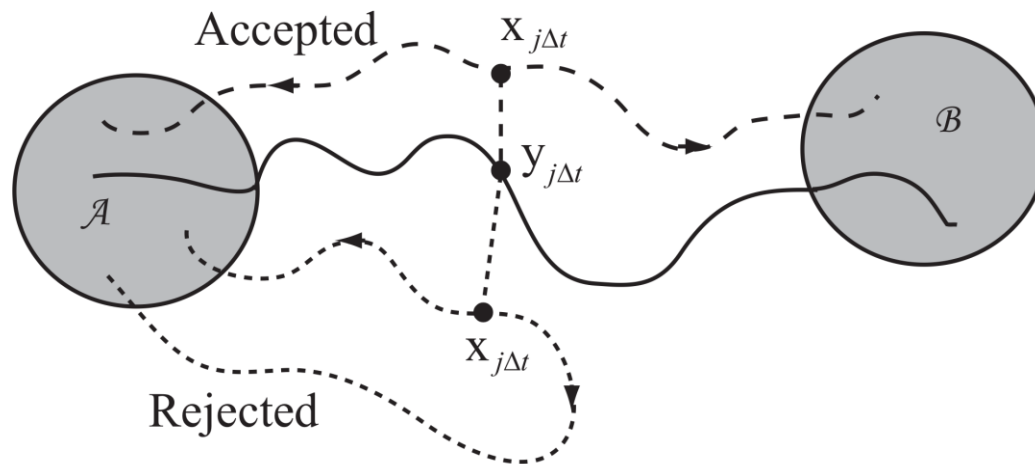# *Kinetics Computation*

2024 Winter Seminar
February 23th, 2024

Sangmin Lee

Content

Bottlenecks in Kinetics Computation

Transition Path Sampling

Menagerie: Other methods

Content

Bottlenecks in Kinetics Computation

Transition Path Sampling

Menagerie: Other methods

Theoretical foundation of molecular simulation

In classical mechanics, the equation of motion is integrated to generate the trajectory.

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{r}}_i}\right) - \frac{\partial \mathcal{L}}{\partial \mathbf{r}_i} = 0. \qquad \dot{q}_\alpha = \frac{\partial \mathcal{H}}{\partial p_\alpha}, \quad \dot{p}_\alpha = -\frac{\partial \mathcal{H}}{\partial q_\alpha}.$$

Energy is conserved in classical mechanics!

By solving F=ma, we sample the **microcanonical** ensemble (constant E) for ensemble averages.

$$\langle a \rangle = \frac{\int \mathrm{d}\mathbf{x}\, a(\mathbf{x})\delta(\mathcal{H}(\mathbf{x}) - E)}{\int \mathrm{d}\mathbf{x}\, \delta(\mathcal{H}(\mathbf{x}) - E)} = \lim_{\mathcal{T} \to \infty} \frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} \mathrm{d}t\, a(\mathbf{x}_t) \equiv \bar{a}.$$
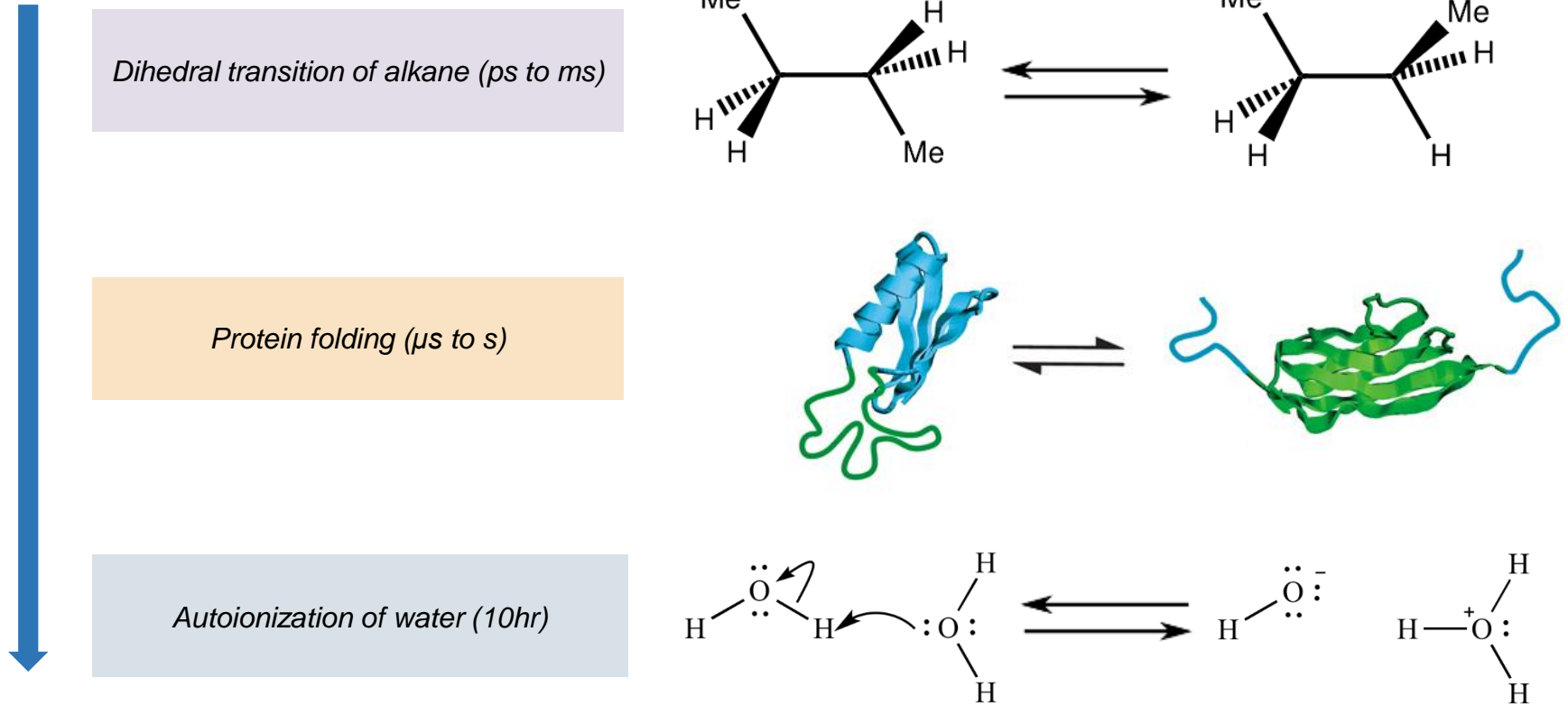
We need to sample the x_t in microcanonical ensemble, which we call it **trajectory**.(.dcd)

In here, we introduce the time discretization parameter dt, known as the **time step.**

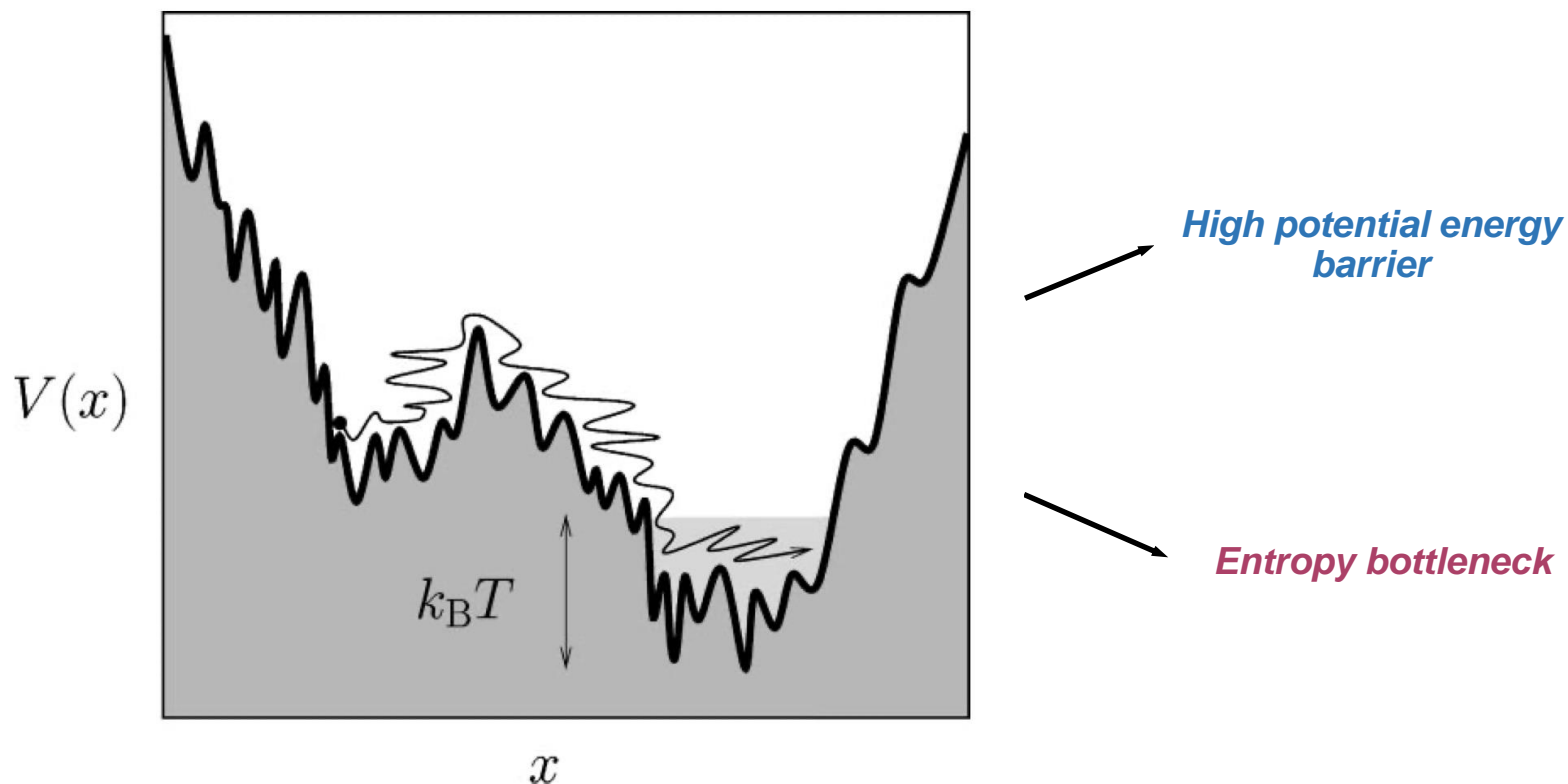Starting with the initial cond x_0, x_dt, x_2dt, x_3dt are generated by applying the integrator iteratively.

$$A = \langle a \rangle = \frac{1}{M} \sum_{n=1}^{M} a(\mathbf{x}_{n\Delta t}) \equiv \bar{a}.$$

Rare events: Disparity of timescales in MD

Dihedral transition of alkane (ps to ms)

Protein folding (μs to s)

Autoionization of water (10hr)

Conventional MD simulations(~1fs) cannot be used to study rare events

Why 'rare' event?

$V(x)$

$k_B T$

x

High potential energy barrier

Entropy bottleneck

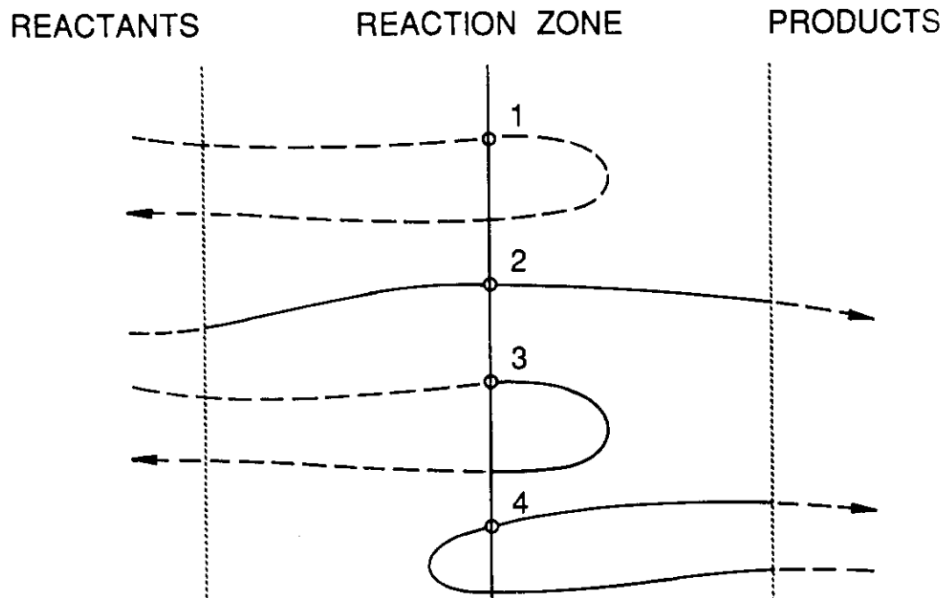*Fluctuation to surmount energy barrier higher than $k_B T$ is a rare event where equipartitioning holds*

6

Transition State Theory (TST)

■ If the TS location is **known**, we may construct a scheme for rare event study
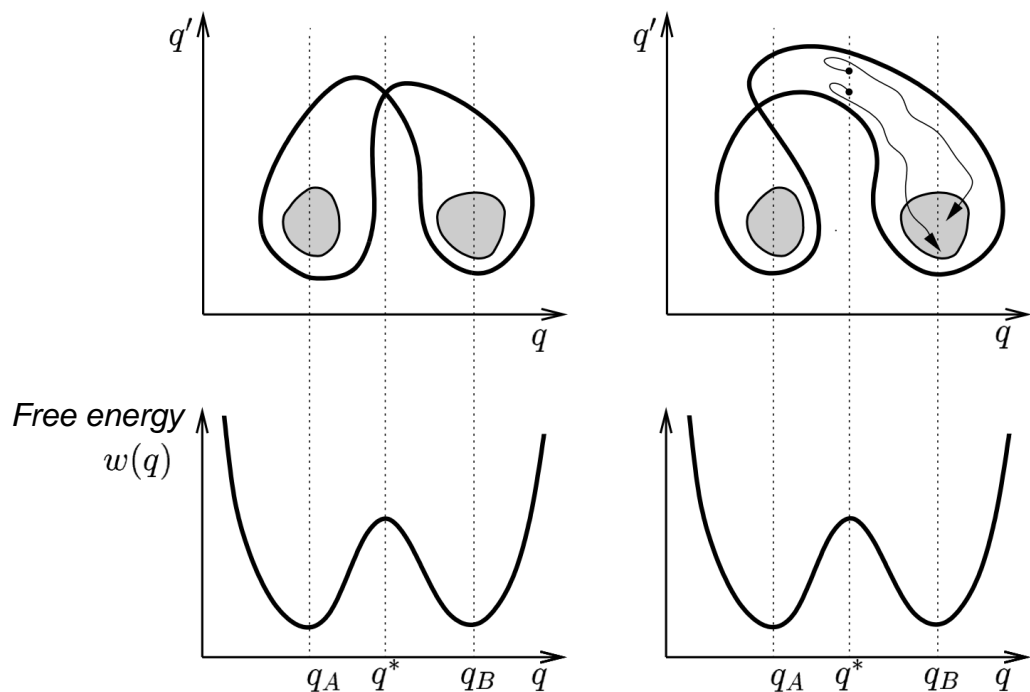

Transition State Theory

1. Calculate the reversible work and thus the probability for reaching TS

2. Initiate many fleeting trajectories from the TS

3. Determine the probability for successfully crossing the threshold

Transition from reactant to TS takes long time,

but transition from TS to reactant or product takes much shorter time!

Adv. Chem. Phys., **2007**, 381-431.

# Concept of 'order parameter' and 'reaction coordinate'

- **Order parameter** discriminates configurations belonging to the reactant and product
- **Reaction coordinate** describes the reaction progress from reactant to product (order parameter+TS)

Is q the desirable reaction coordinate and order parameter in both cases?

Bottleneck: Difficulty of identifying transition state surfaces

- TST procedure presupposes knowledge of the transition state.

- For the complex systems, **TS(saddle point in PES) is hard to be characterized**
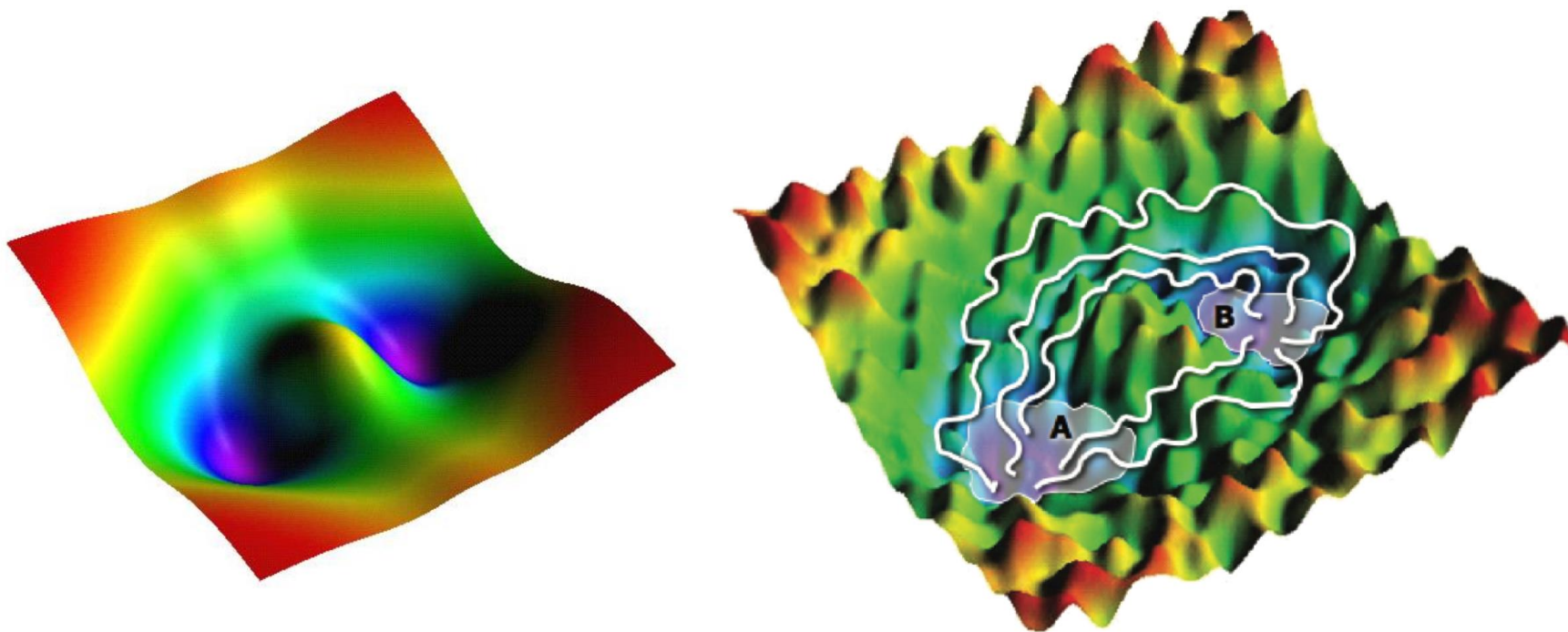
*We need kinetic computation algorithms that do not explicitly dependent upon the choice of reaction coordinate!*

9

Content

Bottlenecks in Kinetics Computation

Transition Path Sampling
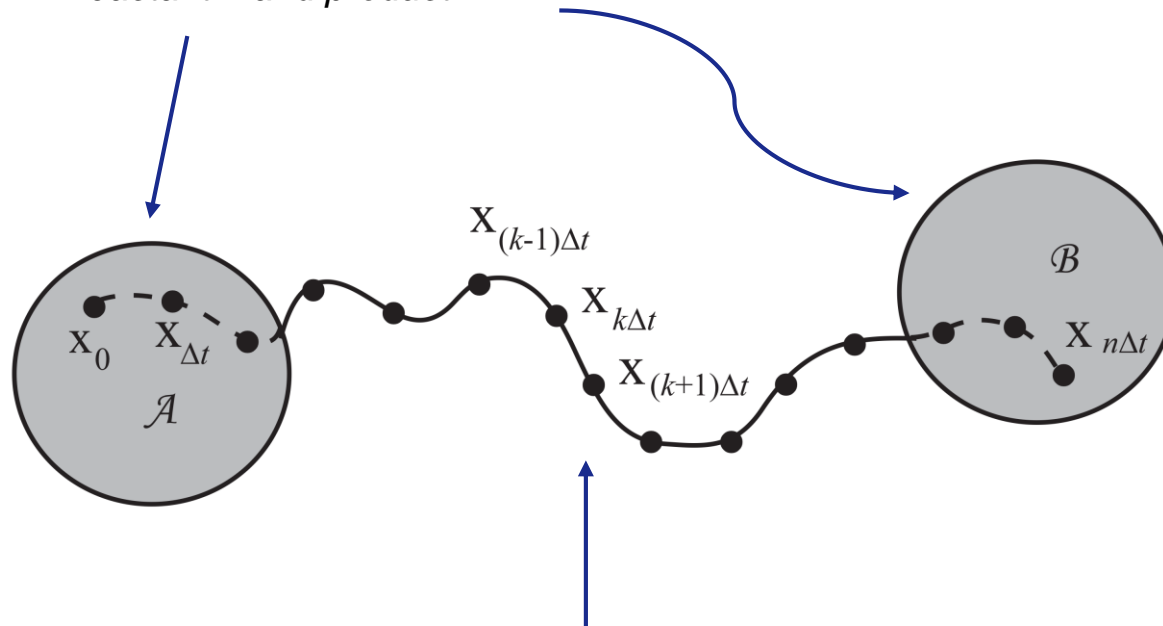
Menagerie: Other methods

Transition path formalism

- Transition path ensemble: trajectories that begin in A and end in B

- Goal: harvest the **transition path ensemble** of true dynamical bottlenecks without
1. a priori knowledge of TS location and 2. timescale issue
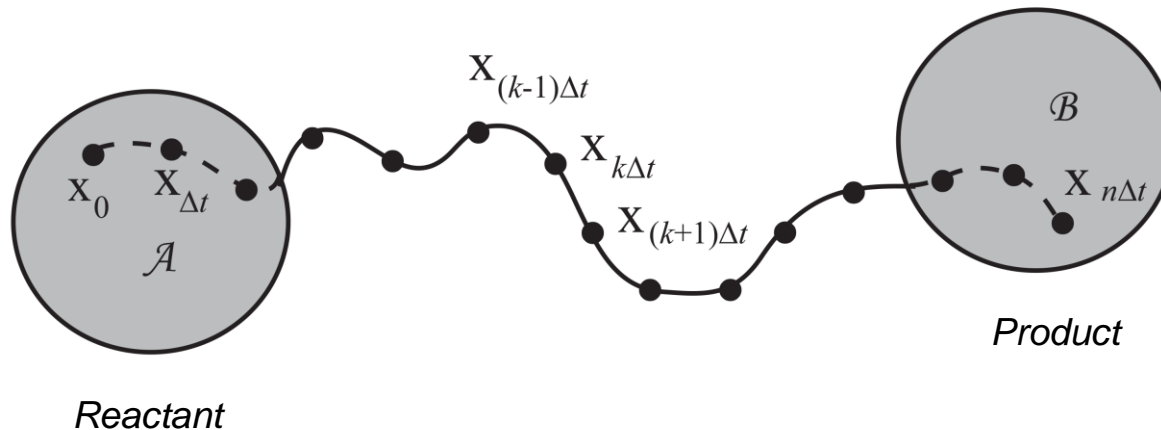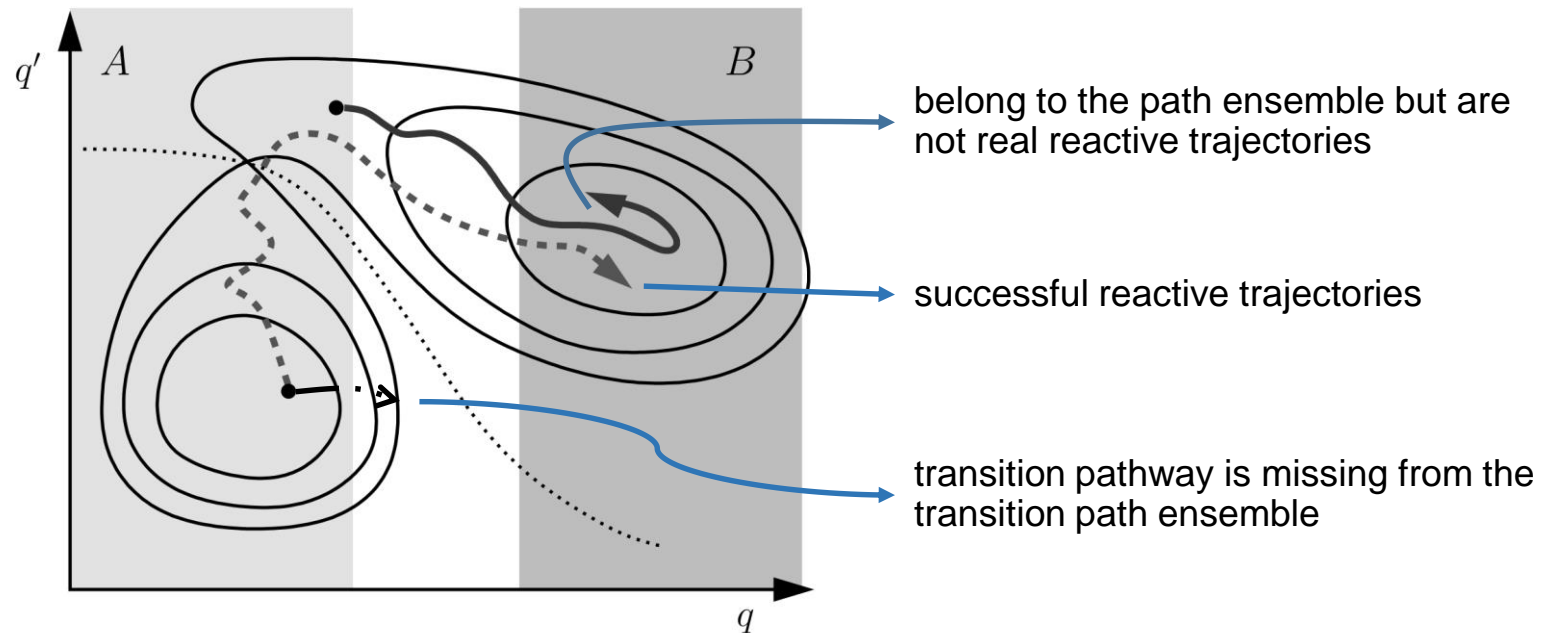
Questions about transition path

Q. How can we define the region of reactant A and product B?

$$\mathbf{X}_{(k-1)\Delta t}$$

$$\mathbf{X}_{k\Delta t}$$

$$\mathcal{B}$$

$$\mathbf{X}_0 \quad \mathbf{X}_{\Delta t}$$

$$\mathbf{X}_{(k+1)\Delta t}$$

$$\mathbf{X}_{n\Delta t}$$

$$\mathcal{A}$$

Q. How can we sample & generate the transition path?

Defining the stable states A and B

■ Regions A and B are characterized by the value of order parameter *q*

belong to the path ensemble but are not real reactive trajectories

successful reactive trajectories

transition pathway is missing from the transition path ensemble

■ A and B have to be large enough to accommodate eq fluctuations

■ A should not overlap with the basin of attraction of B and vice versa

C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.

*Importance Sampling: **Monte Carlo Dynamics***

■ Goal: collect reactive trajectories according to their weight in the transition path ensemble

Generate the initial transition
trajectory Y from A to B

⬇

Based on Y, generate the new
candidate reactive trajectory X

↙ ↘

X is not reactive   X is reactive

⬇   ⬇

Reject X   Calculate the probability P(Y->X) with
                                      detailed balance condition

                                                              ⬇

                                      Accept the new path if a random number
                                      [0,1] is smaller than P(Y->X), o.w. reject X

M(RT)^2 algorithm [N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller (1953)]

■ Goal: collect reactive trajectories according to their weight in the transition path ensemble

How can we generate the initial 'reactive' trajectory?

⬇

Based on Y, generate the new
candidate reactive trajectory X

↙   ↘

X is not reactive   X is reactive

⬇   ⬇

Reject X   Calculate the probability P(Y->X) with
   detailed balance condition

                                    ⬇

                          Accept the new path if a random number
   [0,1] is smaller than P(Y->X), o.w. reject X

M(RT)² algorithm [N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller (1953)]   15

Generating an initial path

■ Running a long MD simulation is ruled out by the rarity of the event!

$$X_{(k-1)\Delta t}$$
$$X_{k\Delta t}$$
$$X_{(k+1)\Delta t}$$
$$X_0 \quad X_{\Delta t} \quad \mathcal{A} \qquad \mathcal{B} \qquad X_{n\Delta t}$$
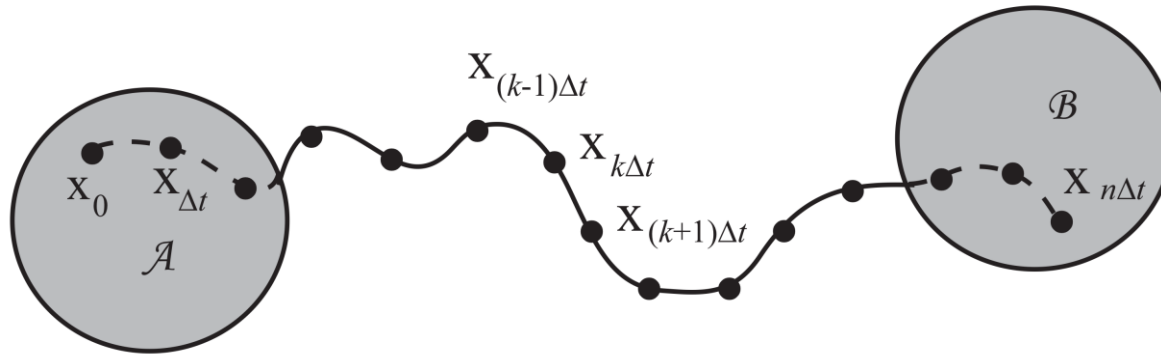
■ Start from an artificial pathway(not a true dynamical trajectory) and believe the eq power of MC?

■ High temperature pathways to accelerate a process

■ Starting a B and letting it evolve to A (i.e., it's much easier to induce the protein to unfold than fold)

Warning: Sampling can be trapped in local regions under poor initial path

■ Goal: collect reactive trajectories according to their weight in the transition path ensemble

Generate the initial transition
trajectory Y from A to B

⬇

How to generate new trajectory for sampling?

↙   ↘

X is not reactive   X is reactive

⬇   ⬇

Reject X   Calculate the probability P(Y->X) with
                                     detailed balance condition

                                     ⬇

                                     Accept the new path if a random number
                                     [0,1] is smaller than P(Y->X), o.w. reject X

M(RT)^2 algorithm [N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller (1953)]

Shooting and shifting

■ Generate the new possible reactive pathways by shooting on the initial reactive trajectory

Accepted $\mathbf{x}_{j\Delta t}$

$\mathbf{y}_{j\Delta t}$

$\mathcal{A}$

$\mathcal{B}$

$\mathbf{x}_{j\Delta t}$

Rejected
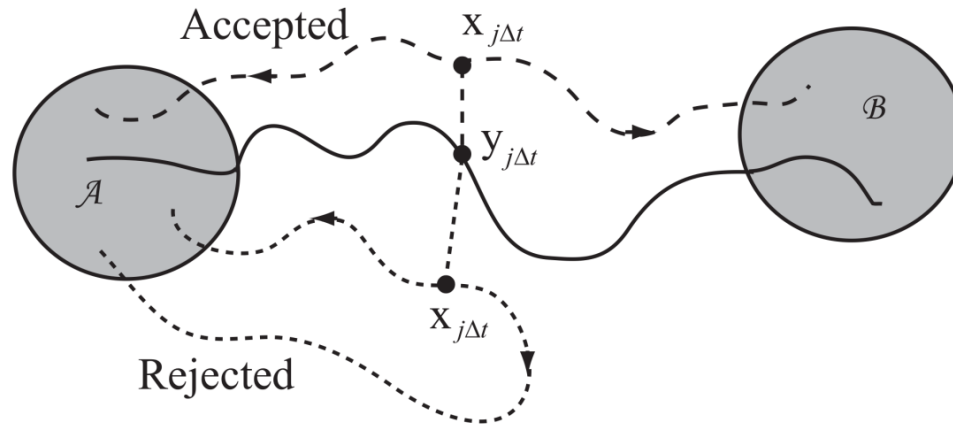
David Chandler

(1944-2017)

*"This is metaphorically akin to **throwing ropes** over rough mountain passes, in the **dark**. "Throwing ropes" in the sense that one shoots short trajectories, attempting to reach one stable state from another. "In the dark" because high-dimensional systems are so complex."*

Shooting and shifting
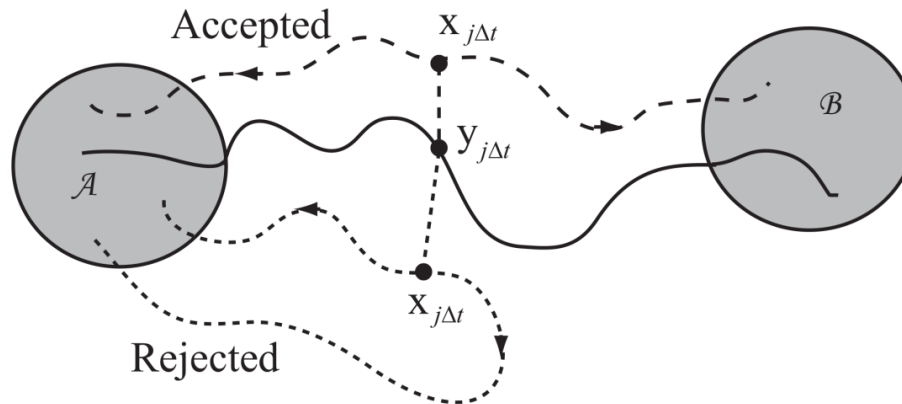
Shooting algorithm

1. Choose an index j randomly on the old reactive trajectory Y

2. Generate a random phase displacement (usually Δ=(0,dp)) to generate the new shooting point $x_{j\Delta t}$ from the old point $y_{j\Delta t}$

3. Integrate the eqn of motion backwards and forwards in time from shooting point to initial point and final point

4. If the initial/final point is not in the A and B region, reject the trial move

5. If the initial/final point is in A and B, accept the move with acceptance probability

■ In some cases, it is advantageous to change both the configuration and momentum. This **shifting** procedure effectively translates the path such as the reptation motion of a dense melt polymer.

Example of TPS simulation

Initial reactive trajectory
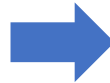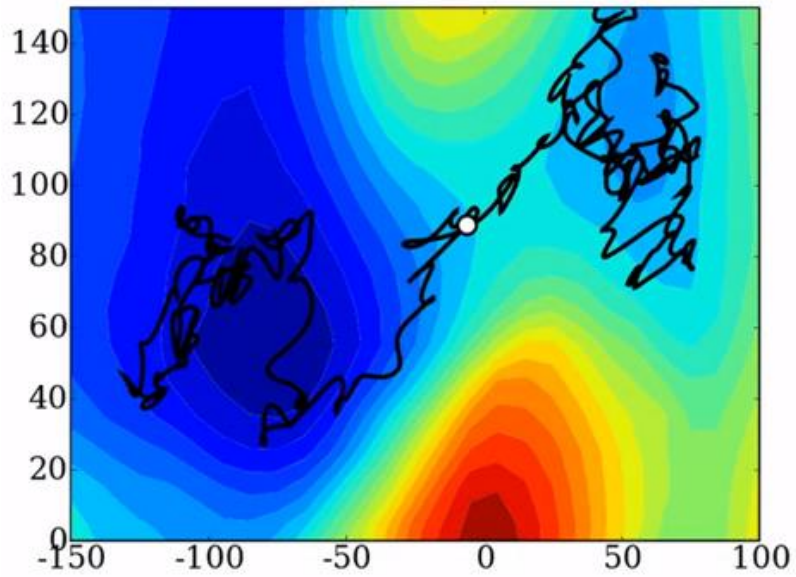
Shooting Forward
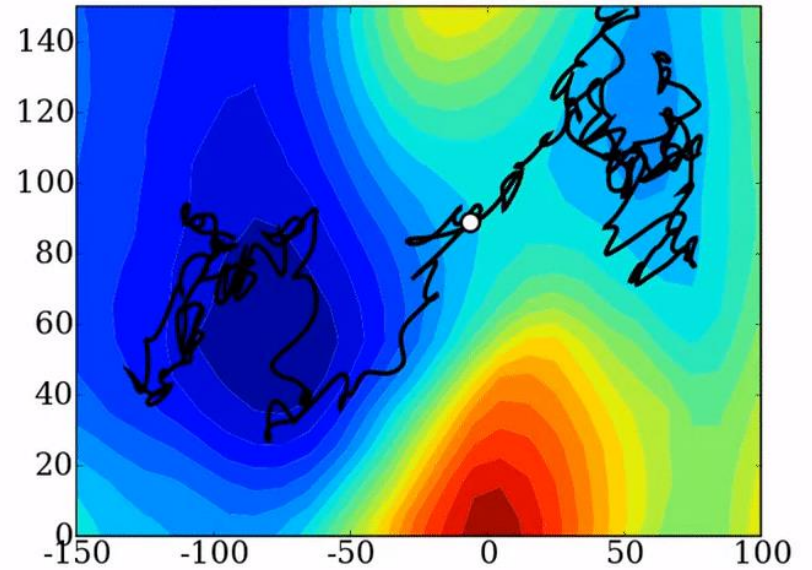
Type of shooting moves
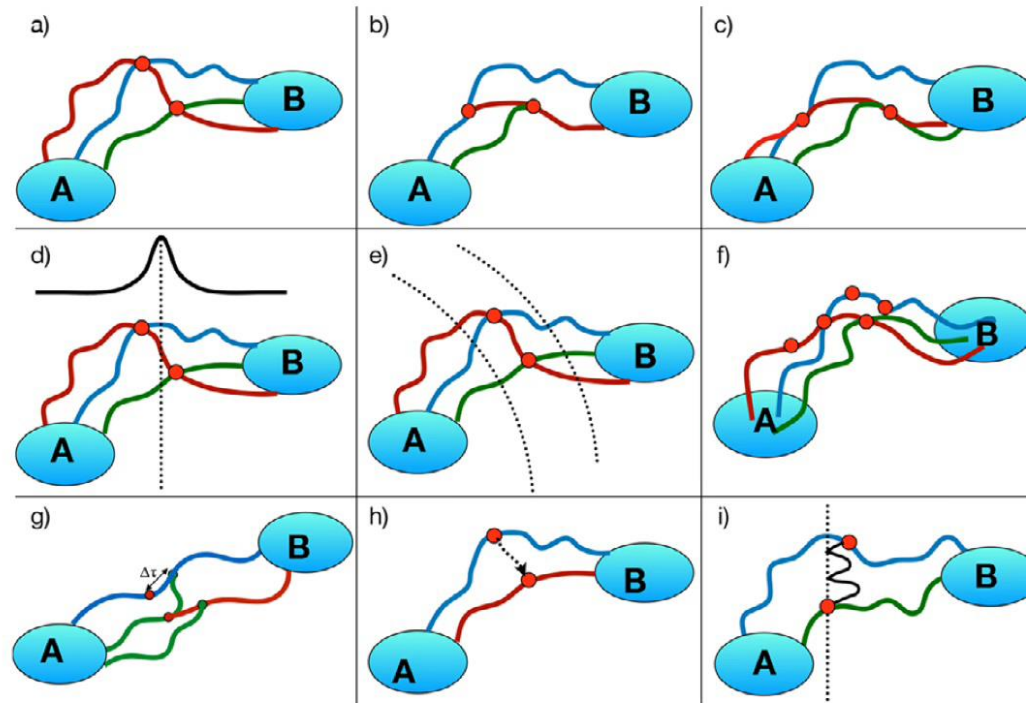
- Aim to increase acceptance by increasing similarity with previous trajectory

- Aim to increase acceptance by shooting from the barrier or making it more likely to generate an acceptable trajectory

- Aim to enhance exploration in path space and achieve better decorrelation

OpenPathSampling: A Python library for path sampling algorithms

OpenPathSampling

1.5.3.dev0

Search docs

- Installation
- Examples
- User Guide Topics
- Videos
- Command Line Interface
- The OPS Ecosystem

🏠 » OpenPathSampling

View page source

OpenPathSampling

A Python library to facilitate path sampling algorithms.

OpenPathSampling (OPS) makes it easy to perform many variants of transition path sampling (TPS) and transition interface sampling (TIS), as well as other useful calculations for rare events, such as committor analysis and flux calculations. In addition, it is a powerful library to build new path sampling methods.

OPS is independent of the underlying molecular dynamics engine, and currently has support for OpenMM and Gromacs, as well as an internal engine suitable for 2D toy models.

To learn more about what OPS can do, look at our examples. If you want to jump right in, take a look at how easy it is to install!

Installation

```
conda install -c conda-forge openpathsampling-cli
# or
pip install openpathsampling-cli
```

Import

```
import openpathsampling as paths

import openpathsampling.engines.openmm as omm
from simtk.openmm import app
import simtk.openmm as mm
import simtk.unit as unit
```

http://openpathsampling.org/latest/cli.html

22

Reaction rate constants

■ Method for computing unimolecular reaction rate constant

$$A \rightleftharpoons B$$
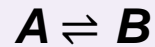
$$C(t) \equiv \frac{\langle h_A(x_0)h_B(x_t)\rangle}{\langle h_A\rangle} = (\langle h_A\rangle + \langle h_B\rangle)k_{AB}t \approx k_{AB}t \ .$$

h_A, h_B = characteristic function of state A and B

Microscopic expressions for the reaction rate constant k_{AB}

Time correlation function C(t) could be calculated using trajectories

Intuitively, it expresses the conditional probability P(B,t | A,0)

C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.

Time correlation function and transition path sampling

■ **C(t)** could be calculated in the transition path sampling framework.

$$C(t) = \frac{\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)]h_B(x_t)}{\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)]} = \frac{Z_{AB}(t)}{Z_A} = F_{AB}(t) - F_A$$

$$\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)] \quad \Longrightarrow$$

Partition function for the set of pathways that start in A and end elsewhere

$$\int \mathcal{D}x(t)\, h_A(x_0)\mathcal{P}[x(t)]h_B(x_t) \quad \Longrightarrow$$

Partition function for the set of pathways that start in A and end in B

Computing rate constant is now reduced to computing free energy.

C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.

Free energy as a reversible work

■ Free energy difference ΔF can also viewed as the reversible work **$W_{AB}(t)$**

$$W_{AB}(t) \equiv -\ln \frac{Z_{AB}(t)}{Z_A} \qquad C(t) = \exp[-W_{AB}(t)]$$
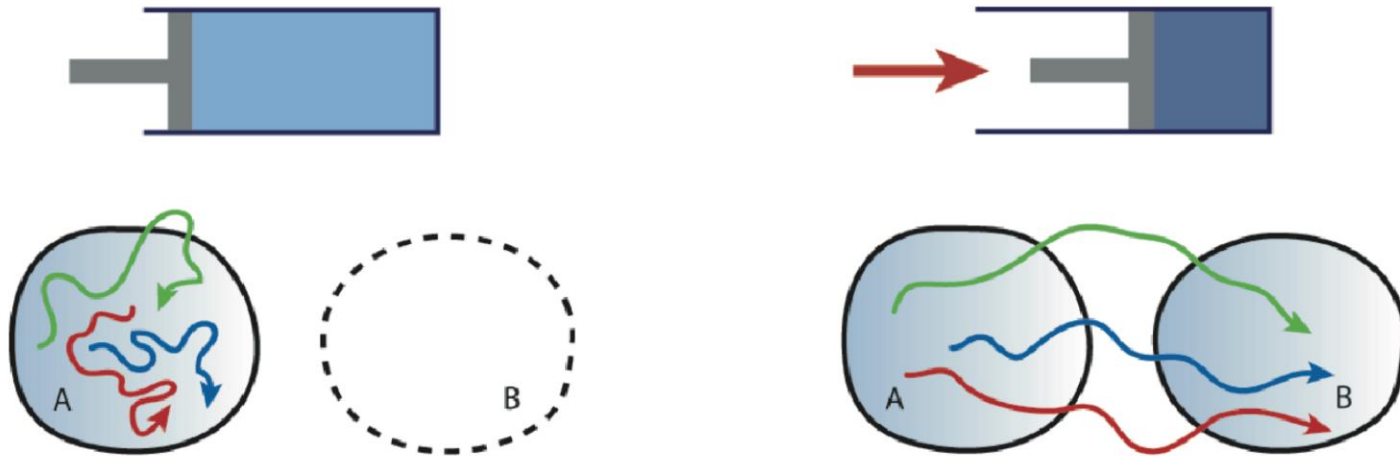
Work we need to "compress" the path ensemble and confine the endpoints to B

C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.

Committor for reaction mechanism study

Committor: for a given configuration r, the fraction of trajectories started in A that reach B after time t

$$p_B(r,t) \equiv \frac{\int \mathcal{D}x(t)\ \mathcal{P}[x(t)]\delta(r_0 - r)h_B(x_t)}{\int \mathcal{D}x(t)\ P[x(t)]\delta(r_0 - r)}$$

Implications

■ Direct indicator of the reaction progress, gives a criterion for ideal reaction coordinate

■ Statistical measure for how *commited* a given configuration is to the product state

$$p_B(r,t) \approx \frac{1}{N}\sum_{i=1}^{N} h_B(x_t^{(i)}) \equiv p_B^{(N)}(r,t)$$

■ Used to define the configuration r of the transition state (How?)
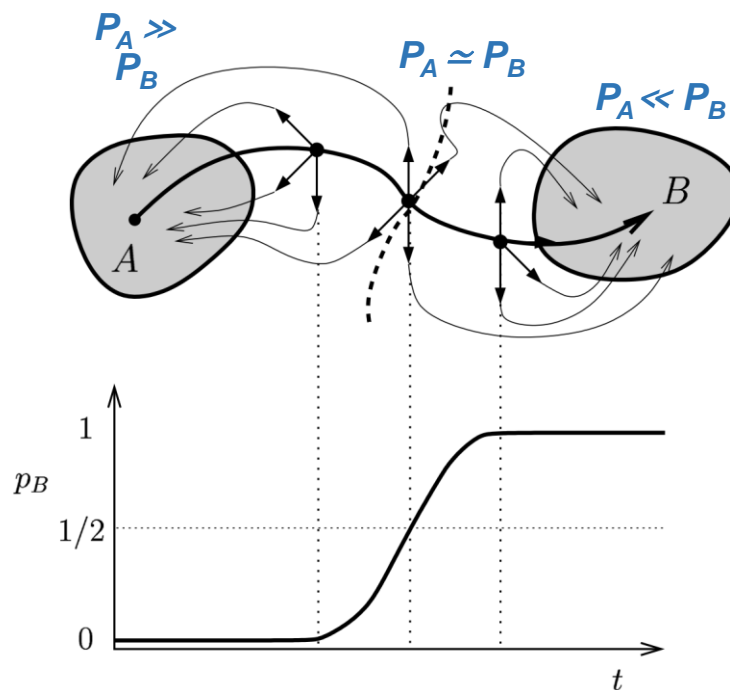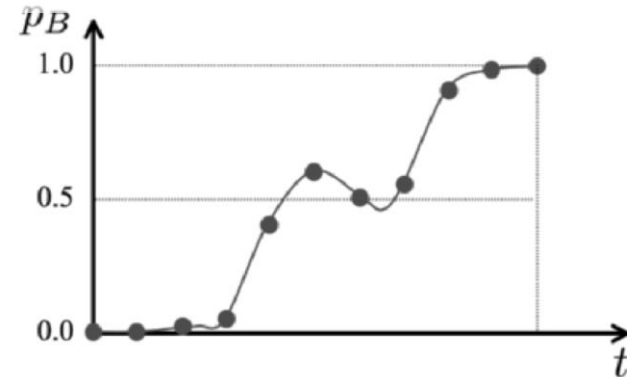
C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.
Annu. Rev. Phys. Chem., **2002**, 53, 291-318.

Transition State Ensemble

■ r is a transition state(TS) if $p_A(r) = p_B(r) = 0.5$

■ TS defined in this statistical way do not coincide with the saddle point of the PES, and **entropic contributions** play an important role in determining TS

$P_B = 0.5$ ➡ *Transition state ensemble*
(isocommittor ½ surface)
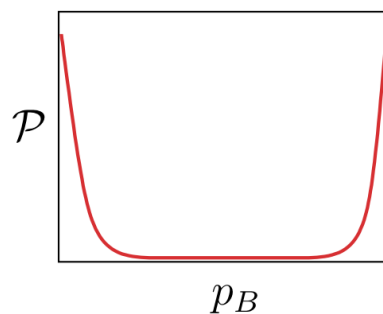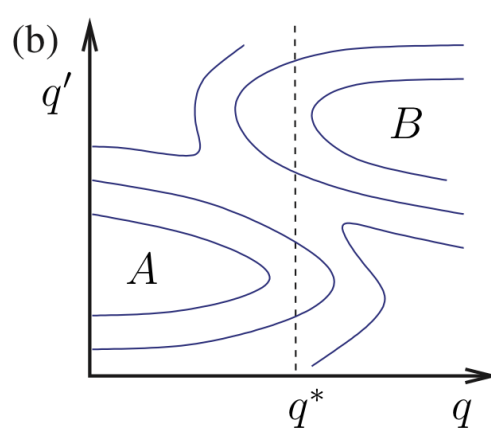
C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.
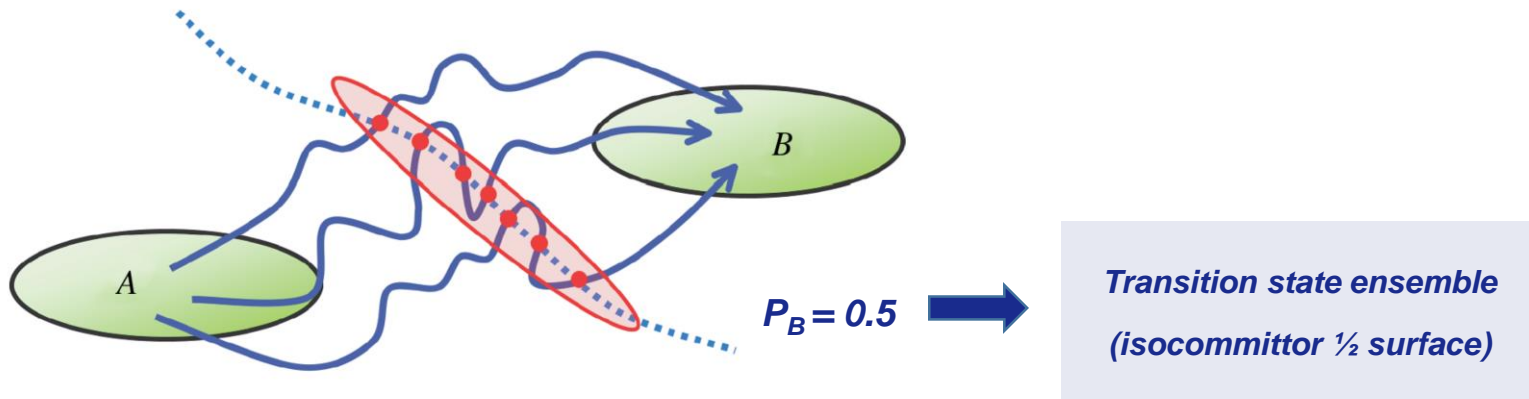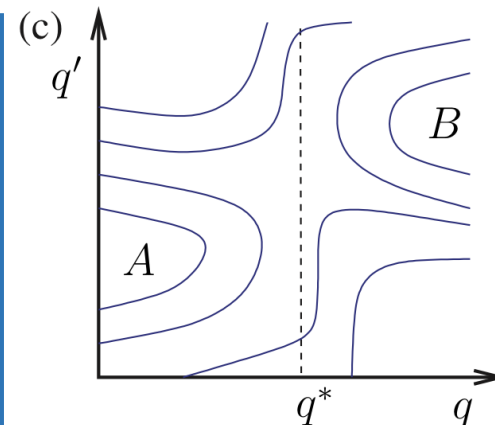
Three scenarios leading to different committor distributions

- Committor distribution is used to choose the desirable reaction coordinate for transition
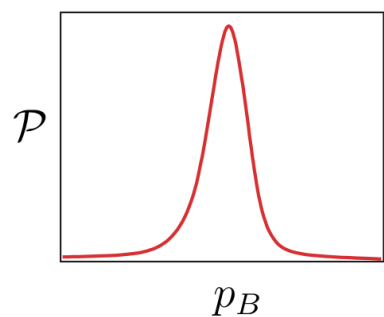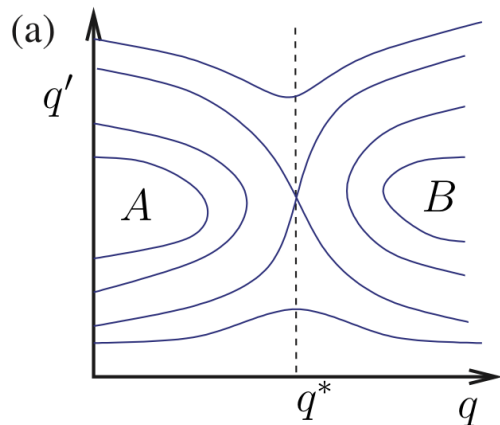
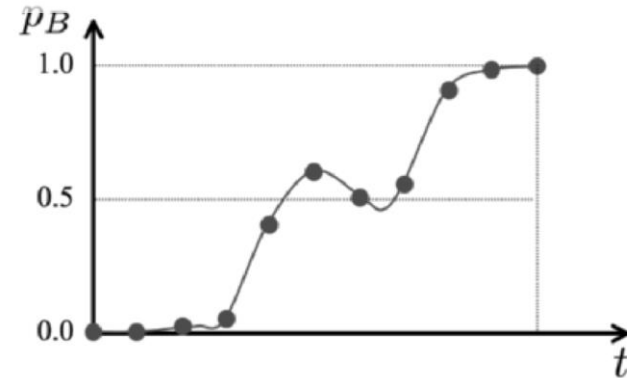(a) *q is a good reaction coordinate*
Transition state at $P_B = 0.5$

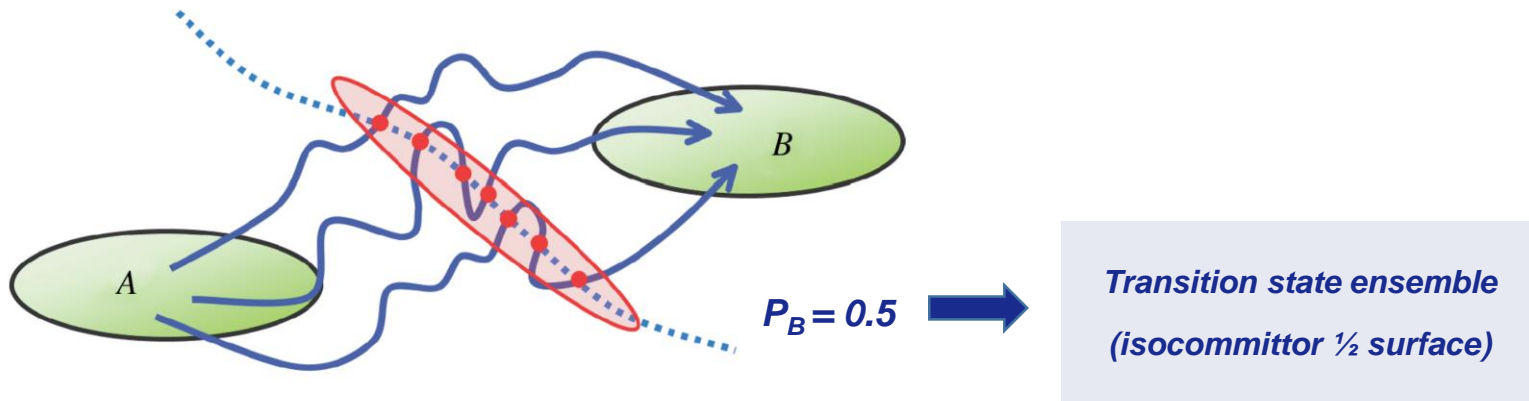(b) *q is insufficient to describe rxn*
Variable q' must be considered
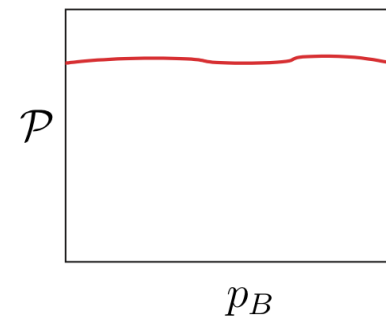
(c) *Transition occurs diffusively in q'*
Committor distribution is flat

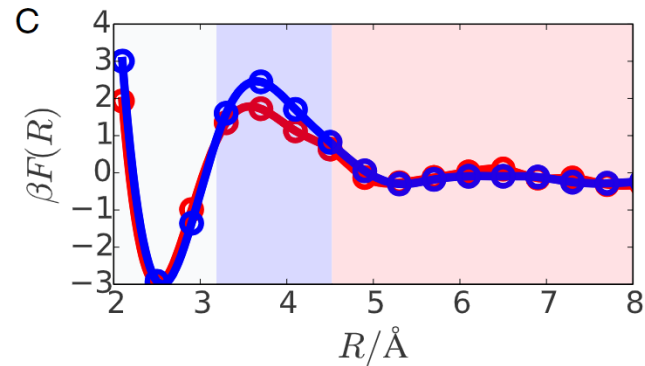C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.
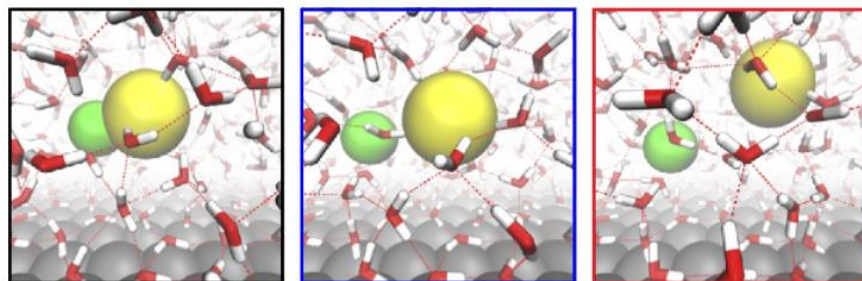
Transition State Ensemble

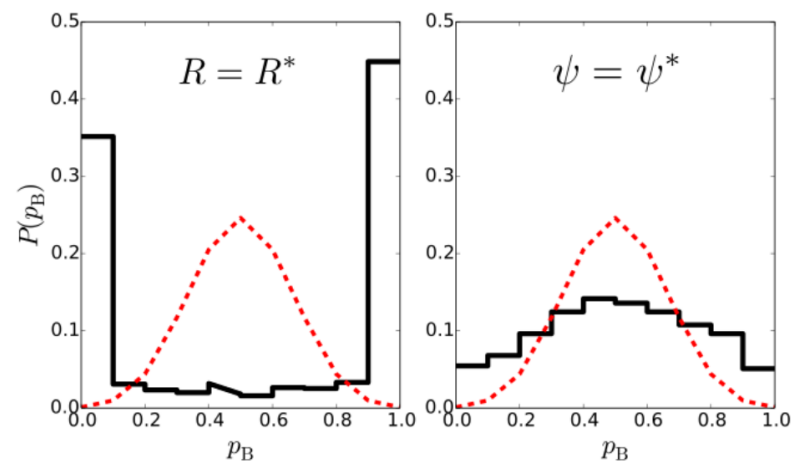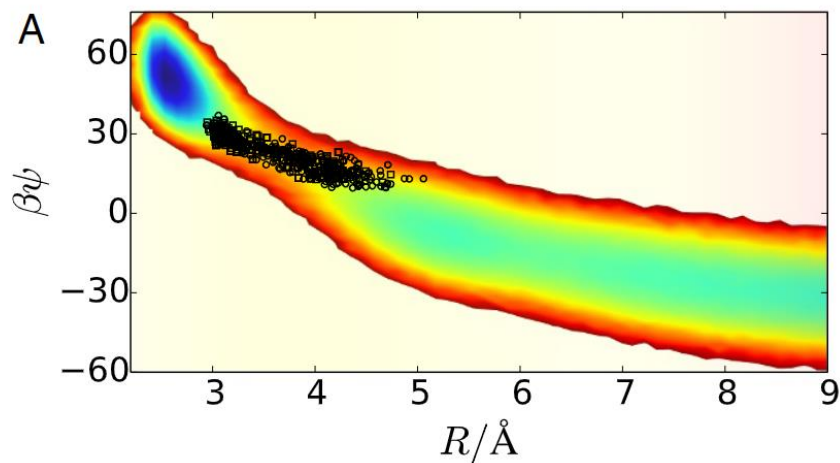■ r is a transition state(TS) if $p_A(r) = p_B(r) = 0.5$

■ TS defined in this statistical way do not coincide with the saddle point of the PES, and **entropic contributions** play an important role in determining TS

$P_B = 0.5$ ➡ *Transition state ensemble*
(isocommittor ½ surface)

C. Dellago et al. *Lect. Notes Phys.* **2006**, 703, 349-391.

Example of committor distribution

$$\psi_{\pm} = \frac{q_{\pm}}{4\pi\epsilon_0} \sum_{i}^{N_{\text{H}_2\text{O}}} \frac{q_i}{|\mathbf{r}_i - \mathbf{r}_{\pm}|}$$

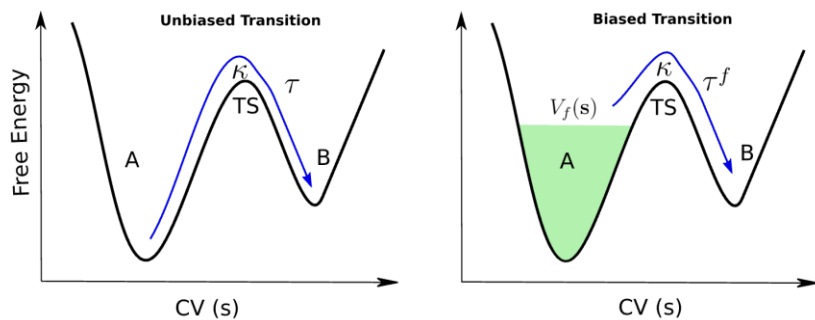# *Content*

Bottlenecks in Kinetics Computation
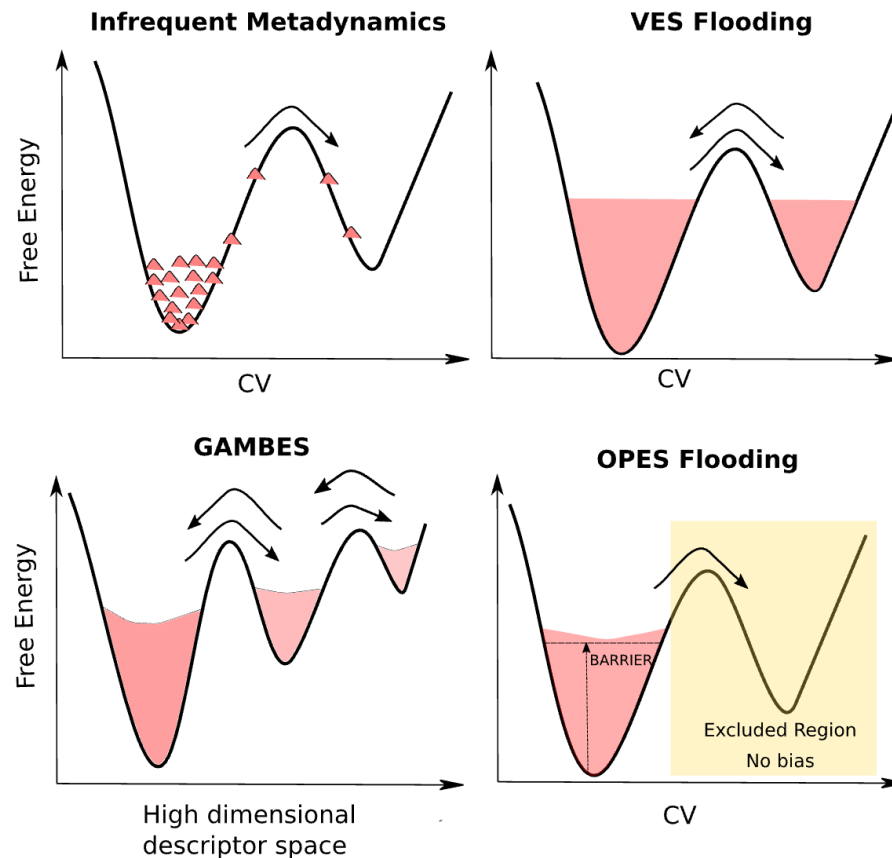
Transition Path Sampling
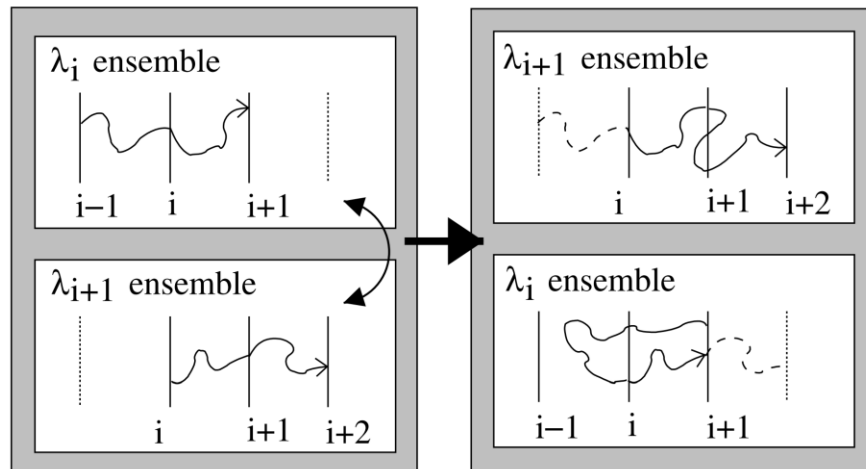
Menagerie: Other methods

Kinetics from Metadynamics

- **No bias deposition on the TS region** is the most important idea

Unbiased Transition — with κ, τ, TS, A, B; CV (s)

Biased Transition — with $V_f(\mathbf{s})$, κ, τ^f, TS, A, B; CV (s)

$$\frac{\tau}{\tau^f} = \langle \exp(\beta V_f(\mathbf{s})) \rangle_{U+V_f}|_{\lambda < \lambda^*}$$

Infrequent Metadynamics — Free Energy vs CV

VES Flooding — CV

GAMBES — Free Energy vs High dimensional descriptor space

OPES Flooding — CV; BARRIER; Excluded Region No bias

Transition Interface Sampling (TIS)
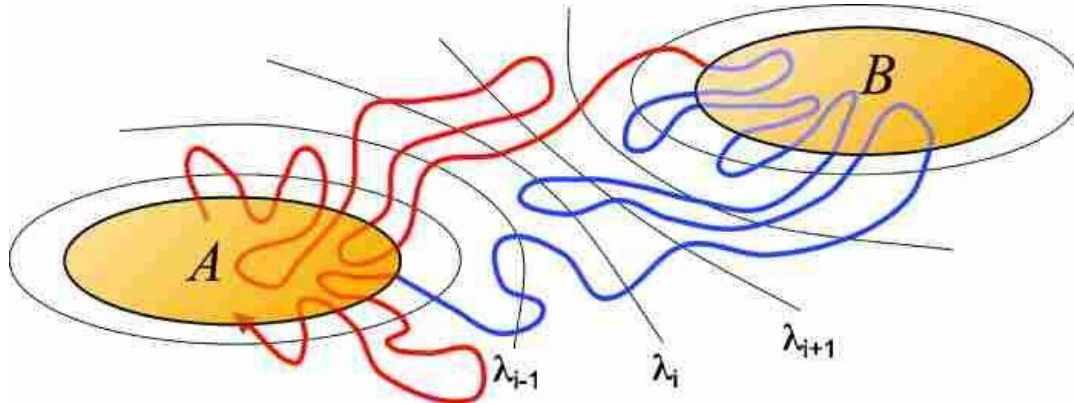
Takeaways

■ Simulating rare event requires a more sophisticated algorithms that could resolve the energy/entropy bottleneck issues.

■ More importantly, algorithms that do not resort to the choice of collective variables are highly warranted for reliable kinetic computation.

■ One way is to harvest the transition path ensemble, thus computing rate constant and postulate the proper reaction coordinate via committor distribution.

Q&A