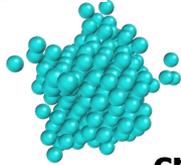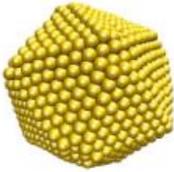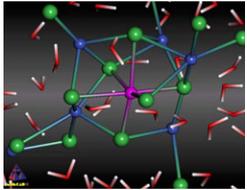# *Rare Event Simulations*

Transition state theory 15.1-15.2

Bennett-Chandler Approach 15.2
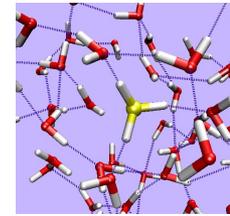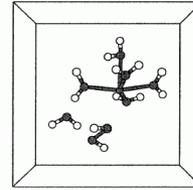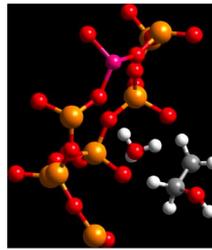
Transition path sampling15.4

Outline

- Part 1
  - Rare events and reaction kinetics
  - Microscopic expression from Linear Response
  - Transition state theory
  - Bennet Chandler approach

- Part 2
  - Two ended methods
  - Transition path sampling
  - Rate constants with path sampling
  - Reaction coordinate analysis
  - Application to crystallization
  - Open Path Sampling

crystallisation

catalysis   solution reactions
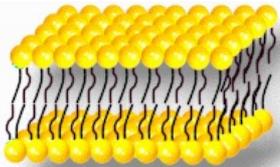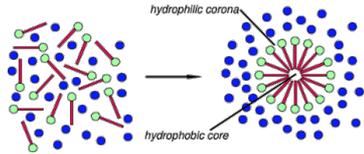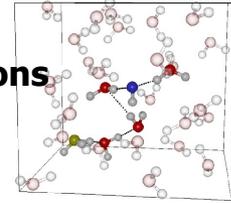
enzyme reactions

complex fluids

RARE
EVENTS

folding & binding

solvent effects

isomerization

DNA-Polymerase-beta

fingers   palm

thumb

8-kD

hydrophilic corona

hydrophobic core
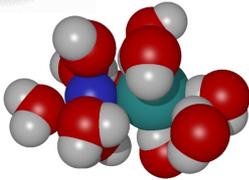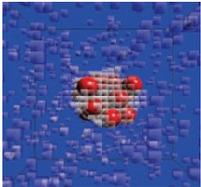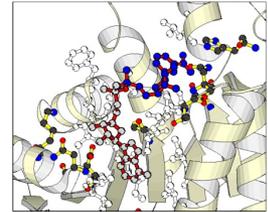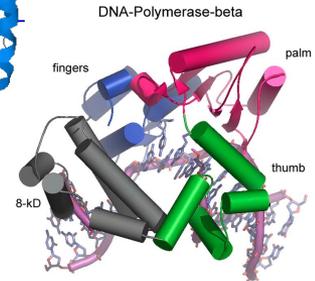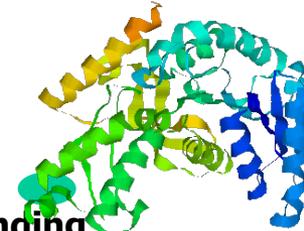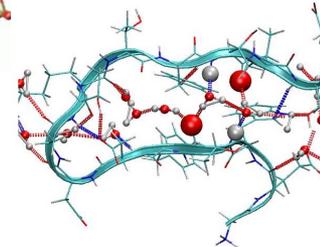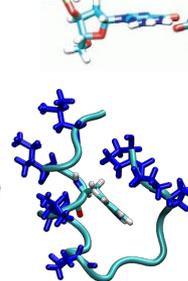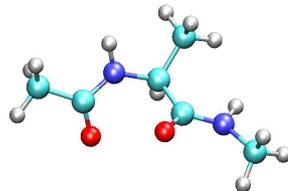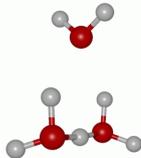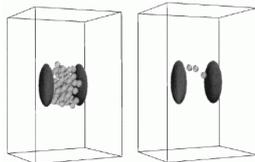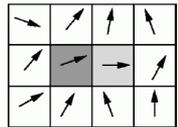
Rare events

Interesting transitions in complex systems

  – solution chemistry

  – protein folding

  – enzymatic reactions

  – complex surface reactions

  – diffusion in porous media

  – nucleation

These reactions happen on a long time scale compared to the molecular timescale
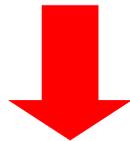
dominated by collective, rare events
Straightforward MD very inefficient
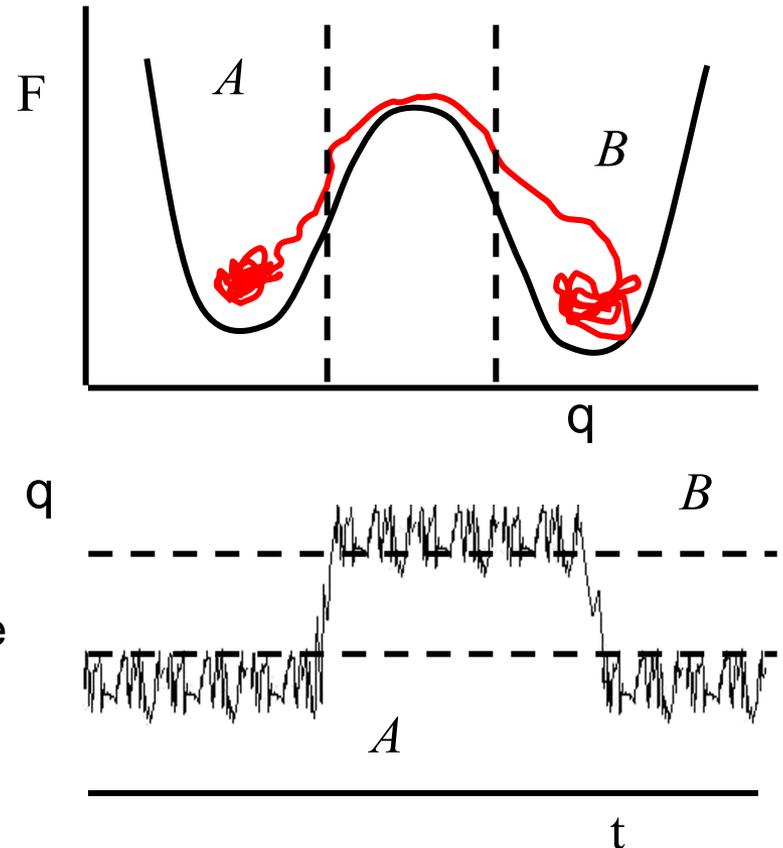
$$\tau_{mol} \ll \tau_{stable}$$

Example: Diffusion in porous material
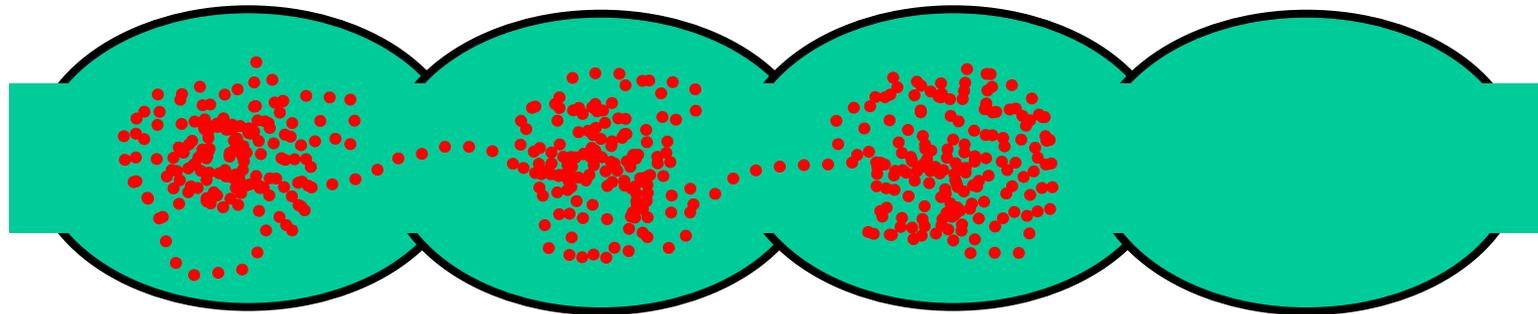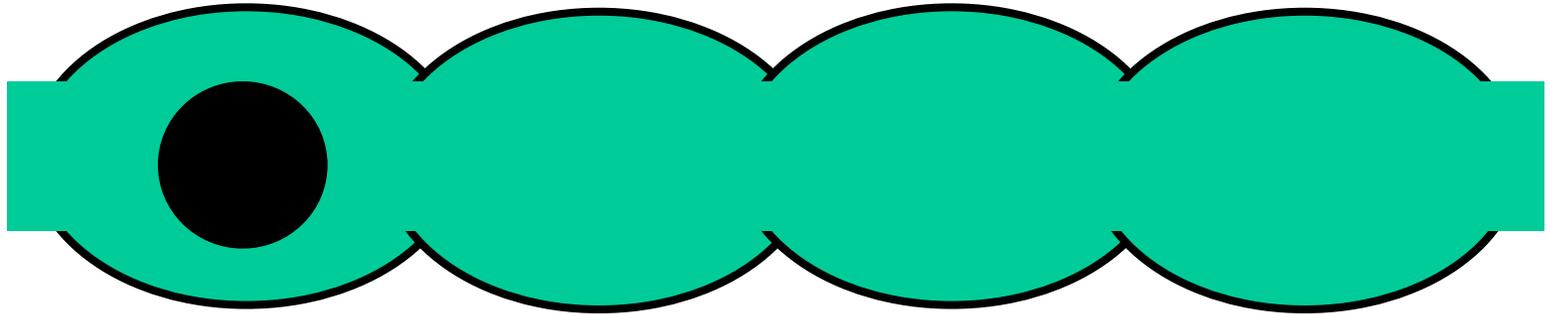
Phenomenological reaction kinetics

A rare event can be seen as a chemical reaction between reactant A and product B

$$A \leftrightarrow B$$
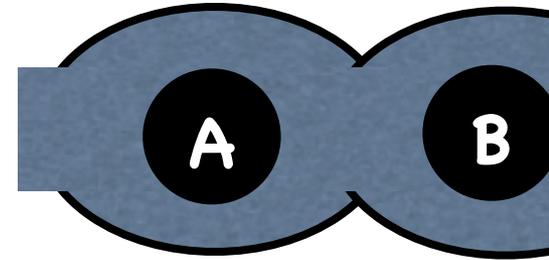
The change in population c(t) is (0<c<1)

$$\frac{\mathrm{d}c_A(t)}{\mathrm{d}t} = -k_{A \to B} c_A(t) + k_{B \to A} c_B(t) \qquad \frac{\mathrm{d}c_B(t)}{\mathrm{d}t} = +k_{A \to B} c_A(t) - k_{B \to A} c_B(t)$$

Total number change in population

$$\frac{\mathrm{d}\left[c_A(t) + c_B(t) \right]}{\mathrm{d}t} = 0$$

Equilibrium:

$$\dot{c}_A(t) = \dot{c}_B(t) = 0$$

This gives a relation between equilibrium population and reaction rates

$$\frac{\langle c_A \rangle}{\langle c_B \rangle} = \frac{k_{B \to A}}{k_{A \to B}}$$

Relaxation time

Let us make a perturbation of the equilibrium populations, e.g by applying an external field.

When releasing the field, the system will relax to the original equilibrium

For state A $c_A(t) = \langle c_A \rangle + \Delta c_A(t)$ For state B: $c_B(t) = \langle c_B \rangle - \Delta c_A(t)$

We can rewrite the kinetics in terms of the perturbation Δc:

$$\frac{d\Delta c_A(t)}{dt} = -k_{A \to B} \Delta c_A(t) - k_{B \to A} \Delta c_A(t)$$

$$\Delta c_A(t) = \Delta c_A(0) \exp\left[-\left(k_{A \to B} + k_{B \to A}\right)t\right]$$

$$= \Delta c_A(0) \exp\left[-t/\tau\right]$$

With relaxation time

$$\tau = \left(k_{A \to B} + k_{B \to A}\right)^{-1} = k_{A \to B}^{-1}\left(1 + \langle c_A \rangle / \langle c_B \rangle\right)^{-1} = \frac{\langle c_B \rangle}{k_{A \to B}}$$

$$c_A(t) + c_B(t) = 1$$

1

population

C$_A$(t)

C$_B$(t)

time
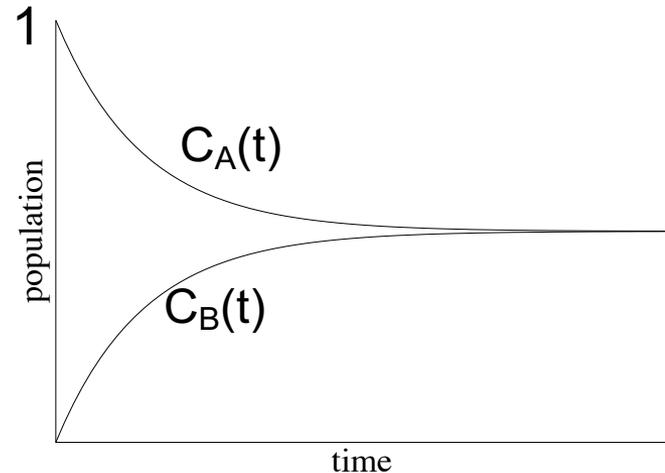
Microscopic theory

Microscopic description of the progress of a reaction

q
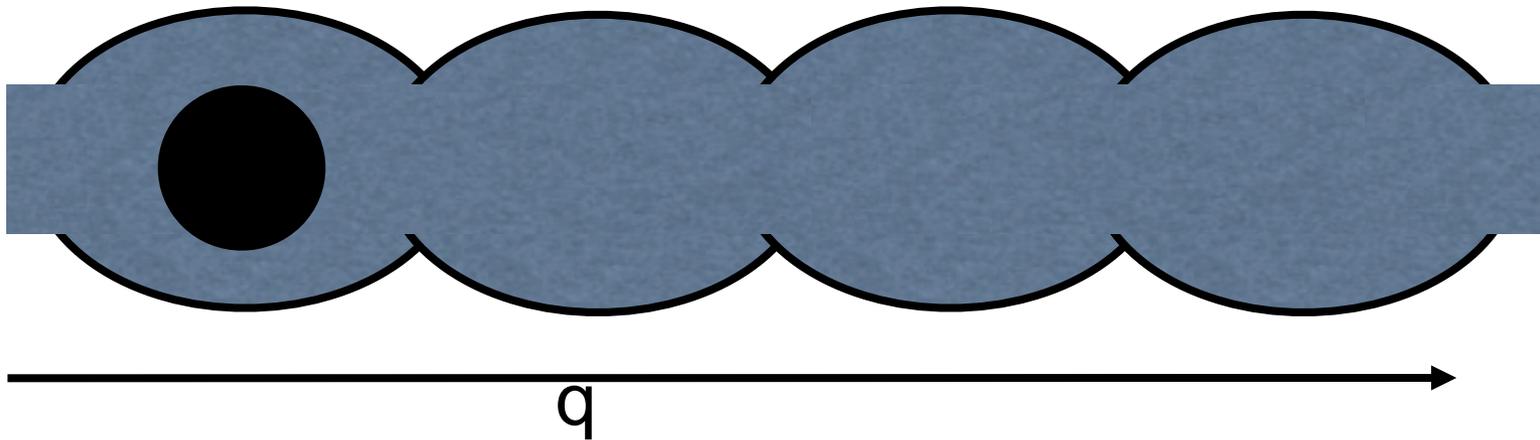
Reaction coordinate: in this case the z-coordinate of the particle

We need to write the kinetics of the reaction in terms of this microscopic reaction coordinate q

Reaction coordinate

Reactant A: $\quad q < q*$

Product B: $\quad q > q*$

Transition state: $\quad q = q*$

Heaviside θ-function

$$\theta(q - q*) = \begin{cases} 0 & q - q* < 0 \\ 1 & q - q* > 0 \end{cases}$$
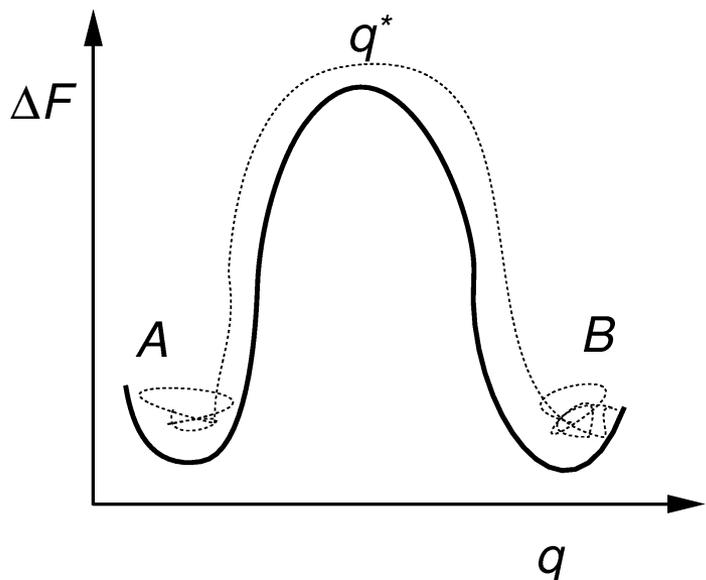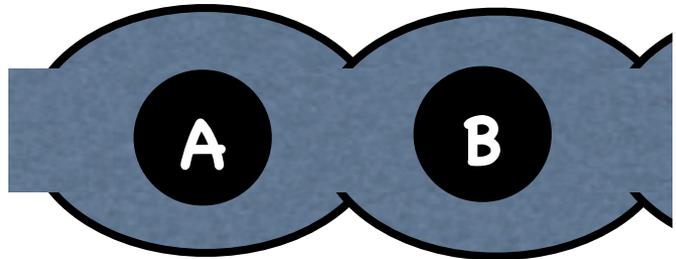
Let us introduce the function g_A:

$$g_A(q - q*) = 1 - \theta(q - q*) = \theta(q* - q)$$

With this function we write for the **probability** $c_A(t)$ the system is in state A:

$$\left\langle c_A(t) \right\rangle = \left\langle g_A(t) \right\rangle$$
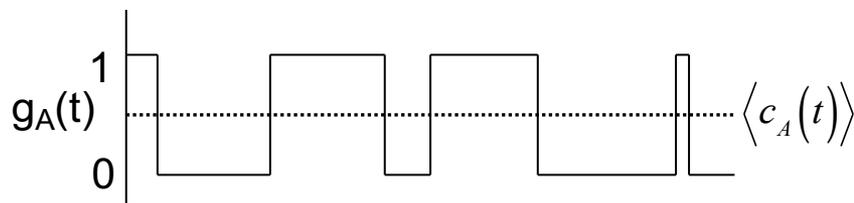
Microscopic theory

Is going to give us the macroscopic relaxation in terms of a microscopic time correlation function

$$\exp\left[-t/\tau\right] = \frac{\left\langle \Delta g_A(0)\,\Delta g_A(t)\right\rangle}{\left\langle c_A\right\rangle\left\langle c_B\right\rangle}$$
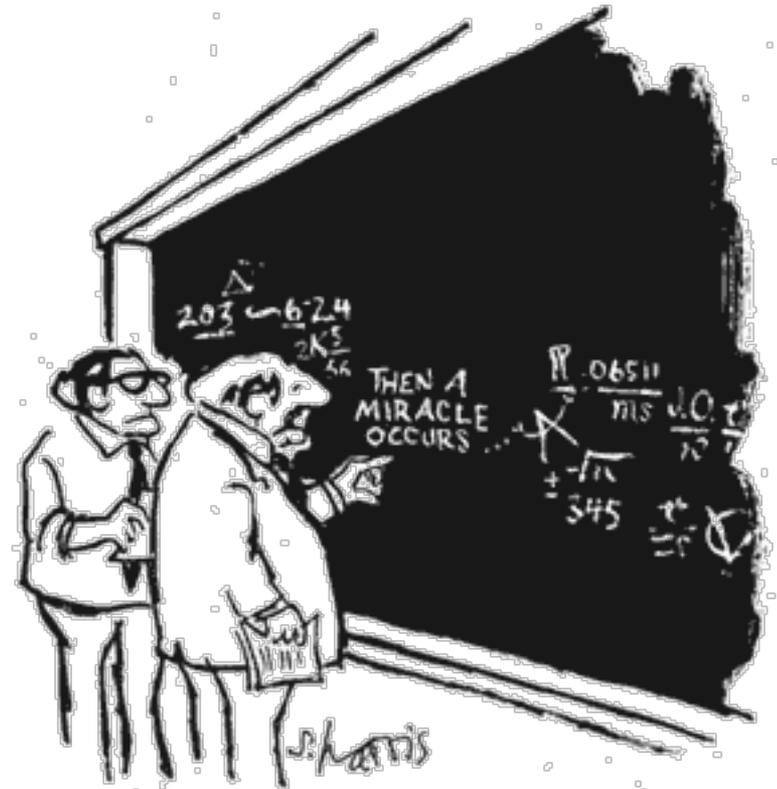
"I THINK YOU SHOULD BE MORE EXPLICIT HERE IN STEP TWO."

This needs linear response theory

Perturbed Hamiltonian

Let us consider the effect of a static perturbation:

$$H = H_0 - \varepsilon g_A (q - q*)$$

This external potential increases the concentration of A

For the equilibrium concentration as a function of ε:

$$\Delta c_A = \langle c_A \rangle_\varepsilon - \langle c_A \rangle_0 \quad = \langle g_A \rangle_\varepsilon - \langle g_A \rangle_0$$

We need to compute the ensemble average in the form of :

$$\langle A \rangle_0 = \frac{\int d\Gamma A \exp[-\beta H_0]}{\int d\Gamma \exp[-\beta H_0]}$$

Linear Response theory (static)

The original Hamiltonian (H_0) is perturbed by εD:

$$H = H_0 - \varepsilon D$$

This gives as change in the expectation value of A:

$$\langle \Delta A \rangle = \langle A \rangle - \langle A \rangle_0$$

with

$$\langle A \rangle = \frac{\int d\Gamma A \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right]}{\int d\Gamma \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right]} \qquad \langle A \rangle_0 = \frac{\int d\Gamma A \exp\left[-\beta H_0\right]}{\int d\Gamma \exp\left[-\beta H_0\right]}$$

If the perturbation is small we can write $\quad \langle A \rangle = \langle A \rangle_0 + \dfrac{\partial \langle A \rangle_0}{\partial \varepsilon} \varepsilon$

For such a small perturbation

$$\langle \Delta A \rangle = \frac{\partial \langle A \rangle_0}{\partial \varepsilon} \varepsilon = \left\langle \frac{\partial A}{\partial \varepsilon} \right\rangle_0 \varepsilon$$

with

$$\left\langle \frac{\partial A}{\partial \varepsilon} \right\rangle = \frac{\partial \langle A \rangle}{\partial \varepsilon} = \frac{\int d\Gamma \beta A D \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right] \int d\Gamma \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right]}{\left\{ \int d\Gamma \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right] \right\}^2}$$

$$- \frac{\int d\Gamma A \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right] \int d\Gamma \beta D \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right]}{\left\{ \int d\Gamma \exp\left[-\beta\left(H_0 - \varepsilon D\right)\right] \right\}^2}$$

Evaluated for $\varepsilon = 0$

$$\left\langle \frac{\partial A}{\partial \varepsilon} \right\rangle_0 = \frac{\int d\Gamma \beta A D \exp\left[-\beta H_0\right]}{\left\{ \int d\Gamma \exp\left[-\beta H_0\right] \right\}} - \frac{\int d\Gamma A \exp\left[-\beta H_0\right]}{\int d\Gamma \exp\left[-\beta H_0\right]} \times \frac{\int d\Gamma \beta D \exp\left[-\beta H_0\right]}{\int d\Gamma \exp\left[-\beta H_0\right]}$$

Giving:

$$\left\langle \frac{\partial A}{\partial \varepsilon} \right\rangle_0 = \beta \left\{ \langle AD \rangle_0 - \langle A \rangle_0 \langle D \rangle_0 \right\}$$

If we apply this result for c_A: $\quad \left\langle \dfrac{\partial A}{\partial \varepsilon} \right\rangle_0 = \beta \left\{ \langle AD \rangle_0 - \langle A \rangle_0 \langle D \rangle_0 \right\}$

with $\quad H = H_0 - \varepsilon g_A (q - q*)$

$$\Delta c_A = \langle g_A \rangle_\varepsilon - \langle g_A \rangle_0$$

$$\frac{\partial \Delta c_A}{\partial \varepsilon} = \beta \left(\left\langle (g_A)^2 \right\rangle_0 - \langle g_A \rangle_0^2 \right)$$

Since $g_A = 0$ or *1*:
g_A *(x)* g_A *(x)* = g_A *(x)*

$$= \beta \left(\langle g_A \rangle_0 \left(1 - \langle g_A \rangle_0 \right) \right)$$

$$= \beta \left(\langle c_A \rangle_0 \left(1 - \langle c_A \rangle_0 \right) \right) = \beta \langle c_A \rangle_0 \langle c_B \rangle_0$$

Giving: $\quad \boxed{\Delta c_A = \beta \langle c_A \rangle_0 \langle c_B \rangle_0 \, \varepsilon}$

Linear Response theory (dynamic)

Let us now switch off the perturbation at t=0

$$H = H_0 - \varepsilon D \qquad\qquad H = H_0 \quad \text{at t>0:}$$

Let us see how the system relaxes to equilibrium (dynamical perturbation)

We take <A>$_0$=0

$$\left\langle \Delta A(t) \right\rangle = \left\langle A(t) \right\rangle - \left\langle A \right\rangle_0 = \left\langle A(t) \right\rangle$$

Similar as for the static case for small values of ε, we have

$$\left\langle \frac{\partial A(t)}{\partial \varepsilon} \right\rangle_0 = \frac{\int d\Gamma \beta A(t) D \exp\left[-\beta H_0 \right]}{\left\{ \int d\Gamma \exp\left[-\beta H_0 \right] \right\}} = \beta \left\langle D(0) A(t) \right\rangle$$

Giving:
$$\left\langle \Delta A(t) \right\rangle = \beta \varepsilon \left\langle D(0) A(t) \right\rangle$$

$$\langle \Delta A(t) \rangle = \beta \varepsilon \langle D(0) A(t) \rangle$$

If we apply this result to $\quad D = \Delta g_A \quad$ and $\quad A = \Delta g_A$

We obtain: $\langle \Delta c_A(t) \rangle = \beta \varepsilon \langle \Delta g_A(0) \Delta g_A(t) \rangle$

From static perturbation: $\qquad \beta \varepsilon = \dfrac{\Delta c_A(0)}{\langle c_A \rangle \langle c_B \rangle}$

$$\langle \Delta c_A(t) \rangle = \Delta c_A(0) \frac{\langle \Delta g_A(0) \Delta g_A(t) \rangle}{\langle c_A \rangle \langle c_B \rangle}$$

Compare linear response expression with the macroscopic expression

$$\Delta c_A(t) = \Delta c_A(0) \exp\left[-t/\tau\right]$$

Make use of linear response theory

$$\left\langle \Delta c_A(t) \right\rangle = \Delta c_A(0) \frac{\left\langle \Delta g_A(0) \Delta g_A(t) \right\rangle}{\left\langle c_A \right\rangle \left\langle c_B \right\rangle}$$

Linear Response theory: see book Appendix C2

$$\Delta g_A(t) = g_A(t) - \left\langle g_A \right\rangle$$

Compare linear response expression with the macroscopic expression

$$\Delta c_A(t) = \Delta c_A(0) \exp\left[-t/\tau\right]$$

$$\exp\left[-t/\tau\right] = \frac{\left\langle \Delta g_A(0) \Delta g_A(t) \right\rangle}{\left\langle c_A \right\rangle \left\langle c_B \right\rangle}$$

Macroscopic relaxation time is determined by microscopic autocorrelation function.

Microscopic rate expression

$$\exp\left[-t/\tau\right] = \frac{\left\langle \Delta g_A(0) \Delta g_A(t) \right\rangle}{\left\langle c_A \right\rangle \left\langle c_B \right\rangle}$$

Δ has disappeared because of the derivative

Derivative

$$-\frac{1}{\tau}\exp\left[-t/\tau\right] = \frac{\left\langle g_A(0)\dot{g}_A(t) \right\rangle}{\left\langle c_A \right\rangle \left\langle c_B \right\rangle} = -\frac{\left\langle \dot{g}_A(0) g_A(t) \right\rangle}{\left\langle c_A \right\rangle \left\langle c_B \right\rangle}$$

Stationary (t is arbitrary, only depends on τ)

$$\frac{d}{dt}\left\langle A(t) B(t+\tau) \right\rangle = 0$$

$$\left\langle A(t) \dot{B}(t+\tau) \right\rangle + \left\langle \dot{A}(t) B(t+\tau) \right\rangle = 0$$

$$\left\langle A(0) \dot{B}(\tau) \right\rangle = -\left\langle \dot{A}(0) B(\tau) \right\rangle$$

We have
$$\frac{1}{\tau}\exp\left[-t/\tau\right] = \frac{\left\langle \dot{g}_A(0)g_A(t) \right\rangle}{\left\langle c_A \right\rangle \left\langle c_B \right\rangle}$$

Using
$$\tau = k_{A\to B}^{-1}\left(1 + \left\langle c_A \right\rangle / \left\langle c_B \right\rangle\right)^{-1} = \frac{\left\langle c_B \right\rangle}{k_{A\to B}}$$

For sufficiently short t, we obtain

$$\boxed{k_{A\to B}(t) = \frac{\left\langle \dot{g}_A(0)g_A(t) \right\rangle}{\left\langle c_A \right\rangle}}$$

Using the definition of g_A in terms of q we can write

$$\dot{g}_A(q-q^*) = \dot{q}\frac{\partial g_A(q-q^*)}{\partial q} = -\dot{q}\frac{\partial g_B(q-q^*)}{\partial q}$$

$$k_{A\to B}(t) = \frac{\left\langle \dot{q}(0)\dfrac{\partial g_B\left(q(0)-q^*\right)}{\partial q}g_B(t) \right\rangle}{\left\langle c_A \right\rangle}$$

We now have an expression that relates the macroscopic reaction rate to microscopic properties

$$k_{A \rightarrow B}(t) = \frac{\left\langle \dot{q}(0) \dfrac{\partial g_B\left(q(0) - q*\right)}{\partial q} g_B(t) \right\rangle}{\left\langle c_A \right\rangle}$$

Let us look at the different terms in this equation

$$g_B(t) = \theta\left(q(t) - q^*\right)$$

Only when the system is in the product state we get a contribution to the ensemble average

$$\frac{\partial g_B\left(q(0) - q*\right)}{\partial q} = \frac{\partial \Theta\left(q(0) - q*\right)}{\partial q}$$
$$= \delta\left(q(0) - q*\right)$$

Only when the system starts at the transition state, we get a contribution to the ensemble average

$$\dot{q}(0)$$

Velocity at t=0

$$\left\langle c_A \right\rangle = \left\langle \Theta\left(q^* - q\right) \right\rangle$$

Concentration of A

$$k_{A \rightarrow B}(t) = \frac{\left\langle \dot{q}(0) \delta\left(q(0) - q*\right) \theta\left(q(t) - q*\right) \right\rangle}{\left\langle \theta\left(q* - q\right) \right\rangle}$$

Transition state theory

$$k_{A \to B}(t) = \frac{\left\langle \dot{q}(0)\delta\left(q(0) - q*\right)\theta\left(q(t) - q*\right)\right\rangle}{\left\langle \theta\left(q* - q\right)\right\rangle}$$

We can rewrite this expression as a product by inserting 1

$$k_{A \to B}(t) = \frac{\left\langle \dot{q}(0)\delta\left(q(0) - q*\right)\theta\left(q(t) - q*\right)\right\rangle}{\left\langle \delta\left(q(0) - q*\right)\right\rangle} \times \frac{\left\langle \delta\left(q(0) - q*\right)\right\rangle}{\left\langle \theta\left(q* - q\right)\right\rangle}$$

Conditional "probability" to find a particle on the top of the barrier with a positive velocity

Ratio of probabilities to find particle on top of the barrier and in the state A

$$k_{A \to B}(t) = \left\langle \dot{q}(0)\theta\left(q(t) - q*\right)\right\rangle_{q = q*} \times \frac{\left\langle \delta\left(q(0) - q*\right)\right\rangle}{\left\langle \theta\left(q* - q\right)\right\rangle}$$

Free energy barrier

$$\frac{\left\langle \delta\left(q(0)-q*\right)\right\rangle}{\left\langle \theta\left(q*-q\right)\right\rangle}$$

Ratio of the probabilities to find a particle on top of the barrier and in the state A

Probability to be on top of the barrier:

$$\left\langle \delta\left(q*-q\right)\right\rangle = C\int dq\,\delta\left(q-q*\right)\exp\left(-\beta F(q)\right) = C\exp\left(-\beta F(q*)\right)$$

Probability to be in state A:

$$\left\langle \Theta\left(q*-q\right)\right\rangle = C\int dq\,\Theta\left(q-q*\right)\exp\left(-\beta F(q)\right) = C\int_{q<q*} dq\exp\left(-\beta F(q)\right)$$

This gives:

$$\frac{\left\langle \delta\left(q(0)-q*\right)\right\rangle}{\left\langle \theta\left(q*-q\right)\right\rangle} = \frac{\exp\left(-\beta F(q*)\right)}{\int_{q<q*} dq\exp\left(-\beta F(q)\right)}$$

We need to determine the free energy as a function of the order parameter

$$\left\langle \dot{q}(0)\theta\big(q(t)-q*\big)\right\rangle_{q=q*}$$

Conditional "probability" to find a particle on the top of the barrier with a positive velocity

$$\dot{q}(0)$$

Assume that on top of the barrier the particle is in equilibrium: use Maxwell-Boltzmann distribution to generate this velocity

$$\dot{q}(0)\theta\big(q(t)-q*\big)$$

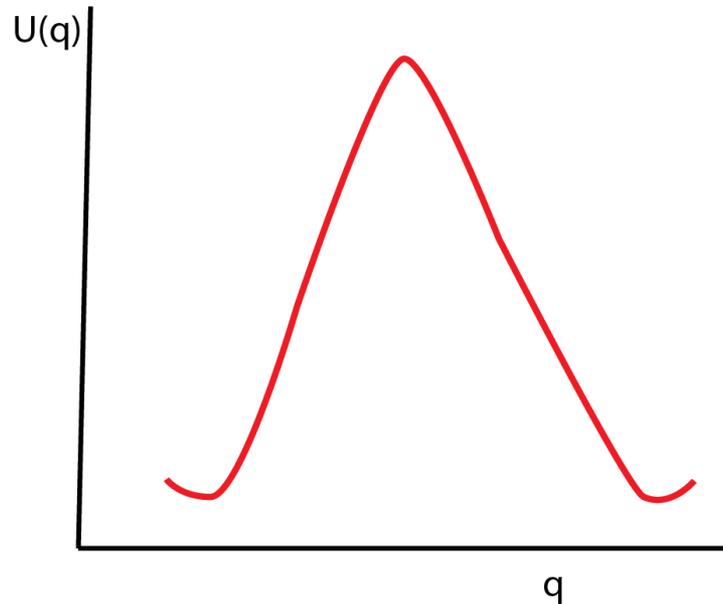Only particles with a positive velocity end up in the product state. We assume that once in the product state they stay there.

$$\lim_{t\to 0^+} \dot{q}(0)\theta\big(q(t)-q*\big) = \dot{q}(0)\theta\big(\dot{q}(0)\big) = 0.5\big|\dot{q}(0)\big|$$

$$k_{A\to B}^{TST} = \lim_{t\to 0^+} \left\langle \dot{q}(0)\theta\big(q(t)-q*\big)\right\rangle_{q=q*} \times \frac{\left\langle \delta\big(q(0)-q*\big)\right\rangle}{\left\langle \theta\big(q*-q\big)\right\rangle}$$

$$k_{A\to B}^{TST} = 0.5\big|\dot{q}(0)\big|\frac{\exp\big(-\beta F(q*)\big)}{\int\limits_{q<q*} dq\,\exp\big(-\beta F(q)\big)}$$

Eyring's TST

1-D ideal gas particle on a hill

$$k_{A \to B}^{TST} = 0.5 \left| \dot{q}(0) \right| \frac{\exp\left(-\beta F(q^*)\right)}{\int\limits_{q < q^*} dq \exp\left(-\beta F(q)\right)}$$
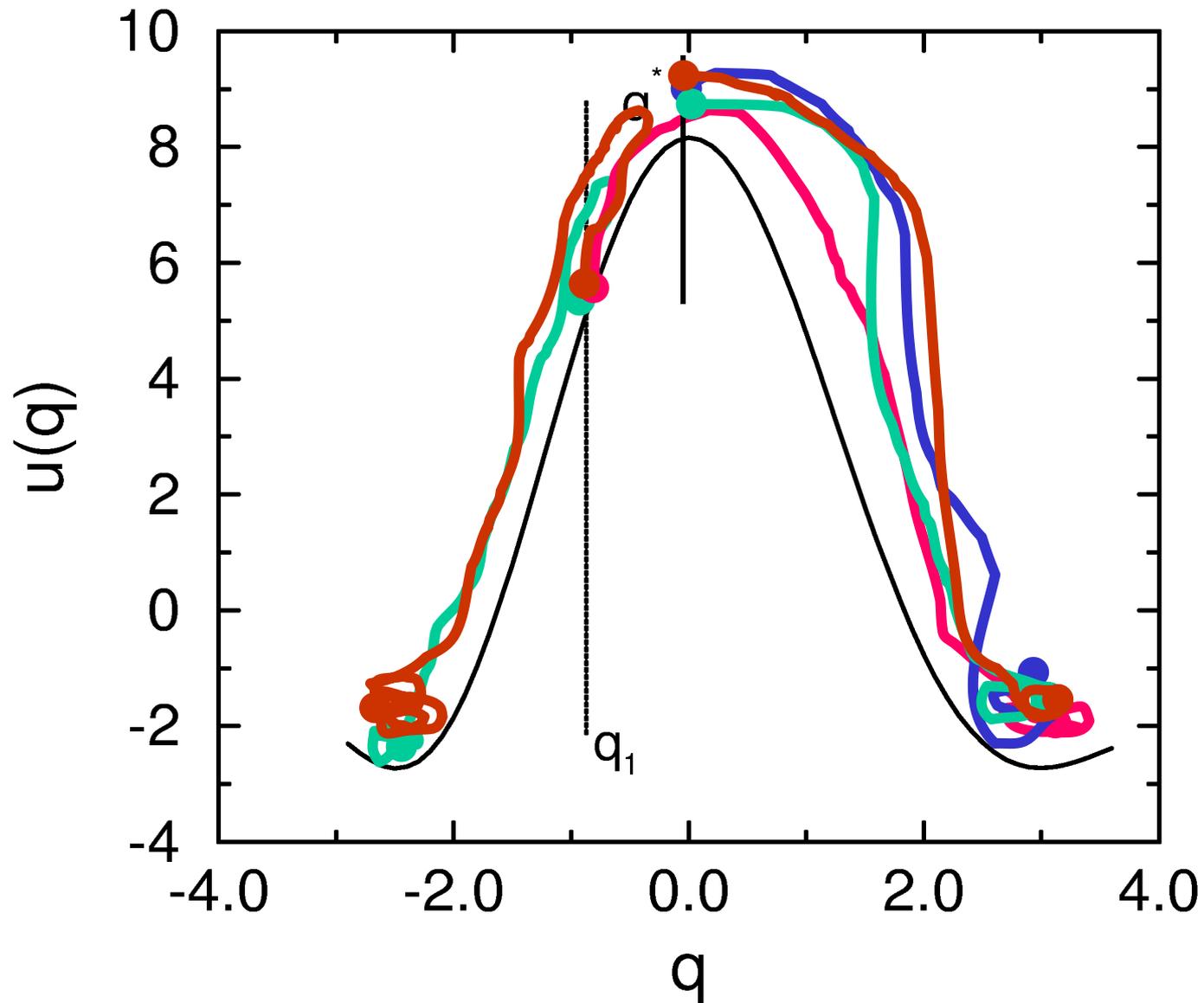
Maxwell-Boltzmann: $\left| \dot{q}(0) \right| = \sqrt{\dfrac{2k_B T}{\pi m}}$

This gives for the hopping rate

$$k_{A \to B}^{TST} = \sqrt{\frac{k_B T}{2\pi m}} \frac{\exp\left(-\beta U(q^*)\right)}{\int\limits_{q < q^*} dq \exp\left(-\beta U(q)\right)}$$

Ideal gas particle on a not-so-ideal hill

q^* is the true transition state

q_1 is the estimated transition state
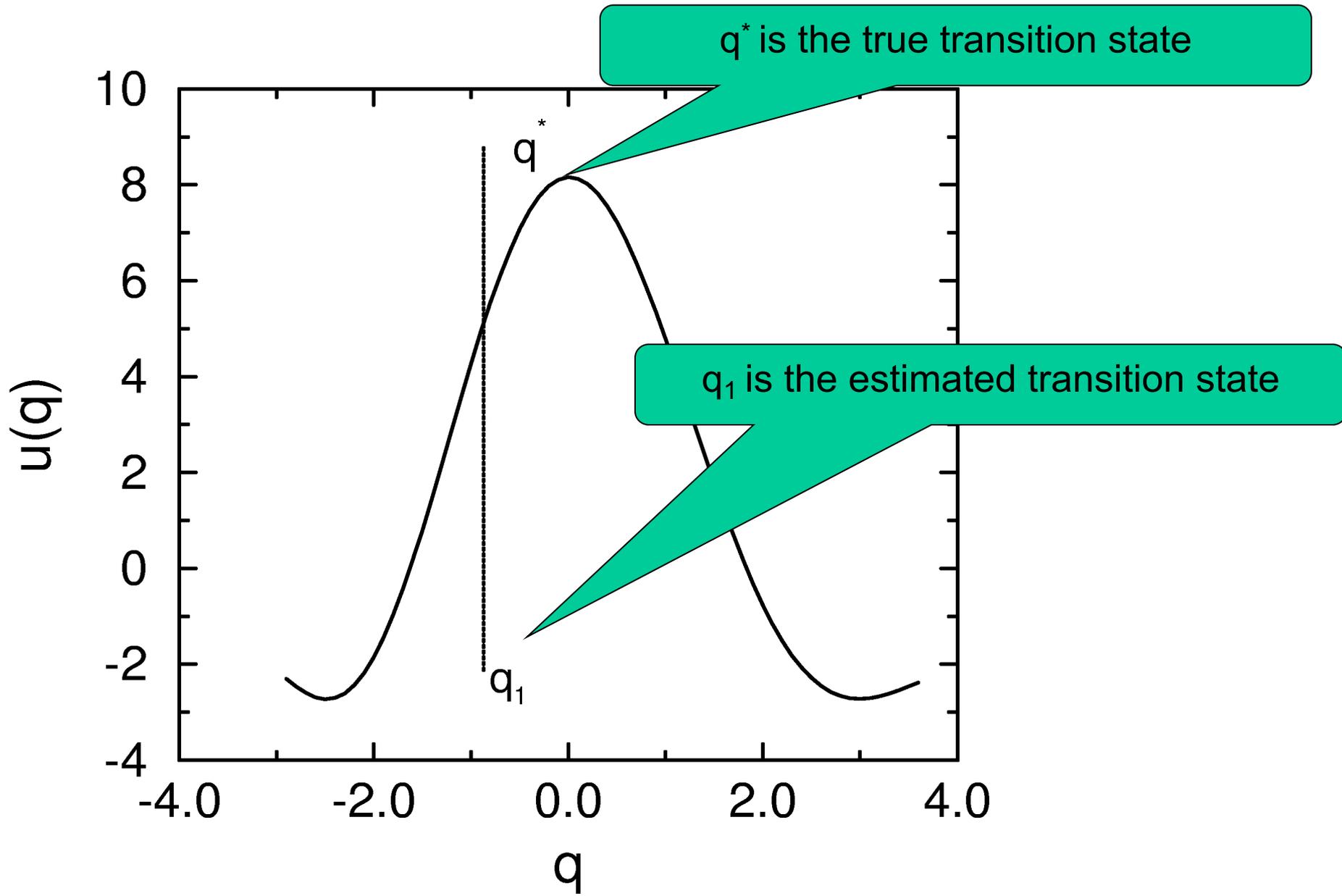
For this case transition state theory will overestimate the hopping rate

Transition state theory

- One has to know the free energy accurately (MC/MD), using umbrella sampling, thermodynamic integration, metadynamics etc.

- Gives only an upper bound to the reaction rate

- Assumptions underlying transition theory should hold: *no recrossings*

$$k_{A \to B}(t) = \frac{\left\langle \dot{q}(0) \delta\big(q(0) - q*\big) \theta\big(q(t) - q*\big) \right\rangle}{\left\langle \theta\big(q* - q\big) \right\rangle}$$

lower value because of recrossings

k_{TST}

$k(t)$

$\sim \exp(-t/\tau_{rxn})$

τ_{mol}

t
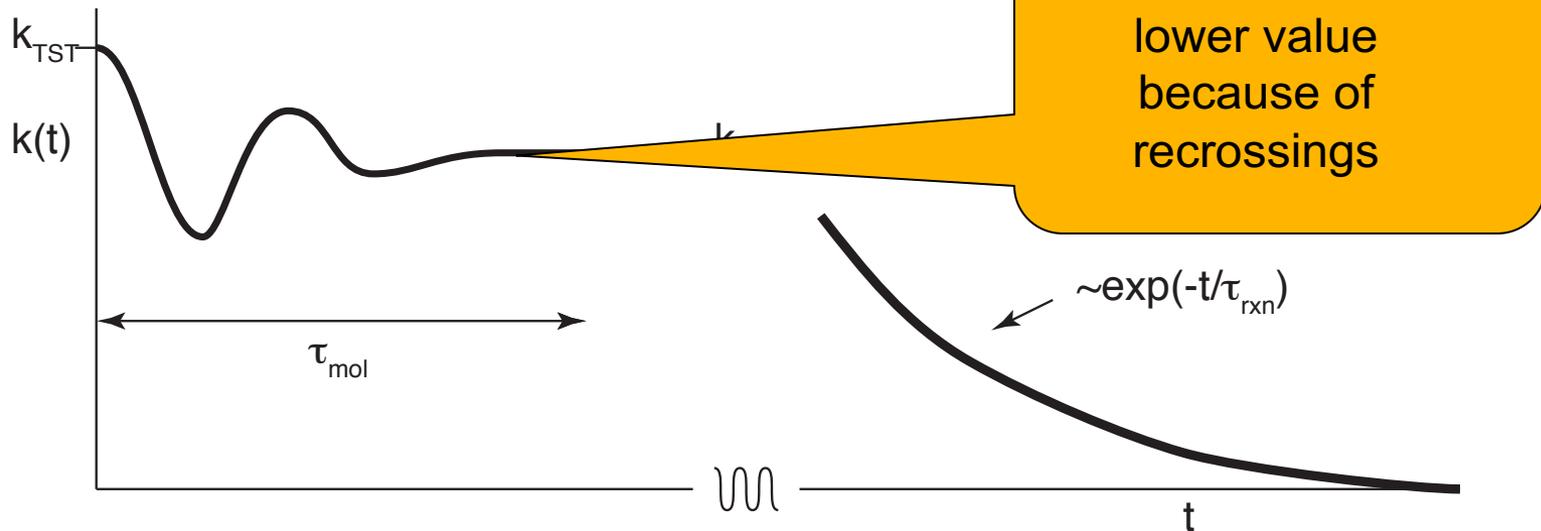
Recrossings lower the rate

trajectories that seem to overcome the barrier but in fact bounce back
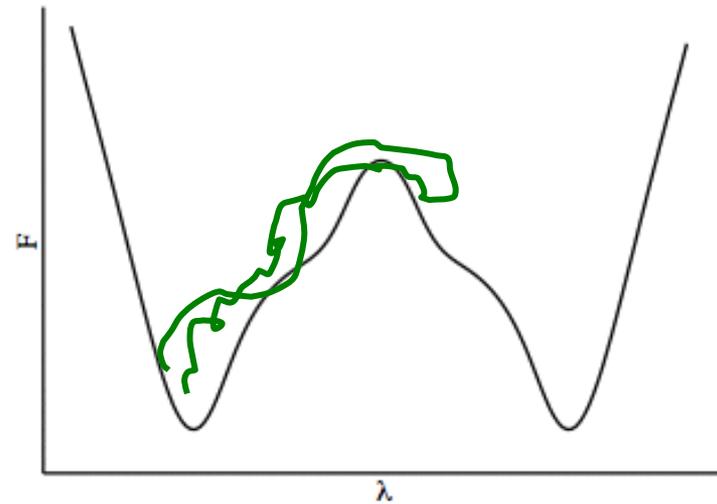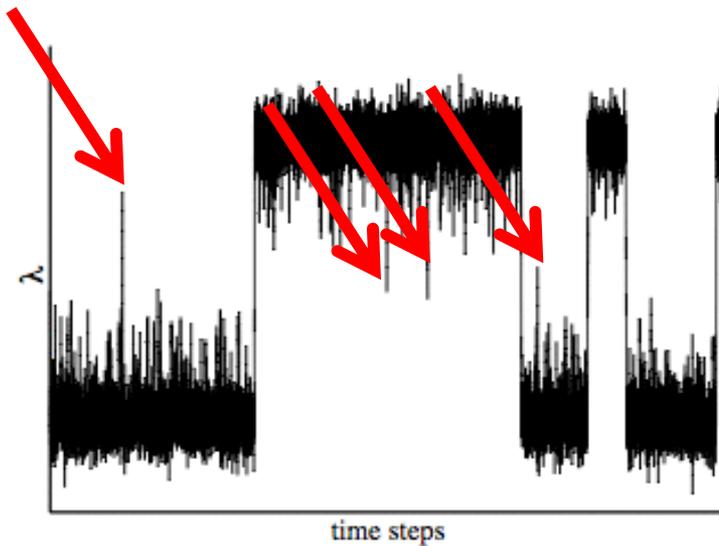
Bennett-Chandler approach

$$k_{A \to B}(t) = \frac{\left\langle \dot{q}(0)\delta\big(q(0)-q*\big)\theta\big(q(t)-q*\big)\right\rangle}{\left\langle \theta\big(q*-q\big)\right\rangle}$$

$$k_{A \to B}(t) = \frac{\left\langle \dot{q}(0)\delta\big(q(0)-q*\big)\theta\big(q(t)-q*\big)\right\rangle}{\left\langle \delta\big(q(0)-q*\big)\right\rangle} \times \frac{\left\langle \delta\big(q(0)-q*\big)\right\rangle}{\left\langle \theta\big(q*-q\big)\right\rangle}$$

Computational scheme:

> Jocelyne Vreede will discuss
> FE methods on Thursday

1. Determine the probability from the free energy using MC or MD, e.g. by umbrella sampling, thermodynamic integration, metadynamics or other free energy methods
2. Compute the conditional average from a MD simulation

Bennett-Chandler approach

$$k_{A \to B}^{TST}(t) = \frac{\left\langle \dot{q}(0)\delta(q(0)-q_1)\theta(\dot{q})\right\rangle}{\left\langle \delta(q(0)-q_1)\right\rangle} \times \frac{\left\langle \delta(q(0)-q_1)\right\rangle}{\left\langle \theta(q_1-q)\right\rangle}$$

$$k_{A \to B}(t) = \frac{\left\langle \dot{q}(0)\delta(q(0)-q_1)\theta(q(t)-q_1)\right\rangle}{\left\langle \delta(q(0)-q_1)\right\rangle} \times \frac{\left\langle \delta(q(0)-q_1)\right\rangle}{\left\langle \theta(q_1-q)\right\rangle}$$

Transmission coefficient

MD simulation to correct the transition state result!

$$\kappa(t) \equiv \frac{k_{A \to B}(t)}{k_{A \to B}^{TST}}$$

$$= \frac{\left\langle \dot{q}(0)\delta(q(0)-q_1)\theta(q(t)-q_1)\right\rangle}{0.5\left|\dot{q}(0)\right|}$$

MD simulation:
1. At t=0 $q=q_1$
2. Determine fraction at product state weighted with initial velocity
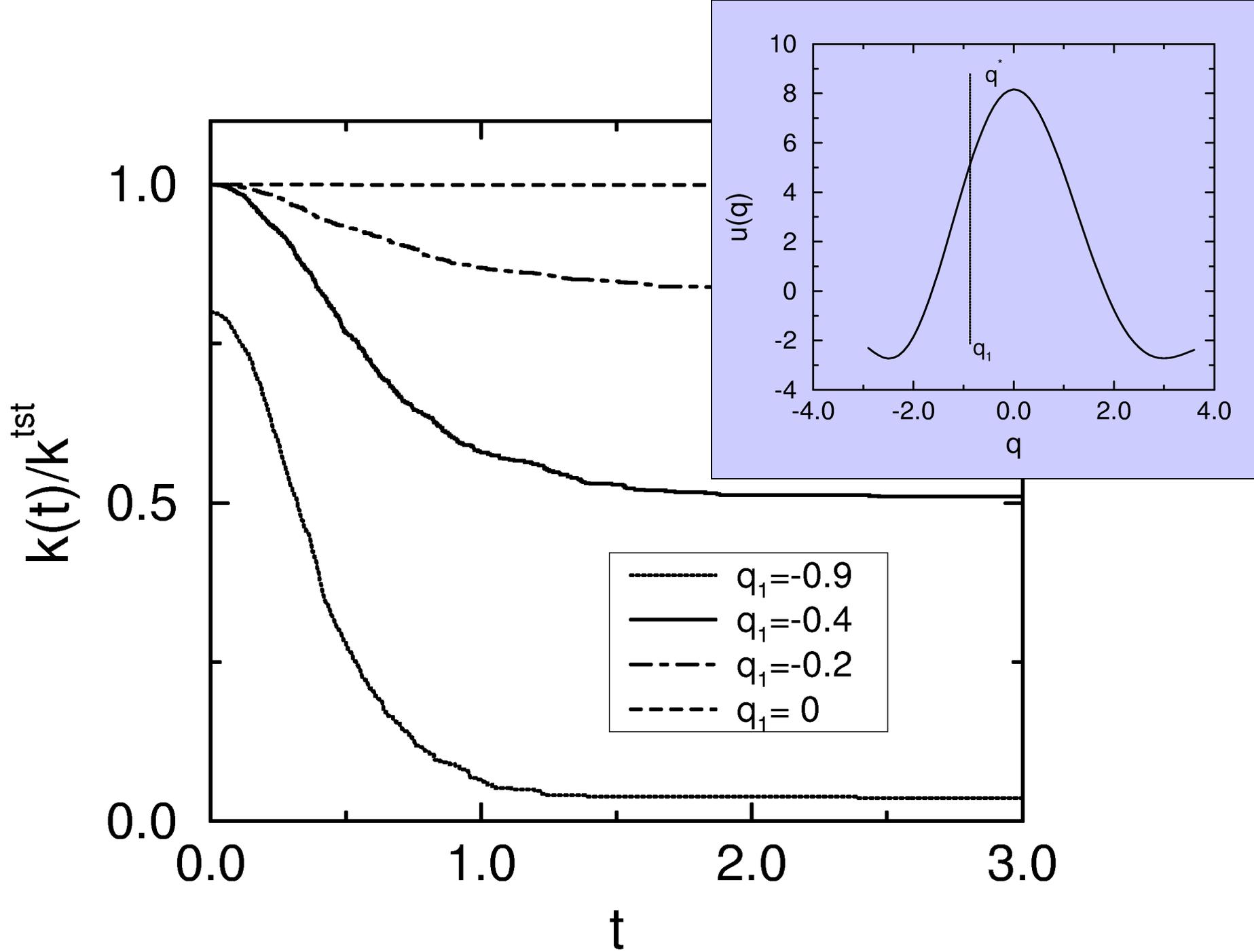
Example diffusion in zeolite

- Zeolites important class of materials

- Diffusion of alkanes in matrix is poorly described

- Approach
  - molecular simulation of alkanes in fixed zeolite frame
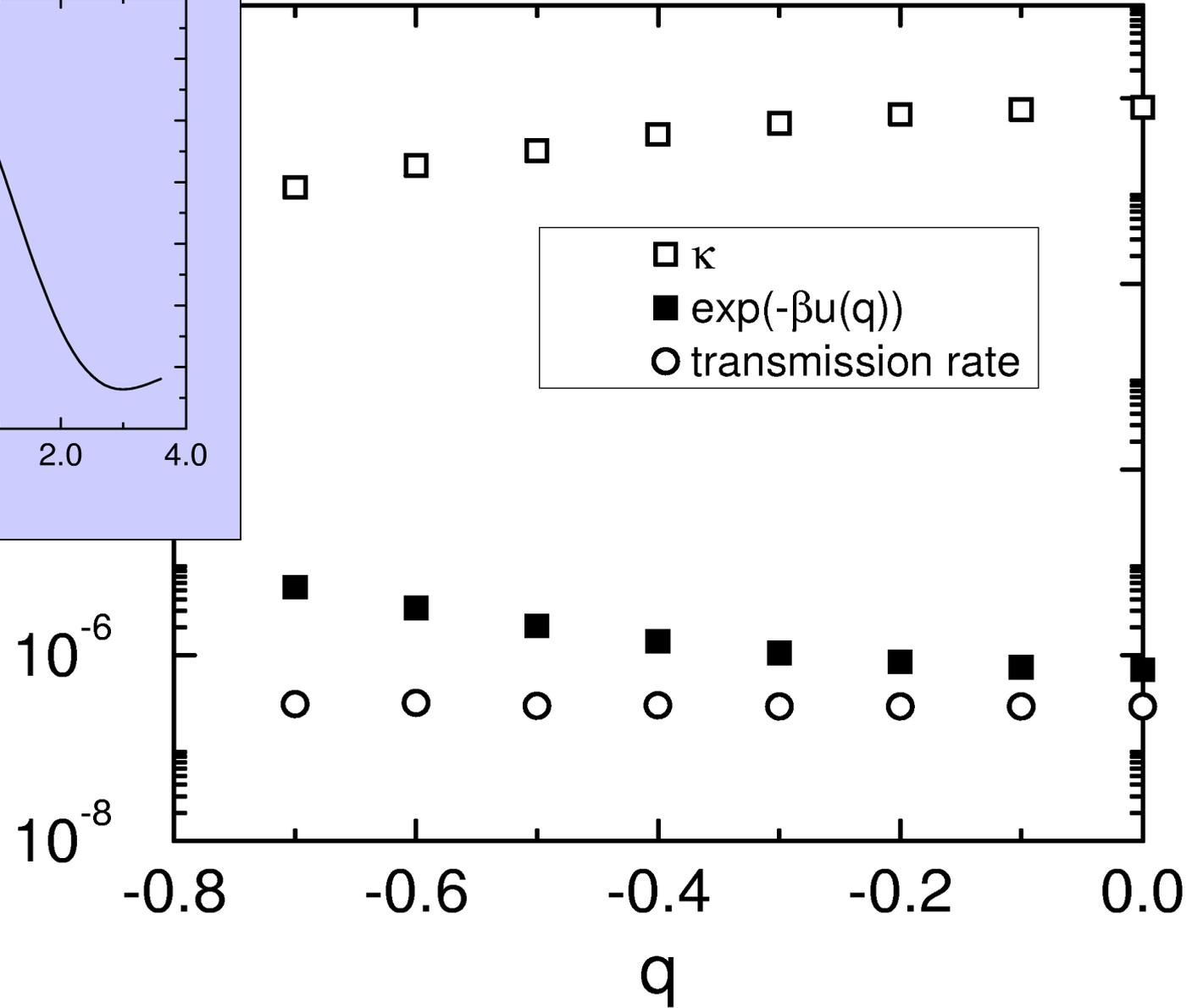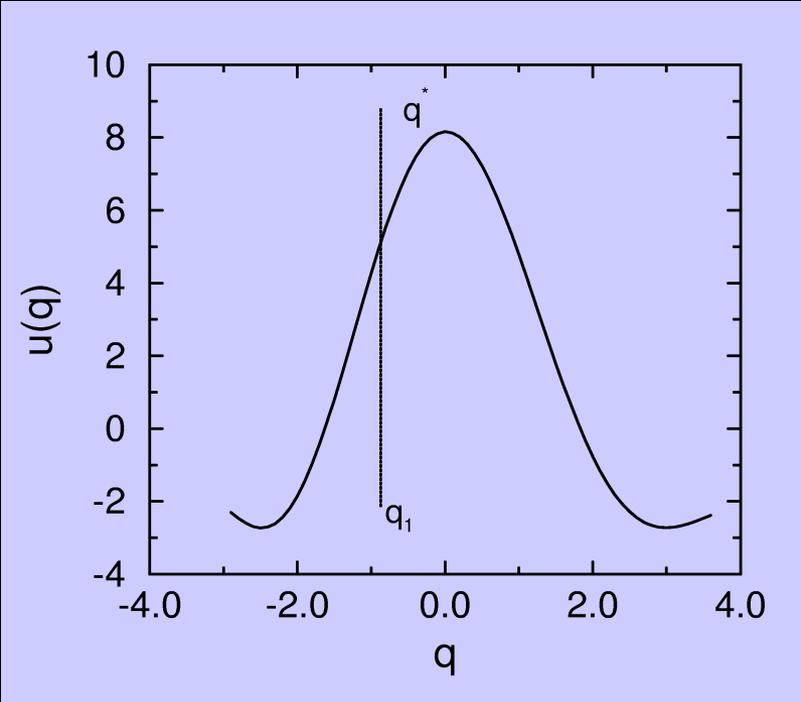 - Unified atom FF by Dubbeldam et al.

D. Dubbeldam, et al., J. Phys. Chem. B, 108, 12301, 2004

Reaction coordinate

cage   window   cage   cage   window   cage

g^* g^*

q q
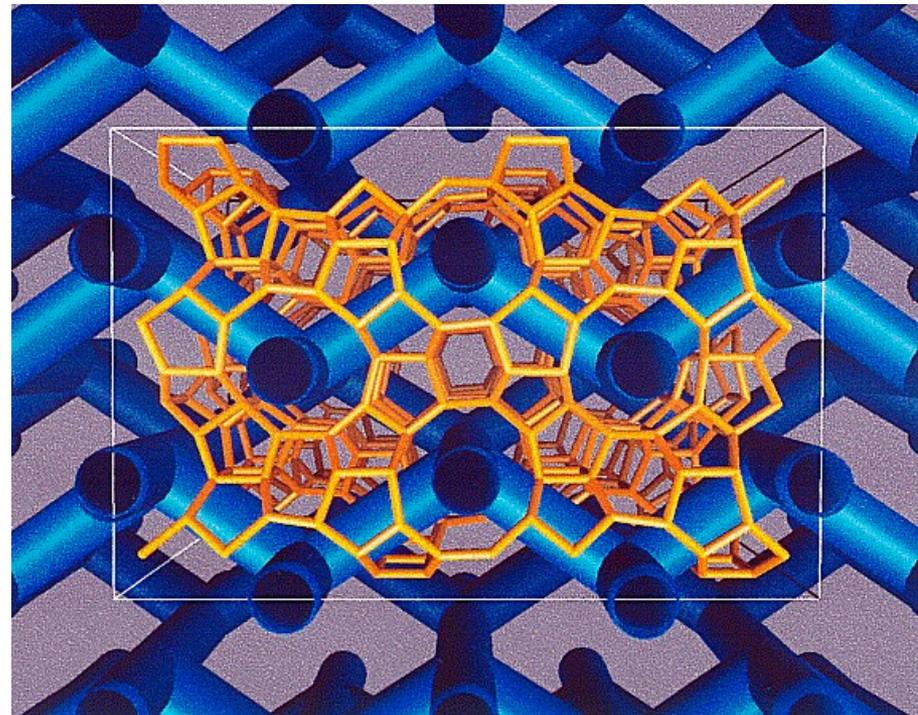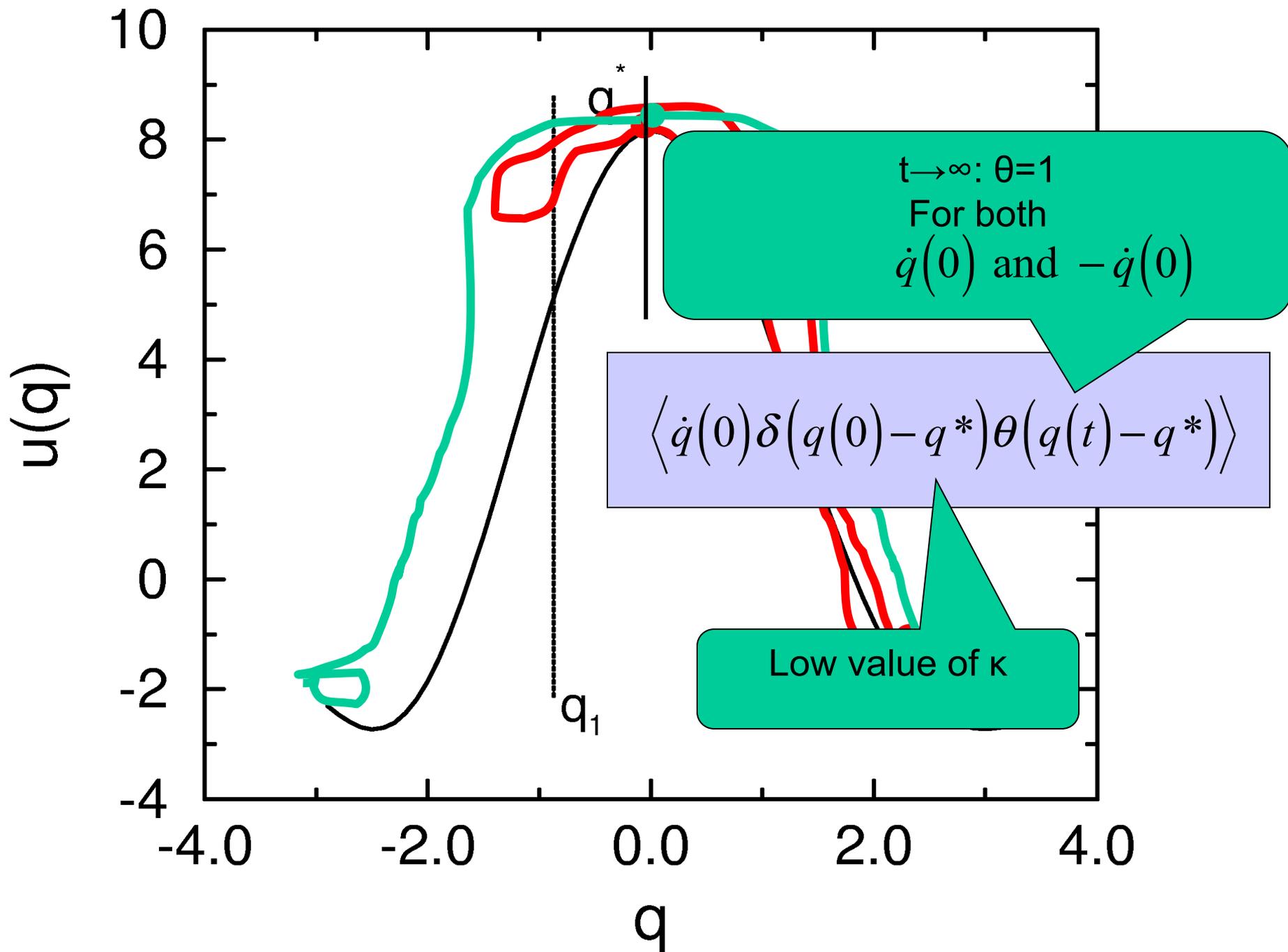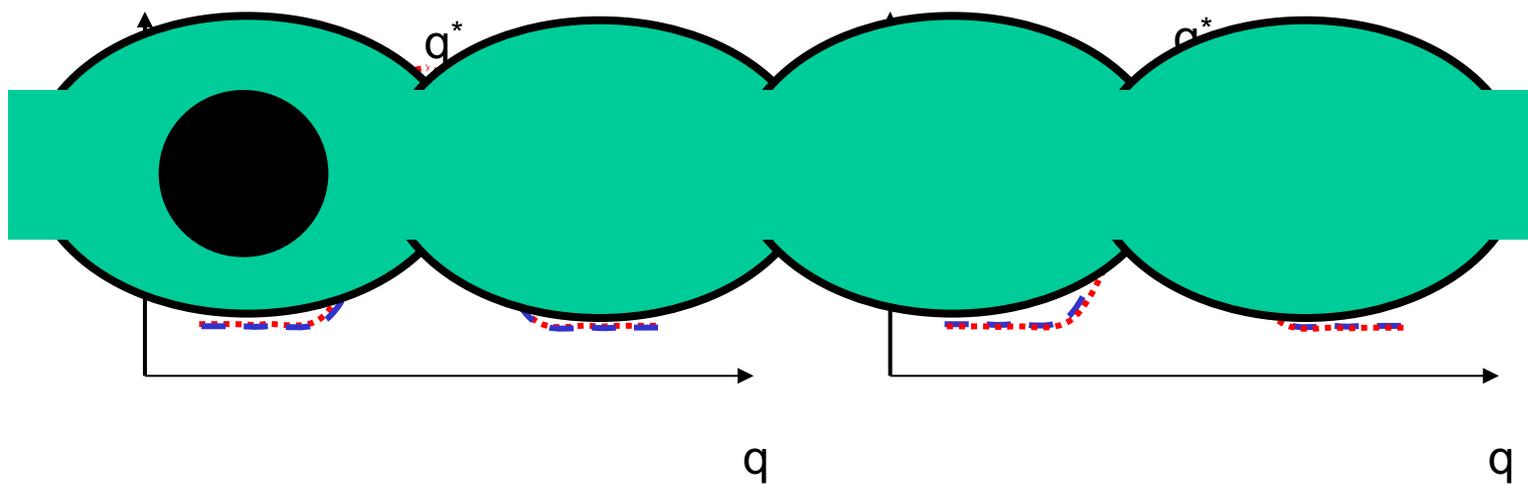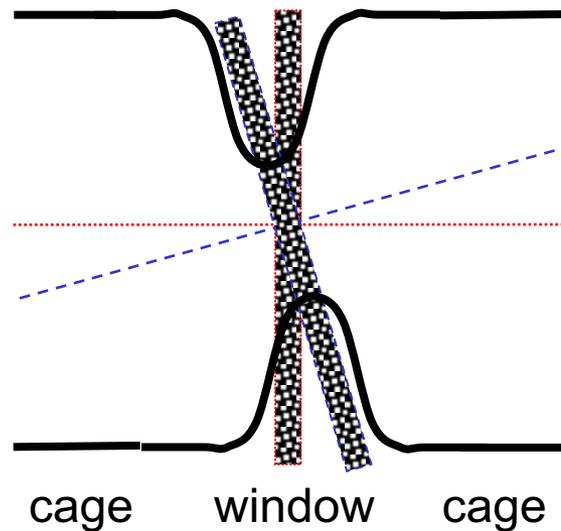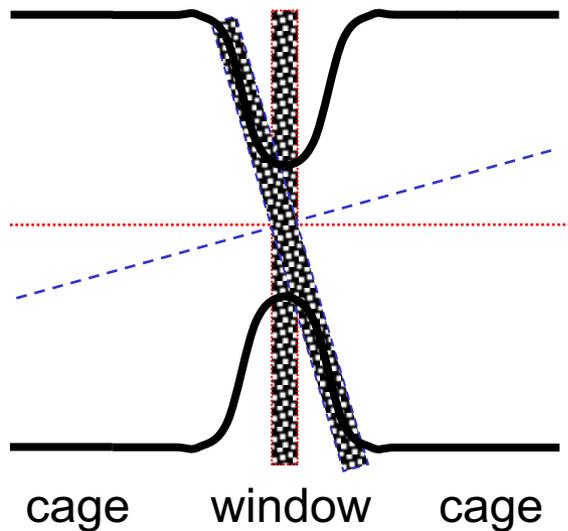
Rate based on free energy can be orders of magnitude wrong!

Outline

- Part 1
  - Rare events and reaction kinetics
  - Microscopic expression from Linear Response
  - Transition state theory
  - Bennet Chandler approach

- Part 2
  - Two ended methods
  - Transition path sampling
  - Rate constants with path sampling
  - Reaction coordinate analysis
  - Application to crystallization
  - Open Path Sampling

Barriers on smooth and rough energy landscapes

- Clearly, barrier is most important for rare event
- But how to obtain this barrier?
- In multidimensional energy landscapes barrier is saddle point
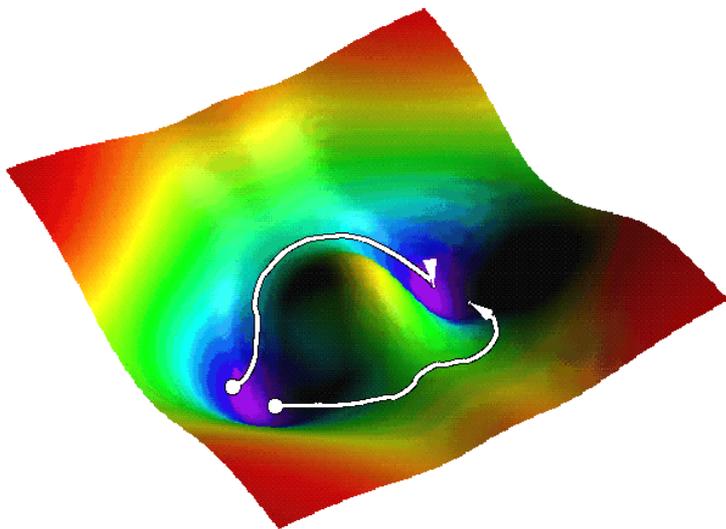
- # saddle points limited
- determined by potential energy
- use eigenvectors or Hessian to find them

- # saddle points uncountable
- entropy important, many pathways
- determined by free energy
- exploring requires sampling schemes

Dellago logo™
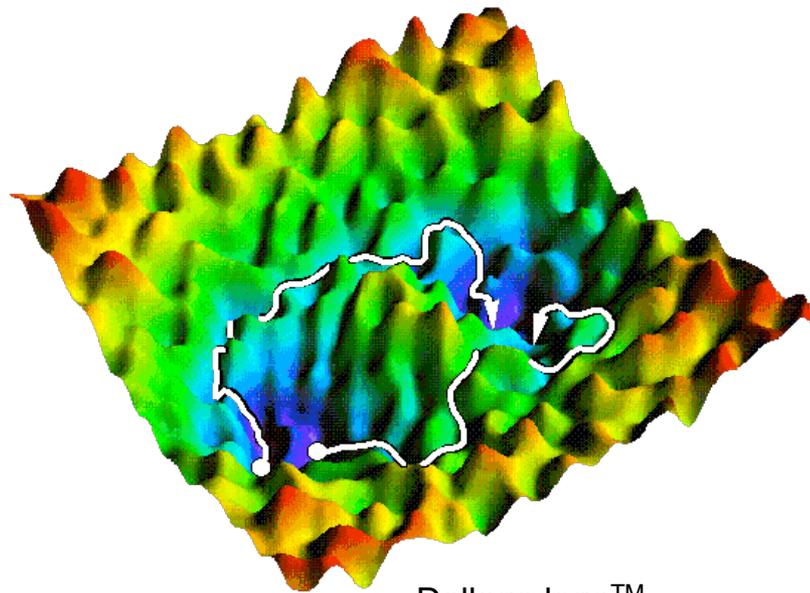
Breakdown of BC approach

kappa can become immeasurable low if the reaction coordinate is
at the wrong value (i.e. the reaction coordinate is wrongly chosen)
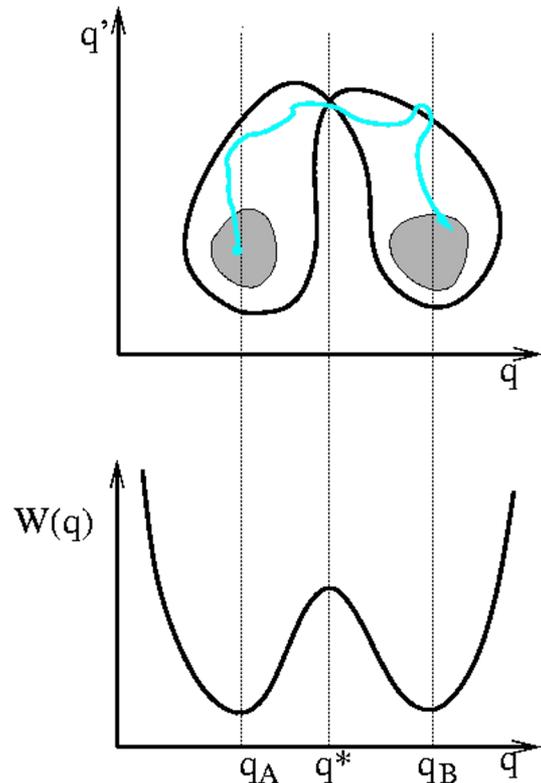
$$W(q) = -kT \ln \int dq' \exp\{-\beta E(q,q')\}$$

If the reaction coordinate is not known, the
wrong order parameter can lead to wrong
transition states, mechanism and rates

Two ended methods

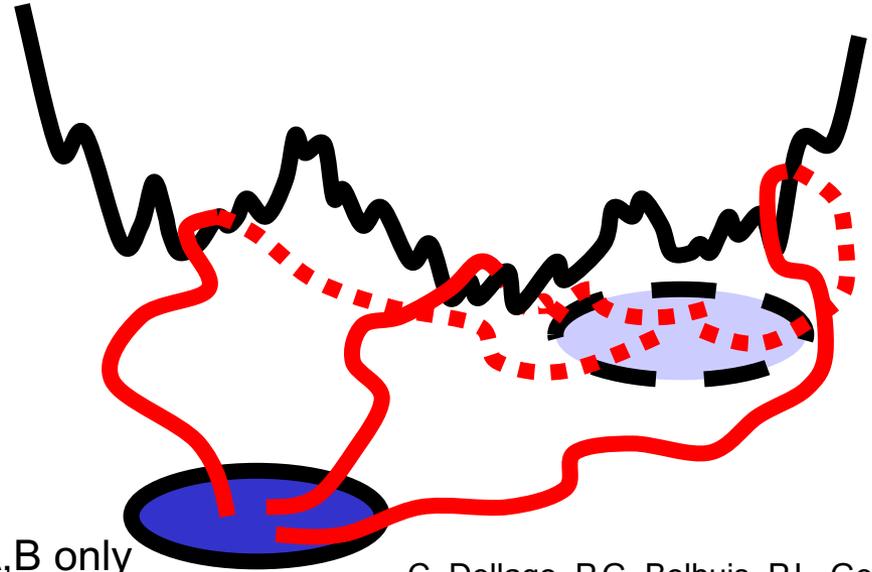Methods that take the entire path
and fix the begin and end point
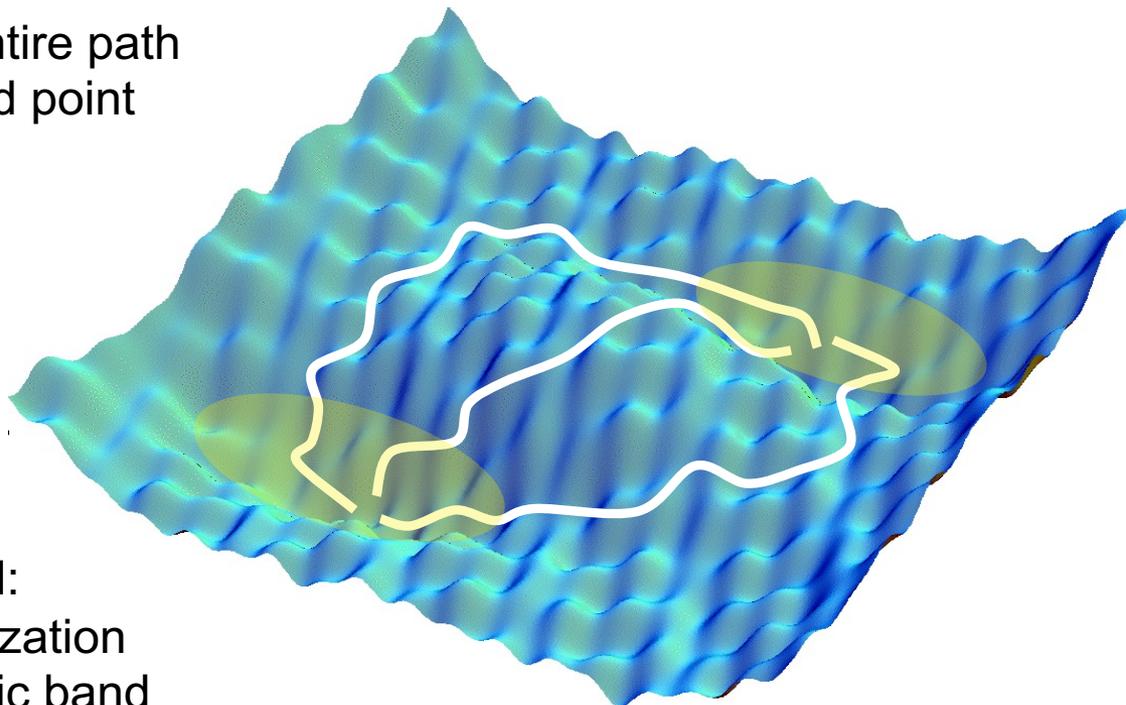
Many methods proposed:
        Action minimization
        Nudged elastic band
        String method
        Path metadynamics
        Milestoning
        Transition path sampling
        ....

Transition path sampling

Samples the path ensemble:

all trajectories that lead over barrier

- Sampling by Monte Carlo
- Requires definition of stable states A,B only
- Results in ensemble of pathways
- Reaction coordinate is a result of simulation not an input
- Allows for calculation of rate constants

C. Dellago, P.G. Bolhuis, P.L. Geissler
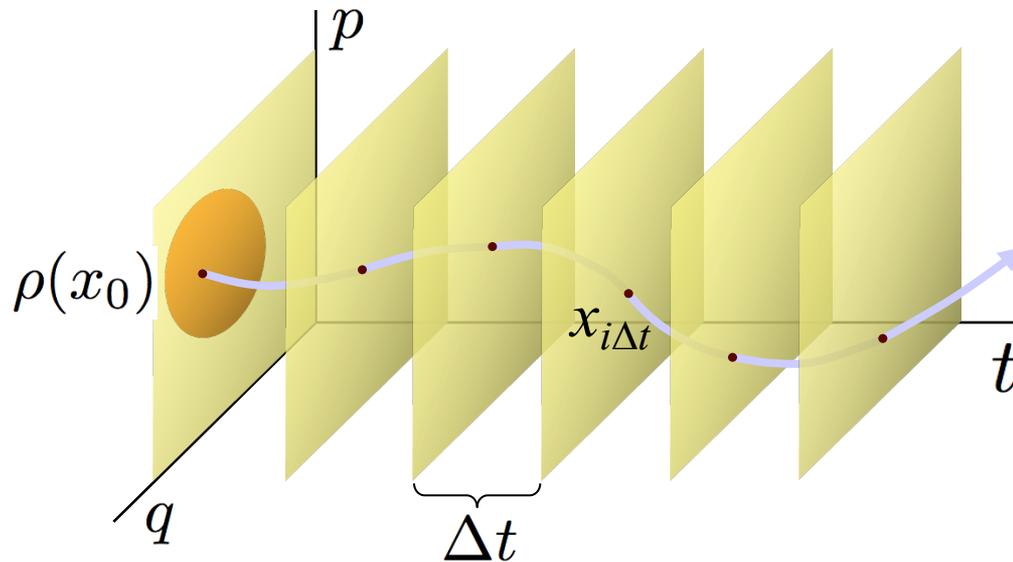
Adv. Chem. Phys. **123**, 1 2002

Apply when process of interest
- – is a rare event
- – is complex and reaction coordinate is not known

Examples: nucleation, reactions in solution, protein folding

Path probability density

$$x(\mathcal{T}) \equiv \{x_0, x_{\Delta t}, x_{2\Delta t}, \ldots, x_{\mathcal{T}}\}$$

Path = Sequence of states

$$\mathcal{P}[x(\mathcal{T})] = \rho(x_0) \prod_{i=0}^{\mathcal{T}/\Delta t - 1} p(x_{i\Delta t} \to x_{(i+1)\Delta t})$$

Transition path ensemble
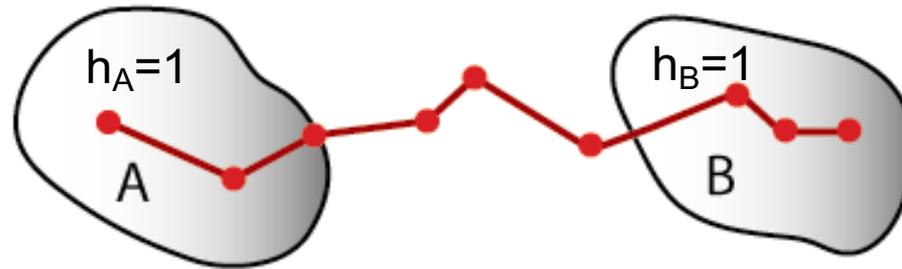
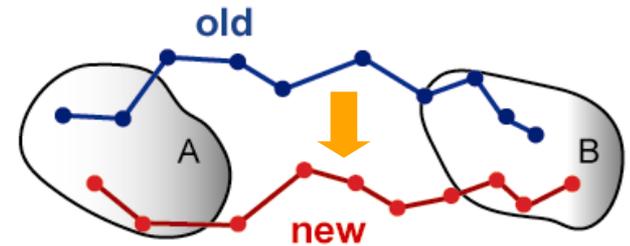$$\mathcal{P}_{AB}[x(\mathcal{T})] \equiv h_A(x_0)\mathcal{P}[x(\mathcal{T})]h_B(x_{\mathcal{T}})/Z_{AB}(\mathcal{T})$$

$$Z_{AB}(\mathcal{T}) \equiv \int \mathcal{D}x(\mathcal{T})\, h_A(x_0)\mathcal{P}[x(\mathcal{T})]h_B(x_{\mathcal{T}})$$

$$\int \mathcal{D}x(\mathcal{T}) \equiv \int \cdots \int dx_0 dx_{\Delta t} dx_{2\Delta t} \cdots dx_{\mathcal{T}}$$

Metropolis MC of pathways

1. Generate **new** path from **old** one

$$x^{(o)}(\mathcal{T}) \longrightarrow x^{(n)}(\mathcal{T})$$

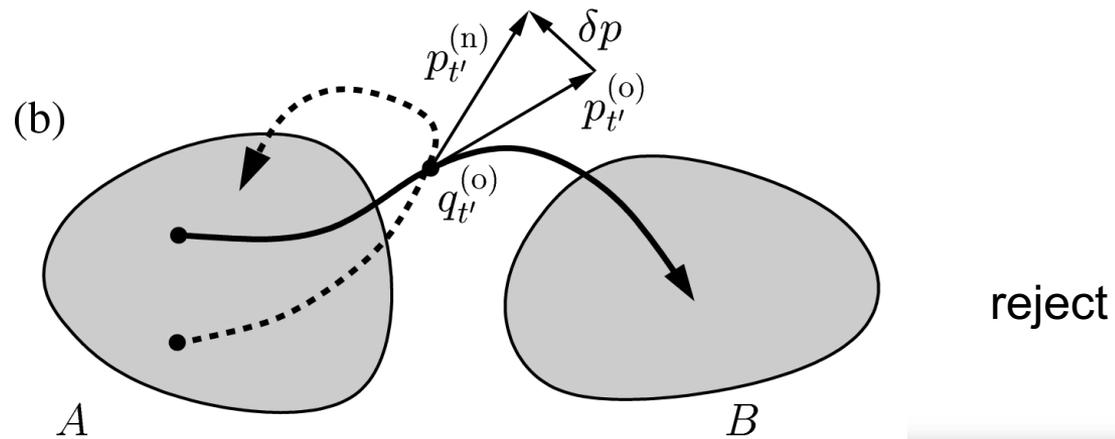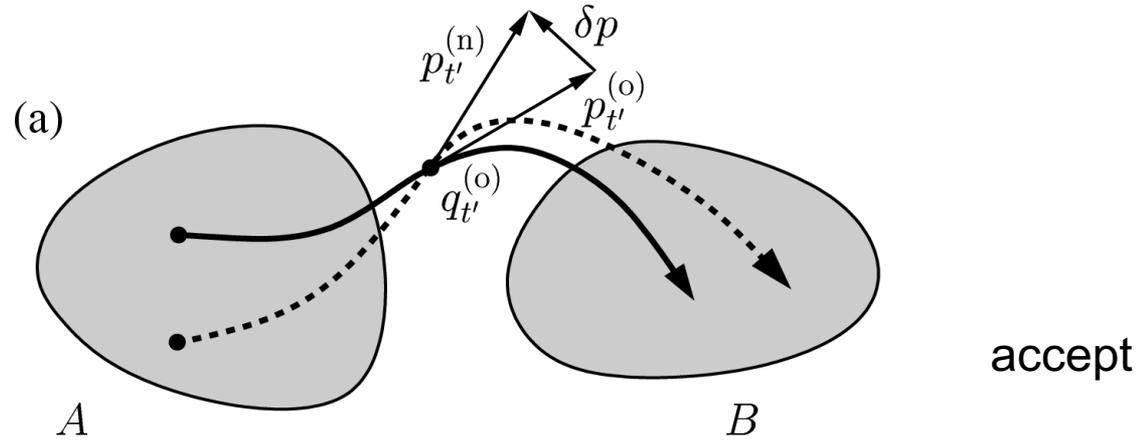2. Accept **new** path according to **detailed balance**:

$$\mathcal{P}_{AB}[x^{(o)}(\mathcal{T})]\pi[x^{(o)}(\mathcal{T}) \to x^{(n)}(\mathcal{T})] = \mathcal{P}_{AB}[x^{(n)}(\mathcal{T})]\pi[x^{(n)}(\mathcal{T}) \to x^{(o)}(\mathcal{T})]$$

$$\pi[x^{(o)}(\mathcal{T}) \to x^{(n)}(\mathcal{T})] = P_{\text{gen}}[x^{(o)}(\mathcal{T}) \to x^{(n)}(\mathcal{T})] \times P_{\text{acc}}[x^{(o)}(\mathcal{T}) \to x^{(n)}(\mathcal{T})]$$

3. Satisfy detailed balance with the **Metropolis rule**:

$$P_{\text{acc}}[x^{(o)}(\mathcal{T}) \to x^{(n)}(\mathcal{T})] = h_A[x_0^{(n)}]h_B[x_{\mathcal{T}}^{(n)}] \min\left\{1, \frac{\mathcal{P}[x^{(n)}(\mathcal{T})]P_{\text{gen}}[x^{(n)}(\mathcal{T}) \to x^{(o)}(\mathcal{T})]}{\mathcal{P}[x^{(o)}(\mathcal{T})]P_{\text{gen}}[x^{(o)}(\mathcal{T}) \to x^{(n)}(\mathcal{T})]}\right\}$$

Shooting moves

(a)

$p_{t'}^{(n)}$ δp

$p_{t'}^{(o)}$

$q_{t'}^{(o)}$

A B

accept

(b)

$p_{t'}^{(n)}$ δp

$p_{t'}^{(o)}$

$q_{t'}^{(o)}$

A B

reject

$$P_{acc}[x^{(o)}(T) \rightarrow x^{(n)}(T)] = h_A(x_0^{(n)}) h_B(x_T^{(n)})$$

$$h_A(t) = \begin{cases} 1 & if \quad x_t \in A \\ 0 & if \quad x_t \notin A \end{cases}$$

Shooting algorithm

$$P_{\text{gen}}^{\text{f}}[x^{\text{o}}(\mathcal{T}) \to x^{\text{n}}(\mathcal{T})] = \prod_{i=t'/\Delta t}^{\mathcal{T}/\Delta t - 1} p\left(x_{i\Delta t}^{(n)} \to x_{(i+1)\Delta t}^{(n)}\right)$$

$$P_{\text{gen}}^{\text{b}}[x^{\text{o}}(\mathcal{T}) \to x^{\text{n}}(\mathcal{T})] = \prod_{i=1}^{t'/\Delta t} \bar{p}\left(x_{i\Delta t}^{(n)} \to x_{(i-1)\Delta t}^{(n)}\right)$$

$$P_{\text{gen}}[x^{(\text{o})}(\mathcal{T}) \to x^{(\text{n})}(\mathcal{T})] = p_{\text{gen}}[x_{t'}^{(\text{o})} \to x_{t'}^{(\text{n})}] \prod_{i=t'/\Delta t}^{\mathcal{T}/\Delta t - 1} p\left(x_{i\Delta t}^{(n)} \to x_{(i+1)\Delta t}^{(n)}\right) \times \prod_{i=1}^{t'/\Delta t} \bar{p}\left(x_{i\Delta t}^{(n)} \to x_{(i-1)\Delta t}^{(n)}\right)$$

$$P_{\text{acc}}[x^{(\text{o})}(\mathcal{T}) \to x^{(\text{n})}(\mathcal{T})] = h_A[x_0^{(n)}]h_B[x_{\mathcal{T}}^{(n)}] \min\left[1, \frac{\rho\left(x_0^{(n)}\right)}{\rho\left(x_0^{(o)}\right)} \prod_{i=0}^{t'/\Delta t - 1} \frac{p\left(x_{i\Delta t}^{(n)} \to x_{(i+1)\Delta t}^{(n)}\right)}{\bar{p}\left(x_{(i+1)\Delta t}^{(n)} \to x_{i\Delta t}^{(n)}\right)} \times \frac{\bar{p}\left(x_{(i+1)\Delta t}^{(o)} \to x_{i\Delta t}^{(o)}\right)}{p\left(x_{i\Delta t}^{(o)} \to x_{(i+1)\Delta t}^{(o)}\right)}\right]$$
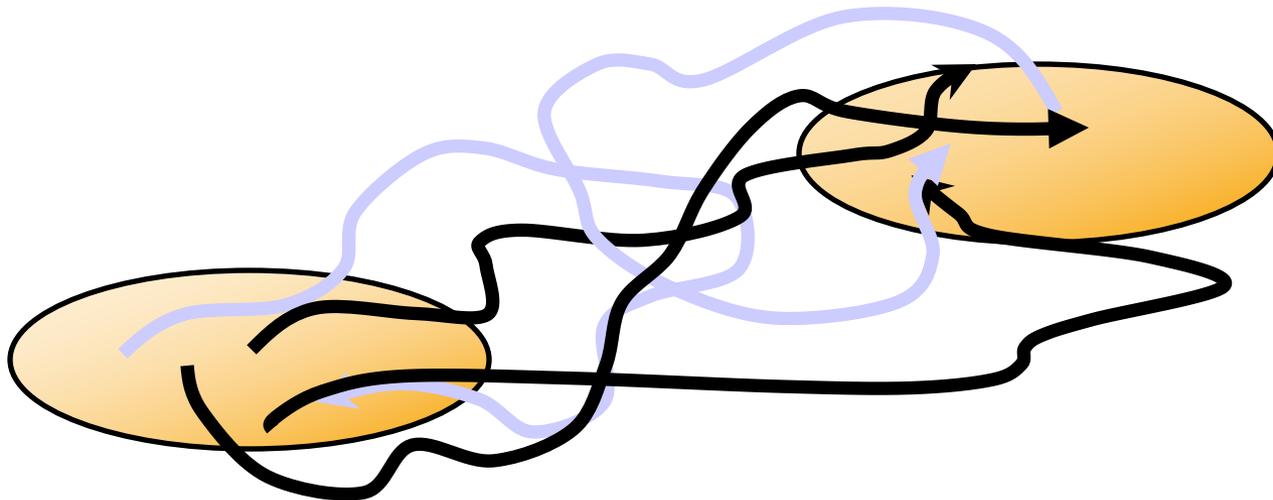
$$\frac{p(x \to y)}{\bar{p}(y \to x)} = \frac{\rho_0(y)}{\rho_0(x)}$$

$$P_{\text{acc}}[x^{(\text{o})}(\mathcal{T}) \to x^{(\text{n})}(\mathcal{T})] = h_A[x_0^{(n)}]h_B[x_{\mathcal{T}}^{(n)}] \min\left[1, \frac{\rho(x_{t'}^{(n)})}{\rho(x_{t'}^{(o)})}\right]$$
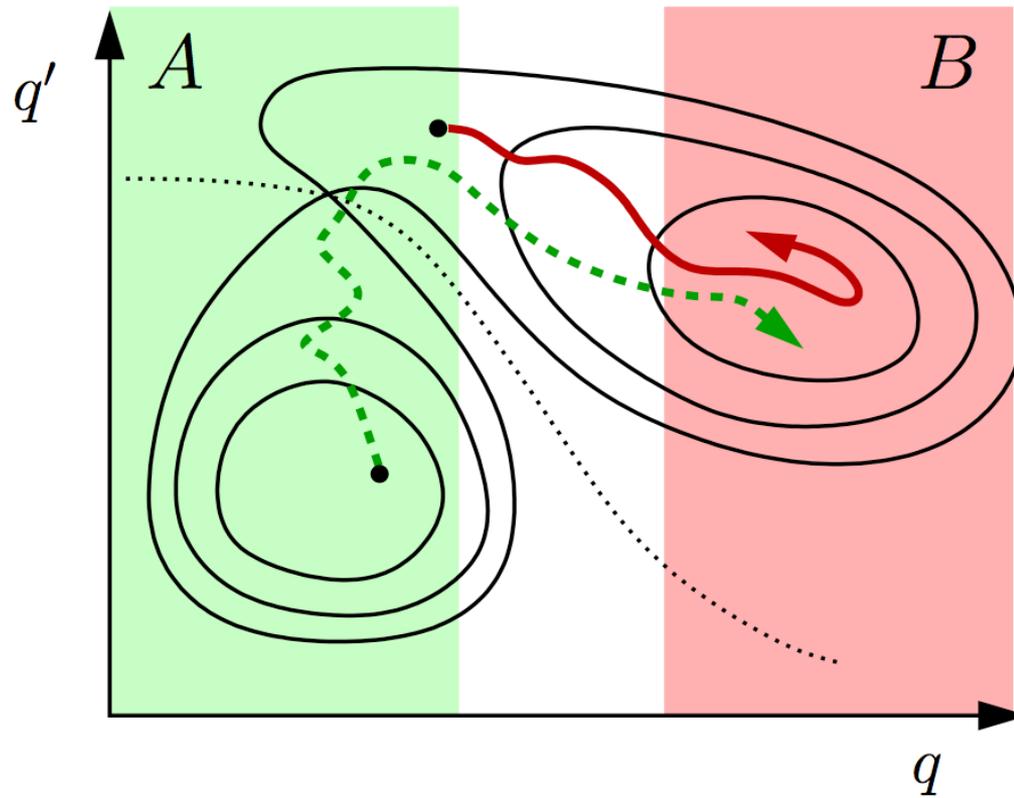
$$P_{\text{acc}}[x^{(\text{o})}(\mathcal{T}) \to x^{(\text{n})}(\mathcal{T})] = h_A[x_0^{(n)}]h_B[x_{\mathcal{T}}^{(n)}]$$
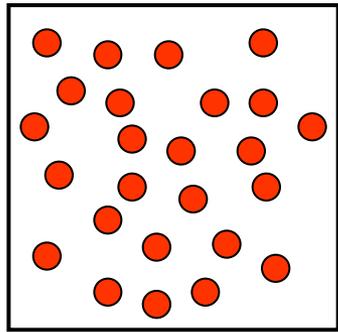
Standard TPS algorithm

- take existing path
- choose random time slice *t*
- change momenta slightly at *t*
- integrate forward and backward in time to create new path of length L
- accept if A and B are connected, otherwise reject and retain old path
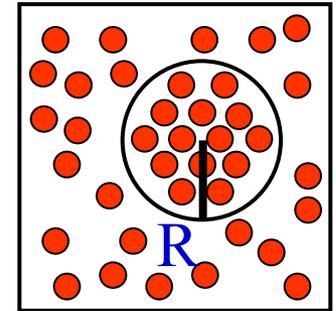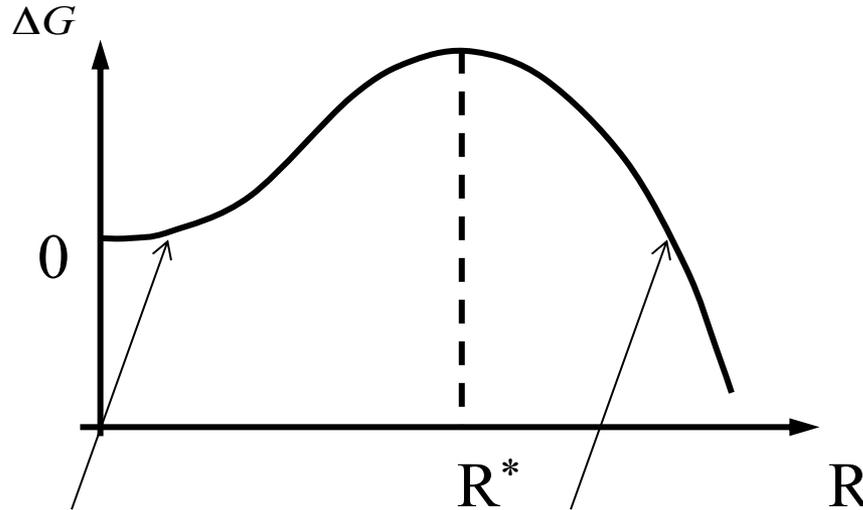- calculate averages
- repeat

Definition of the stable states

Classical nucleation (1926)

Liquid

ΔG

0

R^*

R

Crystal nucleus

R

surface

bulk

$$\Delta G = 4\pi R^2 \gamma - \frac{4}{3}\pi R^3 \rho \Delta\mu_{ls}$$

–How does the crystal form?

–What is the structure of the critical nucleus

–Is classical nucleation theory correct?

   •What is the barrier?

   •Rate constant

γ : surface tension

$\Delta\mu$: chem. pot difference

ρ: density

Path sampling of nucleation

TIS in NPH ensemble, as density and temperature change
N=10000, P=5.68 H=1.41 (25 % undercooling)

D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, *Phys. Rev. Lett.* **94**, 235703 (2005)
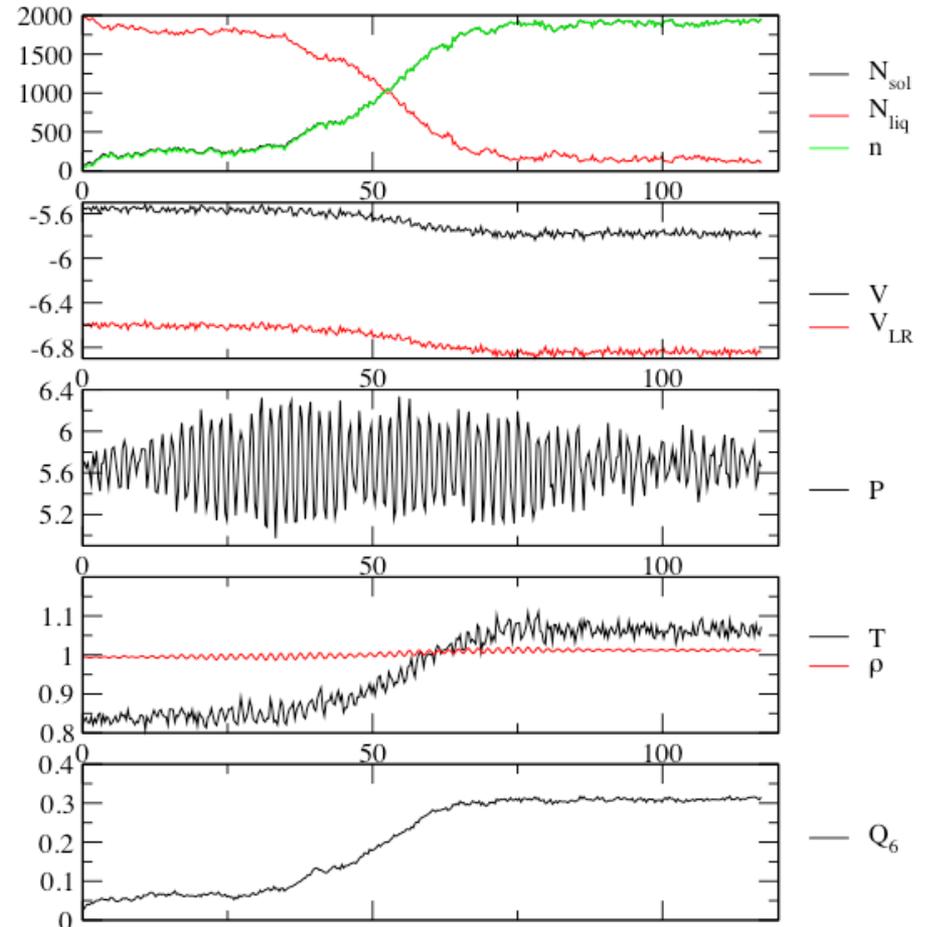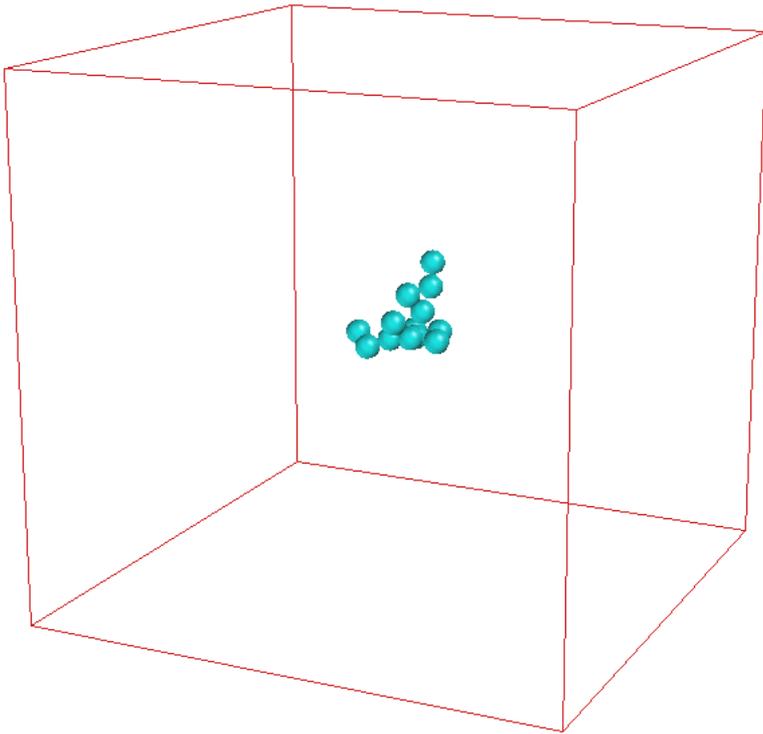
Sampling paths is only the beginning

- *Eugene Wigner:* **"It is nice to know that the computer understands the problem. But I would like to understand it too.** ”

- Path ensemble needs to be further explored to obtain:
  - Rate constants
  - Free energy
  - Transition state ensembles
  - Mechanistic picture
  - Reaction coordinate

- Illustrative example: crystal nucleation

Transition interface sampling
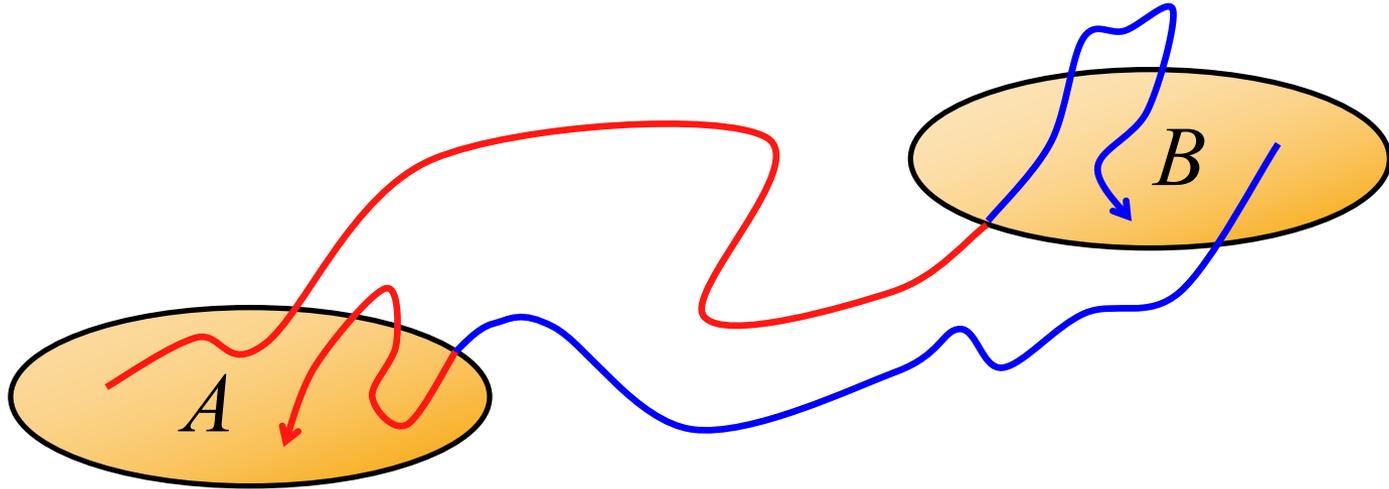
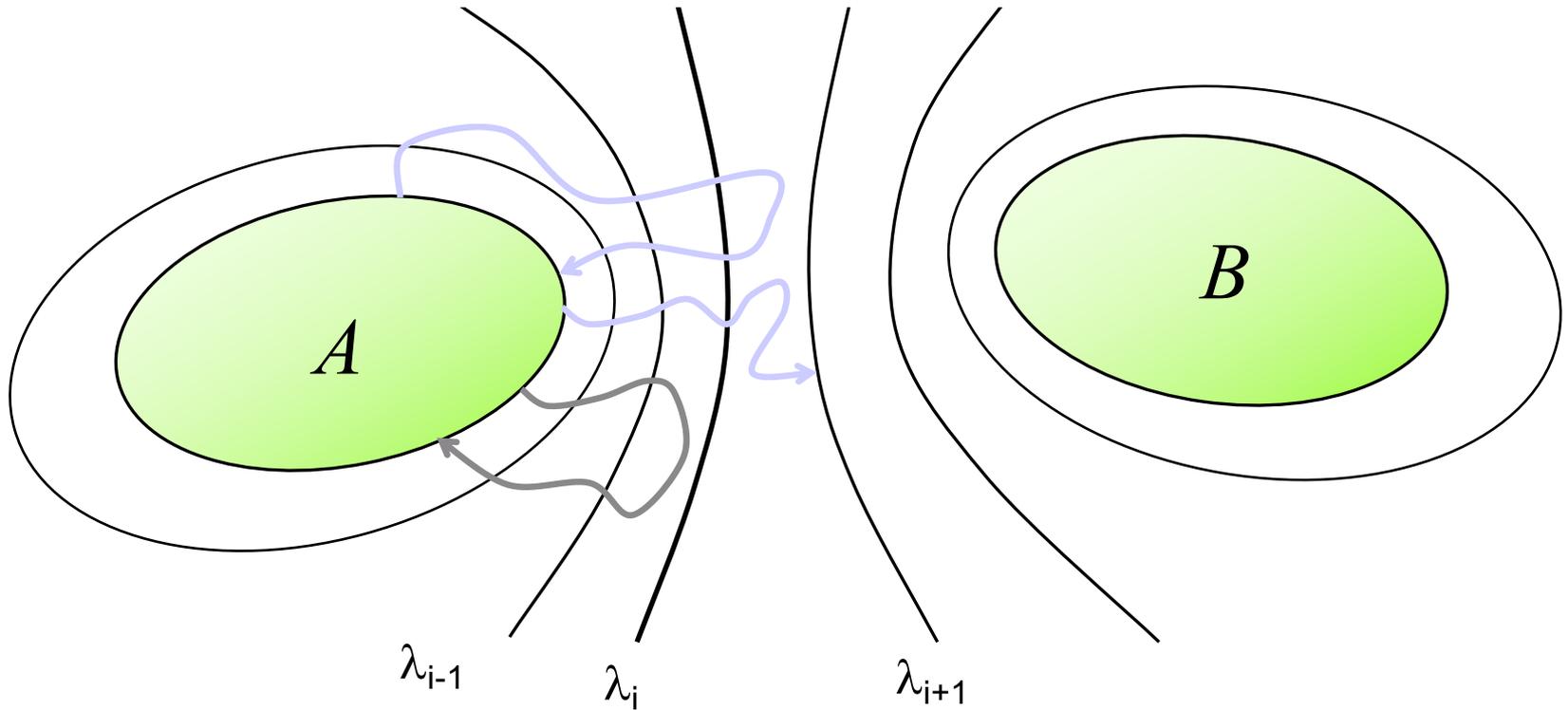**Overall states** in phase space:

$\mathcal{A}$ directly coming from **A**

$\mathcal{B}$ directly coming from **B**

$$\mathcal{C}(t) \equiv \frac{\langle h_{\mathcal{A}}(x_0) h_{\mathcal{B}}(x_t) \rangle}{\langle h_{\mathcal{A}} \rangle}$$

$$k_{AB} = \frac{\langle h_{\mathcal{A}}(x_0) \dot{h}_{\mathcal{B}}(x_0) \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle}$$

T. S. van Erp, D. Moroni and P. G. Bolhuis, *J. Chem. Phys.* **118** , 7762 (2003)
T. S. van Erp and P. G. Bolhuis, *J. Comp. Phys.* **205**, 157 (2005)

$P_A(\lambda_{i+1} \mid \lambda_i)$= probability that path crossing i for first time after leaving A reaches i+1 before A

$$k_{AB}^{TIS} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle} \prod_{i=1}^{n-1} P_A(\lambda_{i+1} \mid \lambda_i) = \Phi_A \prod_{i=1}^{n-1} P_A(\lambda_{i+1} \mid \lambda_i)$$
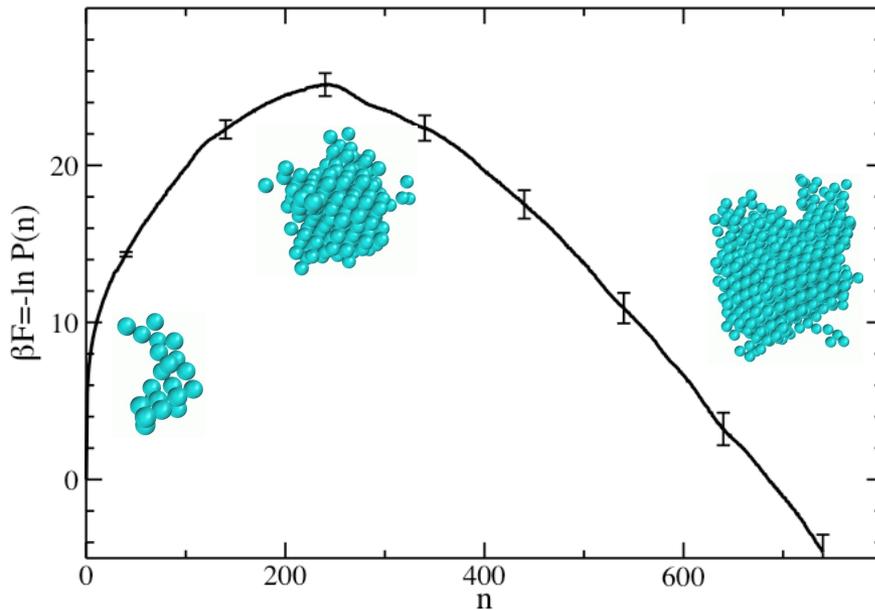
TIS results for nucleation
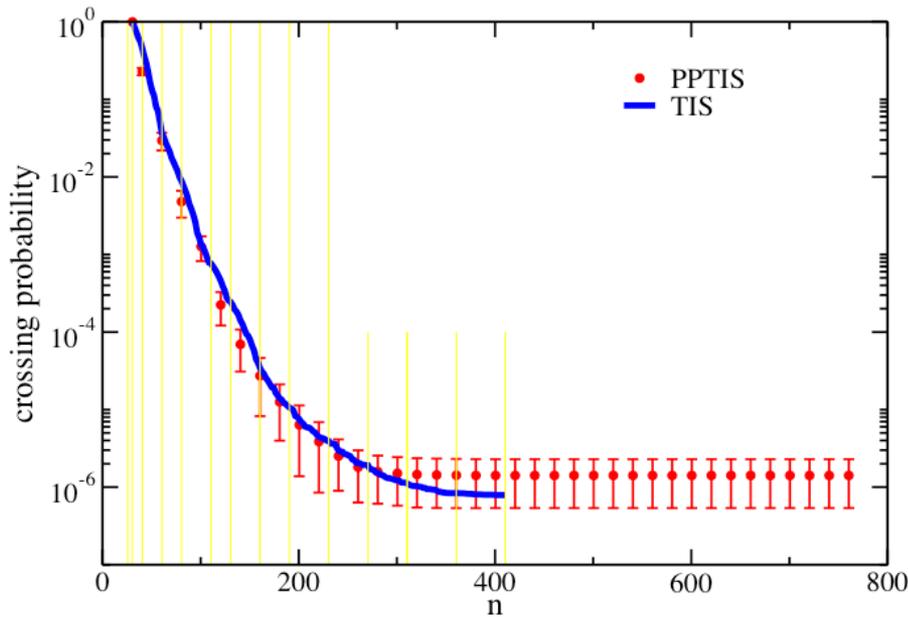
$$\mathcal{P}_A(B|1) = \prod_{i=1}^{n-1} \mathcal{P}_A(i+1|i) = 8 \times 10^{-7}$$

$$\frac{\langle \phi_{A1} \rangle}{\langle h_{\mathcal{A}} \rangle} = 1.29$$

$$k_{AB} = (1.0 \pm 0.8) \times 10^{-6}$$

Free energy follows directly
Moroni, van Erp, Bolhuis, PRE, 2005

Structural analysis?

Committor
(aka p-fold, splitting probability)

$p_B(r, t) =$ Probability that a trajectory initiated at r relaxes into B
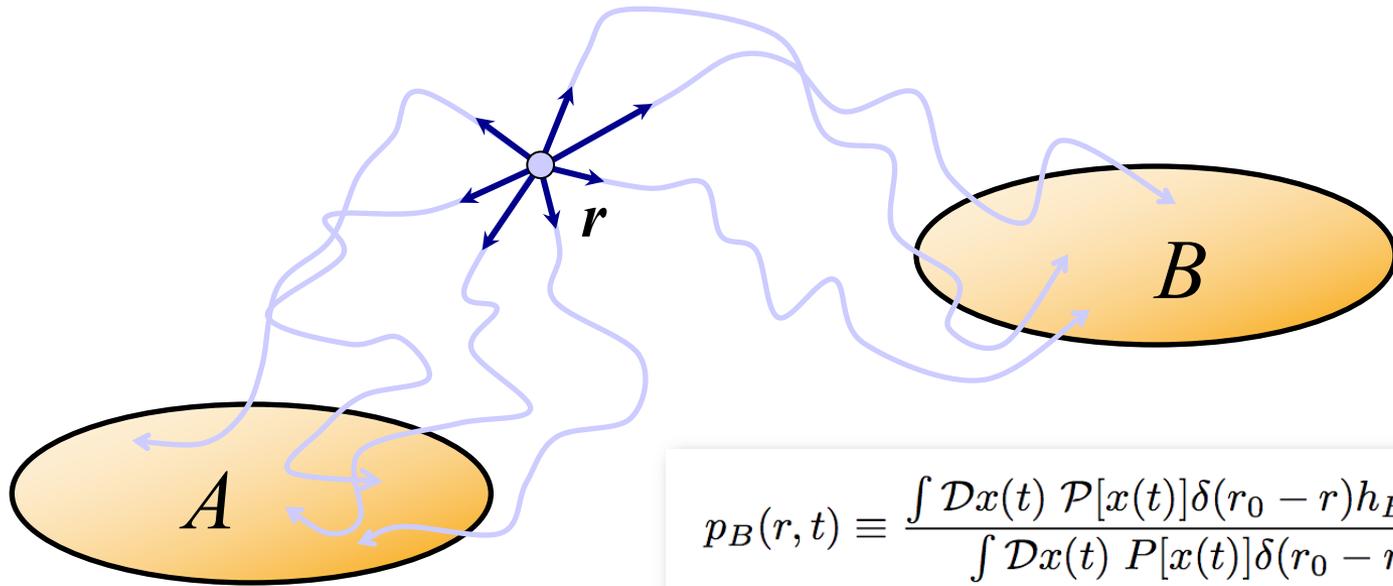
$$p_B(r, t) \equiv \frac{\int \mathcal{D}x(t) \, \mathcal{P}[x(t)]\delta(r_0 - r)h_B(x_t)}{\int \mathcal{D}x(t) \, P[x(t)]\delta(r_0 - r)}$$

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

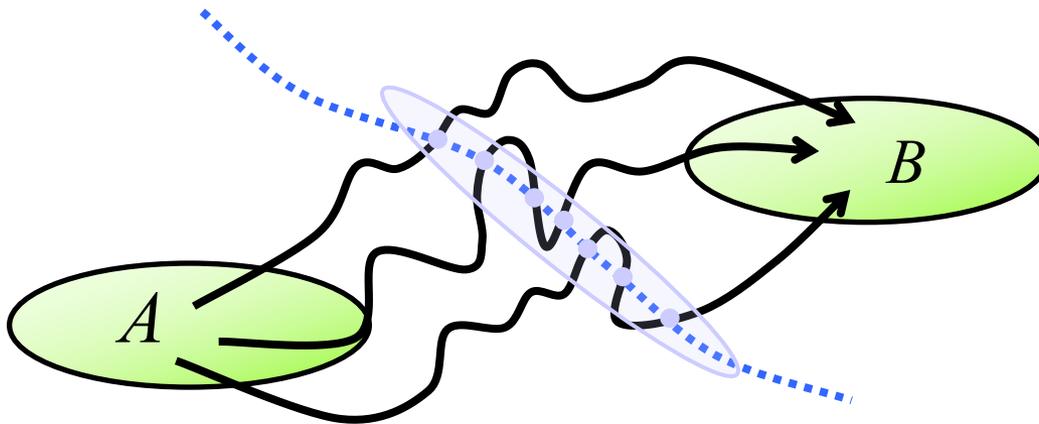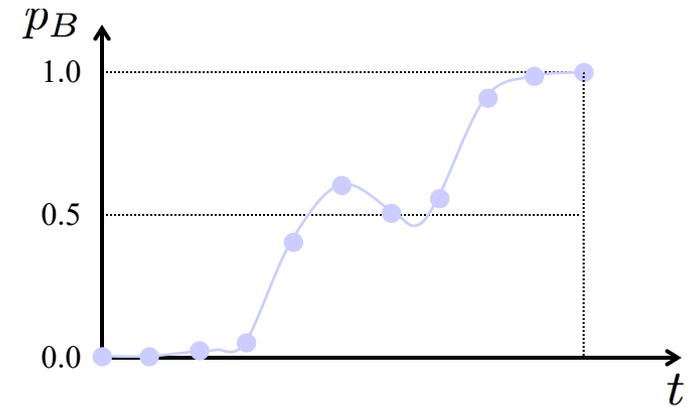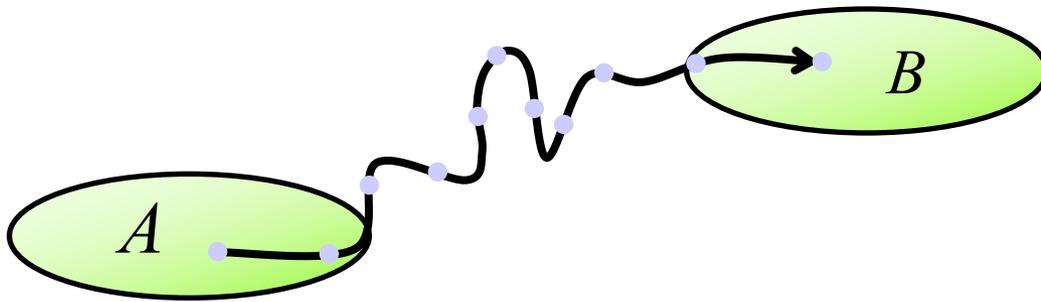$$\sigma = \sqrt{\langle (p_B^{(N)} - p_B)^2 \rangle} = \sqrt{\frac{p_B(1 - p_B)}{N}}$$

L. Onsager, *Phys. Rev.* **54**, 554 (1938).
M. M. Klosek, B. J. Matkowsky, Z. Schuss, *Ber. Bunsenges. Phys. Chem.* **95**, 331 (1991)
V. Pande, A. Y. Grosberg, T. Tanaka, E. I. Shaknovich, *J. Chem. Phys.* **108**, 334 (1998)
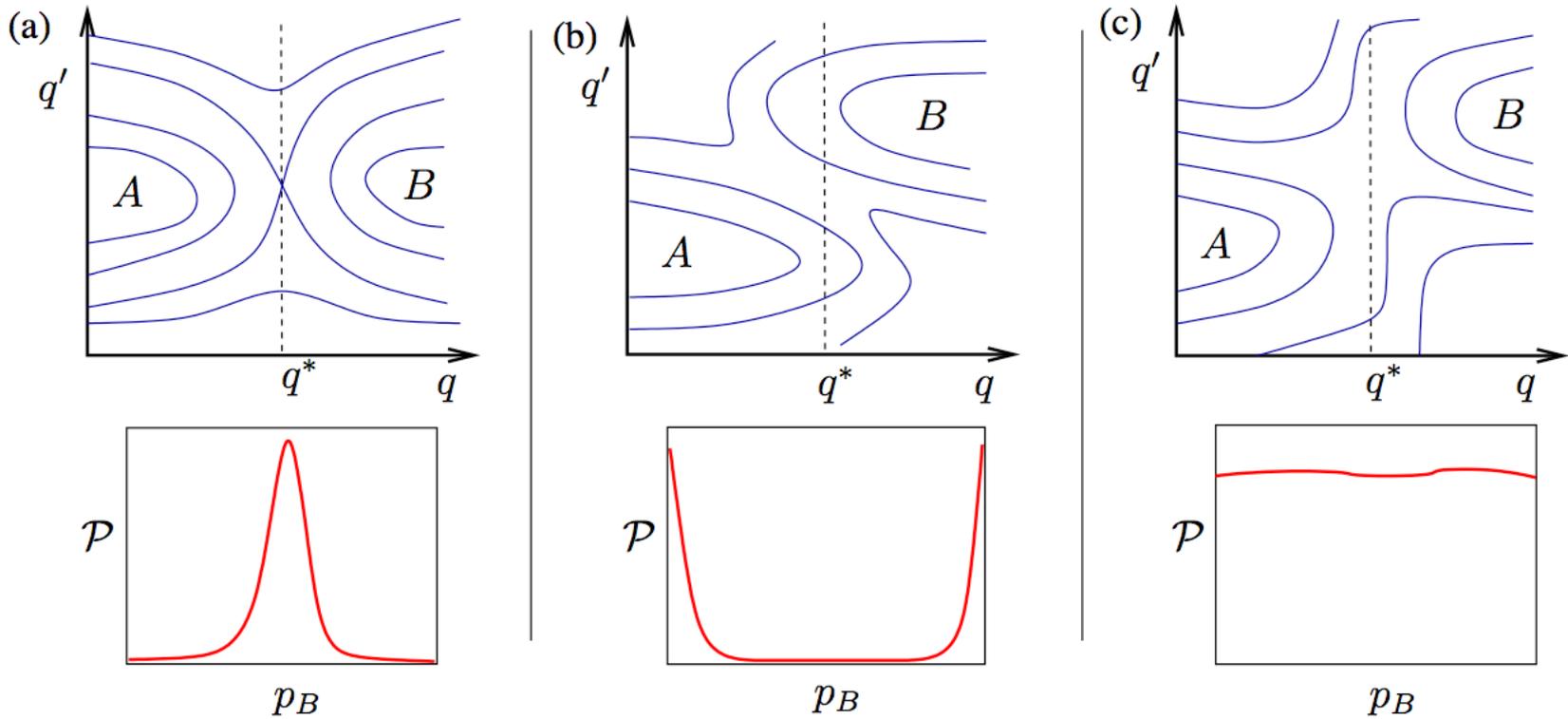
Transition state ensemble

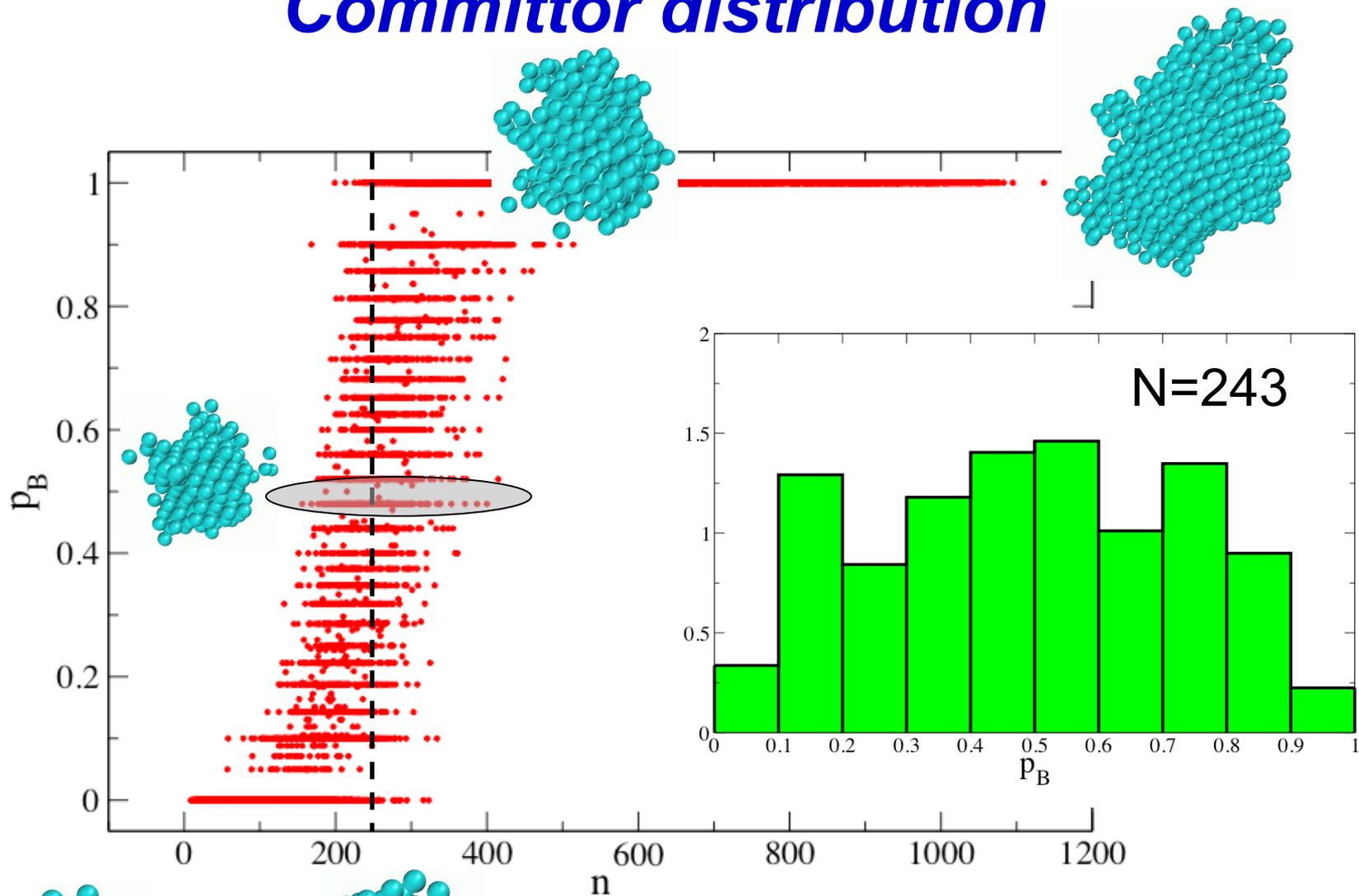r is a **transition state** (TS) if $p_B(r) = p_A(r) = 0.5$

TSE:

Intersections of transition pathways with the $p_B = 1/2$ surface
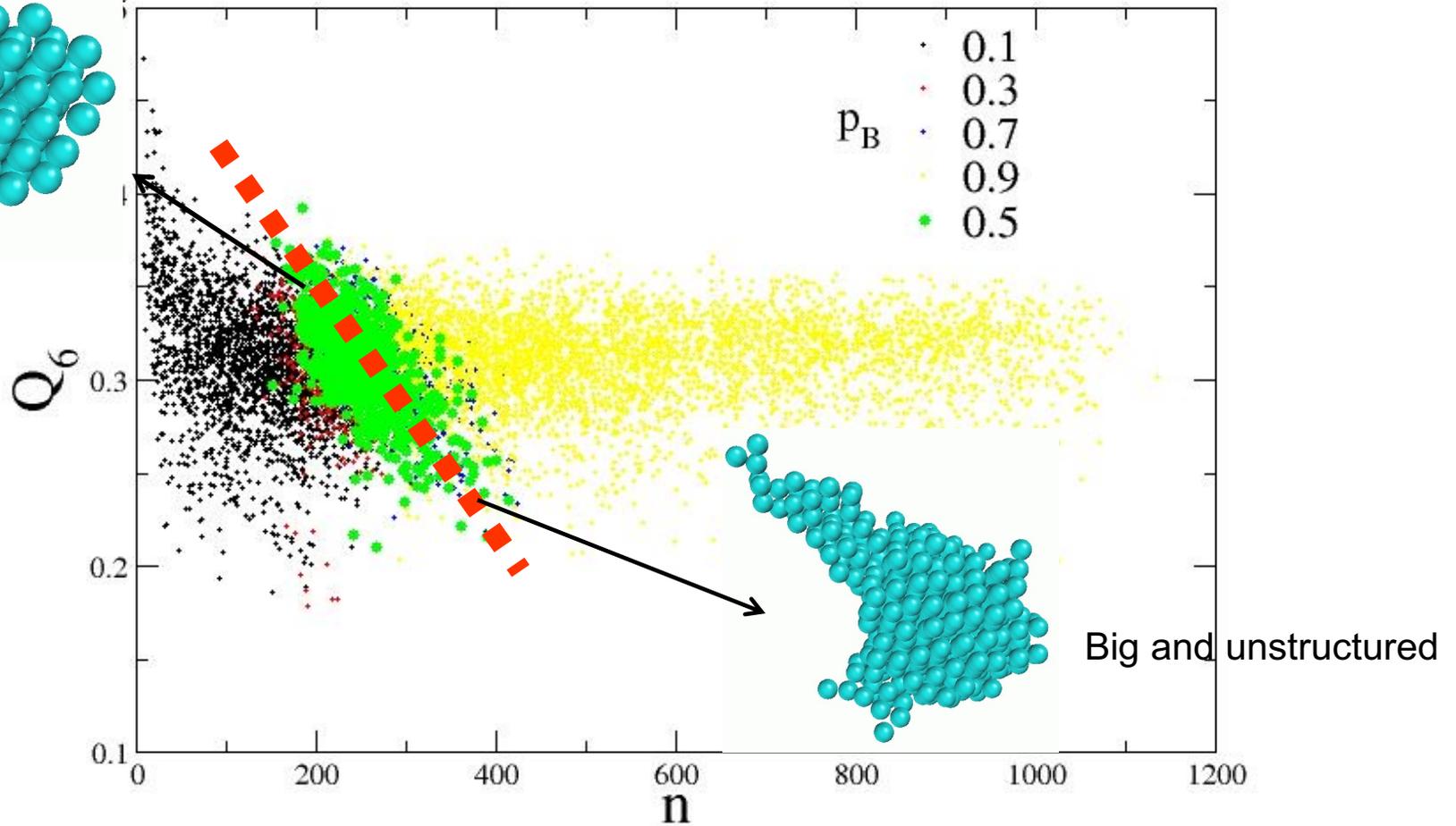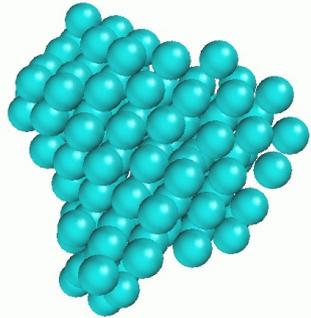
Committor distributions

Committor distribution

N=243

Clearly, n is not entire story

Structure

Small and structured

Big and unstructured

Committor analysis gives valuable insight

The OpenPathSampling package

- a python toolkit to run path sampling algorithms
    - works with OpenMM and simple dynamics
    - Gromacs, Lammps support
    - uses MdTraj, OpenMM, MSMbuilder
- OPS allows flexible definition of
    - states
    - trajectory ensembles
    - sets of interfaces
    - networks of transition
- OPS provides algorithms for sampling
    - TPS
    - TIS (see also T. van Erp's presentation)
    - MSTIS
    - RETIS (SRTIS)
- OPS provides analysis tools
    - crossing probabilities
    - rates, free energies, path densities
    - ….

OpenPathSampling

www.openpathsampling.org
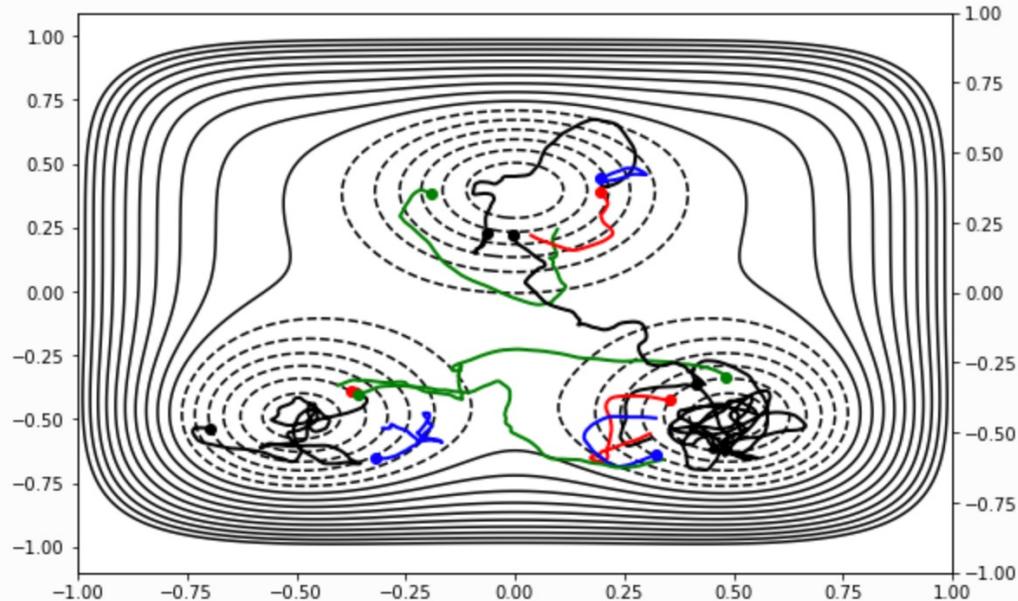
```
DONE! Completed 100 Monte Carlo cycles.
```

In RETIS, there are several different move types (shooting, replica exchange, etc.), and each move type can have a different probability of being selected. Moreover, different move types may have different numbers of specific moves (ensembles affected) within them.

This means that if you wanted to run enough steps that, on average, each shooting mover ran 1000 times, you would need to figure out how many trials (of any type) that corresponds to. OPS has tools to make that easy. First, we select a mover (we'll take the first shooting mover as a representative of all shooting movers), and then we use `scheme.n_steps_for_trial` to get the expected number of steps to get that many trials of that mover. Of course, this is only a expected value; the exact number of trials in the simulation will vary.

```
[25]: representative_mover = scheme.movers['shooting'][0]
      n_steps = int(scheme.n_steps_for_trials(representative_mover, 1000))
      print(n_steps)

      20100
```

Finally, let's run for a lot longer to get enough statistics. Note that this time, we'll run the simulation using `run_until` which picks up from where we left off, and finishes when a total of `n_steps`

Path sampling of clathrate nucleation

2944 TIP4P/ice + 512 CH4
NPT 500 bar 280 K,
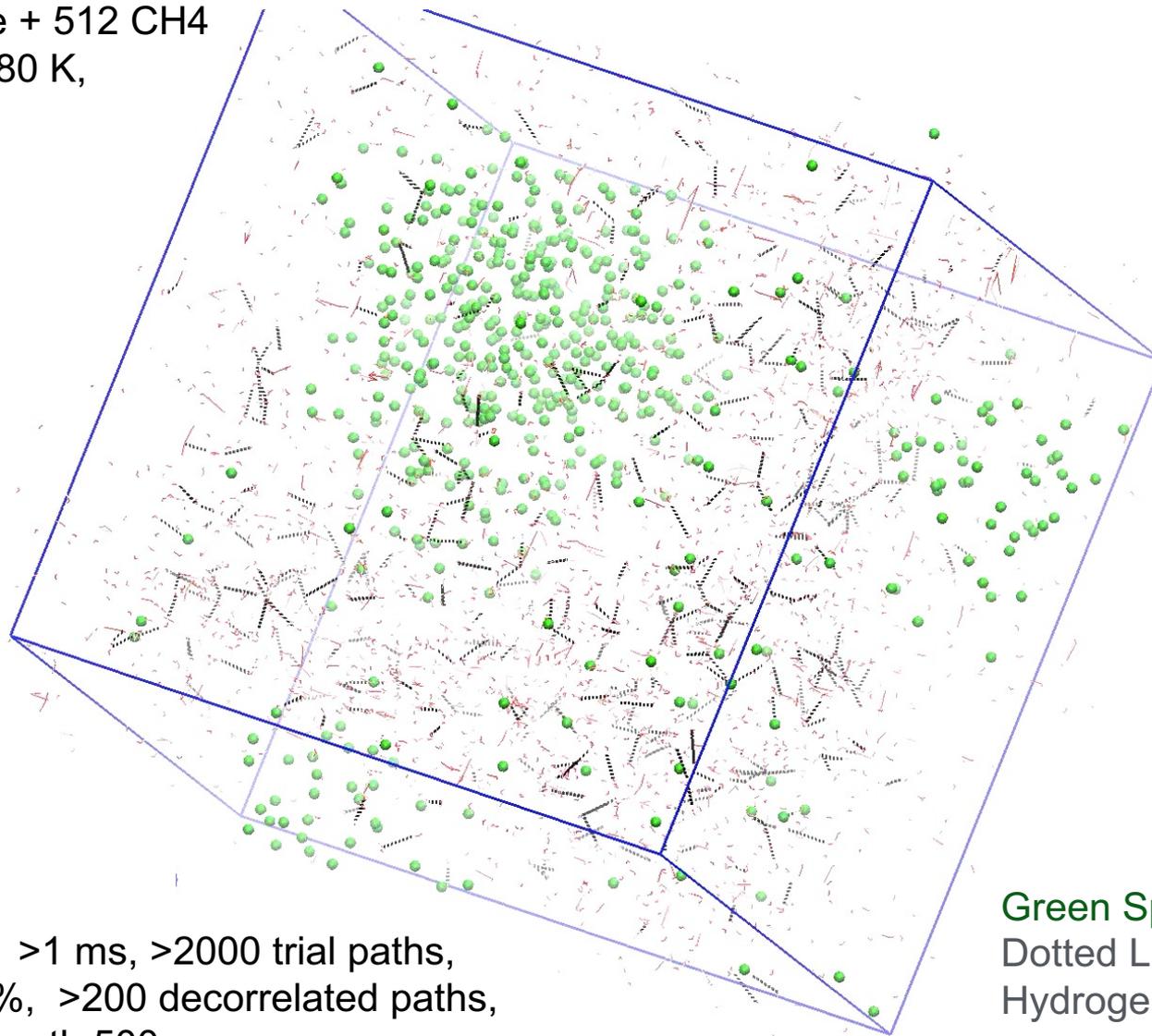
Green Spheres – Methane
Dotted Lines – Water
Hydrogen Bonds

simulation time >1 ms, >2000 trial paths,
acceptance 33%, >200 decorrelated paths,
average path length 500 ns
induction time > 30 kyears

Arjun, Berendsen, PGB, PNAS, 2019

The end

"Mr. Osborne, may I be excused? My brain is full."