependent state trajectory is inside the starting set S, no information is gained about the rare event statistics. Only at that point of time, when the process leaves the starting set, a piece of information about the exit rate is collected. If S, however, is a fuzzy set given by a membership function, then there might be additional information of the kind "the process is about to leave the set". However, how to define an exit rate from a fuzzy set?

1 Introduction

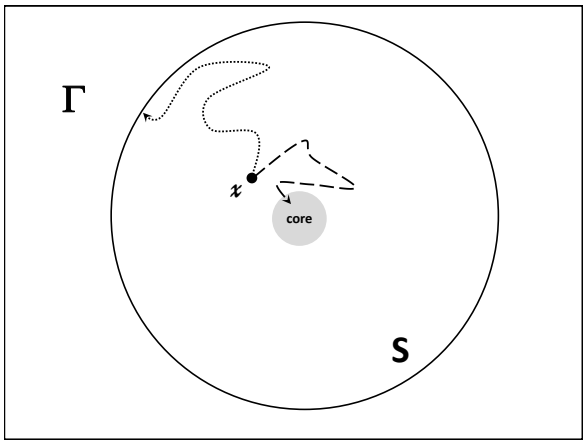

Figure 1. In the state space Γ (rectangle) there is a subset S of starting points (big circle). If we start trajectories of a stochastic process at a starting point $x \in S$ of a fixed time-length t, then there is a certain probability that we get a trajectory that is always within S (dashed line) and that does not leave S (dotted line). We are interested in how this probability depends on the time-length t and on the starting point x. The "core" is a subset of S which is very likely to be reached by trajectories starting in S.

In molecular simulation we are often faced with the situation that is depicted in Fig. 1. Let us assume, that we run a Brownian dynamics (potential-driven diffusion, overdamped Langevin) simulation guided by a given potential energy. This is mathematically formulated in Eq.(1) below. Brownian dynamics trajectories that start in low-energy regions S of the potential energy surface will leave these regions (dotted line) only with a very small probability. It is more likely to observe that the trajectory will approximate (dashed line) and dominantly sample states within the lowest energy part of S. This part is denoted as "core" in Fig. 1. That means that it is very probable to observe trajectories to approximate the core set rather than trajectories which leave the low-energy region S. This "exit event" is, thus, a rare event. The (biological or chemical) stability of many molecular systems is given by the rareness of these events. A straight forward molecular simulation approach to estimate the exit rate is the following: We start several trajectories from $x \in S$ and determine the time they need to exit from S. This approach will be extremely inefficient, because the statistics will depend on rare events.

Many approaches in literature try to overcome the sampling problem of rare events. Here are some examples:

- An uncountable number of molecular simulation methods try to accelerate the transitions between the molecular conformations. Girsanov's Theorem can be applied in order to reweight the accelerated samplings towards the original slow process [1]. However, evaluating the reweighting formula and finding proper acceleration methods can be difficult.

⋆ e-mail: weber@zib.de

- Computing rare events by milestoning [2] discretizes the exit path of the molecular system and determines the (fast) rates between the milestones. However, this approach is based on a discretization scheme for high-dimensional spaces.
- Parallel Trajectory Splicing [3] is a method that extensively makes use of parallel computation facilities on supercomputers. Since this method is not based on high-dimensional discretization schemes it is a promising way, if supercomputers are available to run many trajectories in parallel. Parallel Trajectory Splicing, however, is depending on (set-based) defining different states of the system. It also depends on a certain eigenvalue gap between two consecutive eigenvalues of the infinitesinaml generator $\mathcal{L}^*$. This gap controls the error of the artificially generated long-time simulation (from many short-time trajectories).

What all approaches have in common, is that they try to estimate an exit rate out of a given (metastable) set S. Reaching the boundary of that set S during a Brownian dynamics simulation is a rare event. If S is a set, then there is in principle no preferences between "approaching the core" (Fig. 1, dashed line) or "being about to leave the set S" in these approaches. There is no further discrimmination between those trajectories which stay inside S. There is only a discrimmination between trajectories which either leave the set or stay in S.

This article will start with defining a membership function (fuzzy set) χ instead of S by applying the Robust Perron Cluster Analysis [4, 5] (PCCA+) to the eigenfunctions of some infinitesimal generator. This fuzzyness will provide the discussed discrimmination.

2 Theoretical Background

2.1 The stochastic differential equation

The presented theory is based on reversible, ergodic, and stochastic equations of motion [6]. The motion is a diffusion in an energy landscape $V : \Gamma \to \mathbb{R}$, i.e., the realizations of the dynamics lead to trajectories $x_t \in \Gamma, t \in [0, \infty)$, with

$$d\,x_t = -\nabla_x V(x_t)\,dt + \sigma\,dB_t, \tag{1}$$

where B_t denotes Brownian motion and σ the constant diffusion parameter. The non-linear diffusion equation (1) generates trajectories of states. If instead of a single state x_t a density of states π_t is propagated with the above dynamics, then the corresponding linear Fokker-Planck equation describes this propagation of densities [7]:

$$\frac{\partial \pi_t}{\partial t} = \mathcal{L}\pi_t = -\mathrm{div}_x(\nabla_x V(x)\,\pi_t) + \frac{\sigma^2}{2}\Delta_x \pi_t, \tag{2}$$

where $\mathcal{L}$ is the corresponding Fokker-Planck operator and where Δ_x denotes the Laplacian operator. We assume ergodicity of this process, which leads to the invariant density π defined by $\mathcal{L}\pi = 0$. π gives rise to a weighted scalar product $\langle \cdot, \cdot \rangle_\pi$. Note that, although the diffusion equation (1) is non-linear, the corresponding propagation of densities is given by a linear equation (2). In the following, the adjoint operator $\mathcal{L}^+$ of $\mathcal{L}$, i.e., the infinitesimal generator of the stochastic process in the form

$$\mathcal{L}^+ = -\nabla_x V(x) \cdot \nabla_x + \frac{\sigma^2}{2}\Delta_x,$$

will play an important role. We will use $\mathcal{L}^* = -\mathcal{L}^+$. From this infinitesimal generator the transfer operator $\mathcal{P}^\tau$ of the process for the time-length τ can be derived as

$$\mathcal{P}^\tau = \exp(-\tau \mathcal{L}^*). \tag{3}$$

The transfer operator is a widely used tool to compute transition probabilities between molecular conformations [8]. Its spectral properties [9] are strongly connected to the spectral properties of $\mathcal{L}$ and $\mathcal{L}^*$.

2.2 Defining membership functions

The algorithm PCCA+ [4, 5] is commonly used for identifying metastable membership functions χ of molecular systems. The way this algorithm works is the following: Eigenfunctions of the transfer operator $\mathcal{P}^\tau$ are computed which correspond to eigenvalues close to the eigenvalue $\lambda = 1$ of $\mathcal{P}^\tau$. PCCA+ defines membership functions as a linear combination of these eigenfunctions. Due to (3), the eigenfunctions of $\mathcal{P}^\tau$ correspond to the eigenfunctions of $\mathcal{L}^*$. The constant function $\mathbb{1}$ is always the first (dominant) eigenfunction. In the easiest case, χ is a non-trivial linear combination of $\mathbb{1}$ and an eigenfunction $f : \Gamma \to \mathbb{R}$ of $\mathcal{L}^*$. Let f be an eigenfunction of $\mathcal{L}^*$ corresponding to some eigenvalue $\bar{\epsilon} \neq 0$. From the reversibility of the process one can derive that $\{\mathbb{1}, f\}$ are orthonormal functions (with regard to the scalar product $\langle \cdot, \cdot \rangle_\pi$). Then χ is a linear combination of these two functions with $\chi = \bar{\alpha} f + \bar{\beta}\mathbb{1}$ with real numbers $\bar{\alpha}, \bar{\beta}$. This provides:

$$f = \frac{1}{\bar{\alpha}}\chi - \frac{\bar{\beta}}{\bar{\alpha}}\mathbb{1}.$$

Thus,

$$\mathcal{L}^*\chi = \bar{\alpha}\bar{\epsilon} f = \bar{\epsilon}\chi - \bar{\epsilon}\bar{\beta}\mathbb{1} = \alpha\chi + \beta, \tag{4}$$

where $\alpha = \bar{\epsilon}$ and $\beta = -\bar{\epsilon}\bar{\beta}$.

2.3 Implication of PCCA+ result

Equation (4) can be understood in the following sence. The action of the infinitesimal generator $\mathcal{L}^*$ in the situation of Sec. 2.2 is like shifting and scaling the function χ. If equation (4) holds, what does it imply? Equation (4) is equivalent to

$$\mathcal{L}^*\chi = \epsilon_1\chi - \epsilon_2(1 - \chi), \tag{5}$$

where $\epsilon_1 = \alpha + \beta$ and $\epsilon_2 = -\beta$ are fixed numbers defined by the PCCA+ algorithm. By multiplying (5) with the expression $e^{-\epsilon_1 t}$, and by defining a function $p_\chi(x, t) := \chi(x)\,e^{-\epsilon_1 t}$, we equivalently get:

$$-\mathcal{L}^* p_\chi - \epsilon_2 \frac{1 - \chi}{\chi}\,p_\chi = -\epsilon_1 p_\chi, \tag{6}$$

for all $x \in \Gamma$ with $\chi(x) \neq 0$. The definition of p_χ can be expressed by an ordinary differential equation:

$$\frac{\partial p_\chi}{\partial t} = -\epsilon_1 p_\chi \quad \text{and} \quad p_\chi(x,0) = \chi(x). \tag{7}$$

Combining (7) and (6) leads to the following differential equation:

$$\frac{\partial p_\chi}{\partial t} = -\mathcal{L}^* p_\chi - \epsilon_2 \frac{1-\chi}{\chi} p_\chi, \quad p_\chi(x,0) = \chi(x). \tag{8}$$

The equation (8) is the solution of the following conditional expectation value problem according to the Feynman-Kac formula (Equations III.1 and III.2 in [10]):

$$p_\chi(x,t) = \mathbb{E}\left[\chi(x_t) \cdot \exp\Big(-\epsilon_2 \int_0^t \frac{1-\chi(x_r)}{\chi(x_r)}\, dr\Big)\right]_{x_0=x}. \tag{9}$$

In this expression, $x_t \in \Gamma$ are realizations of the stochastic differential equation starting in $x_0 = x$.

2.4 Holding probability of membership functions

If we assume that $\exp(-\infty) = 0$, then the equation (9) turns into

$$p_{\mathbb{1}_S}(x,t) = \mathbb{E}\left[\mathbb{1}_S(x_t) \cdot \delta_0\Big(\int_0^t (1 - \mathbb{1}_S(x_r))\, dr\Big)\right]_{x_0=x}. \tag{10}$$

Here, we replaced χ with a characteristic function $\mathbb{1}_S$ of a set S. The characteristic function $\mathbb{1}_S : \Gamma \to \{0,1\}$ of the set S, is only 1 for $x \in S$. In equation (10), the function $\delta_0 : \mathbb{R} \to \{0,1\}$ is only 1 if the integral is zero. Equation (10) is the well-established way of defining holding probabilities. Each realization of the stochastic process provides a trajectory $x_t \in \Gamma$. The starting set $S \subset \Gamma$ is a connected, open subset in the state space Γ. The *holding probability* $p_{\mathbb{1}_S}(x,t)$ of the set S is the percentage of realizations of the stochastic process starting in $x_0 = x$ which have never left the set S until time t, i.e., $x_r \in S$ for all $r \in [0,t]$. The longer the time t the more trajectories will leave S, thus, the smaller $p_{\mathbb{1}_S}(x,t)$. Equation (9) can, therefore, be seen as the definition of a *χ-holding probability* of a membership function χ. The term χ-holding probability is different from the set-based definition of a holding probability. In this article we will define further quantities of this fuzzy-set-based kind. Under the assumptions of Sec. 2.3 the χ-holding probability decreases exponentially with time

$$p_\chi(x,t) = \chi(x)\, e^{-\epsilon_1 t}. \tag{11}$$

From the structure of equations (9) and (11), we can derive that $\epsilon_1 > 0$ is the *χ-exit rate* of the membership function χ. Note that, equation (11) is consistent with the definition of $\chi(x) = p_\chi(x,0)$, because $\chi(x)$ is the probabilty to classify x as a starting point of the diffusion process. Usually, one can not show that the set-based holding probability depends exponentially on the simulation time t. The following would just be an approximation:

$$p_{\mathbb{1}_S}(x,t) \approx \mathbb{1}_S(x)\, e^{-\epsilon t}. \tag{12}$$

Since (12) is not an equality, the definition of a set-based exit rate ϵ is often understood as a fitting parameter of exit time distributions [11], as an optimization quantity [12], or as an asymptotic value, e.g., in the large deviation principle [13]. The membership-based χ-holding probability decreases exponentially, if the conditions of Sec. 2.3 are satisfied.

2.5 Formula for exit rates

According to the previously identified relations between the defined quantities ϵ_1 and $p_\chi(x,t)$, we can define what an exit rate out of a fuzzy set should be.

Definition: Given a membership function (fuzzy set) $\chi : \Gamma \to [0,1]$ for the starting points of a potential-based diffusion process with infinitesimal generator $\mathcal{L}^*$ such that $\mathcal{L}^*\chi = \alpha\chi + \beta\mathbb{1}$, then $\epsilon_1 = \alpha + \beta$ is the χ-exit rate out of χ, i.e., the χ-holding probability meets $p_\chi(x,t) = \chi(x)\, e^{-\epsilon_1 t}$.

2.6 Formula for exit paths

If one wants to compute the exit path direction from a given state $x \in \Gamma$, then there are in principle two ways to define them in the χ-context. First, the exit path is connected to a decreasing holding probability. Given the χ-holding probability $p_\chi(x,t)$, the exit path direction can be defined as the negative gradient of p_χ, because this is the direction starting in x in which the holding probability decreases the most. Due to the formula of the exit rate, this *direction* is given by $-\nabla_x\chi(x)$. Second, given a membership function χ, the exit path direction can be defined by a decreasing χ-value. Both definitions lead to the same direction $-\nabla_x\chi(x)$ of the path. Following the negative gradient of χ from a given starting point x provides the χ-exit path. In the PCCA+ context, note that the gradient of χ and the gradient of the eigenfunction f are linearly dependent, which means that also the eigenfunctions of $\mathcal{L}^*$ provide the exit paths in the situation of Sec. 2.2.

2.7 Mean holding time

According to the theory [14] of stocahstic differential equations, the mean holding time of a process can be expressed by the integral of the holding probability. In our case, a corresponding definition of a *χ-mean holding time* $t_1(x)$ depending on the initial state $x_0 = x$ of the process would be the following integral:

$$t_1(x) = \int_0^\infty p_\chi(x,t)\, dt = \chi(x)\, \frac{1}{\epsilon_1}.$$

Due to the choice of the sign of $\mathcal{L}^*$, one would further expect from theory [14] that a set-based mean holding time $t(x)$ meets $\mathcal{L}^* t(x) = 1$ inside the open set S. We replaced S by the fuzzy set χ. For the χ-mean holding time we get:

$$\mathcal{L}^* t_1(x) = \frac{1}{\epsilon_1}\mathcal{L}^*\chi(x) = \chi(x) - \frac{\epsilon_2}{\epsilon_1}(1 - \chi(x)).$$

This is indeed what we expect, if $x \in S$ corresponds to $\chi(x) \approx 1$.

2.8 Summarizing the main equations

In the center of discussions there is a membership function χ which satisfies the "almost" eigenvalue equation

$$\mathcal{L}^*\chi = (\epsilon_1 + \epsilon_2)\chi - \epsilon_2,$$

with $\epsilon_2 \ll \epsilon_1$ and some infinitesimal generator $\mathcal{L}^*$. The exit paths out of those fuzzy sets χ are given by following the gradients $-\nabla\chi(x)$. On the basis of χ, a new quantity denoted as χ-holding probability is defined as

$$p_\chi(x,t) = \mathbb{E}\left[\chi(x_t)\cdot \exp\Big(-\epsilon_2\int_0^t \frac{1-\chi(x_r)}{\chi(x_r)}\,dr\Big)\right]_{x_0=x}.$$

Under certain conditions this quantity depends exponentially on time, i.e.,

$$p_\chi(x,t) = \chi(x)e^{-\epsilon_1 t}.$$

The positive number $\epsilon_1 > 0$ is the χ-exit rate, leading to the χ-mean holding time $t_1(x) = \chi(x)/\epsilon_1$, which is result of a partial differential equation:

$$\mathcal{L}^* t_1(x) = \chi(x) - \frac{\epsilon_2}{\epsilon_1}(1-\chi(x)).$$

This equation is "almost" like the equation $\mathcal{L}^* t(x) = 1$ for computing set-based mean holding times in the interior of a given set S.

3 Approximations

3.1 Implications of $\chi(x) \approx 1$

The definition of a χ-holding probability (9) is different from the set-based definition (10). The random variable in the set-based definition can only be 0 or 1. Thus, only the "exit event" defines the χ-exit rate. In the case of the χ-holding probability the function χ is not constantly 1 inside the fuzzy set of starting points. The χ-exit rate could origin from the exponential penalty term

$$\exp\Big(-\epsilon_2\int_0^t \frac{1-\chi(x_r)}{\chi(x_r)}\,dr\Big).$$

Thus, maybe not the "exit event" produces the χ-exit rate, but it stems from the fuzzy definition of χ. We have to answer the question of the problematic time-scale, i.e., at what time t_2 the exponential expression starts to dominate the function χ? In order to compute this, we will assume that χ is constant. The question is now, when will

$$\exp\Big(-\epsilon_2\, t_2\,\frac{1-\chi}{\chi}\Big) \le \chi?$$

This is the case for

$$t_2 \ge \frac{-\ln(\chi)\chi}{(1-\chi)\epsilon_2}.$$

If we assume, that χ is nearly 1 and take the limit of that expression for $\chi \to 1$, then the result is $t_2 \ge \epsilon_2^{-1}$ according to the rule of De L'Hospital. This means, that the mean holding time of χ should be smaller than ϵ_2^{-1}, such that the "exit event" dominates the definition of p_χ. The mean holding time t_1 (identical to the mean first exit time) is given by $t_1 = \chi(x)\,\frac{1}{\epsilon_1}$. Again we assume $\chi(x) \approx 1$. Thus, if $\epsilon_1 > \epsilon_2$, then the "exit event" dominates the definition of the holding probability. In this case, the definition of the χ-exit rate is consistent with our physical interpretation. A lemma (Lemma 3.6 in [5]) provides the following connection between $\bar\beta$ in equation (4) and the statistical weight π_χ of χ:

$$\bar\beta = \pi_\chi := \frac{\int_\Gamma \chi(x)\,\pi(x)\,dx}{\int_\Gamma \pi(x)\,dx},$$

where π is the invariant density of the stochastic process. Using the easy formula of Sec 2.5, the χ-exit rate is computed as $\epsilon_1 = \bar\epsilon(1-\bar\beta) = \bar\epsilon(1-\pi_\chi)$. Furthermore, $\epsilon_2 = \bar\epsilon\pi_\chi$. The condition $\epsilon_1 > \epsilon_2$ means that the concept of the χ-holding probability and the corresponding χ-exit rate is only physically meaningful, if $\pi_\chi < 0.5$, i.e., it is meaningful for starting points out of a small "subset" χ of the state space Γ.

3.2 Approximating χ

Section 2.5 provides an equation for computing χ-exit rates, which are physically meaningful under certain conditions mentioned in the last section. However, if we apply Sec. 2.5 in the PCCA+ context, then computing the χ-exit rate is given by $\epsilon_1 = \bar\epsilon(1-\pi_\chi)$. In order to compute one (let us denote it as) local property ϵ_1, one needs two global properties of the system, namely one eigenvalue $\bar\epsilon$ of $\mathcal{L}^*$ and the statistical weight of χ. These two quantities are correlated via the eigenfunction f. In practise, the problem of computing a χ-exit rate would turn into a function approximation problem for f in high-dimensional spaces Γ. Function approximation has in general a non-polynomial complexity. A lot of effort has been spent in order to circumvent this "curse of dimensionality". Approximations of eigenfucntions of the transfer operator $\mathcal{P}^\tau$ have been computed using Markov State Modeling [15–17], diffusion maps [18], the variational principle [9], committor functions [19], and many other mathematical tools. Note that the core set approach [20] approximates eigenfunctions of $\mathcal{P}^\tau$ based only on committor values which can be sampled by generating an ensemble of trajectories according to (1).

One result of Sarich [21] (Theorem 13) is the following: The subspace spanned by two eigenfunctions $\{\mathbb{1}, f\}$ and the subspace spanned by a committor function ξ and $\mathbb{1}-\xi$ almost coincide using suitably defined core sets. This is especially the case, if the stochastic process between the core sets is metastable. Thus, instead of computing f, one can compute a committor function ξ. Since ξ is a function between 0 and 1, the PCCA+ result on the basis of this approximation space $\{\xi, \mathbb{1}-\xi\}$ would be $\chi = \xi$. Thus, the membership function defined in Sec. 2.2 is almost identical to a committor function ξ.

However, how can we estimate a committor function, if there is only one core set? We invert the argument of the introduction and note that if we generate a trajectory

of certain length T (much smaller than the mean exit time) starting in x and reach the pre-defined core within or before that time, then we expect $x \in S$. If the trajectory does not reach the core, then we expect $x \notin S$. Thus, by estimating the probability $\chi(x)$ to reach the core in a certain time-span T starting in x we get a membership function. This membership value is very close to a committor function value, if the core sets are assumed to be absorbing. The reason is, that a process starting in S would quickly find the core. A process that needs "too much" time, probably found another core set and is trapped.

Besides committor functions, there are other ideas to access χ efficiently. All methods which compute reaction coordinates and reaction paths [22] (as paths in high dimensional spaces) can also be used to approximate χ, if we use the result of Sec. 2.6 that the holding probability decreases the most in the direction of that path.

3.3 Time discretization

According to the ideas of the last section, one can find methods to evaluate $\chi(x)$ which are not based on a linear combination of ansatz functions. The computation is based on reaction paths or on sampling from trajectories starting in x. Thus, the function evaluation $\chi(x)$ is a result of a simulation, i.e., of a discretization of time and not of a discretization of Γ and, thus, circumvents the curse of dimensionality. We will apply this idea in order to find the paramters α and β such that $\alpha\chi + \beta\mathbb{1}$ approximates $\mathcal{L}^*\chi$ as good as possible. However, even if it is possible to evaluate $\chi(x)$ pointwise by defining a core set and running simulations or, alternatively, by computing reaction paths starting in x, the pointwise computation of $\mathcal{L}^*\chi$ is not that straight forward. To solve this, we will exploit the fact that $\mathcal{P}^\tau = \exp(-\tau\mathcal{L}^*)$. Starting with the condition that we want to acchieve,

$$\mathcal{L}^*\chi = \alpha\chi + \beta,$$

we get that $\mathcal{L}^*$ is a scale-shift-operator for χ. For $i > 0$, an iterative application of that operator leads to

$$\frac{(-\tau\mathcal{L}^*)^i}{i!}\chi = \frac{(-\tau\alpha)^i}{i!}\chi + \frac{(-\tau)^i\alpha^{i-1}\beta}{i!}.$$

By taking the sum for $i = 1, \ldots, \infty$ and adding χ on both sides

$$\mathcal{P}^\tau\chi = e^{-\tau\alpha}\chi + \frac{\beta}{\alpha}(e^{-\tau\alpha} - 1). \tag{13}$$

This has the following algorithmic consequences. Instead of $\mathcal{L}^*\chi$ we can evaluate $\mathcal{P}^\tau\chi$ pointwise: Given the point $x \in \Gamma$ for which we want to evaluate $\mathcal{P}^\tau\chi$, we start M trajectories in x of time-length τ. For all the end points $x_\tau^{(k)}, k = 1, \ldots, M$ of those trajectories we average over the values $\chi(x_\tau^{(k)})$. This provides the value of $\mathcal{P}^\tau\chi(x)$. We will also evaluate $\chi(x)$ at the starting point x and, after that, solve the linear regression problem

$$\min_{\gamma_1,\gamma_2} \|\mathcal{P}^\tau\chi(\cdot) - \gamma_1\chi(\cdot) - \gamma_2\|, \tag{14}$$

where every starting point x generates one entry $\mathcal{P}^\tau\chi(x) - \gamma_1\chi(x) - \gamma_2$ of the vector. If the regression problem is exactly solvable, then

$$\gamma_1 = e^{-\tau\alpha}, \quad \gamma_2 = \frac{\beta}{\alpha}(e^{-\tau\alpha} - 1) \tag{15}$$

according to Eq. (13). The proposed computation of the χ-exit rate is depicted in Algorithm 1.

Algorithm 1 Computing χ-exit rates

1. Determine a finite set X of points in Γ.

2. For every point $x \in X$ evaluate $\chi(x)$ and $\mathcal{P}^\tau\chi(x)$ using simulations of (1).

3. Solve the linear regression problem (14) for the computation of γ_1 and γ_2.

4. With the aid of Eq. (15) compute the values

$$\alpha = -\frac{1}{\tau}\ln(\gamma_1)$$

and

$$\beta = \frac{\alpha\gamma_2}{\gamma_1 - 1}.$$

5. The χ-exit rate is $\epsilon_1 = \alpha + \beta$.

3.4 Square-Root-Approximation of $\mathcal{L}^*$

The sqrt-approximation is only needed for the artificial, illustrative examples in order to can compute analytical (non-statistical) results. A spatial discretization is, in principle, not needed for the application of the above theory. $\mathcal{L}^*$ is a continuous operator. For some easy examples, we will use a matrix $L^* \in \mathbb{R}^{n\times n}$ instead of $\mathcal{L}^*$. A possible, heuristic discretization scheme is available [23]. Note that $-L^*$ can be regarded as a transition rate matrix. If we assume a discretization of the state space Γ into n subsets, then the transition rate between neighboring subsets i and j is given by $-L^*_{ij} = \sqrt{\pi_j/\pi_i}$, where $\pi_i = \exp(-\frac{1}{k_bT}V(i))$ is the Bolzmann weight and $V(i)$ is the potential energy value at the center of box i. k_b is the Bolzman factor, T is the temperature. The diagonal elements of L^* are adjusted such that the row sum of L^* is zero. This type of defining L^* leads to a reversible process with a stationary distribution given by the Bolzmann distribution. We will use this square root approximation for the numerical examples below.

4 Illustrative examples

4.1 Idea 1: χ-exit rates from eigenvalues and eigenfunctions

The first example demonstrates how χ-exit rates can be computed if a non-trivial eigenfunction and its eigenvalue $\bar{\epsilon} \neq 0$ of $\mathcal{L}^*$ are known. For this purpose the following 2-dimensional potential energy function $V : \mathbb{R}^2 \to \mathbb{R}$ is analyzed:

$$V(x) \quad = \quad 3\,\exp\big(-(4x_1 - 2)^2 - (4x_2 - \frac{7}{3})^2\big)$$

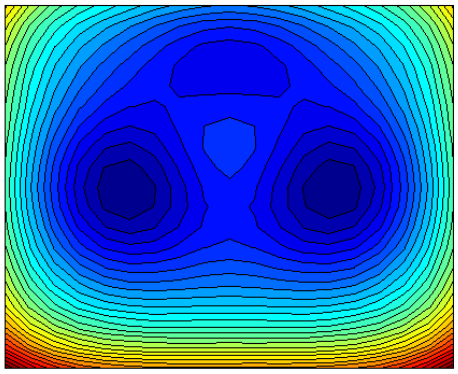

Figure 2. The potential energy function (16) has two deep minima (dark blue, left and right) and a less deep minimum at the top. High values are indicated by red color, whereas low values are indicated by blue color.

$$
\begin{aligned}
&-3\,\exp\big(-(4x_1-2)^2-(4x_2-\frac{11}{3})^2\big)\\
&-5\,\exp(-(4x_1-3)^2-(4x_2-2)^2)\\
&-5\,\exp(-(4x_1-1)^2-(4x_2-2)^2)\\
&+0.2\,(4x_1-2)^4+0.2\,(4x_2-\frac{7}{3})^4. \qquad (16)
\end{aligned}
$$

This function is depicted in Fig. 2. In order to easily construct a discretized infinitesimal generator $-L^*$ on that potential, we generated a regular 50×50 box discretization on the definition set $[0,1] \times [0,1]$. The transition rates between neighboring boxes i and j were set to be $-L^*_{ij} = \sqrt{\pi_j/\pi_i}$, where $\pi_i = \exp(-V(i))$ is the Bolzmann weight and $V(i)$ is the potential energy value at the center of box i. This construction is according to the proposed square root approximation of infinitesimal generators [23].

With the aid of this discretization scheme, we can approximate one eigenfunction f, which has the third lowest eigenvalue $\bar{\epsilon} = 0.0086$. This eigenfunction is shown in Fig. 3. According to what has been said in Sec. 2.6 one can already derive the exit paths out of the top minimum of the potential energy surface from the negative gradient of the eigenfunction. Given a point x in the definition set Γ, the holding probability decreases the most in the direction of $-\nabla f(x)$. The highest value of the eigenfunction is $\max_i f(i) = 0.0543$ whereas the minimal value is $\min_i f(i) = -0.0133$. Given one eigenfunction of $\mathcal{L}^*$, the computation of the membership function $\chi = \bar{\alpha} f + \bar{\beta}\mathbb{1}$ based on PCCA+ is unique [4] with

$$\bar{\alpha} = \frac{1}{\max_i f(i) - \min_i f(i)}$$

and

$$\bar{\beta} = \frac{-\min_i f(i)}{\max_i f(i) - \min_i f(i)}.$$

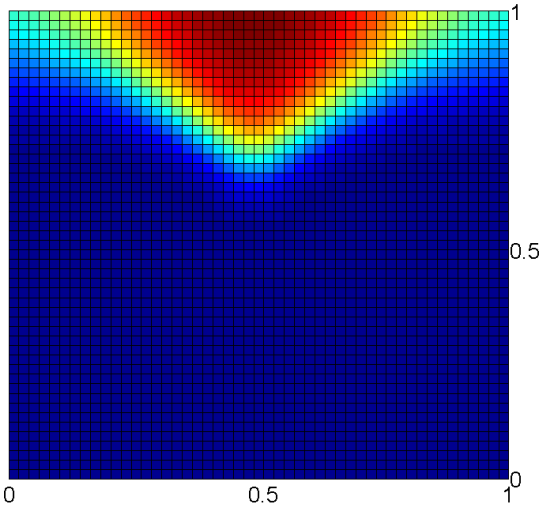

Figure 3. An approximation of one eigenfunction of $\mathcal{L}^*$ using the discretized operator L^*. This eigenfunction "corresponds" to the less deep minimum of the potential energy function.

These quantities are sufficient to calculate the χ-exit rate $\epsilon_1 = \bar{\epsilon}(1-\bar{\beta}) = 0.0069$, also the statistical weight $\pi_\chi = \bar{\beta} = 0.1965$, and the penalty parameter of the holding probability which is $\epsilon_2 = \bar{\epsilon}\pi_\chi = 0.0017$. The χ-exit rate is physically meaningful according to Sec. 3.1, because $\epsilon_2 \ll \epsilon_1$.

4.2 Comparison: fuzzy vs. set

The mean holding time for a set S is zero at the boundary of the set. The χ-mean holding time $t_1(x) = \chi(x)/\epsilon_1$ is only zero for $\chi(x) = 0$. Therefore, the set-based holding time $t(x)$ computed from the partial differential equation $\mathcal{L}^* t(x) = 1$ must be different from $t_1(x)$. Note that in the molecular simulation setting, a simulated process has not reached the core of another conformation if the trajectory is at the boundary of the set S. If we want to compare the results of the χ-mean holding time with a set-based approach, the fuzzyness of the answer to the question "where in Γ do we reach another conformation?" plays an important role. If we, e.g., define S in the situation of Sec. 4.1 to be that part of Γ which is defined by

$$S = \{x \in \Gamma \,|\, \chi(x) > 0.22\},$$

then a set-based holding probability will be zero at the boundary of S, while the χ-holding probability will be $\frac{0.22}{\epsilon_1} = 31.88$. This is a huge difference. In Fig. 4, we compare the χ-mean holding time with the computed set-based mean holding time for the 2500 cells of the discretization of $\mathcal{L}^*$. There is a clear correlation between these two quantities, except for the fact, that the χ-mean holding time $t_1(x)$ does not have the zero-plateau. For a definition of S on the basis of the condition $\chi(x) > 0.5$, the set-based holding time $t(x)$ would be much smaller than $t_1(x)$. For a definition of S on the basis of the condition $\chi(x) > 0.1$,

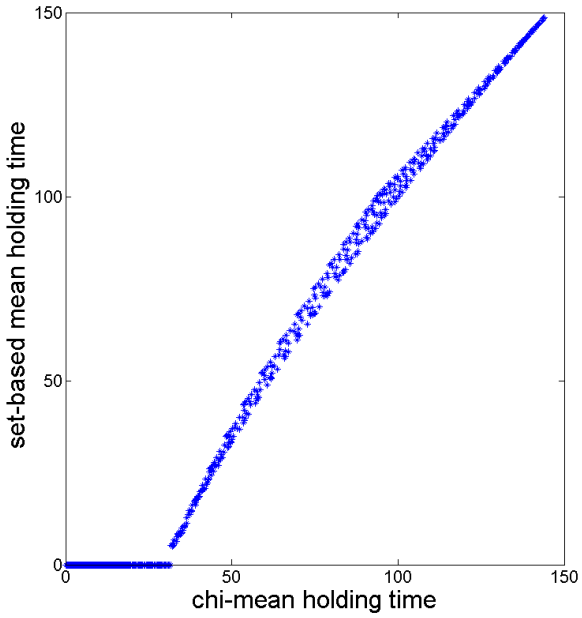

Figure 4. In this plot the x-axis is the χ-mean holding time $t_1(x)$ and the y-axis is the set-based mean holding time $t(x)$ compared for the values of the 2500 discretization boxes in the situation of Sec. 4.1 and for $S = \{x \in \Gamma \,|\, \chi(x) > 0.22\}$. The two quantities are mainly identical, except for a different behavior at the boundary of S.

the set-based holding time would mostly be much higher than $t_1(x)$. The χ-mean holding time is, thus, like a "compromise" in that sense. It is like a "mean" mean holding time for different possible choices of the boundary of the metastable set S.

4.3 Idea 2: Linear regression of PCCA+ results

The case, that has been described in Sec. 4.1 is artificial. Usually, not every single eigenfunction of $\mathcal{L}^*$ can be interpreted as a membership function. Sometimes the membership function χ has to be composed as the linear combination of several eigenfunctions. This situation is shown in Fig. 5. In this situation

$$\chi = 0.4452 \cdot \mathbb{1} + 17.7865 \cdot f_2 - 4.1266 \cdot f_3,$$

where f_2 and f_3 are approximated eigenfunctions of $\mathcal{L}^*$ corresponding to the eigenvalues 0.0025 and 0.0086, respectively. From the first factor, we can directly extract $\pi_\chi = 0.4452$ which is less than 0.5, thus, we will have the good case that $\epsilon_2 < \epsilon_1$. In the given situation it is possible to compute $\mathcal{L}^*\chi$ analytically:

$$\mathcal{L}^*\chi = 17.7865 \cdot 0.0025 \cdot f_2 - 4.1266 \cdot 0.0086 \cdot f_3.$$

In the case of several eigenfunctions, $\mathcal{L}^*\chi$ is not a linear combination of $\mathbb{1}$ and χ any more. But if we solve the linear regression problem of minimizing the norm $\|\mathcal{L}^*\chi - \alpha\chi - \beta\mathbb{1}\|$, then the result is $\alpha = 0.0028$ and $\beta = -0.0014$. Thus, the χ-exit rate is $\epsilon_1 = \alpha + \beta = 0.0014$.

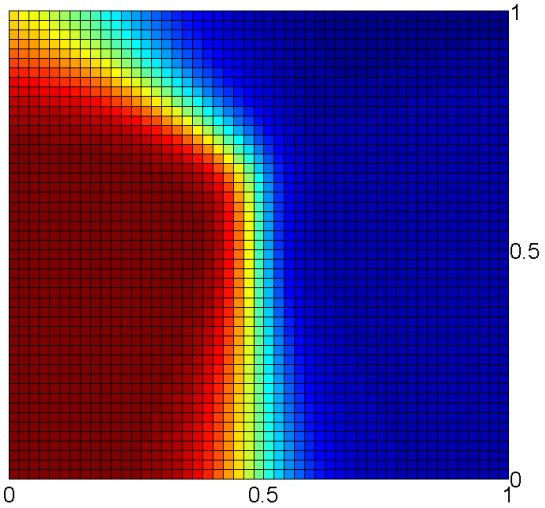

Figure 5. The membership function χ which represents the starting point assignment for the left deep minimum of the potential energy surface of Fig. 2. This function is a linear combination of three eigenfunctions of $\mathcal{L}^*$.

4.4 Idea 3: Committor functions as approximation space

Given the square root approximation L^* (with $k_bT = 1$) one can easily compute the discretized committor function of the process between the left and the right deep minimum. This committor function will now serve as an approximation for the membership function, it will, therefore, also be denoted as χ. For the committor function between the two minima, two core sets are needed. The core sets are based on the 50×50 discretization of Γ. All discretization boxes having a statistical weight higher than 0.0025 are assigned to one of the core sets. After computing χ, the propagated values $P^\tau\chi$ with $\tau = 100$ are determined. Note that $P^\tau = \exp(-\tau L^*)$. In Fig. 6 the χ-values are plotted against the $P^\tau\chi$-values for solving the regression problem (14). Depending on the regression norm, the regression results may differ. Taking the $\|\cdot\|_2$-norm, the results are $\gamma_1 = 0.8201$ and $\gamma_2 = 0.900$. Thus, $\alpha = 0.0020$ and $\beta = -0.0010$. The χ-exit rate is approximated to be $\epsilon_1 = 0.0010$, which is lower than the result of Sec. 4.3.

4.5 Idea 4: Short-time simulations for estimating committor functions

In the three illustrative examples above, a discretized version of the infinitesimal generator was given. Thus, we were able to compute (approximative) eigenfunctions and eigenvalues of $\mathcal{L}^*$ that can be used to define membership functions and χ-exit rates out of those corresponding fuzzy sets χ. The need for a discretization is a drawback of that method in high-dimensional conformational spaces of molecular systems. In Sec. 3.2 it has been discussed, whether there is a way to estimate the (committor) function χ by running simulations only. If this is possible, then

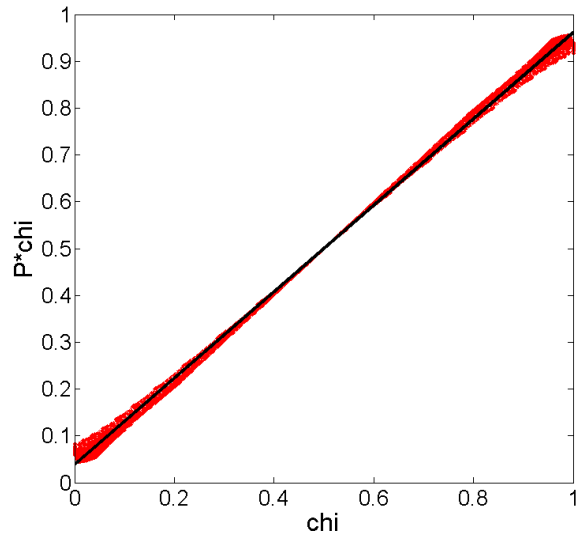

Figure 6. Computing χ as a committor function of the process and $P^\tau\chi$ for $\tau = 100$. The red points correspond to the plot of the 2500 χ-values against their propagated values $P^\tau\chi$. The black line is the $\|\cdot\|_2$- regression result. Many possible lines can fit the given data depending on the regression norm.

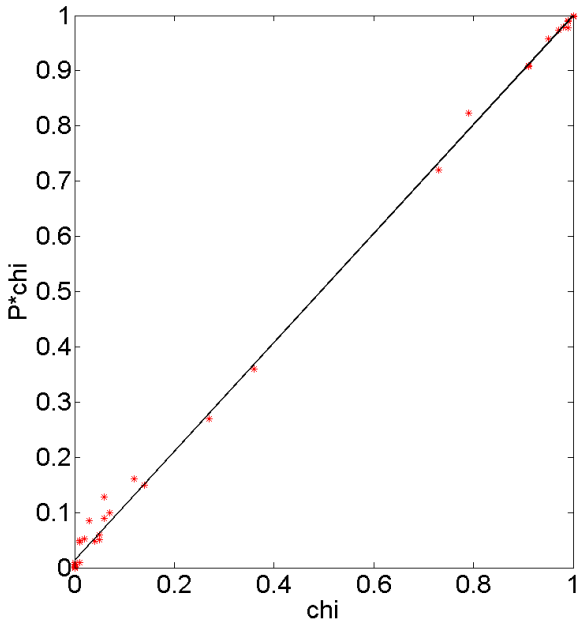

Figure 7. "Evaluating" χ and $\mathcal{P}^\tau\chi$ at 50 randomly chosen points $x \in \Gamma$. The relation betwen these two quantities should be linear. By linear regression a line is fitted into the data providing $\gamma_1 = 0.9870$ and $\gamma_2 = 0.0128$.

the Algorithm 1 would also provide χ-exit rates. In order to illustrate how this algorithm practically works, we took again the potential energy function of Fig. 2. This time we will apply a Brownian dynamics simulation of $\mathcal{P}^\tau$ instead of the square root approximation of $\mathcal{L}^*$. The Brownian dynamics is chosen to be "faster" than the square-root approximation: For computing the value of $\chi(x)$ for each $x \in X$, we start 100 trajectories in x according to (1) with $\sigma = 0.8$ and an Euler-Maruyama time discretization of $\delta t = 0.001$. $\chi(x)$ is defined as the percentage of those trajectories which have reached a certain core set within less than 100 integration steps. The core set is reached, if the x_1-coordinate is in the interval $[0.2, 0.3]$ and the x_2-coordinate is in the interval $[0.4, 0.5]$. By this procedure we get a χ-function which is very similar to that in Fig. 5. Different from the situation of Sec. 4.3, χ is only given point-wise. 50 randomly chosen points x in the box $[0, 1]^2$ have been used for this χ-function evaluation. $\mathcal{P}^\tau\chi(x)$ has also been evaluated in a similar way. 100 trajectories with 50 integration steps only (total time length is $\tau = 0.05$) have been generated to propagate x. At those propagated points $x^{(k)}$ the χ-function has been evaluated as described above. $\mathcal{P}^\tau\chi(x)$ is given by the averaged χ-value at the propagated points. In Fig. 7 the $\mathcal{P}^\tau\chi$-values are plotted against the χ-values at the 50 chosen points.

If χ had been a linear combination of an eigenfunction of $\mathcal{L}^*$ and $\mathbb{1}$, then this plot would show a line. From the axis intercept γ_2 and the slope γ_1 of that line, the χ-exit rate is estimated. Small deviations from the perfect line due to sampling errors, however, lead to high relative errors in ϵ_1. By fitting a line to the data points in Fig. 7, we estimated the χ-exit rate $\epsilon_1 = 0.0042$.

5 Molecular example

Algorithm 1 can easily be applied to molecular systems. We will demonstrate this algorithm for the simulation of an n-pentane molecule, shown in Fig. 8. This molecule has 17 atoms. Thus, its state space $\Gamma = \mathbb{R}^{3\cdot 17}$ is 51-dimensional.

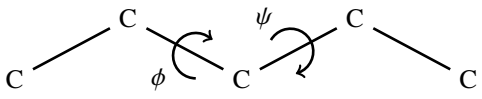

Figure 8. The pentante molecule consists of 5 carbon atoms and 12 hydrogen atoms (not shown). Two torsional angles ϕ and ψ determine the conformation of that molecule. The depicted conformation corresponds to $\phi = \psi = 180°$, which is the most stable conformation. For this conformation we aim to compute the exit rate.

Defining the core set. The definition of the core set depends on the kind of rare event statistics which we want to estimate. If we want to figure out the slow diffusion of the pentane molecule in the 3-dimensional space, then we would project the 51-dimensional coordinates, e.g., onto the center of mass of pentane and define the core set as a ball in this 3-dimensional space. However, chemists are more interested in the *internal* transitions of the molecule (i.e., its conformations). It is well-known that

the pentane molecule has several metastable low-energy conformations. Those can be determined by considering two torsional angles ϕ and ψ. Each torsion angle is defined by 4 consecutive carbon atoms, see Fig 8. A long-term simulation of pentane at 700K (with fast transitions between the conformations) reveals that there are nine different peaks of the stationary (Boltzmann) distribution in the ϕ-ψ-digram shown in Fig. 9 (cyan circles and dotted boxes). If we want to estimate the exit rate from the

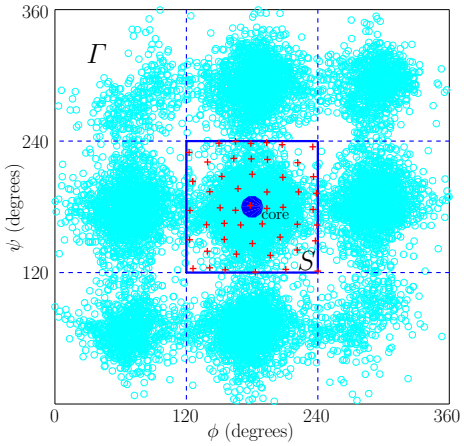

Figure 9. Torsion angle distribution of the two torsion angles of pentane at 700K, for the interval of 0 to 360 degrees (cyan circles). Sampling starting points $x \in X$ for the Algorithm 1 from the box $[120, 240]^2 =: S$ (red crosses). Total number of starting points is 50.

central conformation of pentane, then the core set can be defined as all states $x \in \Gamma$ of pentane which have a pair of torsion angles approximately at $(\phi, \psi) \approx (180, 180)$. This core set is indeed a non-convex, connected, unbound set in $\Gamma = \mathbb{R}^{51}$. In the ϕ-ψ-plane it is a circle, see the blue circle in the center of Fig. 9. Although the described core sets can easily be projected to a 3- or 2-dimensional space, neither the potential energy V nor the membership function χ is a 3- or 2-dimensional function. This example of pentane is indeed 51-dimensional with $\chi : \mathbb{R}^{51} \to [0, 1]$.

Application of Algorithm 1. Instead of discretizing the 51-dimensional space for approximating χ, which would lead to a curse of dimensionalty, we will only evaluate $\chi(x)$ and $\mathcal{P}^\tau\chi(x)$ at 50 different points $x \in X$. The projection of those 50 points onto the ϕ-ψ-plane is shown as red crosses in Fig. 9. To compute the $\chi(x)$-value 30 GROMACS stochastic dynamics (SD) simulations for each $x \in X$ with time discretization step $\delta t = 0.001$ps at a temperature of 310K (conformational transitions are rare events at this temperature) were performed. All molecular simulations are performed with GROMACS 5.1.2 [24, 25]. Based on this simulation the percentage of the trajectories which have reached the core within first 0.5ps was calculated. The coordinates at the end points of these simulations are used for computing the value of $\mathcal{P}^\tau\chi$. For each point 30 more simulations with 1000 time

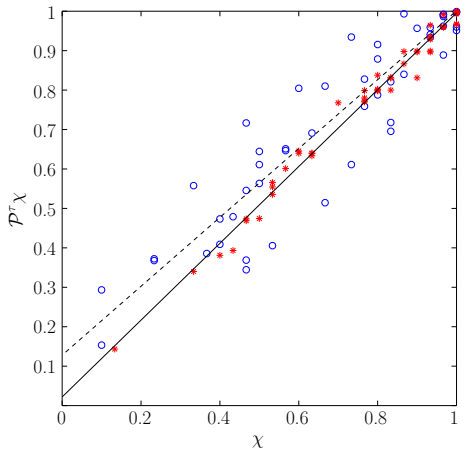

Figure 10. χ and $\mathcal{P}^\tau\chi$ for the indicated 50 starting points in Fig. 9. The red stars correspond to a GROMACS SD simulation. The linear regression of the corresponding χ-$\mathcal{P}^\tau\chi$-correlation is given by a solid line. If an MD simulation is performed (blue circles), then the correlation between χ and $\mathcal{P}^\tau\chi$ is worse (dotted line) and does not lead to an interpretable rate estimation.

steps and the same time step $\delta t = 0.001$ps (total time is 1ps) were done. $\mathcal{P}^\tau\chi(x)$ was calculated as the average of the χ-values for the propagated points. Using linear regression we get $\gamma_1 = 0.9738$, $\gamma_2 = 0.0220$. Therefore, the χ-exit rate is computed to be $\epsilon_1 = 0.01 ps^{-1}$. This result means, that on average it needs about 100ps to exit from the central conformation of n-pentane. However, to yield this result, we used only much shorter trajectories (of 0.5ps or 1ps) which could easily be generated independently in parallel on different processors.

MD versus SD. For showing that the theory depends on a certain kind of equations of motion, we performed molecular dynamics (MD) simulations in which all parameters and starting states were chosen to be equal to the SD simulation. MD simulation is based on Newton's equations of motion and not on (1). For detailed information, please, see chapter 3.8 of the GROMACS manual [24, 25]. In the MD case, the linear regression led to a different result with $\gamma_1 = 0.8731$ and $\gamma_2 = 0.1280$ (see Fig. 10). Thus, $\alpha = 0.2714$ and $\beta = -0.2738$. Therefore, we get a negative value for the exit rate. For the presented theory it is mandatory to use (1) as the equations of motion. MD is not applicable.

Parallelization. For our approach, two different kinds of computational parallelization can be combined. There are already some built-in parallelization schemes in GROMACS to run trajectories. As we additionally know that all 50 start states are independent, it is trivial to parallelize the simulation of those trajectories as well. For the MD/SD simulations and calculation of χ, $\mathcal{P}^\tau\chi$ values all available cores were used. In order to achieve optimal simulation performance the number of starting trajectories

should be divisible by the number of used cores.

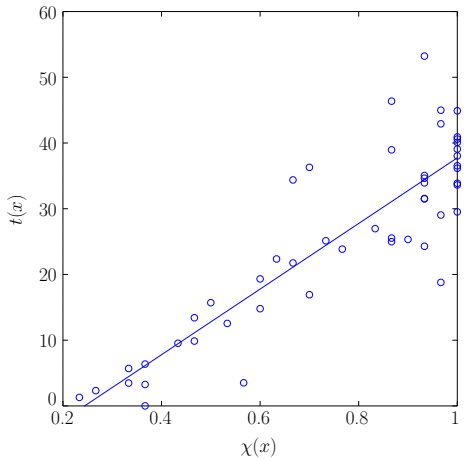

Figure 11. χ and τ for the indicated 50 starting points in Fig. 9. There is a relation between the membership value $\chi(x)$ and the simulated mean exit time $t(x)$.

Validation. For a validation we will compare the χ-exit rate with a set-based exit rate. For this validation we, thus, need to define a starting set S, which makes this comparision difficult, because the quality of our result will depend on the choice of S. We computed the mean first exit time t for leaving the blue box S indicated in Fig. 9. This is an *arbitrarily* chosen starting set according to chemical intuition about the dominant conformation of *n*-pentane. To estimate the exit rate, we again started 30 SD-simulations from the 50 starting points given as red crosses in Fig. 9. These simulations were 200 times longer (100ps) than our simulations for the χ-exit rate estimation. In principle, by this estimation a function $t(x)$ of set-based mean holding times in Γ is approximated point-wise. From Fig. 11 we see that the starting points with $\chi(x) = 1$ have a mean holding time of about 40ps. This would mean, that the set-based exit rate should be about 0.025ps^{-1}, which is in the same order of magnitude as our result 0.01ps^{-1} but 2.5 times higher. According to our presented theory, we additionally know that $t_1(x) = \frac{1}{\epsilon_1}\chi(x)$ is the χ-mean holding time. Like in Fig. 4: If $t_1(x)$ would be a good representation of the set-based mean holding time behavior $t(x)$, then there should be a linear dependence between the simulated exit time and the membership value $\chi(x)$ as it is indeed the case in Fig. 11. Note that the $t(x)$-computation suffers from high variance for larger holding times.

We calculated the mean exit time for each starting point based on the 30 SD-simulations. Only 1 of 50 starting points has the mean exit time less than 1ps, therefore we can say that in this case the exit rate is equal to $\tau = 0.02$.

6 Conclusion

If one question is too complicated to answer, then maybe there is a slightly different question which provides the same kind of technical value and reveals simple relations. We started the article with the observation that a potential-driven diffusion process tends to sample from the core of a metastable set S. Trajectories starting in some point $x \in S$ which are constructed according to Eq. (1) probably quickly reach the core and extremely rarely leave the set S. In Sec. 3.2 we inverted this observation and turned it into a definition of a membership function χ. If a trajectory starts in $x \in \Gamma$ and quickly reaches a pre-defined core, then we expect that x is part of the (implicitly predefined) metastability. The function value $\chi(x)$ is defined as the portion of trajectories which start at x and reach the core quickly. A function χ of that type can be efficiently estimated point-wise by running simulations. There is no curse of dimensionality, if the χ-function is approximated point-wise. Given that the metastability is now a fuzzy set χ, then we need a new definition of what we want to understand to be the *holding probability*. In this article we introduced a definition of a χ-holding probability which provides easy mathematical relations between χ-exit rates, χ-exit paths, and χ-mean first exit times. In principle, the χ-exit rate is given by the slope and the axis intercept of the linear relation between χ and $\mathcal{P}^\tau\chi$. For recovering the linear relation between χ and $\mathcal{P}^\tau\chi$, it is very "useful" that χ has values ranging from 0 to 1, i.e., the information of "how much does a state x belong to the metastability" is exploited.

Acknowledgement. The work has partially been financed by the CRC-1114 "Scaling Cascades in Complex Systems", project A05. Data and m-scripts can be found at http://www.zib.de/ext-data/soft_exit_rate.

References

[1] W. Zhang, H. Wang, C. Hartmann, M. Weber, C. Schütte, SIAM J. Sci.Comput. **36**, A2654 (2014)

[2] E. Vanden-Eijnden, M. Venturoli, J. Chem. Phys. **130**, 194101 (2009)

[3] D. Perez, E.D. Cubuk, A. Waterland, E. Kaxiras, A.F. Voter, J. Chem. Theory Comput. **12**, 18? (2016)

[4] P. Deuflhard, M. Weber, Linear Algebra and its Applications **161** (2005), 398 Special issue on matrices and mathematical biology

[5] M. Weber, Ph.D. thesis, FU Berlin (2006)

[6] C. Schütte, M. Sarich, *Metastability and Markov State Models in Molecular Dynamics: Modeling, Analysis, Algorithmic Approaches*, Vol. 24 of *Courant Lecture Notes* (American Mathematical Society, 2013), `http://publications.mi.fu-berlin.de/1304/`

[7] C. Schütte, W. Huisinga, *Biomolecular conformations can be identified as metastable sets of molecular dynamics*, in *Handbook of Numerical Analysis* (Elsevier, 2003), pp. 699–744

[8] C. Schütte, *Conformational dynamics: Modelling, theory, algorithm, and application to biomolecules* (1998), habilitation Thesis

[9] F. Nüske, B.G. Keller, G. Pérez-Hernández, A. Mey, F. Noé, J. Chem. Theory Comput. **10**, 1739 (2014)

[10] H. Gzyl, Journal of Mathematical Analysis and Applications **142(1)**, 77 (1989)

[11] S. Srebnik, I. Yungerman, G. Kohav, M. Sheintuch, Chemical Engineering Science **62**, 2242 (2007)

[12] C. Hartmann, R. Banisch, M. Sarich, T. Badowski, C. Schütte, Entropy **16**, 350 (2014)

[13] H. Cramér, *Sur un nouveau theoreme-limite de la theorie des probabilites* (Colloque consacre a la theorie des probabilites, Part 3, Actualites scientifiques et industrielles, 1938)

[14] G.A. Pavliotis, *Stochastic Processes and Applications* (Springer, 2014)

[15] K.A. Beauchamp, G.R. Bowman, T.J. Lane, L. Maibaum, I.S. Haque, V.S. Pande, J Chem Theor Comput (2011)

[16] J. Chodera, N. Singhal, V.S. Pande, K. Dill, W. Swope, Journal of Chemical Physics **126**, 155101 (2007)

[17] M. Sarich, F. Noé, C. Schütte, SIAM Multiscale Model. Simul. **8**, 1154 (2010)

[18] B. Nadler, L. Stephane, R.R. Coifman, I.G. Kevrekidis, Neural Information Processing Systems (NIPS) Proceedings (2005)

[19] C. Schütte, F. Noé, J. Lu, M. Sarich, E. Vanden-Eijnden, J. Chem. Phys. **134**, 204105 (2011)

[20] N. Djurdjevac, M. Sarich, C. Schütte, Proceedings of the International Congress of Mathematicians Hyderabad, India pp. 3105–3131 (2010)

[21] M. Sarich, Ph.D. thesis, Free University Berlin (2011)

[22] W. Li, A. Ma, J. Chem. Phys. **143**, 174103 (2015)

[23] H.C. Lie, K. Fackeldey, M. Weber, SIAM Journal on Matrix Analysis and Applications **34**, 738 (2013)

[24] *Gromacs reference manual version 5.1.2*, ftp://ftp.gromacs.org/pub/manual/manual-5.1.2.pdf, accessed: 2017-04-04

[25] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, J. Chem. Theory Comp. 4 **2**, 435 (2008)