

On the Relation Between Projections of the Reweighted Path Ensemble

Peter G. Bolhuis · Wolfgang Lechner

Received: 1 April 2011 / Accepted: 16 August 2011 / Published online: 9 September 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract We derive several distribution functions for the recently introduced reweighted path ensemble [Rogal et al. in J. Chem. Phys. 133:174109, 2010]: the configurational and path densities, the reactive current, and the generalized committors for the different path types. We relate these distributions to the free energy and to the expressions for the rate constant in the transition state theory, the reactive flux method, the transition path (interface) sampling framework, and the Bayesian path statistics. In addition, we compute the transmission coefficient (distribution) from the reweighted path ensemble. Finally, we derive the path sampling shooting point distributions. For a simple two dimensional Langevin model we illustrate how these novel distributions can be used as analysis tools in rare event simulations.

Keywords Rare event simulations · Transition path sampling · Rate constants

1 Introduction

Chemical reactions, nucleation, protein conformation changes, and permeation of ions through membranes are a few examples of physical processes that occur on a time scale which is very long with respect to the underlying molecular timescale. To investigate the kinetic mechanism of such rare events with a molecular simulation technique such as Molecular Dynamics (MD), one would, in principle, like to obtain a very long trajectory which traverses the barrier many times. Such a long MD run not only yields the dynamics and kinetic rate constants, but also gives unbiased insight in the reaction coordinates. However, such an approach is highly inefficient due to the long times spent in the reactant or product states, and the seldom occurrence of actual transitions between these states. These long time scales often correspond to the presence of high free energy barriers between reaction and product. Many special simulation techniques have been devised to overcome and

P.G. Bolhuis (✉) · W. Lechner
van 't Hoff Institute for Molecular Sciences, University of Amsterdam, PO Box 94157, 1090 GD
Amsterdam, The Netherlands
e-mail: bolhuis@uva.nl

sample such barriers. Examples are umbrella sampling [31], flooding [18], local elevation, hyper-dynamics [32, 33], meta-dynamics [24], Wang-Landau sampling [36], adaptive force bias [7], temperature accelerated dynamics [30, 34]. These methods apply to the system an artificial biasing potential as a function of predefined collective variables that pushes it over the barrier. While the free energy as a function of these collective variables is correctly recovered, the dynamics is altered due to the applied bias. The transition state theory based reactive flux methods corrects this bias by computing a transmission coefficient [1, 6, 23]. However, as the bias potential is defined in terms of collective variables, prior knowledge of the correct reaction coordinate is needed, to avoid problems like hysteresis and a very low transmission coefficient.

The transition path sampling (TPS) method [5, 8–10] collects short unbiased dynamics trajectories that connect the reactant with the product. TPS therefore avoids the requirement of prior knowledge of the reaction coordinate. Instead, it requires a reasonable definition of the stable states. The constraint that TPS imposes is on the trajectory level rather than the instantaneous configurations along the trajectories. This allows for both insight in the unbiased dynamical mechanisms, as well as a reaction rate constant computation. Insight in the mechanism can be summarized by the reaction coordinate. As follows from transition path theory [12, 13], the optimal reaction coordinate is the committor: the probability to reach the final state from a configuration \mathbf{r} . However, this is a high dimensional function, whereas usually, a reaction coordinate is viewed as a low dimensional combination of collective variables that best describes the committor. By employing a likelihood maximization one can analyze the path ensemble and find the combination of such variables [27, 28].

The rate constants can be efficiently computed using the transition interface sampling (TIS) framework [15, 16], in which a collective variable is used to describe a foliation of hyper-surfaces or interfaces. A TIS simulation collects trajectories that leave the reactant stable state, and cross a specific interface. TIS gives access to the flux of trajectories through each of the interfaces, and thus the rate. The combination of TIS with replica exchange allows for an even more efficient sampling [3, 14].

Recently we showed how to reweight the paths in TIS in order to obtain the total path ensemble which can be viewed as an approximation to a very long unbiased MD run [29]. The reweighted path ensemble gives access to the averaged committor functions, the free energy landscapes, the rate, as well as the mechanism. In Ref. [25] we developed a likelihood maximization string method to obtain the most optimal low dimensional reaction coordinate that describes the process.

In this paper we explore the relation between the reweighted path ensemble and the committor function, rate constant and free energy in more detail. Moreover, we make the connection with the transition path theory of E and Vanden-Eijnden [12, 13], and with the Bayesian path statistics of Hummer [19]. The paper is organized as follows. In Sect. 2 we briefly recapitulate the derivation of the reweighted path ensemble, and how the configurational density, free energy, averaged committor, path density and reactive current follow from it. We then (re)introduce the concept of generalized averaged committors [26] and path densities. Next, we make the connection between these concepts and the TIS rate constant expression, and establish their relation with the transmission coefficient in the reactive flux methods, as well as with the Bayesian path statistics [19]. Finally, we show that the generalized committor functions also describe the shooting point distributions for the various shooting algorithms. In Sect. 3 we illustrate our novel concepts on the simple potential introduced in Ref. [29]. We end with concluding remarks.

2 Theory and Methods

2.1 The Reweighted Path Ensemble

In principle, one can obtain the complete path ensemble from a long straightforward MD simulation, but this would be very inefficient as transitions between states are extremely rare. In Ref. [29] we showed how to compute the complete path ensemble efficiently within the transition interface sampling (TIS) [16] methodology by reweighting the interface ensembles using the well-known Ferrenberg-Swendsen histogram reweighting [17]. TIS introduces $n + 1$ non-intersecting hyper-surfaces or interfaces between states A and B described by a progress variable $\lambda(\mathbf{x})$, which is a function of the phase space point \mathbf{x} . Each point $\mathbf{x} \equiv \{\mathbf{r}^N, \mathbf{p}^N\}$ contains the positions and momenta of all N particles in the system. The $n + 1$ interfaces are defined by an ordered sequence $\lambda_0, \lambda_1, \dots, \lambda_n$ where the first interface λ_0 is the boundary of state A , and the last one λ_n is identical to the boundary of state B . The TIS path ensemble consists of trajectories $\mathbf{x}^L \equiv \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_L\}$, a sequence of phase space points \mathbf{x} . Each point or time slice \mathbf{x} in the trajectories is separated from the previous by a time step Δt , yielding a path length $t_P = L \Delta t$. Paths belonging to the TIS ensemble of interface i start in A , and either end in B or return to A , provided that they have crossed a certain interface λ_i . Defining the region of phase space beyond interface i by $\Lambda_i^+ = \{\mathbf{x}: \lambda(\mathbf{x}) > \lambda_i\}$, the TIS path probability is

$$\mathcal{P}_{AA_i}[\mathbf{x}^L] = \tilde{h}_i[\mathbf{x}^L] \rho_{eq}(\mathbf{x}_0) \prod_{i=0}^{L-1} p(\mathbf{x}_i \rightarrow \mathbf{x}_{i+1}) / Z_{AA_i}, \tag{1}$$

where $\rho_{eq}(\mathbf{x})$ is the (steady state) equilibrium distribution, e.g. the canonical distribution $\rho_{eq}(\mathbf{x}) = e^{-\beta H(\mathbf{x})}$, with H the Hamiltonian, and $\beta = 1/k_B T$ is the reciprocal temperature, with k_B Boltzmann’s constant. $p(\mathbf{x} \rightarrow \mathbf{y})$ denotes the Markovian short time probability to reach state \mathbf{y} from \mathbf{x} within one time interval Δt . The indicator function $\tilde{h}_i[\mathbf{x}^L]$ equals unity for trajectories that start in A , end in A or B , and cross λ_i , and is zero otherwise. The normalizing factor Z_{AA_i} is defined by $\int \mathcal{D}\mathbf{x}^L \mathcal{P}_{AA_i}[\mathbf{x}^L] = 1$ where the path integral runs over all paths of all possible lengths L . The TIS path ensemble can be sampled using the regular shooting algorithm [9]. Sampling efficiency can be improved by employing replica exchange TIS [3, 14].

An expression similar to (1) can be written for the reverse process path probability $\mathcal{P}_{BA_i^-}[\mathbf{x}^L]$ of leaving B and crossing the interface i into region of phase space $\Lambda_i^- = \{\mathbf{x}: \lambda(\mathbf{x}) < \lambda_i\}$. Reweighting the TIS path ensembles by applying histogram reweighting to the crossing histograms [29] leads to the reweighted path ensemble

$$\mathcal{P}[\mathbf{x}^L] = c_A \sum_{j=1}^{n-1} \mathcal{P}_{AA_j}[\mathbf{x}^L] W^A[\mathbf{x}^L] + c_B \sum_{j=1}^{n-1} \mathcal{P}_{BA_j}[\mathbf{x}^L] W^B[\mathbf{x}^L], \tag{2}$$

where the function $W^A[\mathbf{x}^L] = \sum_{i=1}^{n-1} \tilde{w}_i^A \theta(\lambda_{\max}[\mathbf{x}^L] - \lambda_i) \theta(\lambda_{i+1} - \lambda_{\max}[\mathbf{x}^L])$ selects the correct interface weight for each path \mathbf{x}^L based on the maximum λ along the path. Here the θ -function $\theta(x) = 1$ for $x > 0$ and zero otherwise. Similarly $W^B[\mathbf{x}^L] = \sum_{i=1}^{n-1} \tilde{w}_i^B \theta(\lambda_{\min}[\mathbf{x}^L] - \lambda_i) \theta(\lambda_{i+1} - \lambda_{\min}[\mathbf{x}^L])$ selects the weights for paths from B based on the minimum λ along the path. The weights \tilde{w}_i^A and \tilde{w}_i^B can be obtained from histogram reweighting [17] of the forward and reverse crossing probability histograms, respectively [29]. The unknown constants c_A and c_B follow from matching the AB and BA histograms for overlapping interfaces [29].

This reweighted path ensemble only takes into account paths that leave A (or B) and cross an interface, not all the paths that remain in A (or B) and do not cross any interfaces. However, these paths are easily accessible from the TIS flux computation [16], or from the additional first and last interface ensemble $P_{A_1}^-$ (and $P_{A_{n-1}}^+$) in the replica exchange TIS framework [3, 14]. The complete path ensemble is then found by adding $P_{A_1}^-$ and $P_{A_{n-1}}^+$ to (2), where the final and initial parts of the trajectories in $P_{A_1}^-$ between λ_0 and λ_1 and likewise in $P_{A_{n-1}}^+$ between λ_n and λ_{n-1} are removed to avoid double counting. (Note that in many applications $\lambda_0 = \lambda_1$, although this is not required.) The complete reweighted path ensemble finally becomes

$$\begin{aligned} \mathcal{P}_c[\mathbf{x}^L] = c_A & \left[w_1^A P_{A_1}^- + \sum_{j=1}^{n-1} \mathcal{P}_{AA_j}[\mathbf{x}^L] W^A[\mathbf{x}^L] \right] \\ & + c_B \left[w_{n-1}^B P_{A_{n-1}}^+ + \sum_{j=1}^{n-1} \mathcal{P}_{BA_j}[\mathbf{x}^L] W^B[\mathbf{x}^L] \right]. \end{aligned} \tag{3}$$

In summary, the idea of the reweighted path ensemble is to obtain the unbiased path ensemble. While paths crossing the barrier are very rare, they are sampled through TIS, and obtain their corrected weight through applying histogram reweighting.

2.2 Projections of the Reweighted Path Ensemble

Once the reweighted path ensemble is known it can be used to compute all kinds of distributions, for instance, the free energy $F(\mathbf{q})$ as a function of an arbitrary m -dimensional collective variable space $\mathbf{q} = \{q^{(1)}, q^{(2)}, \dots, q^{(m)}\}$. Up to a constant the free energy is equal to the logarithm of the probability density $\rho(\mathbf{q})$ to find a configuration in the unbiased ensemble at a certain point \mathbf{q}

$$F(\mathbf{q}) = -k_B T \ln \rho(\mathbf{q}) + const, \tag{4}$$

where k_B is Boltzmann’s constant. This probability density $\rho(\mathbf{q})$ is given by

$$\rho(\mathbf{q}) = C \int \mathcal{D}\mathbf{x}^L \mathcal{P}_c[\mathbf{x}^L] \sum_{k=0}^L \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q}), \tag{5}$$

where $\delta(\mathbf{z}) = \prod_{i=1}^m \delta(z^{(i)})$ is the Dirac delta function, and $C^{-1} = \int \mathcal{D}\mathbf{x}^L \sum_{k=0}^L \mathcal{P}_c[\mathbf{x}^L]$ is a normalizing constant. Besides the free energy we can project the averaged committor function p_B on arbitrary surfaces by using the indicator function $h_B(\mathbf{x}_L)$

$$p_B(\mathbf{q}) = \frac{\int \mathcal{D}\mathbf{x}^L \mathcal{P}_c[\mathbf{x}^L] h_B(\mathbf{x}_L) \sum_{k=0}^L \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q})}{\int \mathcal{D}\mathbf{x}^L \mathcal{P}_c[\mathbf{x}^L] \sum_{k=0}^L \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q})}, \tag{6}$$

where $h_B(\mathbf{x}) = 1$ if $x \in B$ and zero otherwise. This expression holds because each slice on the path can be seen as a realization of a committor shot [25]. It is important to realize that $p_B(\mathbf{q})$ is an *averaged* committor and not the full committor $p_B(\mathbf{r})$. The full committor $p_B(\mathbf{r})$ is computed for one specific configuration \mathbf{r} , and thus allows for a committor distribution test (see e.g. Ref. [4]), whereas $p_B(\mathbf{q}) = \int d\mathbf{r} \prod_{i=1}^m \delta(q^{(i)}(\mathbf{r}) - q^{(i)}) p_B(\mathbf{r})$ is only the average

value. While, in principle, the reweighted path ensemble contains also the full committor, in practice we do not have access to $p_B(\mathbf{r})$ because there is usually only one shot for each \mathbf{r} , thus only a 0 or a 1. Nevertheless, $p_B(\mathbf{q})$ can yield crucial hints for reaction coordinate analysis [27, 28].

Another interesting quantity is the path density $n(\mathbf{q})$:

$$n(\mathbf{q}) = C \int \mathcal{D}\mathbf{x}^L \mathcal{P}_c[\mathbf{x}^L] h_{\mathbf{q}}(\mathbf{x}^L), \tag{7}$$

where $h_{\mathbf{q}}(\mathbf{x}^L) = \text{sgn}(\int_0^{L\Delta t} dt \delta(\mathbf{q} - \mathbf{q}(\mathbf{x}_t))) \approx \text{sgn}(\sum_{k=0}^L \delta(\mathbf{q} - \mathbf{q}(\mathbf{x}_k)))$. This function equals unity if the path visits \mathbf{q} , and zero otherwise. C is the constant defined above. In other words, n measures the number (density) of paths at point \mathbf{q} . (If subsequent slices are more than one bin apart in the histogram all intermediate bins have to be filled, as the condition is that the path *visits* the position in q -space. To avoid complications, we use an interpolation scheme.) Note that this path density is, in contrast to the configurational density $\rho(\mathbf{q})$, not normalized to unity. The reason for this becomes clear below.

Vanden-Eijnden and coworkers [12, 13] introduced the notion of the current associated with the reactive trajectories, defined as

$$J(\mathbf{q}) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T dt \dot{\mathbf{q}} \delta(\mathbf{q} - \mathbf{q}(t)) \chi_{AB}(\mathbf{x}_t), \tag{8}$$

where $\chi_{AB}(\mathbf{x}_t) = 1$ if \mathbf{x}_t lies on a trajectory connecting A with B , and zero otherwise. The flow lines of this vector field indicate how the reactive trajectories connect A with B on average, while not counting dead end detours [12, 13]. Using the reweighted path ensemble we can express this current as

$$J(\mathbf{q}) = C \int \mathcal{D}\mathbf{x}^L \mathcal{P}_c[\mathbf{x}^L] h_A(\mathbf{x}_0) h_B(\mathbf{x}_L) \sum_{k=0}^L \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q}) \dot{\mathbf{q}}(\mathbf{x}_k). \tag{9}$$

Here $\dot{\mathbf{q}} \approx \frac{\mathbf{q}(x_{k+1}) - \mathbf{q}(x_k)}{\Delta t}$ is the estimated velocity in the projected space \mathbf{q} .

In principle, one can project any other property of the path ensemble in this way.

2.3 Generalized Densities and Committors

Assuming there are no other stable states than A and B , the path ensembles naturally divide in four categories: paths that leave A and return to A , paths that leave A and reach B , and the reverse paths that leave B and reach A , and finally the paths that leave B to return to B . These four types of paths are denoted AA , AB , BA and BB , respectively. For each of these types we can define a configurational probability density

$$\rho_{ij}(\mathbf{q}) = C \int \mathcal{D}\mathbf{x}^L \mathcal{P}_c[\mathbf{x}^L] h_i(\mathbf{x}_0) h_j(\mathbf{x}_L) \sum_{k=0}^L \delta(\mathbf{q}(\mathbf{x}_k) - \mathbf{q}), \tag{10}$$

where i, j denotes one of the $\{AA, AB, BA, BB\}$ possibilities. Note that the sum of the four configurational densities is simply the total configurational density

$$\rho(\mathbf{q}) = \rho_{AA}(\mathbf{q}) + \rho_{AB}(\mathbf{q}) + \rho_{BA}(\mathbf{q}) + \rho_{BB}(\mathbf{q}). \tag{11}$$

From these configuration densities we can directly define generalized (averaged) committors [26] as the relative probability that a certain path type is found for a certain \mathbf{q} ,

$$p_{AA}(\mathbf{q}) = \frac{\rho_{AA}(\mathbf{q})}{\rho(\mathbf{q})} \tag{12a}$$

$$p_{AB}(\mathbf{q}) = \frac{\rho_{AB}(\mathbf{q})}{\rho(\mathbf{q})} \tag{12b}$$

$$p_{BA}(\mathbf{q}) = \frac{\rho_{BA}(\mathbf{q})}{\rho(\mathbf{q})} \tag{12c}$$

$$p_{BB}(\mathbf{q}) = \frac{\rho_{BB}(\mathbf{q})}{\rho(\mathbf{q})}. \tag{12d}$$

These generalized committors naturally add up to unity. These definitions are analogous to that of the average committor (6). In fact, the normal averaged committors $p_A(\mathbf{q})$ and $p_B(\mathbf{q})$ are simply

$$p_A(\mathbf{q}) = p_{AA}(\mathbf{q}) + p_{BA}(\mathbf{q}) = \frac{\rho_{AA}(\mathbf{q}) + \rho_{BA}(\mathbf{q})}{\rho(\mathbf{q})} \tag{13a}$$

$$p_B(\mathbf{q}) = p_{AB}(\mathbf{q}) + p_{BB}(\mathbf{q}) = \frac{\rho_{AB}(\mathbf{q}) + \rho_{BB}(\mathbf{q})}{\rho(\mathbf{q})}. \tag{13b}$$

Similarly, we can define the path densities for the four sub-ensembles,

$$n_{ij}(\mathbf{q}) = C \int \mathcal{D}\mathbf{x}^L \mathcal{P}_c[\mathbf{x}^L] h_i(\mathbf{x}_0) h_j(\mathbf{x}_L) h_{\mathbf{q}}(\mathbf{x}^L), \tag{14}$$

where i, j is again one of the $\{AA, AB, BA, BB\}$ possibilities.

2.4 The TIS Rate Constant

In this section we derive that the rate constants directly follow from the above defined distributions. The forward rate constant k_{AB} in TIS [15, 16] is expressed as

$$k_{AB} = \phi_{01} P_A(\lambda_n | \lambda_1), \tag{15}$$

with ϕ_{01} the flux through interface λ_1 out of state A and crossing probability $P_A(\lambda_n | \lambda_1)$ is the probability that trajectories once they crossed λ_1 will reach B (i.e. interface λ_n) without first returning to A . We can express this crossing probability directly in terms of the path densities $n_{ij}(\mathbf{q})$, by identifying $\mathbf{q} = \lambda$,

$$P_A(\lambda_n | \lambda_1) = \frac{n_{AB}(\lambda_1)}{n_{AA}(\lambda_1) + n_{AB}(\lambda_1)}, \tag{16}$$

which is the number of paths from A to B , divided by all paths leaving A at interface λ_1 .

TIS expresses the rate also as a product over the crossing probabilities for an arbitrary interface λ_i [15, 16]

$$k_{AB} = \phi_{01} P_A(\lambda_n | \lambda_i) P_A(\lambda_i | \lambda_1). \tag{17}$$

The first factor $P_A(\lambda_n|\lambda_i)$ is analogous to (16),

$$P_A(\lambda_n|\lambda_i) = \frac{n_{AB}(\lambda_i)}{n_{AA}(\lambda_i) + n_{AB}(\lambda_i)}. \tag{18}$$

Comparing (15) and (17) gives the expression for the crossing probability

$$P_A(\lambda_i|\lambda_1) = \frac{n_{AA}(\lambda_i) + n_{AB}(\lambda_i)}{n_{AB}(\lambda_i)} \frac{n_{AB}(\lambda_1)}{n_{AA}(\lambda_1) + n_{AB}(\lambda_1)}. \tag{19}$$

As the path density $n_{AB}(\lambda)$ is necessarily constant between λ_1 and λ_n , this simplifies into

$$P_A(\lambda_i|\lambda_1) = \frac{n_{AA}(\lambda_i) + n_{AB}(\lambda_i)}{n_{AA}(\lambda_1) + n_{AB}(\lambda_1)}. \tag{20}$$

Indeed, this can be interpreted as the number of paths that reach λ_i with respect the paths that leave from λ_1 .

Based on a very long MD trajectory a more direct way of defining the rate constant is

$$k_{AB} = \frac{N_{AB}(\mathcal{T})}{\mathcal{T}_A}. \tag{21}$$

Here the rate is equal to $N_{AB}(\mathcal{T})$, the number of transitions from A to B in a time \mathcal{T} , divided by the dwelling time \mathcal{T}_A that the trajectory spends in state A . If one has access to the reweighted path ensemble the number of transitions N_{AB} is proportional to the path density $n_{AB}(\lambda)$. The dwelling time \mathcal{T}_A is proportional to the integral of the configuration density $\Delta t \int_A d\lambda \rho(\lambda)$, where the integral runs over the phase space volume of the basin of attraction of A (i.e. is probably larger than the TIS state definition). The rate constant can be thus be expressed as

$$k_{AB} = \frac{n_{AB}(\lambda)}{\Delta t \int_A d\lambda \rho(\lambda)}. \tag{22}$$

Note that here we have made use of the fact that the path density and the configurational densities in the reweighted path ensemble contain the same constant C . Comparing the above equation to the TIS rate expression (15)

$$k_{AB} = \phi_{01} \frac{n_{AB}(\lambda_1)}{n_{AA}(\lambda_1) + n_{AB}(\lambda_1)} = \frac{n_{AB}(\lambda_1)}{\Delta t \int_A d\lambda \rho(\lambda)}, \tag{23}$$

leads to an expression for the flux

$$\phi_{01} = \frac{n_{AA}(\lambda_1) + n_{AB}(\lambda_1)}{\Delta t \int_A d\lambda \rho(\lambda)}, \tag{24}$$

which is the number of times paths that cross the interface λ_1 in the average dwell time \mathcal{T}_A .

The path density $n_{AB} = n_{BA}$ is constant between the stable states, and is in fact independent of the used order parameter λ (for a one dimensional representation). In contrast, the path densities n_{AA} and n_{BB} are dependent on the choice of λ . This makes n_{AB} a fundamental kinetic property of the system.

2.5 The Transition State Theory Rate and the Transmission Coefficient

The transition state theory gives the following expression for the rate constant [1, 6]

$$k_{AB}^{TST} = \frac{e^{-\beta F(\lambda^*)}}{\int_A e^{-\beta F(\lambda)}} \langle \dot{\lambda} \theta(\dot{\lambda}) \rangle_{\lambda=\lambda^*}, \tag{25}$$

where $F(\lambda)$ is the free energy as a function of λ , λ^* denotes the location of the maximum of the free energy barrier, often identified with the dividing surface or separatrix, and $\dot{\lambda}$ is a time derivative, i.e. the velocity of the trajectories in terms of the order parameter λ . The angular brackets denote an average over phase points that are restricted to the dividing surface λ^* . The first factor is the probability to be on top of the barrier, while the second is an average over phase points at the dividing surface that have a positive velocity $\dot{\lambda}$. Hence, transition state theory assumes that all trajectories that reach the dividing surface or separatrix λ^* with a positive velocity will also reach B and thus contribute to the rate. In reality, however, there can be recrossings, and hence the k_{TST} is an upper limit.

In the reactive flux method [1, 6] the rate constant is corrected by only counting the paths that really reach B after time t :

$$k(t) = \frac{e^{-\beta F(\lambda^*)}}{\int_A e^{-\beta F(\lambda)}} \langle \dot{\lambda}_0 \theta(\lambda_t - \lambda^*) \rangle_{\lambda_0=\lambda^*}. \tag{26}$$

Here, the second factor is an average of the initial velocity $\dot{\lambda}_0$ over trajectories that start at λ^* and are in the final state B at time t (here B is defined as $\lambda > 0$). This function $k(t)$ reaches a plateau value $k(t) = k_{AB}$ for times $\tau_{mol} < t \ll \tau_{rxn}$. The function $R(t) = \langle \dot{\lambda}_0 \theta(\lambda_t - \lambda^*) \rangle_{\lambda_0=\lambda^*}$ then also reaches a constant value R , which is known as the unnormalized transmission coefficient.

Making use of the relation between the free energy $F(\lambda)$ and the configurational density $\rho(\lambda)$ (4), the rate constant can be expressed as

$$k_{AB} = \frac{\rho(\lambda^*)}{\int_A d\lambda \rho(\lambda)} R. \tag{27}$$

Comparing this expression with (23) the unnormalized transmission coefficient is thus simply

$$R = \frac{n_{AB}(\lambda^*)}{\Delta t \rho(\lambda^*)}. \tag{28}$$

The *normalized* transmission coefficient κ relates the transition theory state rate constant estimate to the (exact) reactive flux rate constant through $k_{AB} = \kappa k_{AB}^{TST}$. The expression for the transmission coefficient is thus the plateau value of

$$\kappa = \frac{\langle \dot{\lambda}_0 \theta(\lambda_t - \lambda^*) \rangle_{\lambda_0=\lambda^*}}{\langle \dot{\lambda}_0 \theta(\dot{\lambda}) \rangle_{\lambda_0=\lambda^*}} = \frac{R}{\frac{1}{2} \langle |\dot{\lambda}_0| \rangle_{\lambda_0=\lambda^*}}, \tag{29}$$

where in the last equality we made use of the fact that the distribution of $\dot{\lambda}$ is symmetric. The transmission coefficient counts the fraction of paths that start at dividing surface λ^* with an initial positive velocity (towards state B), and eventually reach B . The maximum value for κ is unity, meaning that all paths starting at the dividing surface are reactive and cross the dividing surface only once.

In principle one has to compute the average over $\dot{\lambda}$ explicitly in a simulation. However, we can connect the transmission coefficient expression to the path densities by introducing the concept of the density $n'_{ij}(\lambda)$ of paths crossing an interface λ including recrossings. (In contrast, the density n_{ij} counts only the effective crossings [15].) Using this density of recrossing we can express the configurational density as

$$\rho_{ij}(\lambda) = n'_{ij}(\lambda) \frac{1}{\langle |\dot{\lambda}| \rangle_{\lambda} \Delta t}. \tag{30}$$

The factor $(\langle |\dot{\lambda}| \rangle_{\lambda} \Delta t)^{-1}$ arises from the fact that the configurational density contains only phase points separated by Δt , while the path density counts the crossings of an interface, hence also points between configurations contribute. Inserting (30) into (29), and using (28) we obtain

$$\begin{aligned} \kappa &= \frac{2n_{AB}(\lambda^*)}{\Delta t \rho(\lambda^*) \langle |\dot{\lambda}_0| \rangle_{\lambda_0=\lambda^*}} \\ &= \frac{2n_{AB}(\lambda^*)}{n_{AA}^r(\lambda^*) + n_{AB}^r(\lambda^*) + n_{BA}^r(\lambda^*) + n_{BB}^r(\lambda^*)}. \end{aligned} \tag{31}$$

Now, since $n'_{ij}(\lambda) \geq n_{ij}(\lambda)$ by necessity, and $n_{AB}(\lambda) = n_{BA}(\lambda)$, the maximum κ is indeed unity, as it should be when all paths at λ^* are reactive and do not recross. This expression of κ can be extended to arbitrary λ interfaces.

While one could explicitly compute the recrossing path densities n'_{ij} in the reweighted path ensemble, the same information can be found in the unnormalized transmission coefficient (28), which is in turn based on the configurational density $\rho(\lambda)$.

The definition of R (28) can also be extended to any interface

$$R(\lambda) = \frac{n_{AB}(\lambda)}{\Delta t \rho(\lambda)}. \tag{32}$$

We can even extend this expression for R to generic multidimensional order parameters \mathbf{q}

$$R(\mathbf{q}) = \frac{n_{AB}(\mathbf{q})}{\Delta t \rho(\mathbf{q})}. \tag{33}$$

However, we must realize that for a multidimensional \mathbf{q} the condition that if q_i and q_j is visited then also all points between them must be visited is not valid. So, the multidimensional expression cannot be used for quantitative predictions of the transmission coefficient. Still, such projections can be useful as an analysis tool, as they indicate the location of the dividing surface (in the \mathbf{q} space).

2.6 Bayesian Path Statistics

In this section we show the relation between the reweighted path ensemble distributions and the Bayesian path statistics introduced by Hummer [19]. The Bayesian relation for the path statistics is given by

$$P(q|\text{TP})P(\text{TP}) = P(\text{TP}|q)P(q), \tag{34}$$

with $P(q) = \rho_{eq}(q) = \rho(q) / \int dq \rho(q)$ the equilibrium distribution. This relation can be used to express the rate, as well as to optimize the reaction coordinate [2, 19]. The analysis

starts by realizing that $P(\text{TP}|q)$, the probability for a configuration with a (one dimensional) order parameter q to lead to a transition path between A and B , is the sum of the generalized committor functions

$$P(\text{TP}|q) = p_{AB}(q) + p_{BA}(q) = 2p_{AB}(q) \approx 2p_A(q)p_B(q). \tag{35}$$

The factor of two arises because a forward and backward path can always be swapped. The approximation is for diffusive dynamics. The probability $P(q|\text{TP})$ is the probability density of the configurations in the transition path ensemble:

$$P(q|\text{TP}) = \frac{\rho_{eq}(q)(p_{AB}(q) + p_{BA}(q))}{\int dq \rho_{eq}(q)(p_{AB}(q) + p_{BA}(q))} = \frac{\rho_{AB}(q)}{\int dq \rho_{AB}(q)}. \tag{36}$$

Thus, the probability $P(\text{TP})$ to be on a transition path between A and B in an infinite trajectory is

$$P(\text{TP}) = \frac{P(\text{TP}|q)P(q)}{P(q|\text{TP})} = \frac{2 \int dq \rho_{AB}(q)}{\int dq \rho(q)}. \tag{37}$$

Using $\langle h_A \rangle + \langle h_B \rangle = 1$ and $\langle h_A \rangle k_{AB} = \langle h_B \rangle k_{BA}$, the rate constant can be written as [19]

$$2(k_{AB}^{-1} + k_{BA}^{-1})^{-1} = 2\langle h_A \rangle k_{AB} \approx P(\text{TP})/\langle t_P \rangle, \tag{38}$$

where $\langle t_P \rangle$ is the average transition path length. The probability $P(\text{TP})$ is, besides by (37), also given by the fraction of time spend on transition paths

$$P(\text{TP}) = \frac{\text{total time on } AB}{\text{total time}} = \frac{2 \int dq \rho_{AB}(q) \Delta t}{\int dq \rho(q) \Delta t}, \tag{39}$$

which is indeed identical to (37). The total time on transition paths divided by the average transition path length $\langle t_P \rangle$ is clearly the number of transition paths $n_{AB} + n_{BA} = 2n_{AB}$ in the total time. Combining this fact with (38) and (39) the rate becomes

$$k_{AB} = \frac{n_{AB}}{\langle h_A \rangle \int dq \rho(q) \Delta t} = \frac{n_{AB}}{\Delta t \int_A dq \rho(q)}, \tag{40}$$

which is indeed equal to the previous rate equations.

2.7 Shooting Point Distributions in the Path Ensembles

The path sampling algorithm selects a random point on the path called the shooting point, from which the new trial trajectory is generated. A successful path, that is, one which is a member of the TPS or TIS ensemble, can then be accepted according to a metropolis scheme. We can investigate the distribution of the trial shooting points P_{trial} and a subset of these, the accepted shooting points P_{acc} , i.e. shooting points that result in trajectories that are accepted. While the trial and accepted shooting points ensembles are not directly physically interpretable quantities, unlike the free energy, committor and path densities, as they depend on the algorithm, they can be expressed in these quantities.

In a two-way shooting transition path sampling simulation, a shooting point \mathbf{x} is accepted with a probability

$$P_{\text{acc}}^{\text{TPS}} = P(\mathbf{x}|\text{TP})(p_{AB} + p_{BA}) = 2P(\mathbf{x}|\text{TP})p_A p_B, \tag{41}$$

and rejected with a probability

$$P_{\text{rej}}^{\text{TPS}} = P(\mathbf{x}|\text{TP})(p_{AA} + p_{BB}) = P(\mathbf{x}|\text{TP})(p_A^2 + p_B^2), \tag{42}$$

where we left out the explicit dependence of the committor functions on \mathbf{x} . $P(\mathbf{x}|\text{TP})$ is the probability to choose a certain shooting point \mathbf{x} from a transition path TP. In the second equalities we assumed diffusive behavior. The trial shooting points are the sum of accepted and rejected points:

$$P_{\text{trial}}^{\text{TPS}} = P_{\text{acc}}^{\text{TPS}} + P_{\text{rej}}^{\text{TPS}} = P(\mathbf{x}|\text{TP}), \tag{43}$$

where we used $p_{AA} + p_{AB} + p_{BA} + p_{BB} = p_A + p_B = 1$. Applying (34) and (35) yields

$$P_{\text{trial}}^{\text{TPS}} = P(\mathbf{x}|\text{TP}) \propto (p_{AB} + p_{BA})P(\mathbf{x}) = 2p_B p_A P(\mathbf{x}). \tag{44}$$

Here, we have used the fact that $P(\text{TP})$ is a constant and $P(\mathbf{x})$ is the equilibrium (Boltzmann) distribution. The accepted shooting points are distributed according to

$$P_{\text{acc}}^{\text{TPS}} = P(\mathbf{x}|\text{TP})(p_{AB} + p_{BA}) \propto 4p_B^2 p_A^2 P(\mathbf{x}). \tag{45}$$

For one-way shooting, the probability to accept a forward shot is

$$P_{\text{acc,for}}^{\text{oneway}} = \frac{1}{2}P(\mathbf{x}|\text{TP})p_B, \tag{46}$$

whereas the probability for a backward shot

$$P_{\text{acc,back}}^{\text{oneway}} = \frac{1}{2}P(\mathbf{x}|\text{TP})p_A. \tag{47}$$

The factor 1/2 comes from the fact that we choose a forward or a backward shot with the same probability. Combining the two we find that the overall distribution of accepted shooting points is

$$P_{\text{acc}}^{\text{oneway}} = \frac{1}{2}P(\mathbf{x}|\text{TP}). \tag{48}$$

Analogous the distributions of the rejected shooting points are

$$P_{\text{rej,for}}^{\text{oneway}} = \frac{1}{2}P(\mathbf{x}|\text{TP})p_A \tag{49}$$

$$P_{\text{rej,back}}^{\text{oneway}} = \frac{1}{2}P(\mathbf{x}|\text{TP})p_B. \tag{50}$$

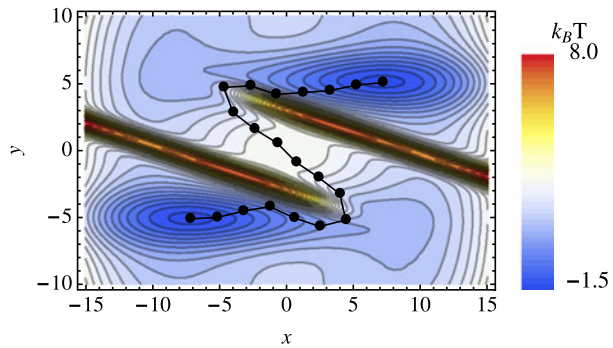
Adding the accepted and rejected distribution together yields

$$P_{\text{oneway}} = P_{\text{acc}}^{\text{oneway}} + P_{\text{rej}}^{\text{oneway}} = P(\mathbf{x}|\text{TP}). \tag{51}$$

This means for one-way shooting, the trial shooting points, the accepted shooting points as well as the rejected shooting points all share the same distribution $P(\mathbf{x}|\text{TP})$. Moreover, the trial distribution of two-way and one-way shooting are identical, and is given by (44).

The different shooting point distributions were investigated for a simple model in Ref. [25].

Fig. 1 The z -potential. The optimized string can act as a series of interfaces in a TIS simulation [29]



3 Results and Discussion

3.1 The z -Potential

In this section we employ the same two dimensional model system as used in Refs. [25, 29]. This potential is given by

$$\begin{aligned}
 V(x, y) = & \frac{x^4 + y^4}{20480} - 3e^{-0.01(x+5)^2 - 0.2(y+5)^2} - 3e^{-0.01(x-5)^2 - 0.2(y-5)^2} \\
 & + \frac{5e^{-0.2(x+3(y-3))^2}}{1 + e^{-x-3}} + \frac{5e^{-0.2(x+3(y+3))^2}}{1 + e^{x-3}} + 3e^{-0.01(x^2+y^2)} \quad (52)
 \end{aligned}$$

and is visualized in Fig. 1. It has two stable states, one at $(-7.2, -5.1)$ labeled *A* and one at $(7.2, 5.1)$ labeled *B*. These two minima are separated by a barrier in the shape of the letter *s* (or a reversed *z*), due to the presence of two high potential ridges. The potential at the origin is $V = 4.28k_B T$ above the minima.

The system consist of a single particle evolving according to Langevin dynamics on this potential. For details on the algorithm see e.g. Ref. [9]. The friction is set to $\gamma = 1$, the temperature is set to $\beta = 4$, where $\beta = 1/k_B T$ is the inverse temperature. The time-step is set to $\Delta t = 0.05$. The *A* and *B* regions itself are defined as ellipsoidal regions around the minima (x_m, y_m) as $\{x, y | (x - x_m)^2 + \frac{1}{16}(y - y_m)^2 < R^2\}$. The radius is set to $R = 0.5$. Dynamical trajectories connecting both states roughly follow the *z*-shape.

3.2 Replica Exchange TIS Using String Interfaces

While the simulation settings are the same as in Refs. [25, 29], we repeat them here for clarity. We perform replica exchange TIS simulations using a string description [11] to parametrize the λ -interfaces. This string consists of 20 equidistant beads connecting the minima of the stable states, and is optimized using a non-linear reaction coordinate analysis [25]. The resulting string is shown in Fig. 1. In the TIS simulation each configuration along the trajectory is assigned to the nearest string bead by a Voronoi construction. Linear interpolation yields a continuous value of $0 \leq \lambda \leq 1$ along the string; zero belonging to state *A*, one to state *B* [25, 29]. The interfaces were divided in two sets: 14 interfaces for the *AB* transition were chosen at $\lambda = \{0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.3, 0.325, 0.35, 0.375, 0.40, 0.425, 0.45, 0.5\}$, while the 14 interfaces for the *BA* transition were located at $\lambda = \{0.5, 0.55, 0.575, 0.6, 0.625, 0.65, 0.7, 0.75, 0.775, 0.8, 0.825, 0.85, 0.875, 0.9\}$. The

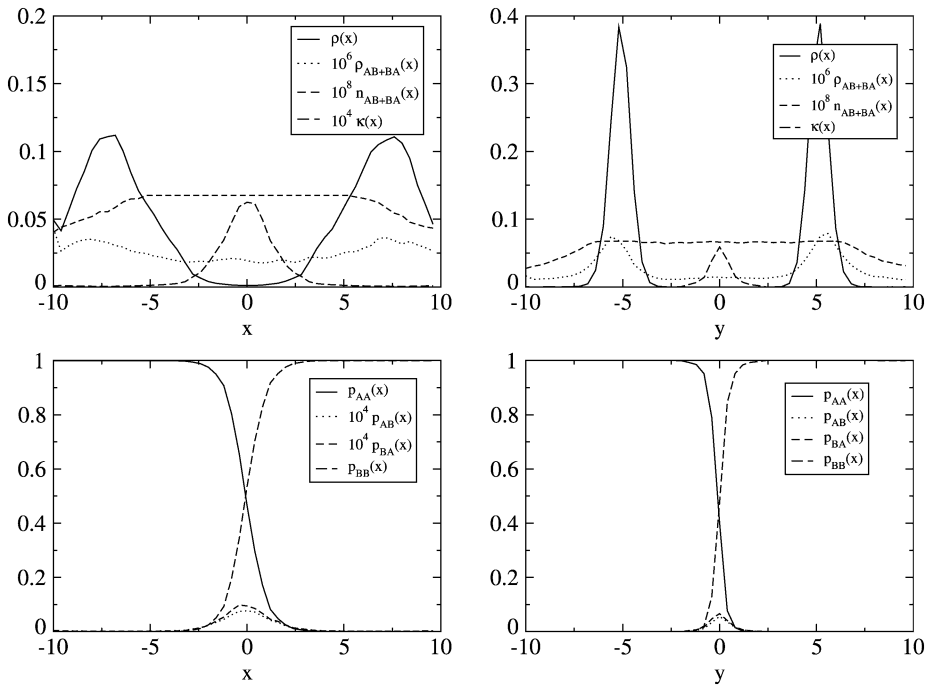


Fig. 2 Distributions projected on the x -axis (left) and y -axis (right). The distributions are scaled to the same range. *Top row*: the total configurational density ρ , the AB and BA configurational densities, the path density n and the transmission coefficient κ . *Bottom row*: generalized committors. Note the large difference between the κ -distribution of the x and y projections, and the large difference of the generalized committors (*bottom row*)

stable state definitions of A and B are such that they fall entirely within the range $\lambda < 0.1$ and $\lambda > 0.9$, respectively. Including an additional first and last interface at $\lambda = 0.1$ and $\lambda = 0.9$ enhances the path sampling and simultaneously allows computing the fluxes ϕ_{01} and $\phi_{n,n-1}$ [3, 14, 29]. The two sets of interfaces share one common interface at $\lambda = 0.5$ where the AB paths can exchange with BA paths. Paths were saved for further analysis each 100th shooting move. Of these paths only every tenth configuration was saved. For more details we refer to Ref. [29].

During the replica exchange TIS simulation we collected the forward and backward crossing probabilities. Subsequent application of histogram reweighting in combination with (3) results in the reweighted path ensemble with the constants c_A and c_B identical due to the symmetry of the potential.

3.3 1D Projections of the Reweighted Path Ensemble

We illustrate the different distributions defined in Sect. 2 for the reweighted path ensemble of the z -potential. In particular, we computed as a function of $\mathbf{q} = x, y$ and $\{x, y\}$, the configurational densities $\rho(\mathbf{q})$ from (10), the generalized committors $p(\mathbf{q})$ from (12a)–(12d), the path densities $n(\mathbf{q})$ from (14), and the transmission coefficient $\kappa(\mathbf{q})$ from (29). For the latter quantity we did compute $\langle \lambda_0 \rangle_{\lambda_0=\lambda} \approx 0.19$, which is approximately constant along the projection on x or y , as expected for a canonical ensemble.

Fig. 3 Generalized committors as a function of $\{x, y\}$. *Top left:* p_{AA} , *top right:* p_{BA} , *bottom left:* p_{AB} , *bottom right:* p_{BB} . The dark area corresponds to $p_{ij} = 0$, the light gray regions to $p_{ij} = 1$. The white area is not sampled

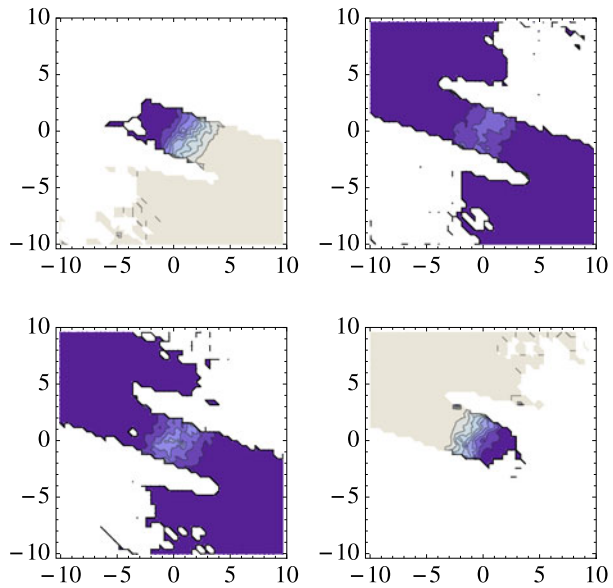
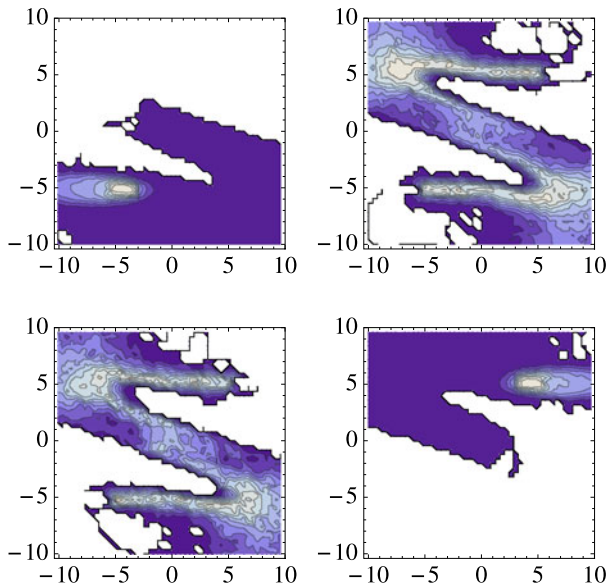


Figure 2 shows the one dimensional projections on x or y . The path density n_{AB} is combined with the n_{BA} , as the paths are completely reversible. Note that these very low path densities are independent of the choice of projection, and in fact constant in the barrier region. The configurational density $\rho(\mathbf{q})$ is very much dependent on the choice of projection. Most noticeable, the minimum density at the barrier location $\rho(x=0) \gg \rho(y=0)$. This is also the case for the free energy $\beta F(\mathbf{q}) = -\ln \rho(\mathbf{q})$ which is not shown. Note that the configurational densities of the AB and BA paths are not proportional to the total configurational density, since all AA and BB paths are missing. This is the reason why the transition path ensemble cannot reproduce the free energy. Most striking in these results is the fact that the transmission coefficient $\kappa(x)$ is four orders of magnitude lower than $\kappa(y)$. This is directly related to the large difference in configurational density, and thus the transition state theory rate constant estimates. It is also an indication that y is a much better reaction coordinate than x [25].

The correct rate constant can also be deduced from these data by applying (22). Because $\Delta t = 10 \times 0.05 = 0.5$ (the path data is only saved every 10 time steps), the $k_{AB} = n_{AB}/\Delta t \int_A d\lambda \rho(\lambda) = 3.28 \times 10^{-10}/(0.5 \cdot 0.5) = 1.3 \times 10^{-9}$. Indeed this is close to the direct TIS results reported in Ref. [29].

The generalized committors shown in the bottom row of Fig. 2 behave as expected. Note that the p_{AB} and p_{BA} are again orders of magnitude smaller for the x projection than for the y projections. Other than that, the two committor projections look very much the same. The quality of the reaction coordinate is thus not so much directly visible in the normal committor $p_B = p_{BA} + p_{BB} \approx p_{BB}$, but rather in the p_{AB} behavior. This was already argued by Hummer in Ref. [19] and by Best and Hummer [2], who defined the quality of the reaction coordinate as the ‘peakedness’ of $P(TP|q) = p_{AB} + p_{BA}$ (see (35)). Hence, also from the committor plots in Fig. 2 it follows that y is the better reaction coordinate of the two.

Fig. 4 The path densities n_{ij} as a function of $\{x, y\}$. *Top left: n_{AA} , top right: n_{BA} , bottom left: n_{AB} , bottom right: n_{BB} .* The dark regions indicated a low density, light gray regions a high density. The white area is not sampled



3.4 2D Projections of the Reweighted Path Ensemble

In addition to the 1D projections, we also computed the distributions as a function of two dimensions $\{x, y\}$. In the two dimensional plots we only focus on the generic behavior.

Figure 3 shows the generalized committors p_{ij} as a function of x and y . Note that the p_{AB} and p_{BA} distributions are only finite around the transition state as one would expect, as there the probability of a connecting path is highest. The maximum of $p_{AB} = p_{BA} \approx 0.5$, is significantly higher than the 1D projections, indicating that the best reaction coordinate is a two dimensional one, e.g. the string that follows the z -shape [25].

Figure 4 shows the path densities n_{ij} as computed by (14). Here, we do not refer to the absolute scale of the density, as these are already given in the 1D figures, but instead focus on the generic features. While the n_{AA} and n_{BB} densities are almost entirely located close to the stable states, n_{AB} and n_{BA} are spanning the entire barrier region, as in the 1D case. Such projections of path densities are thus useful for path analysis purposes, as they give almost equal weight to parts of the paths that are close to the stable states as well as to the transition states. The path density hence traces out a contour of the mechanism, in the \mathbf{q} -space. In contrast, in a regular free energy projection the transition state might be easily swamped by the AA and BB paths. Such path density projections were already used for analysis in Refs. [20–22, 35], albeit based on TPS only.

Figure 5 shows the logarithm of the path densities n_{ij} . Here, it can be seen how tiny the path densities n_{AA} and n_{BB} actually are at the barrier.

In Fig. 6 we plot the 2D unnormalized transmission coefficient $R(\mathbf{q})$ based on (28). Note that only close to the transition state at the origin R and hence κ is finite.

3.5 The Reactive Current

The reactive current $J(\mathbf{q})$ follows from (8), and is shown in Fig. 7. The time derivatives of q were approximated numerically as $\dot{q} \approx \frac{q(x_{k+1}) - q(x_k)}{\Delta t}$. The flow lines of the reactive current clearly follow the free energy landscape, and the current is highest in the stable states regions

Fig. 5 Logarithm of path densities n_{ij} as a function of $\{x, y\}$. *Top left:* $\ln n_{AA}$, *top right:* $\ln n_{BA}$, *bottom left:* $\ln n_{AB}$, *bottom right:* $\ln n_{BB}$. The *dark areas* include now the unsampled areas. The contours are separated by 1 unit, i.e. each contour demarcates a factor of e in density

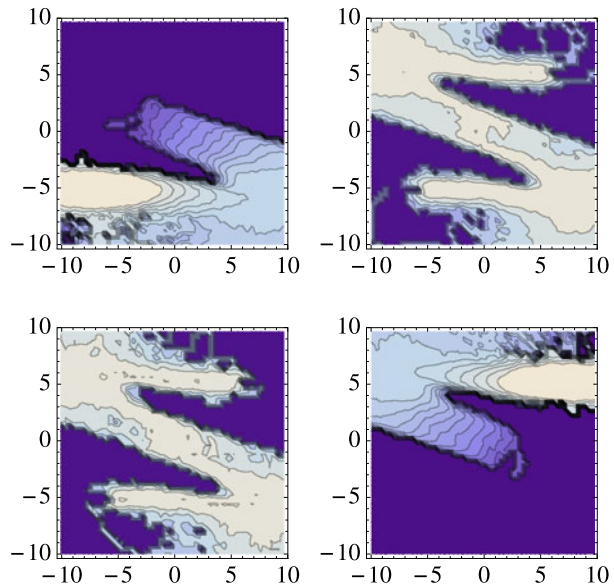
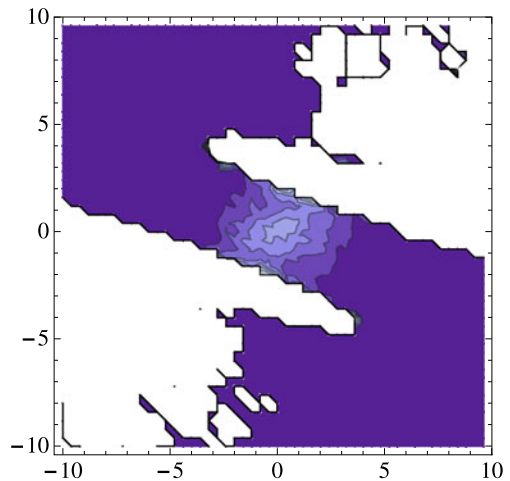


Fig. 6 Unnormalized transmission coefficient $R_{AB} + R_{BA}$ as function of $\{x, y\}$. The *dark regions* indicate a low value, *light gray* a high value. The *white area* is not sampled. Note that the transmission coefficient is highest at the barrier, as expected

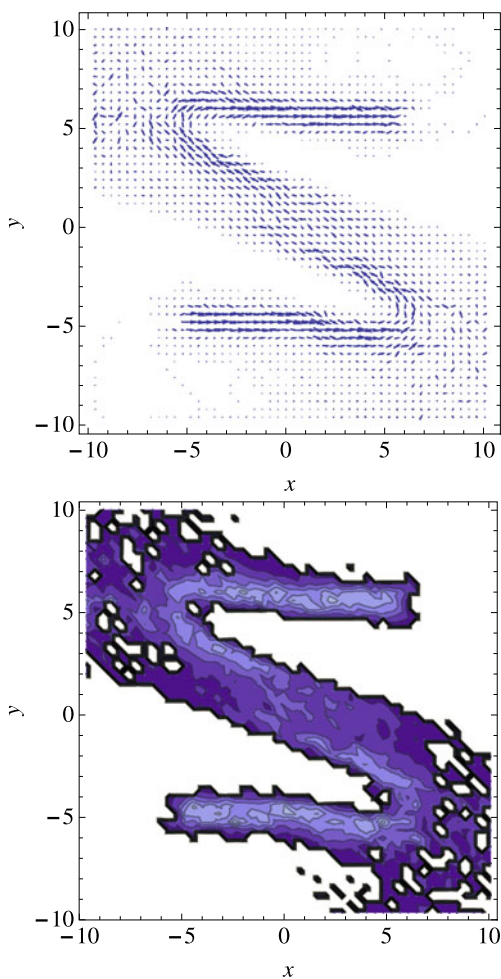


and smaller in the barrier region. In fact, the norm of the current (shown in Fig. 7) is similar to the path densities n_{AB} or n_{BA} in Fig. 4. As E and Vanden-Eijnden point out in Ref. [13], one of the advantages of the reactive current is that dead ends in the free energy landscape are not counted, as the current cancels in such regions. Dead end regions do appear in the path densities. This difference is especially visible in the top left and right bottom corners of the path density Fig. 4 and current norm projection in Fig. 7.

4 Conclusion

We have recapitulated the concepts of the reweighted path ensemble in the (replica exchange) TIS framework, and introduced several distributions that can be useful as an anal-

Fig. 7 *Top*: the reactive current J . *Bottom*: the norm of the current $|J|$. Note the similarity with the path density in Fig. 4



ysis tool. These distributions follow from projections of the reweighted path ensemble and include the configurational density, which is directly related to the free energy, the generalized committers, the reactive current and the novel path density n_{ij} . The path density n_{AB} (and n_{BA}) is a fundamental kinetic quantity that is independent on the choice of reaction coordinate. We related these distributions to the known expressions for the rate constant in the TIS framework as well as in the transition state theory and reactive flux framework. The latter led to a novel expression for the transmission coefficient in terms of the path density and the configurational density. We showed that these concepts are directly related to the Bayesian path statistics of Hummer [19]. In addition, the reweighted path ensemble framework is also closely related to the transition path theory of Vanden-Eijnden et al. [12, 13], as the calculation of the reactive current shows.

We illustrated the novel concepts of the path densities and the location dependent transmission coefficient on the simple 2D potential introduced in Ref. [29]. Such multidimensional projections show the power of the reweighted path ensemble. All static and dynamical information is in principle available in the ensemble of paths, in combination with the

natural path weights. We believe that these projections can be used as tools for analyzing the results of transition mechanisms in more complex systems.

Acknowledgements This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is part of the Netherlands Organization for Scientific Research (NWO).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Bennett, C.H.: Molecular dynamics and transition state theory: the simulation of infrequent events. In: Christofferson, R. (ed.) Algorithms for Chemical Computations. ACS Symposium Series, vol. 46. Am. Chem. Soc., Washington (1977)
2. Best, R., Hummer, G.: Reaction coordinates and rates from transition paths. *Proc. Natl. Acad. Sci. USA* **102**, 6732–6737 (2005)
3. Bolhuis, P.G.: Rare events via multiple reaction channels sampled by path replica exchange. *J. Chem. Phys.* **129**, 144108 (2008)
4. Bolhuis, P.G., Dellago, C., Chandler, D.: Reaction coordinates of biomolecular isomerization. *Proc. Natl. Acad. Sci. USA* **97**(11), 5877–5882 (2000)
5. Bolhuis, P.G., Chandler, D., Dellago, C., Geissler, P.L.: Transition path sampling: throwing ropes over rough mountain passes in the dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002)
6. Chandler, D.: Statistical mechanics of isomerization dynamics in liquids and the transition state. *J. Chem. Phys.* **68**, 2959 (1978)
7. Darve, E., Pohorille, A.: Calculating free energies using average force. *J. Chem. Phys.* **115**(20), 9169–9183 (2001)
8. Dellago, C., Bolhuis, P.G.: Transition path sampling and other advanced simulation techniques for rare events. *Adv. Polym. Sci.* **221**, 167–233 (2009)
9. Dellago, C., Bolhuis, P.G., Csajka, F.S., Chandler, D.: Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **108**, 1964 (1998)
10. Dellago, C., Bolhuis, P.G., Geissler, P.L.: Transition path sampling. *Adv. Chem. Phys.* **123**, 1–78 (2002)
11. E, W., Ren, W.i, Vanden-Eijnden, E.: String method for the study of rare events. *Phys. Rev. B.* **66**, 052301 (2002)
12. E, W., Vanden-Eijnden, E.: Towards a theory of transition paths. *J. Stat. Phys.* **123**, 503–523 (2006)
13. E, W., Vanden-Eijnden, E.: Transition-path theory and path-finding algorithms for the study of rare events. *Annu. Rev. Phys. Chem.* **61**, 391–420 (2010)
14. van Erp, T.S.: Reaction rate calculation by parallel path swapping. *Phys. Rev. Lett.* **98**, 268301 (2007)
15. van Erp, T.S., Bolhuis, P.G.: Elaborating transition interface sampling methods. *J. Comput. Phys.* **205**, 157 (2005)
16. van Erp, T.S., Moroni, D., Bolhuis, P.G.: A novel path sampling method for the sampling of rate constants. *J. Chem. Phys.* **118**, 7762–7774 (2003)
17. Ferrenberg, A.M., Swendsen, R.H.: Optimized Monte Carlo data-analysis. *Phys. Rev. Lett.* **63**(12), 1195–1198 (1989)
18. Grubmüller, H.: Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E* **52**, 2893–2906 (1995)
19. Hummer, G.: From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **120**, 516–523 (2004)
20. Juraszek, J., Bolhuis, P.G.: Sampling the multiple folding mechanisms of Trp-cage in explicit solvent. *Proc. Natl. Acad. Sci. USA* **103**(43), 15859 (2006)
21. Juraszek, J., Bolhuis, P.G.: Rate constant and reaction coordinate of Trp-cage folding in explicit water. *Biophys. J.* **95**, 4246–4257 (2008)
22. Juraszek, J., Bolhuis, P.G.: Effects of a mutation on the folding mechanism of beta-hairpin. *J. Phys. Chem. B* **113**(50), 16184–16196 (2009)
23. Keck, J.C.: Statistical investigation of dissociation cross-sections for diatoms. *Discuss. Faraday Soc.* **33**, 173 (1962)
24. Laio, A., Parrinello, M.: Escaping free-energy minima *Proc. Natl. Acad. Sci. USA* **99**, 12562 (2002)
25. Lechner, W., Rogal, J., Juraszek, J., Ensing, B., Bolhuis, P.G.: Nonlinear reaction coordinate analysis in the reweighted path ensemble. *J. Chem. Phys.* **133**(17), 174110 (2010)

26. Moroni, D.: Efficient sampling of rare event pathways. Ph.D. thesis, Universiteit van Amsterdam (2005)
27. Peters, B., Trout, B.L.: Obtaining reaction coordinates by likelihood maximization. *J. Chem. Phys.* **125**, 054108 (2006)
28. Peters, B., Beckham, G.T., Trout, B.L.: Extensions to the likelihood maximization approach for finding reaction coordinates *J. Chem. Phys.* **127**, 034109 (2007)
29. Rogal, J., Lechner, W., Juraszek, J., Ensing, B., Bolhuis, P.G.: The reweighted path ensemble. *J. Chem. Phys.* **133**(17), 174109 (2010)
30. Sørensen, M.R., Voter, A.F.: Temperature-accelerated dynamics for simulation of infrequent events. *J. Chem. Phys.* **112**, 9599 (2000)
31. Torrie, G.M., Valleau, J.P.: Monte Carlo free energy estimates using non-Boltzmann sampling. *Chem. Phys. Lett.* **28**, 578 (1974)
32. Voter, A.F.: Hyperdynamics: accelerated molecular dynamics of infrequent events. *Phys. Rev. Lett.* **78**, 3908 (1997)
33. Voter, A.F.: A method for accelerating the molecular dynamics simulation of infrequent events. *J. Chem. Phys.* **106**, 4665 (1997)
34. Voter, A.F., Sørensen, M.R.: Accelerating atomistic simulations of defect dynamics: hyperdynamics, parallel replica dynamics, and temperature-accelerated dynamics. *Mater. Res. Soc. Symp. Proc.* **538**, 427 (1999)
35. Vreede, J., Juraszek, J., Bolhuis, P.G.: Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein. *Proc. Natl. Acad. Sci. USA* **107**, 2397–2402 (2010)
36. Wang, F., Landau, D.P.: Efficient multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **86**(10), 2050 (2001)