Optimal reaction coordinates, kinetics, and free energy of rare events by reweighted path sampling

Peter Bolhuis
van 't Hoff institute for Molecular Sciences
University of Amsterdam, The Netherlands

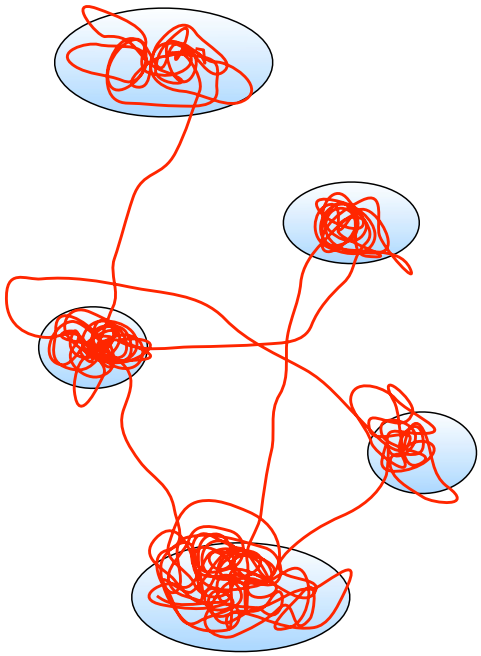
ACMM Amsterdam Center for Multiscale Modeling

Outline

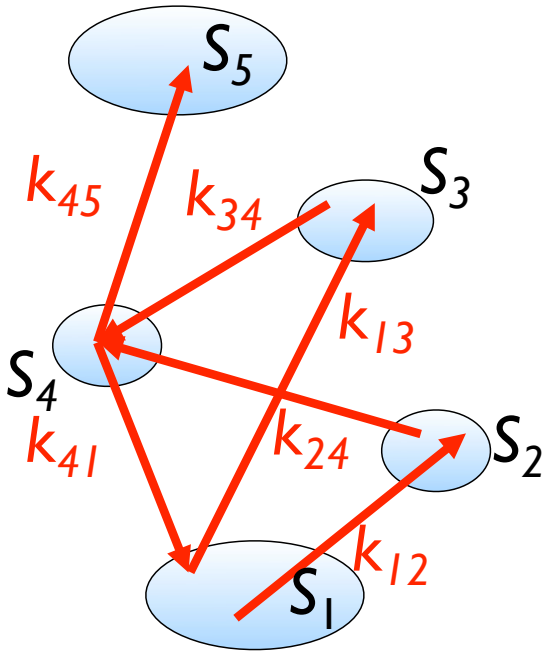
- Introduction
- Path sampling
- Rates with transition interface sampling
- Linear and non-linear reaction coordinate analysis
- Application: crystal nucleation
- Conclusions

Markov approximation

molecular dynamics trajectory



coarse grained trajectory



integrate equations of motion

time step Δt ≈ fs

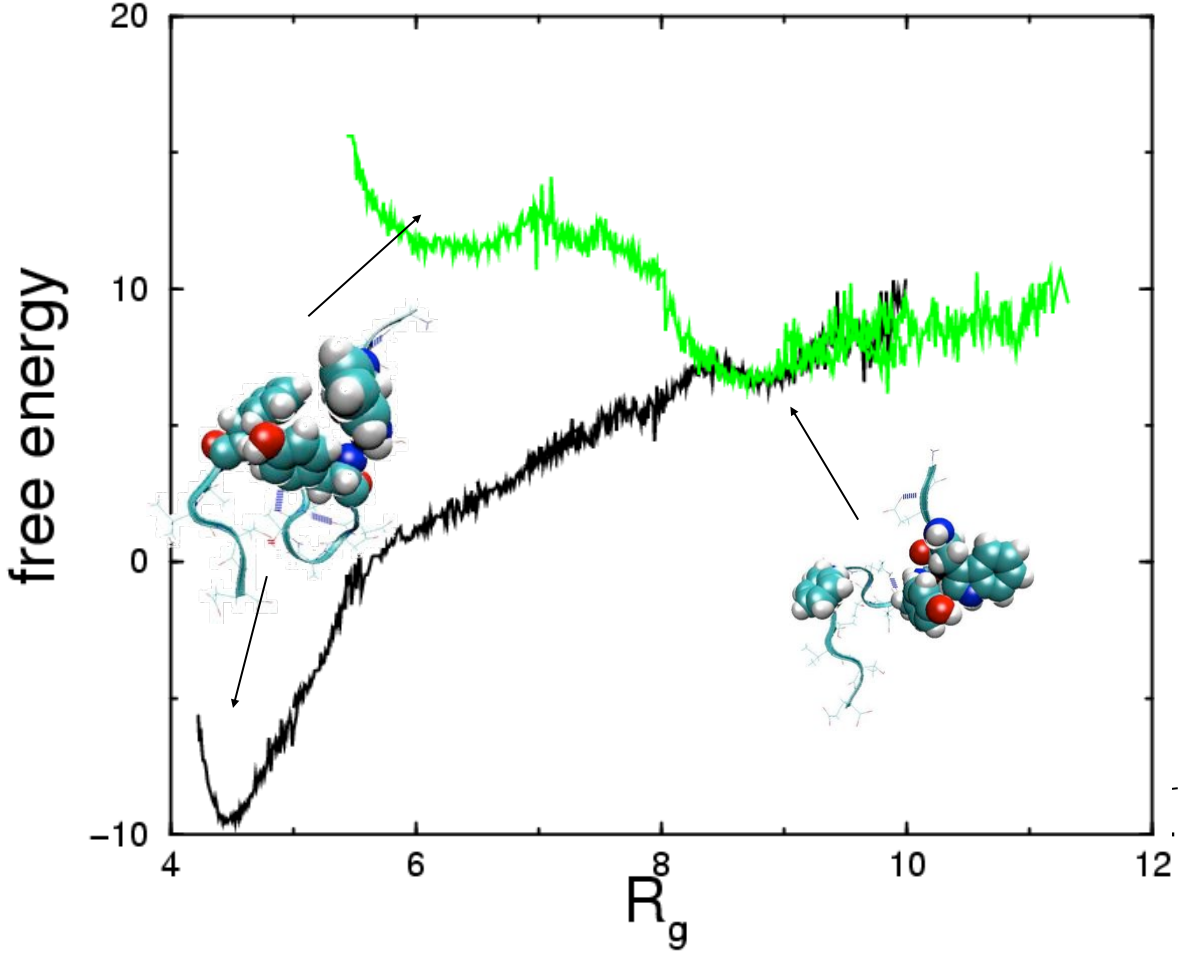
$$\frac{P(S_j,t)}{dt} = -\sum_i k_{ji} P(S_j,t) + \sum_i k_{ij} P(S_i,t)$$

master equation,
solve e.g. by kinetic Monte Carlo
time step set by rates

Biased sampling requires reaction coordinate

Objectives: free energy barrier, rates, transition states and mechanism.

But if reaction coordinate is not correct, all these might be wrong!



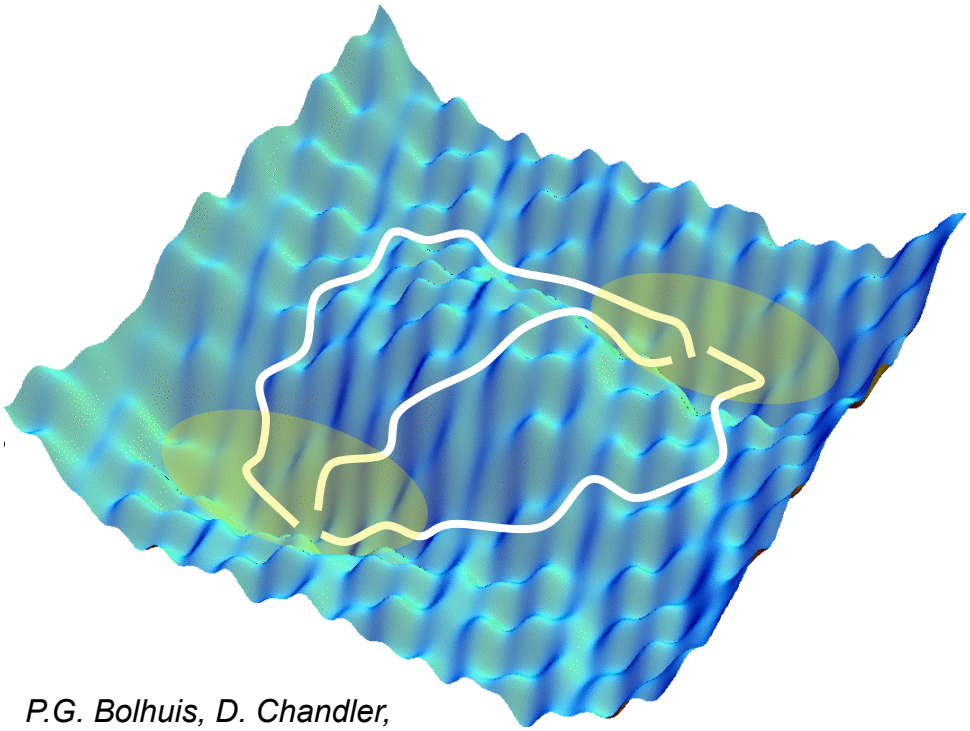
$$\beta W(q) = -\ln \int dq' e^{-\beta F(q,q')}$$

Need for methods that create pathways without prior knowledge of the RC:

Transition path sampling

Transition path sampling

Importance sampling of the rare event path ensemble:
all dynamical trajectories that lead over (high) barrier and connect stable states.



P.G. Bolhuis, D. Chandler,
C. Dellago, P.L. Geissler

Annu. Rev. Phys. Chem 2002

Why TPS?

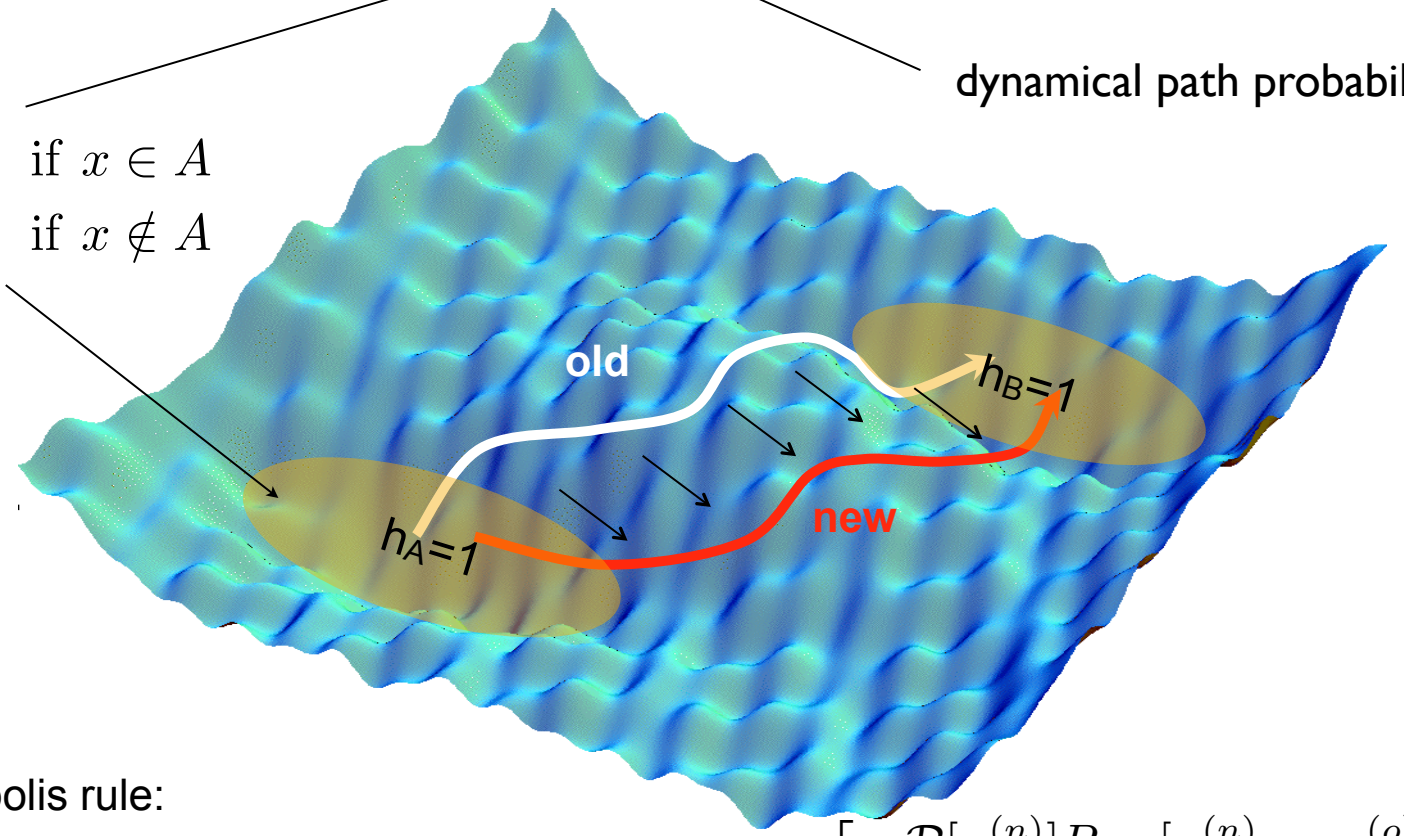
-selects unbiased rare paths
-no reaction coordinate needed
-RC follows from committor analysis
-rate constant

Importance sampling of paths

$$\mathcal{P}_{\mathcal{AB}}[\mathbf{x}(L)] = h_A(x_0)\mathcal{P}[\mathbf{x}(L)]h_B(x_L)/Z_{AB}(L)$$

dynamical path probability

$$h_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$



Metropolis rule:

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

Flexible shooting algorithm

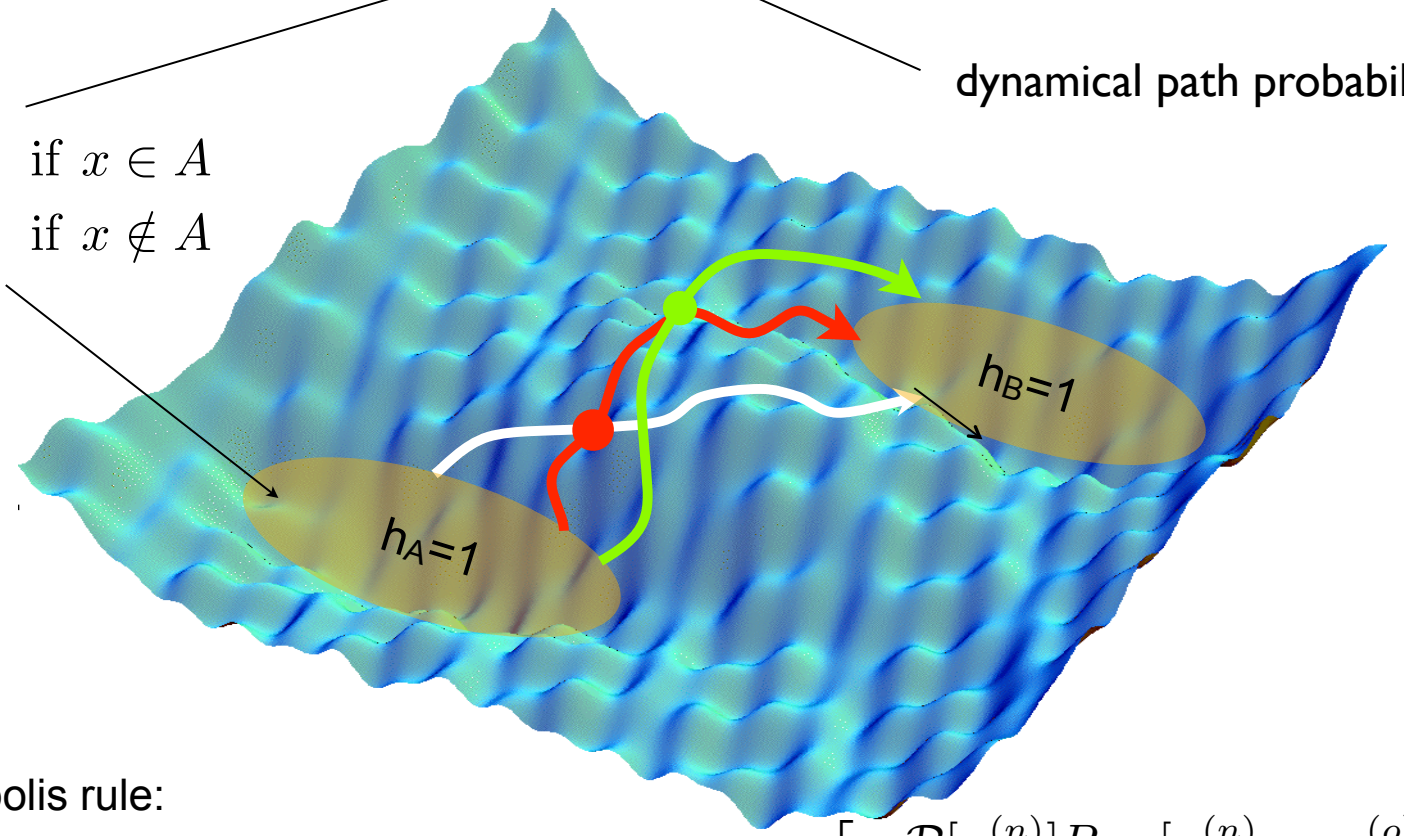
$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)})\min\left(1, \frac{L^{(o)}}{L^{(n)}}\right)$$

Importance sampling of paths

$$\mathcal{P}_{\mathcal{AB}}[\mathbf{x}(L)] = h_A(x_0)\mathcal{P}[\mathbf{x}(L)]h_B(x_L)/Z_{AB}(L)$$

dynamical path probability

$$h_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$



Metropolis rule:

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

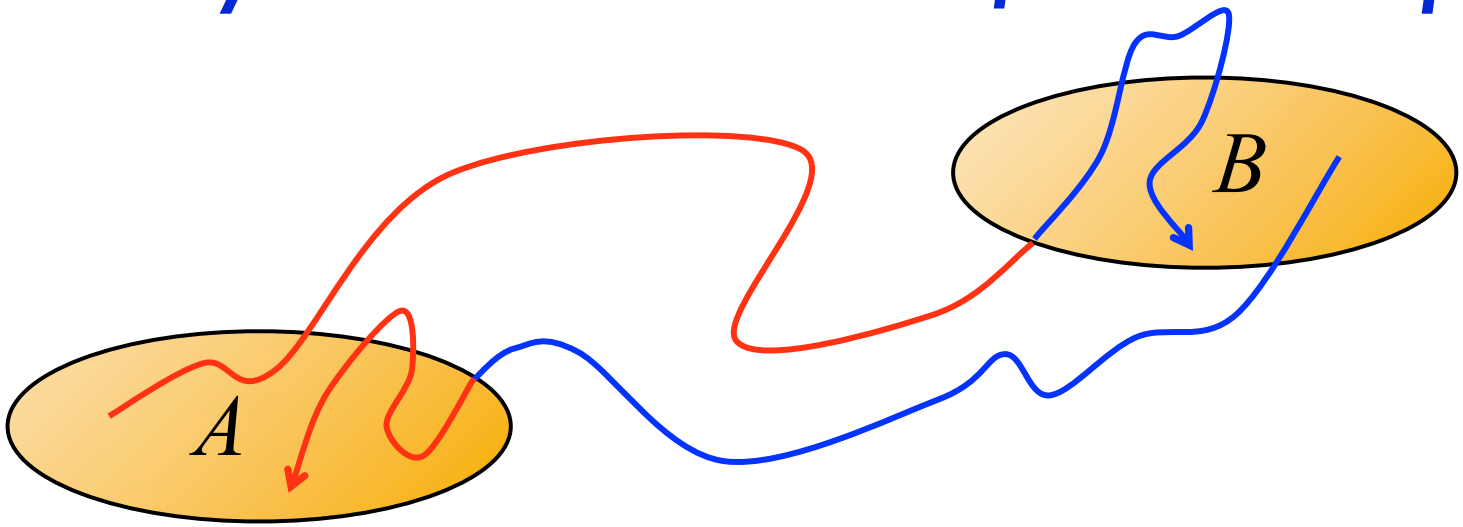
Flexible shooting algorithm

$$P_{acc}[\mathbf{x}^{(o)} \to \mathbf{x}^{(n)}] = h_A(x_0^{(n)})h_B(x_L^{(n)})\min\left(1, \frac{L^{(o)}}{L^{(n)}}\right)$$

Outline

- Introduction
- Path sampling
- **Rates with transition interface sampling**
- Linear and non-linear reaction coordinate analysis
- Application: crystal nucleation
- Conclusions

Rates by transition interface sampling



Overall states in phase space:

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

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

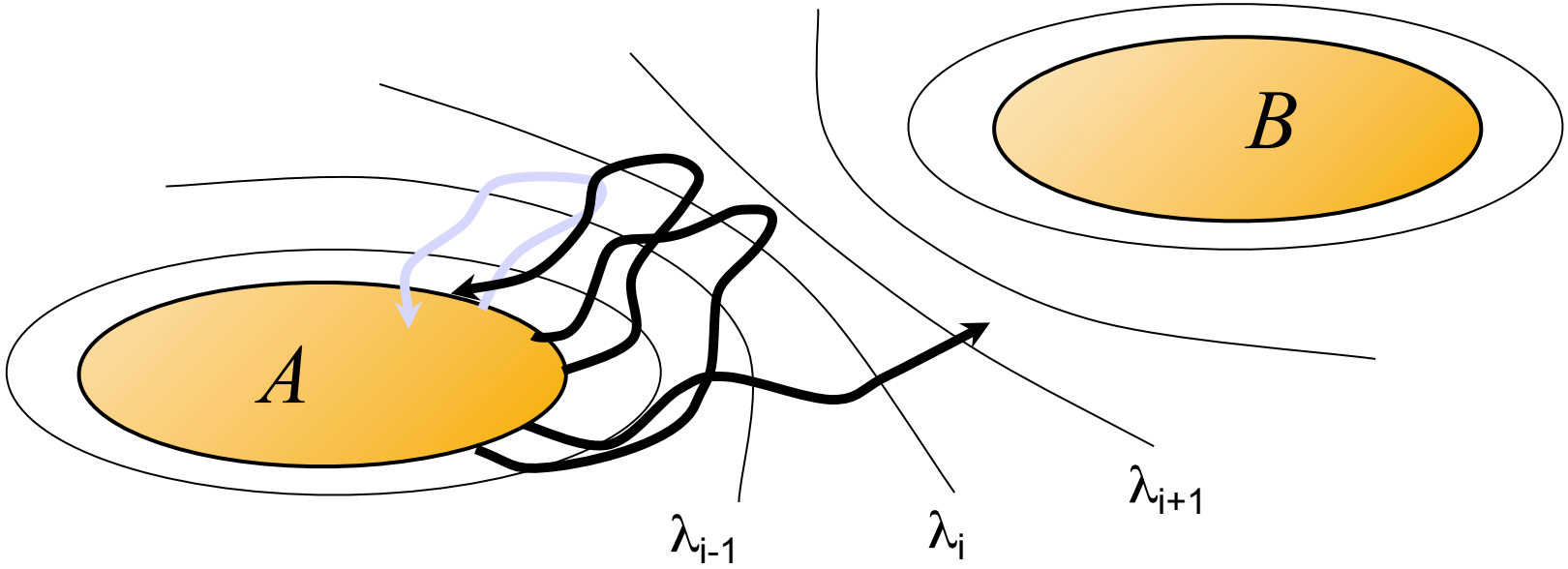
$$C(t) \equiv \frac{\langle h_{\mathcal{A}}(x_0) h_{\mathcal{B}}(x_t) \rangle}{\langle h_{\mathcal{A}} \rangle}$$

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

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)

Rates by transition interface sampling



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

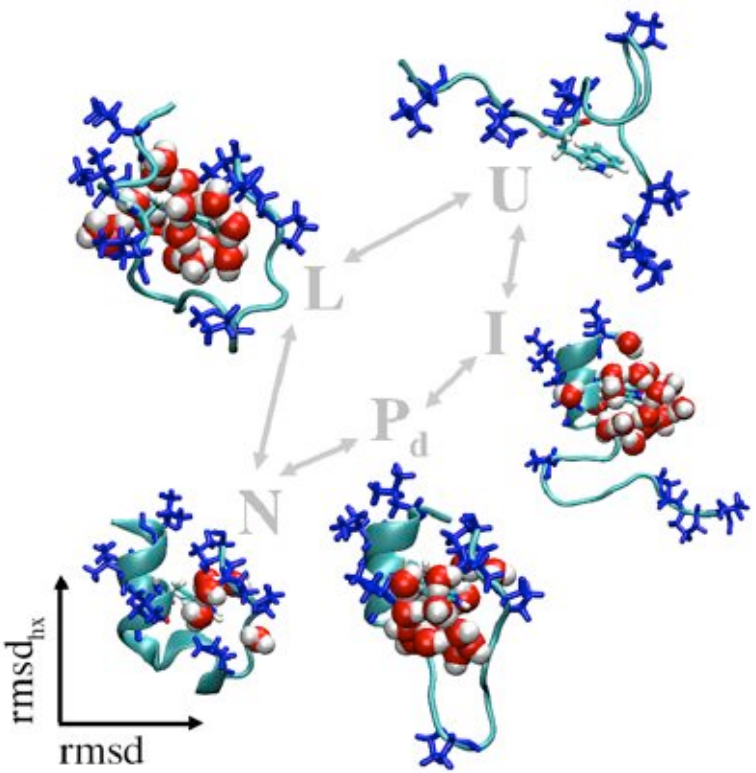
$$k_{AB} = \frac{\langle \phi_{AB} \rangle}{\langle h_{\mathcal{A}} \rangle} = \frac{\langle \phi_A \rangle}{\langle h_{\mathcal{A}} \rangle} P_A(\lambda_B | \lambda_A) = \frac{\langle \phi_A \rangle}{\langle h_{\mathcal{A}} \rangle} \prod_{i=1}^{n-1} P_A(\lambda_{i+1} | \lambda_i)$$

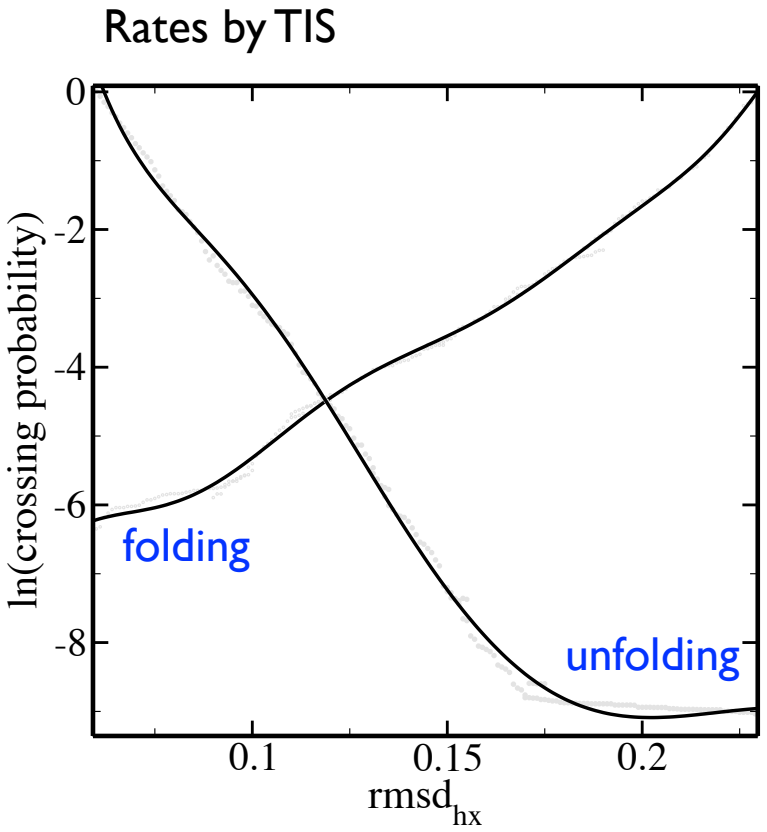
flux $\frac{\langle \phi_A \rangle}{\langle h_A \rangle} = \frac{1}{\Delta t} \frac{N_c^+}{N_{\mathrm{MD}}}$

Also the basis of FFS (ten Wolde et al.)

Sample paths with MC
- -flexible shooting
- -time reversal moves for AA paths
- -interface replica exchange moves
- -first interface exploration

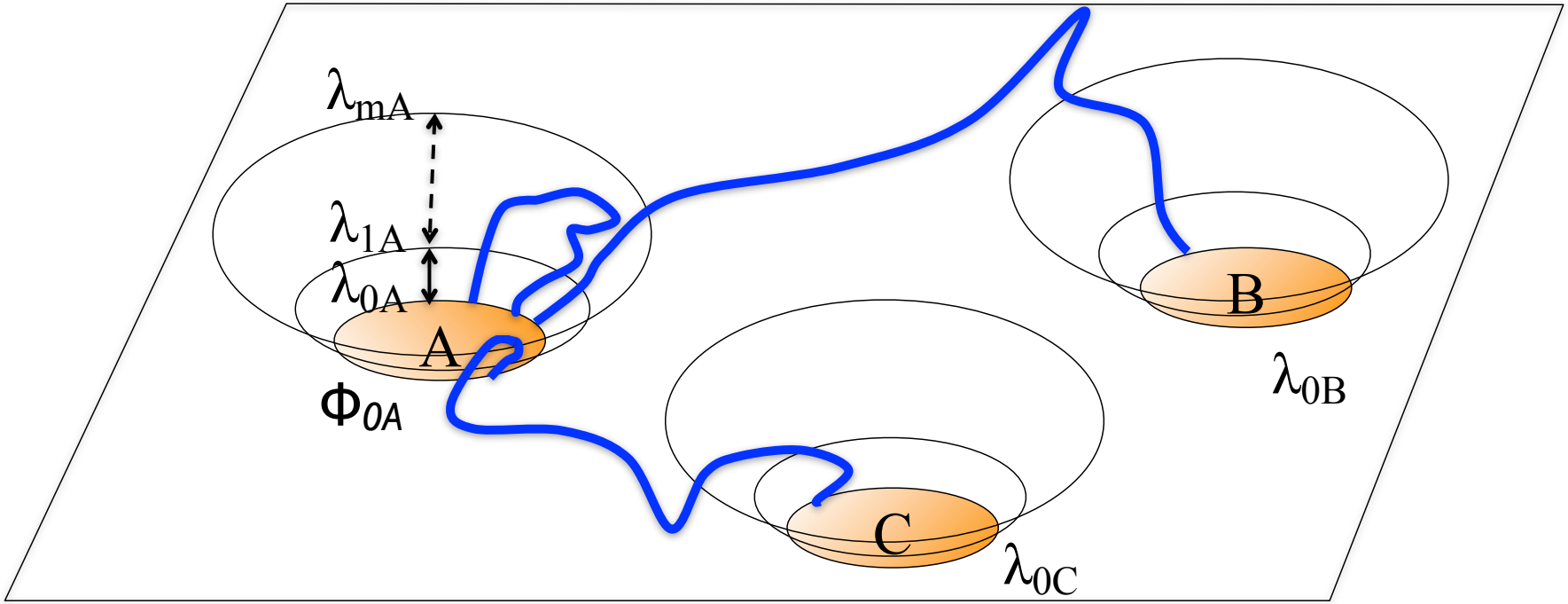
Application to Trp-cage folding





J. Juraszek, PGB PNAS 2006
Biophys. J. 95 4246 (2008)

Markov models by multiple state TIS



$$k_{Ai} = \underbrace{\frac{\langle \phi_{\lambda_{m_A}} \rangle}{\langle h_A \rangle}} \cdot \underbrace{P_A(\lambda_{0_i} | \lambda_{m_A})}$$

TIS:

$$\frac{\langle \phi_A \rangle}{\langle h_A \rangle} \prod_{s=0}^{m-1} P_A(\lambda_{(s+1)_A} | \lambda_{s_A})$$

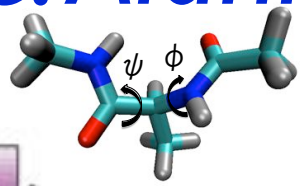
MSTIS:

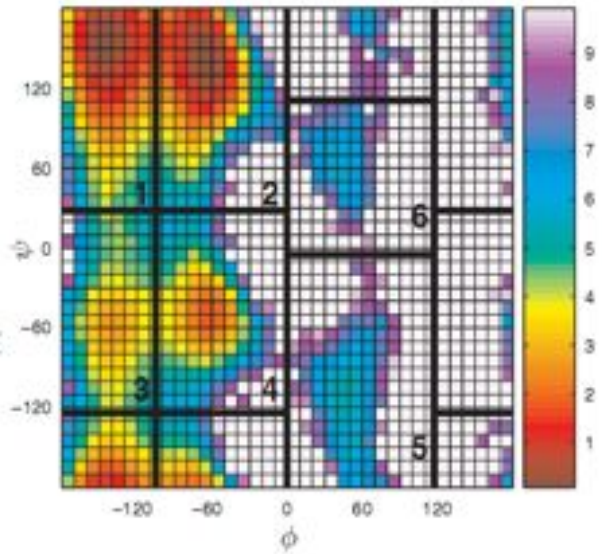
$$\frac{\text{no. of pathways coming from A, cross } \lambda_{mA}, \text{ end i}}{\text{no. of pathways coming from A, cross } \lambda_{mA}}$$

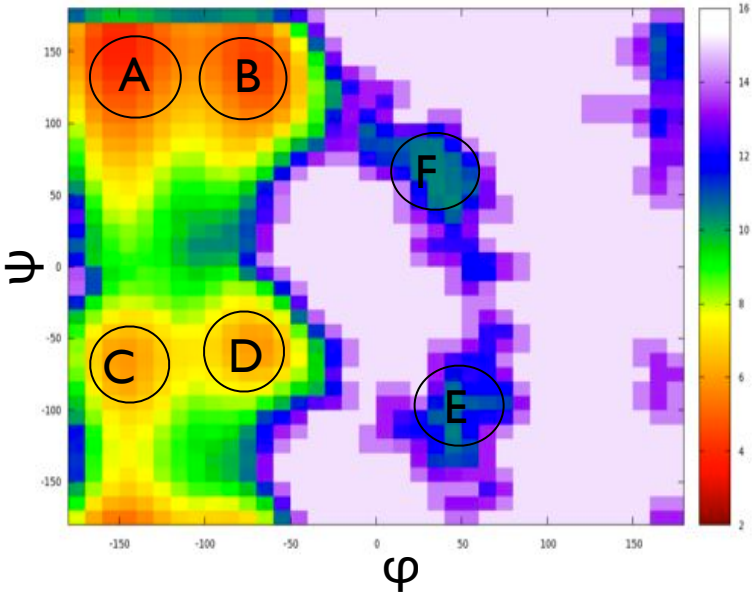
rates can be used in markov state model

J. Rogal, PGB, J. Chem. Phys. (2008).
J. Rogal, PGB, J. Chem. Phys. (2010).

Example: Alanine dipeptide

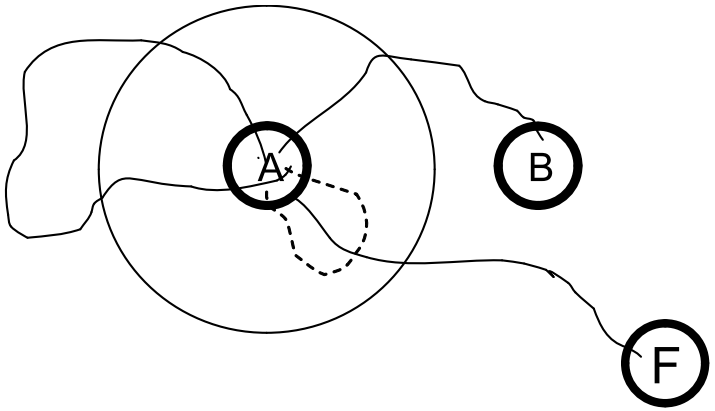


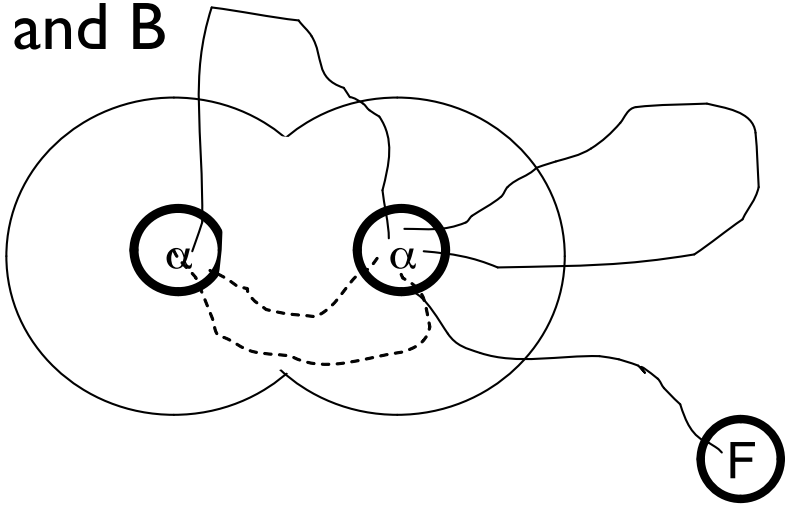




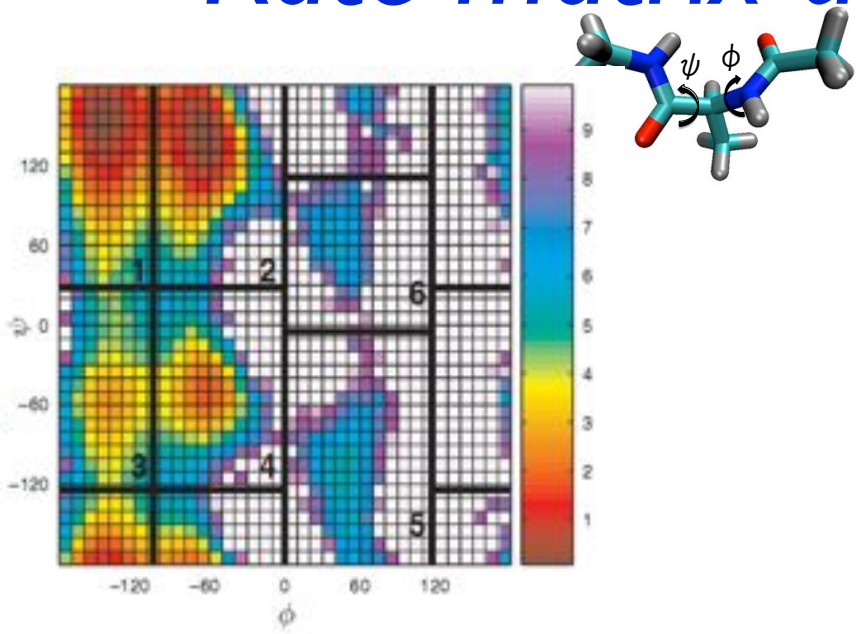
Chodera, Multiscal Model. Simul. 5, 1214-1226 (2009).

Combination of A and B

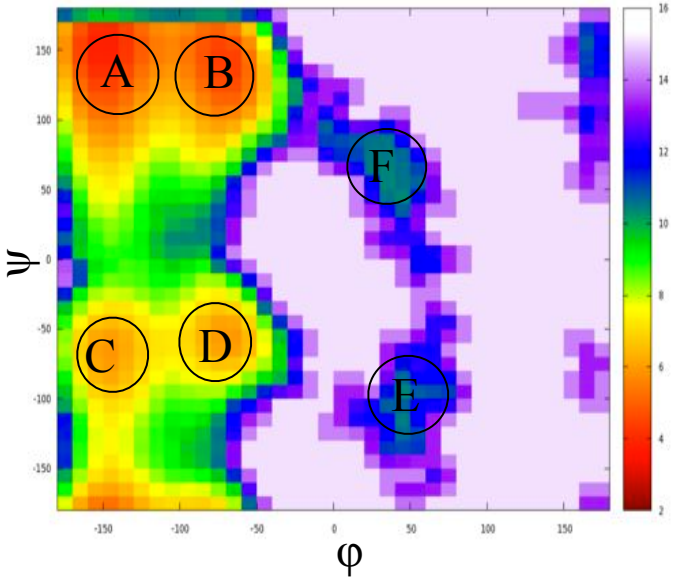




Rate matrix using MSTIS



Chodera, Multiscal Model. Simul. 5, 1214-1226 (2009).



Du, Marino & PGB, JCP 135, 145102 (2011).

From

To

	1 and 2	3 and 4	5	6
1 and 2		0.0335	0.0011	0.073
3 and 4	0.0046		0.018	0
5	0	0.0001		0.023
6	0.0001	0	0.011	

From

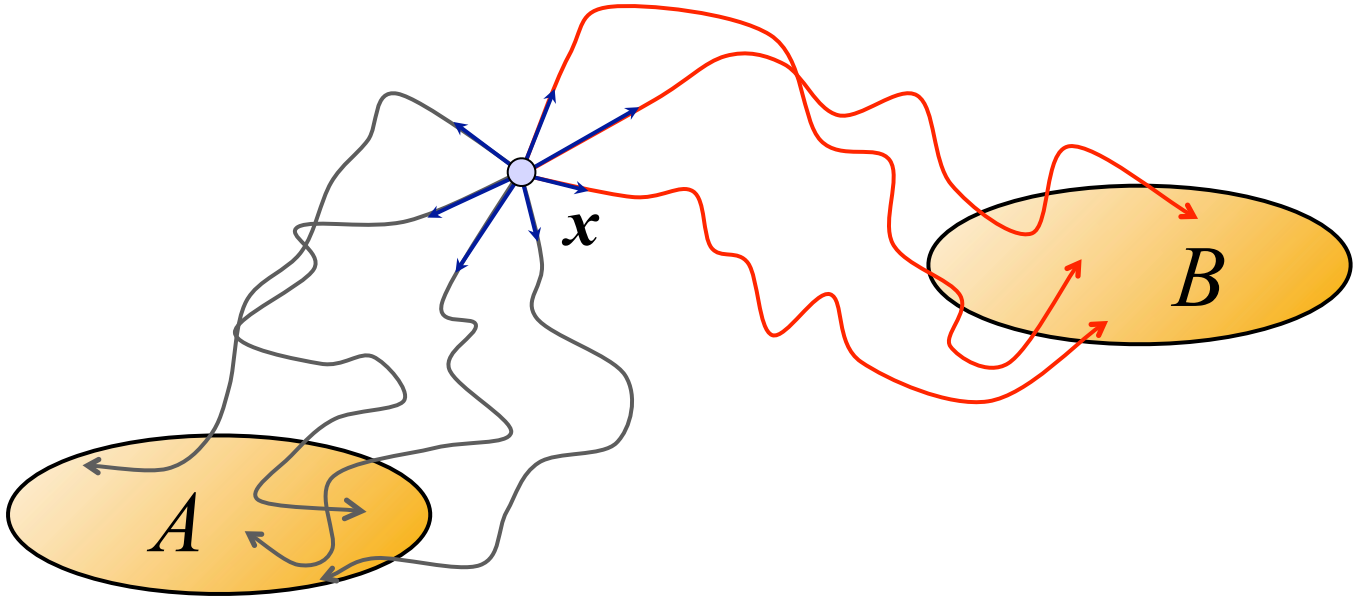
	A & B	C & D	E	F
A & B		0.035	0.0011	0.038
C & D	0.0037		0.017	0.0002
E	0.000006	0.0001		0.01
F	0.0002	0.000004	0.008	

current work: application to Trp cage folding

Outline

- Introduction
- Path sampling
- Rates with transition interface sampling
- **Linear and non-linear reaction coordinate analysis**
- Application: crystal nucleation
- Conclusions

Committor Analysis



$$p_B(x) \cong \frac{1}{N_s}\sum_s h_B \quad \text{where} \quad h_B = \begin{cases} 1 \; if \; t \rightarrow B \\ 0 \; if \; t \rightarrow A \end{cases}$$

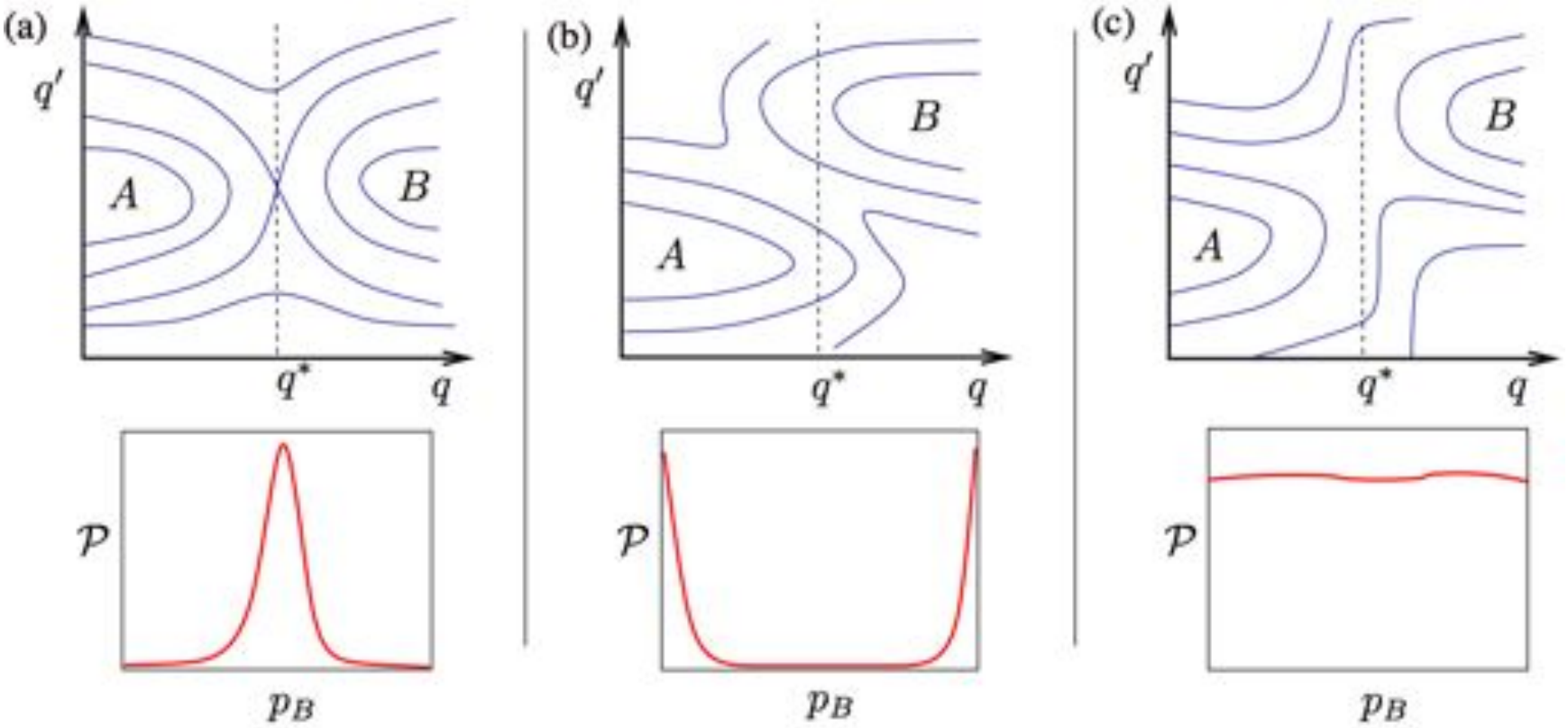
x is a **transition state** if $p_B(x) = 0.5$

pB(x) is **the** reaction coordinate

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) W.E, E. Vanden-Eijnden, J. Stat.Phys,**123** 503 (2006)

Committor analysis

An attempt to find out the reaction coordinate



Good RC

Bad RC

analysis very expensive: requires pB histogram for every q

cheaper approaches:

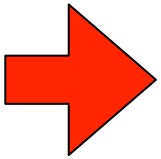
GNN approach. Ma and Dinner, JPC **109** 6769 (2005)

Bayesian approach Best and Hummer, PNAS **102** 7632(2004)

Likelihood Maximization. Peters and Trout, JCP **125**, 054108 (2006)

What is a reaction coordinate?

Different meaning depending on objective.

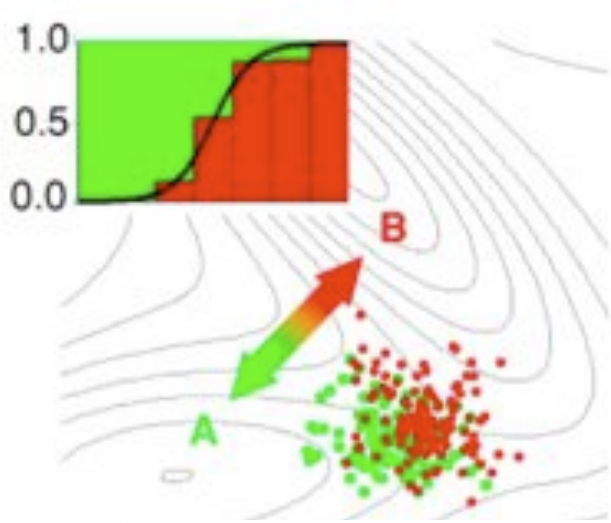


meaning of "reaction coordinate"	objective
committor function	perfect and complete description of transition without physical insight
best low dimensional model of committor	insight in mechanistic detail in terms of meaningful collective variables
best low dimensional model of separatrix by committor or by variational TST	insight in transition state ensemble and direction of reactive flux
good low dimensional model of transition	allows free energy along RC, rates, transmission coefficients.
reasonable order parameters or collective variables	distinguishes reactant from product allows (meta)stable states description

Likelihood maximization

- Each TPS shot can be seen as a committor shot. Based on this look for best model of reaction coordinate *r*
- The probability p(TP|r) to be on a transition path given a structure x with an rc *r* is (for diffusive dynamics)

$$p(TP|r) = 2p_B(r)(1 - p_B(r))$$

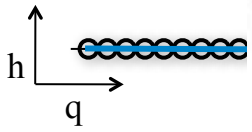


- Assume committor function to be

$$p_B(x) = \frac{1}{2} + \frac{1}{2}\tanh\left[r(q(x)\right]$$

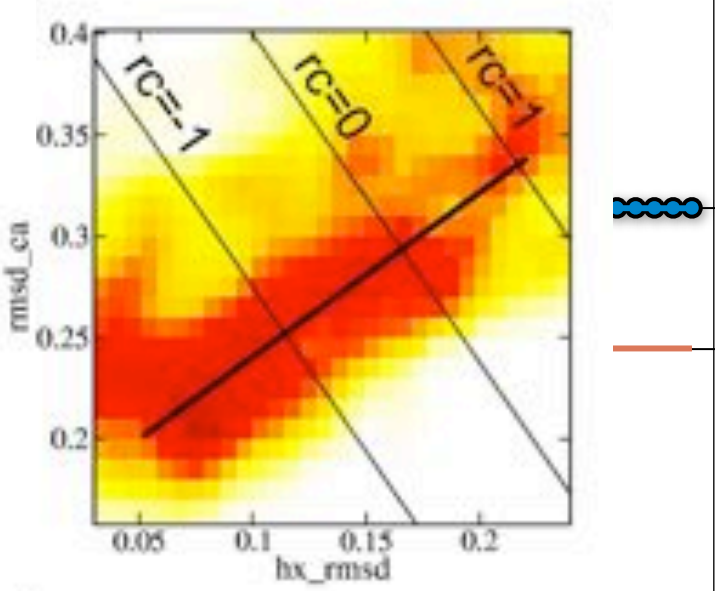
- parametrize r as linear combination of q

$$r(\mathbf{x}) = \sum_i \alpha_i q(\mathbf{x}) + \alpha_0$$



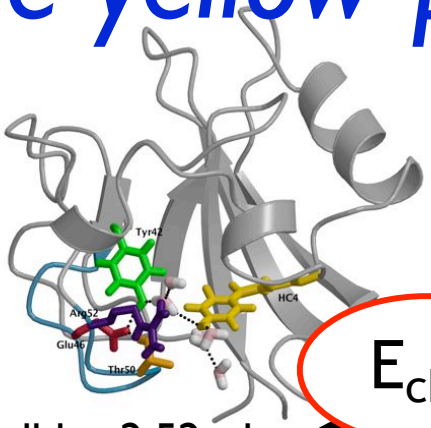
- best r is maximizing likelihood

$$L(\alpha) = \prod_{i=1}^{N_B} p_B(r(q(\mathbf{x}_i^{(B)})) \prod_{i=1}^{N_A} (1 - p_B(r(q(\mathbf{x}_i^{(B)}))))$$



rc = -4.5 + 13 $rmsd_{hx}$ + 8 $rmsd_{ca}$

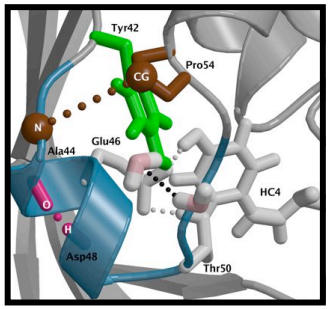
Photoactive yellow protein unfolding



unfolded

E_{chro}

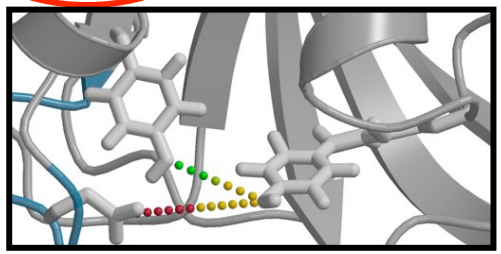
rc =-3.49 + 15.28 $rmsd_{\alpha 3}$+ 5.65 dhb+ 2.52 da_{PA}

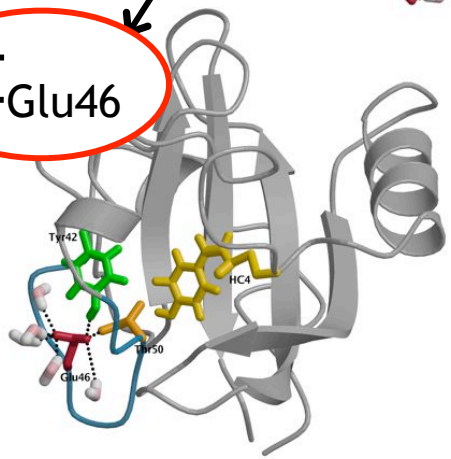


E_{Glu46}

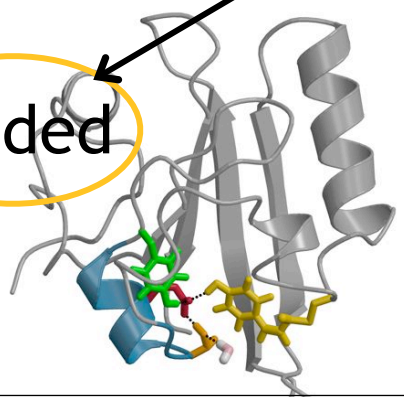
$U_{\alpha 3}$

$I_{\alpha 3}$





folded

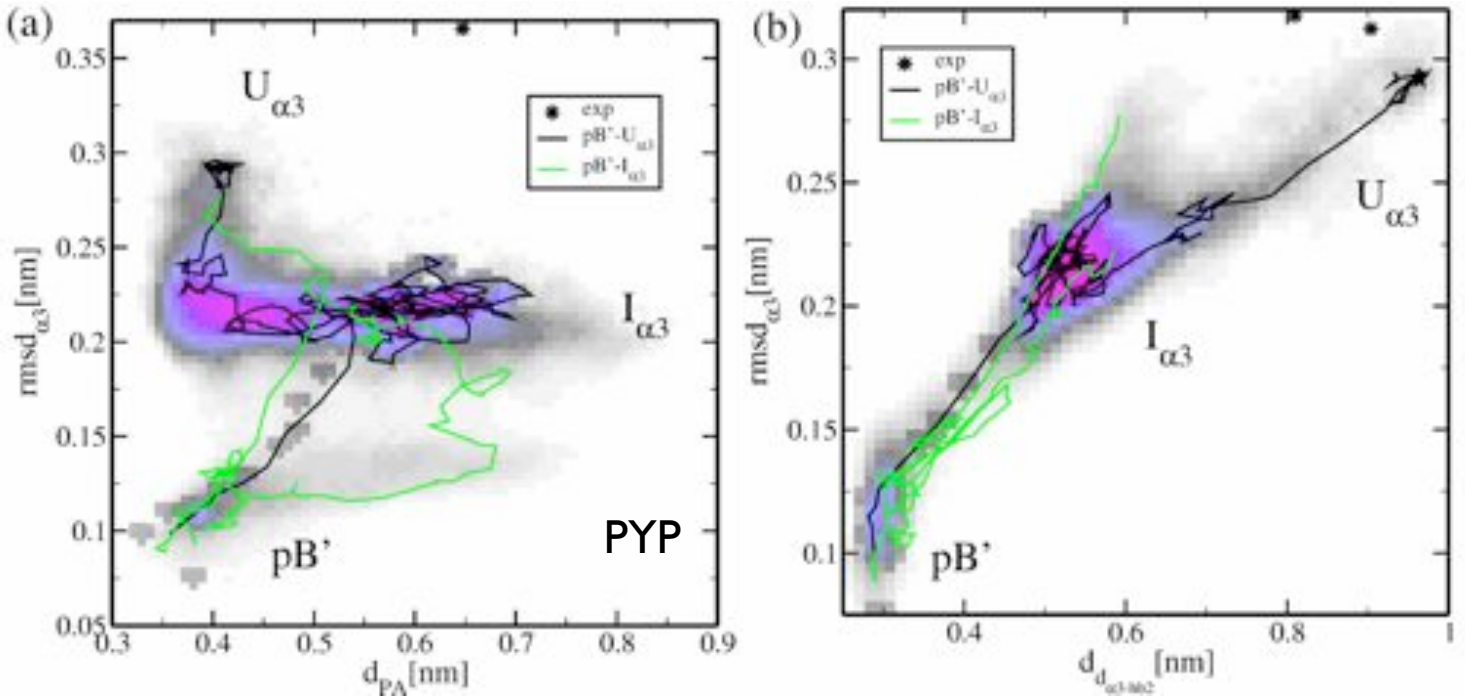


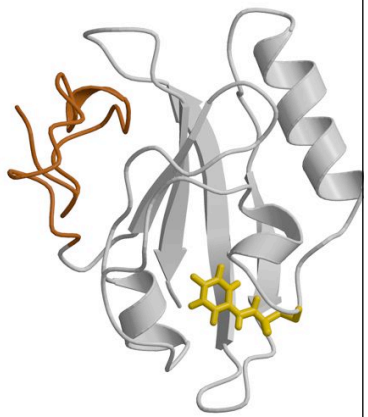
Complex of intermediates

Vreede, Juraszek, PGB, PNAS 2010

Non-linear reaction coordinate

- Need for non-linear RC:
  - Linear RC only valid close to TS.
  - when mechanism away from TSE is different, but collective coordinates are not known

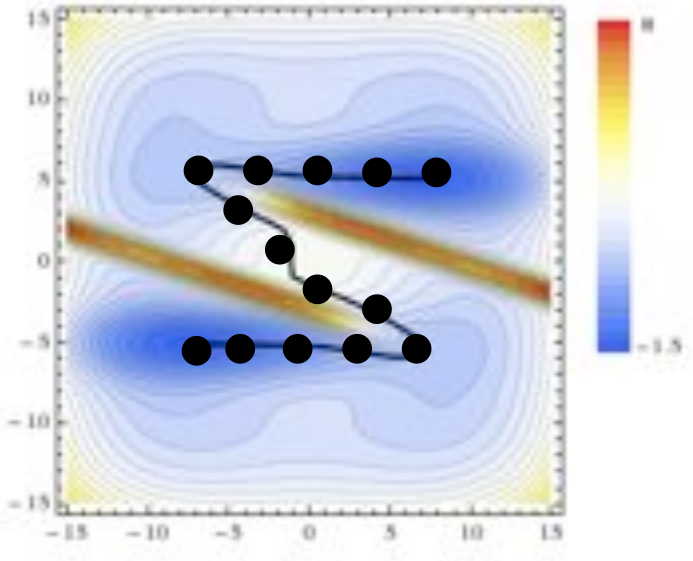




- approximate RC by string in collective variable space (c.f. string-method Eijnden et al.):
  - use likelihood maximization to optimize string

String likelihood maximization

- Algorithm:
  - choose collective variables **q**
  - create string S
  - compute likelihood L
  - maximize L by moving string
  - choose other **q** and repeat
  - maximum likelihood yields best non-linear RC in best **q**-space



$$S = \{\mathbf{q}_0, \mathbf{q}_1 \ldots \mathbf{q}_n\}$$

- Example z potential:
  - x only
  - y only
  - x,y combination

$$\alpha(x) = \alpha(\mathbf{q}(x), S)$$

$$r(x) = f(\alpha(x))$$

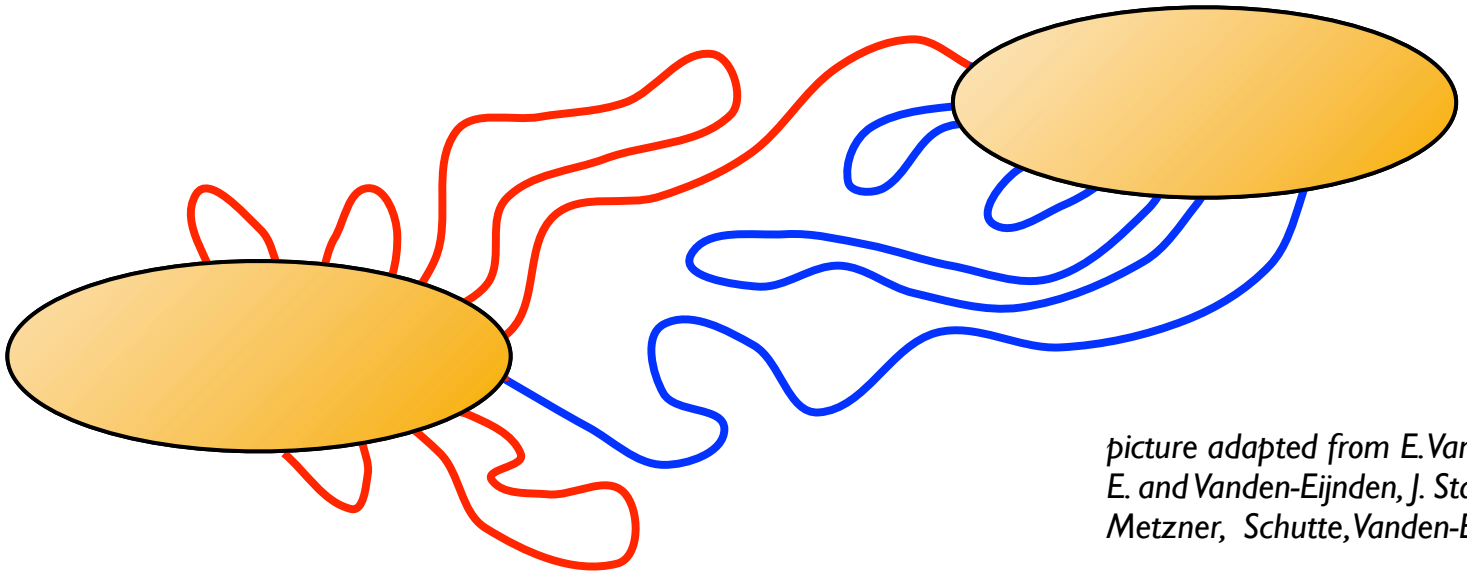
$$p_B(x) = \frac{1}{2}(1 + \tanh(r(x)))$$

- **Problem**: not enough statistics in tails
- **Solution**: use complete path ensemble

$$L = \prod_{x_i \in B} p_B(x_i)) \prod_{x_i \in A} (1 - p_B(x_i))$$

The complete path ensemble

- the complete path ensemble contains all information of the entire pB surface
- includes all paths between A and B, including AA and BB paths.



picture adapted from E. Vanden-Eijnden
E. and Vanden-Eijnden, J. Stat. Phys 2006
Metzner, Schutte, Vanden-Eijnden, JCP 2006

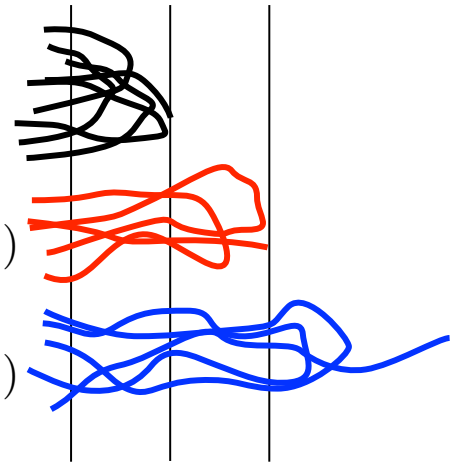
- direct evaluation of the total path ensemble is inefficient, for each reactive (AB) path there are millions of failed (AA) ones.
- But we can compute the weight for each path by replica exchange transition interface sampling (RETIS)

Reweighting paths using WHAM

histograms:

$$P_A(\lambda|\lambda_1) = \sum_{i=1}^{n-1} \bar{w}_i \theta(\lambda_{i+1} - \lambda)\theta(\lambda - \lambda_i) \sum_{j=1}^{i} P_A(\lambda|\lambda_j)$$

$$\bar{w}_i = \frac{1}{\sum_{j=1}^{i} 1/w_j}$$

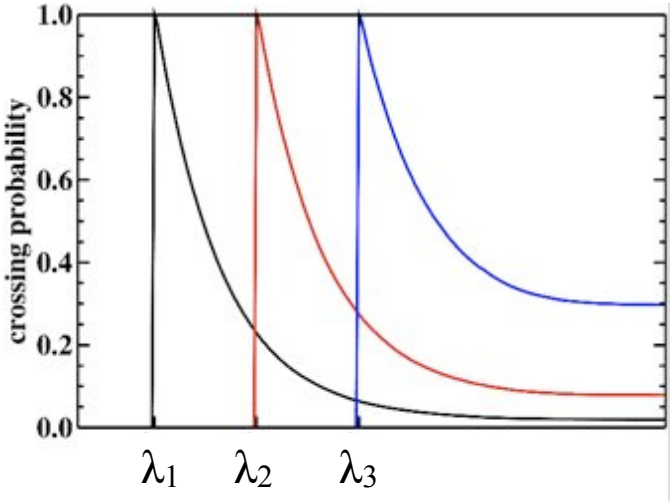


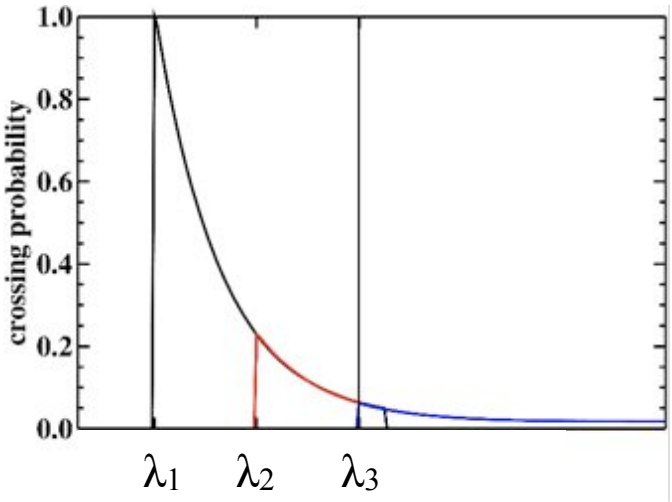
path weights

$\bar{w}_1 = 1$

$\bar{w}_2 = P_A(\lambda_2|\lambda_1)$

$\bar{w}_3 = P_A(\lambda_3|\lambda_1)$





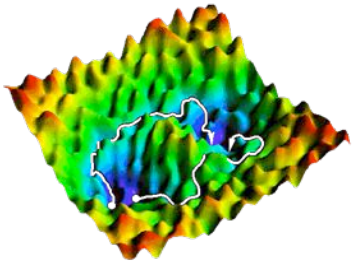
reweighting paths:

$$\mathcal{P}_A[\mathbf{x}(L)] = \sum_{i=1}^{n-1} \bar{w}_i g_i[\mathbf{x}(L)] \sum_{j=1}^{i} \mathcal{P}_{A\Lambda_j}[\mathbf{x}(L)]$$

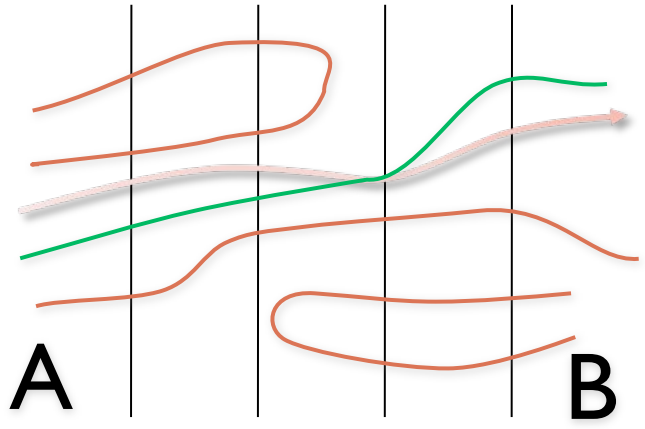
$$g_i(\mathbf{x}(L)) = \theta(\lambda_{max}[\mathbf{x}(L)] - \lambda_i)\theta(\lambda_{i+1} - \lambda_{max}[\mathbf{x}(L)])$$

Rogal et al, JCP 2010
see also Minh and Chodera, JCP 2009

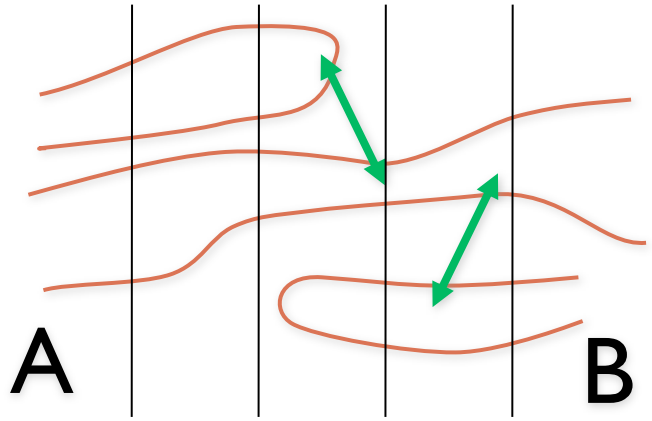
Replica Exchange Transition Interface Sampling



Shooting Move

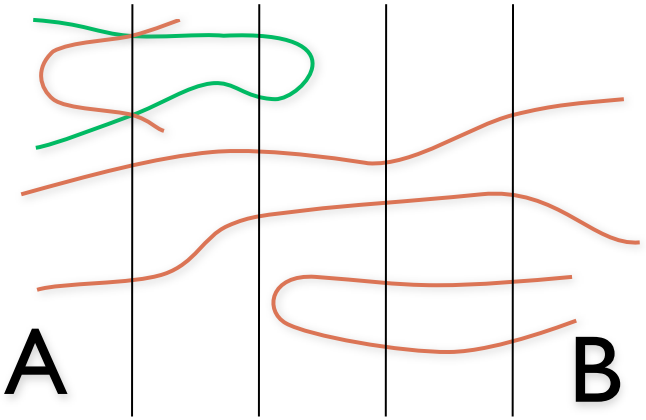


Exchange Move

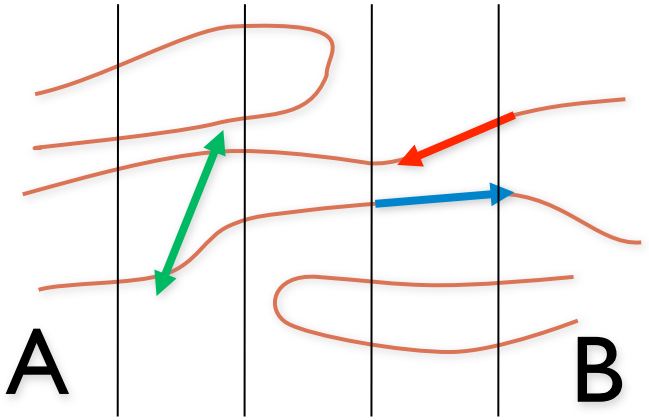


T.S. van Erp, PRL **98**, 268301 (2007)
P.G. Bolhuis, JCP **129**, 114108 (2008)

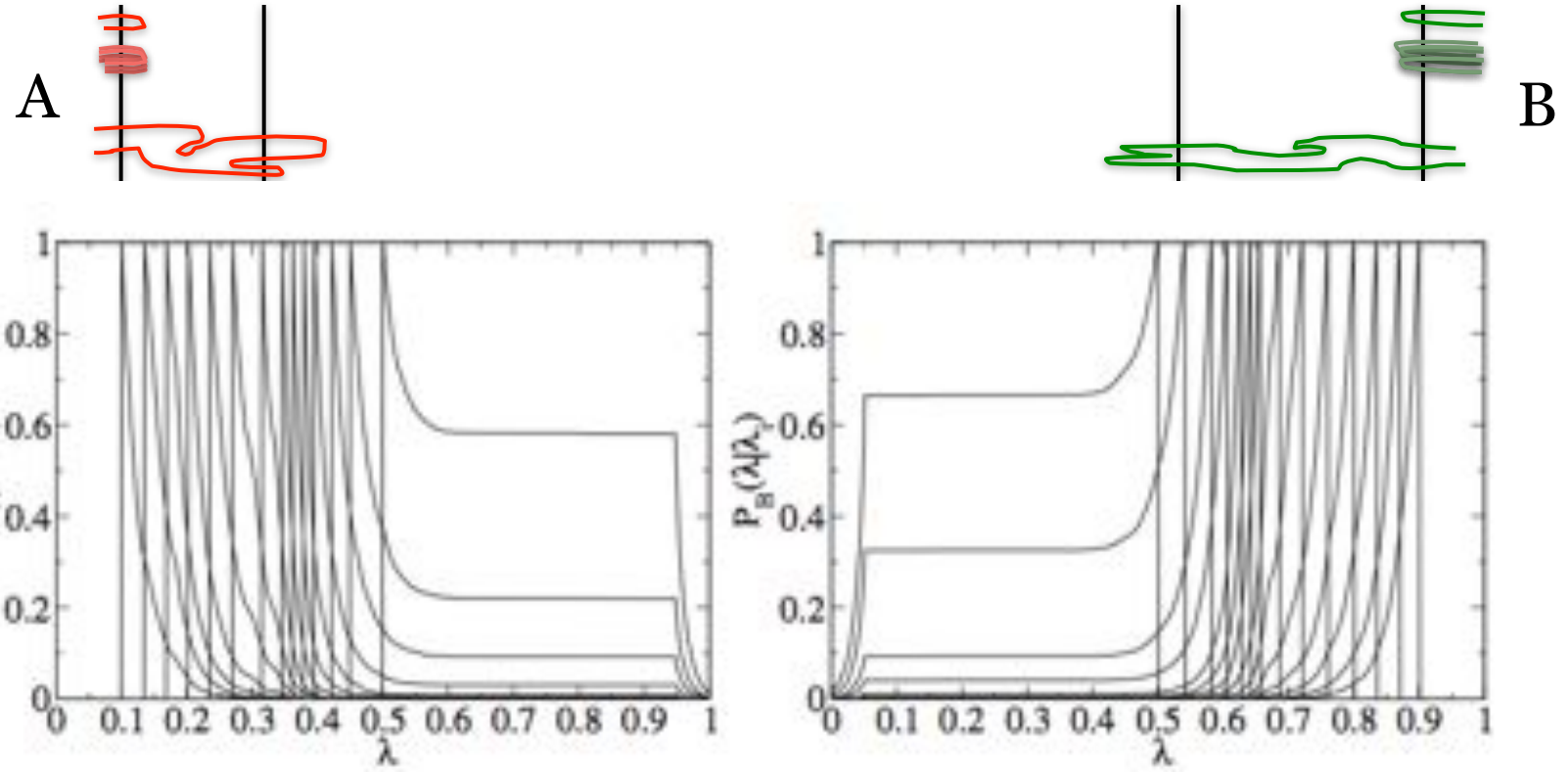
First Interface Move



Forward-Backward Move



The Reweighted Path Ensemble



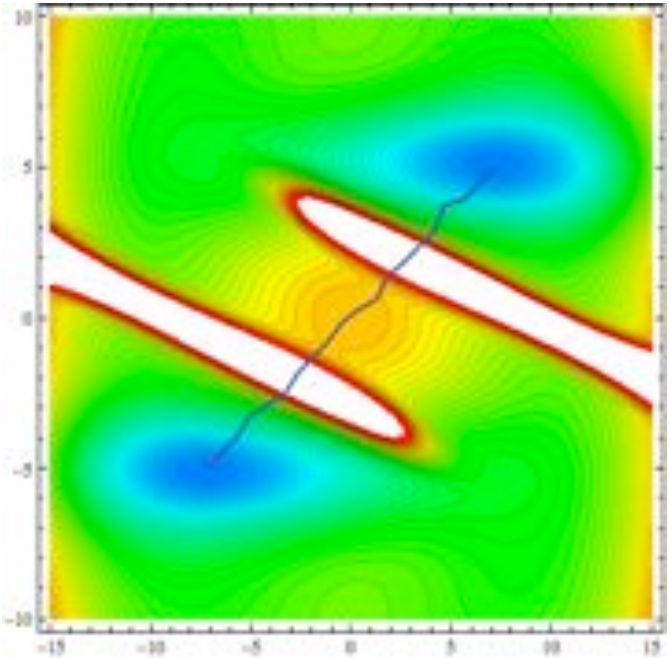
$$\mathcal{P}[\mathbf{x}^L] = c_A \sum_{j=1}^{n-1} \mathcal{P}_{A\Lambda_j}[\mathbf{x}^L] W^A[\mathbf{x}^L] + c_B \sum_{j=1}^{n-1} \mathcal{P}_{B\Lambda_j}[\mathbf{x}^L] W^B[\mathbf{x}^L]$$

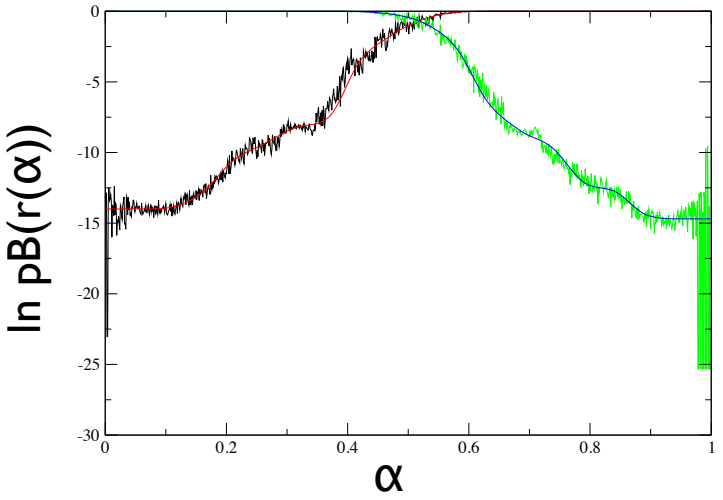
$$W^A[\mathbf{x}^L] = \sum_{i=1}^{n-1} \bar{w}_i^A \theta(\lambda_{max}[\mathbf{x}^L] - \lambda_i)\theta(\lambda_{i+1} - \lambda_{max}[\mathbf{x}^L])$$

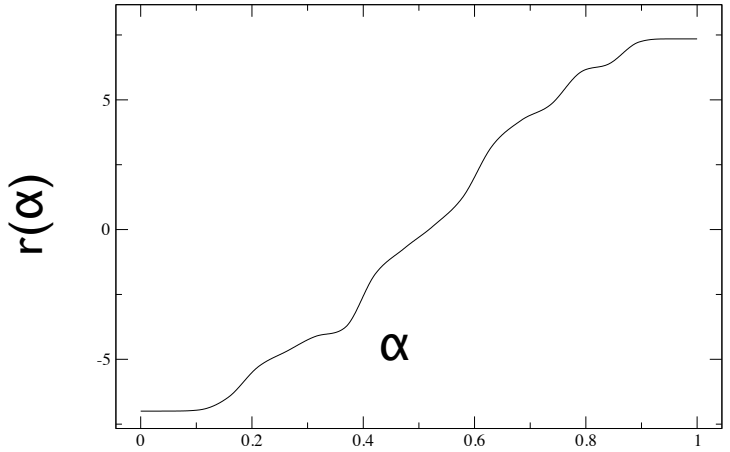
Rogal et al , J. Chem. Phys. 133, 174109 (2010)
Lechner et al .J. Chem. Phys. 133 174110 (2010).

Optimizing string

$$\ln L = \sum_{x_i \in B} \bar{w}_i \ln p_B(x_i) + \sum_{x_i \in A} \bar{w}_i \ln(1 - p_B(x_i))$$



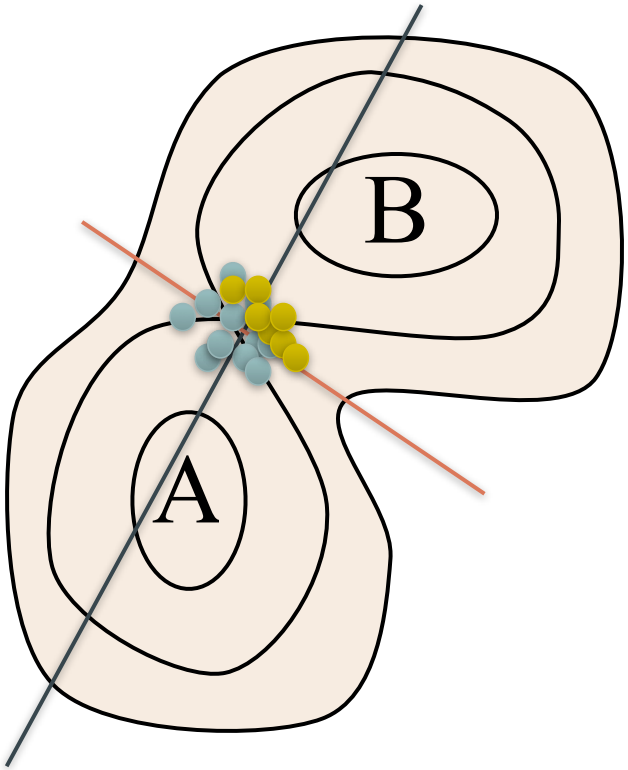


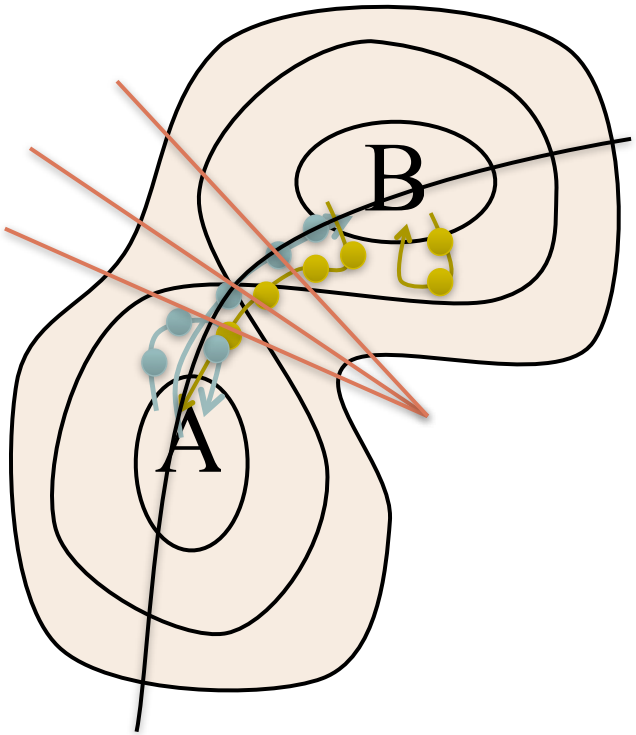


ln L(x only) = -9x10^5
ln L(y only) = -7x10^4
ln L(diagonal) = -5x10^5
ln L(string(x,y)) = -521

String clearly best coordinate

Non-linear reaction coordinate optimization





Mapping: $r[q] = a_0 + \sum a_i q_i$ $r[q] \leftarrow$ *String* + *Mapping*

Data: Shooting Points Reweighted Path Ensemble

Projection of RPE

The reweighted path ensemble

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

with

$$W^A[\mathbf{x}^L] = \sum_{i=1}^{n-1} \bar{w}_i^A \theta(\lambda_{max}[\mathbf{x}^L] - \lambda_i)\theta(\lambda_{i+1} - \lambda_{max}[\mathbf{x}^L])$$

can be used to project the conditional path dependent population density

$$\rho_{AB}(\mathbf{q}) = C \int \mathcal{D}\mathbf{x} \sum_{k=0}^{L} \prod_{i=1}^{m} \delta(q^{(i)}(\mathbf{x}_k) - q^{(i)}) \mathcal{P}[\mathbf{x}^L] h_A(\mathbf{x}_0) h_B(\mathbf{x}_L)$$

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

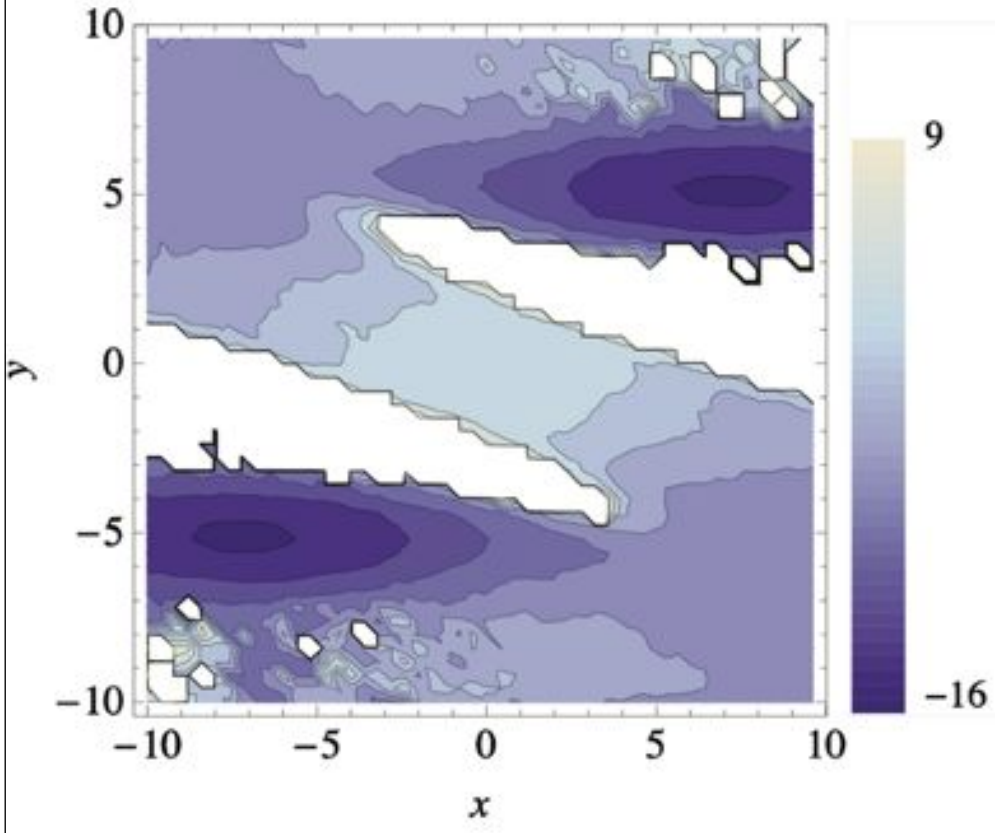
and thus the free energy

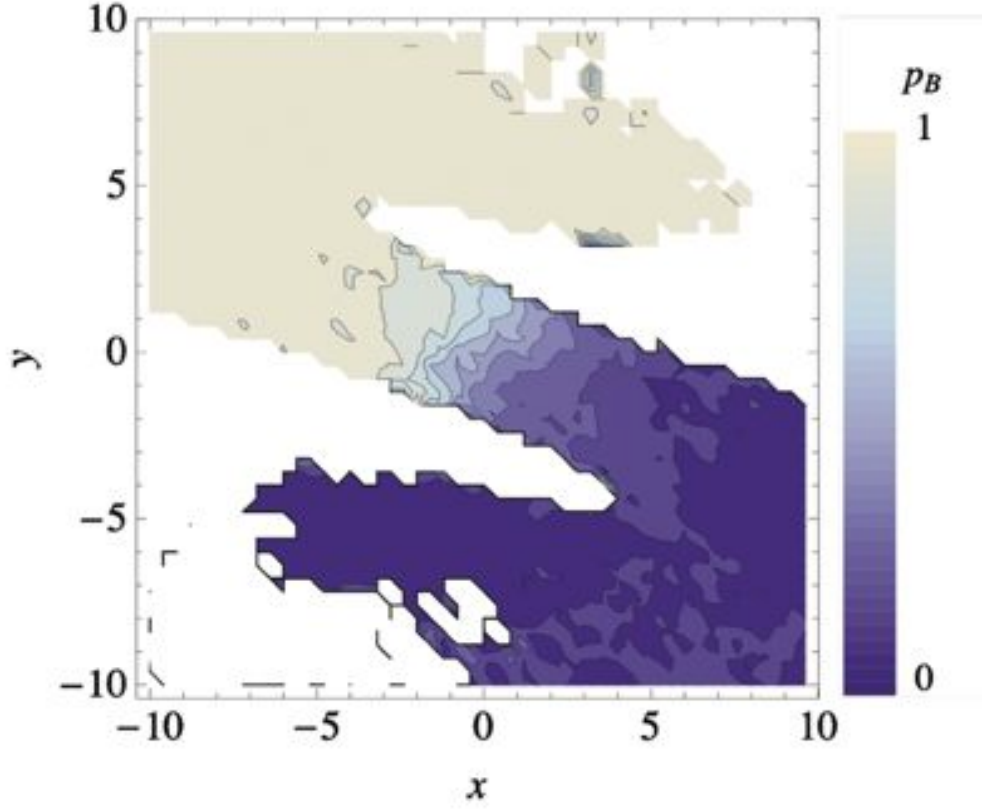
$$F(\mathbf{q}) = -k_B T \ln(\rho(\mathbf{q})) + const,$$

and the committor

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

FE and pB landscape from RPE





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

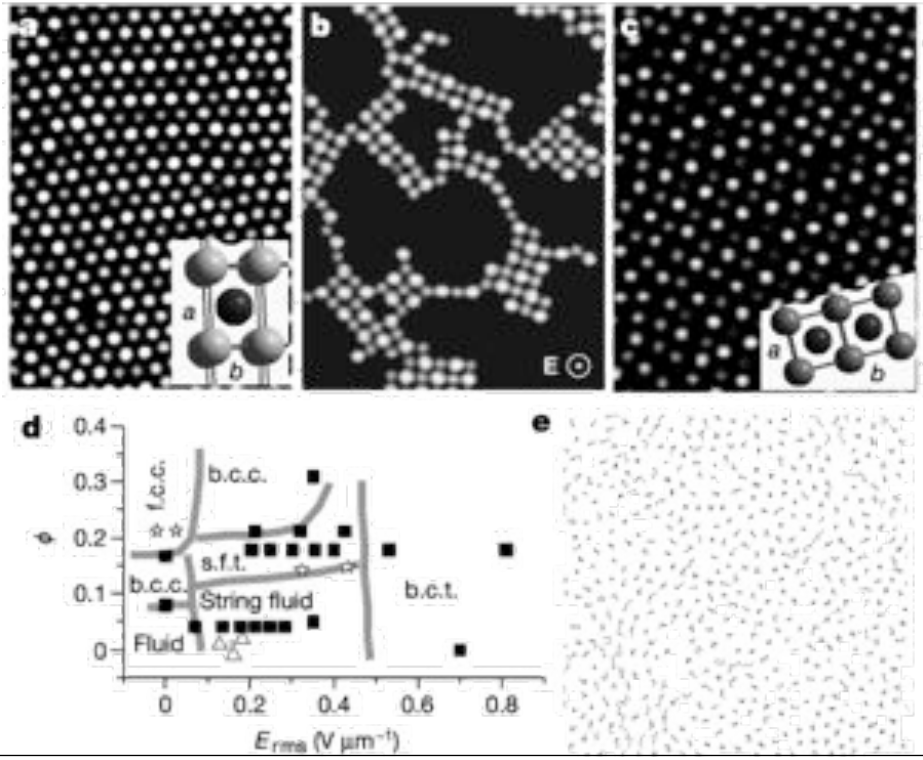
$$p_A(\mathbf{q}) = \frac{\rho_{AA}(\mathbf{q}) + \rho_{BA}(\mathbf{q})}{\rho(\mathbf{q})}$$

Outline

- Introduction
- Path sampling
- Rates with transition interface sampling
- Linear and non-linear reaction coordinate analysis
- **Application: crystal nucleation**
- Conclusions

Crystallization

- Fundamental for all material sciences
- Crucial for protein crystallography
- Recent interest through advances in colloid science
  - 3D real space imaging via confocal microscopy
  - optical tweezers
  - particle tracking

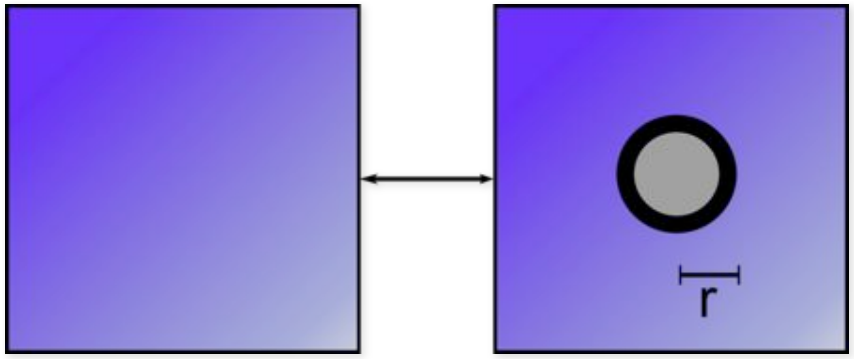


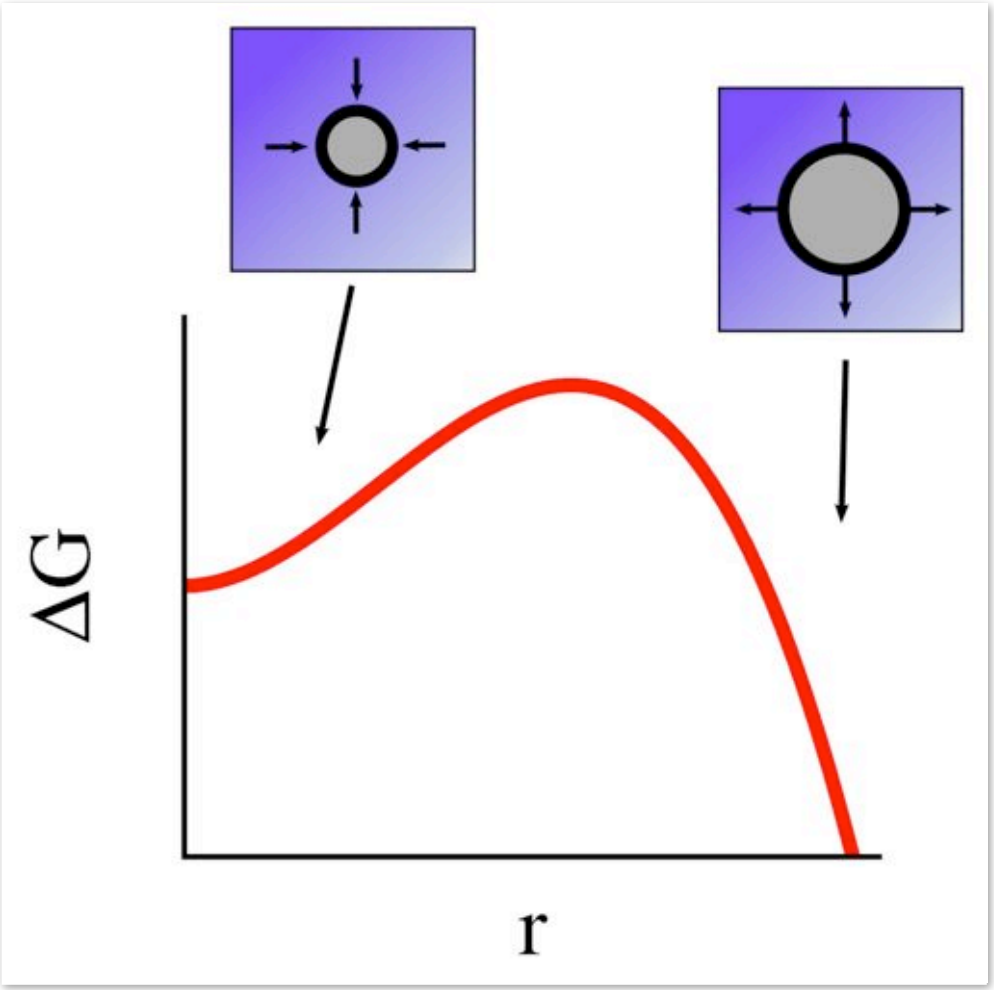
Yethiraj and van Blaaderen
Nature 421, 513-517 (2003)

Why do crystals nucleate?

Volume ↓ Surface ↓

$$\Delta G = \frac{4\pi r^3}{3}\rho_s \Delta\mu + 4\pi r^2 \gamma$$





Assumptions classical nucleation theory

Four basic assumptions

1. the nucleation is governed by the largest solid cluster in the system
2. the surface tension is independent of the nucleus size

3. the nucleation process is independent of the stable crystal structure: related to Ostwald's step rule:
   - crystallization is multistep process
   - instead of forming thermodynamically stable crystal structure first form kinetically most accessible structure
   - governed by kinetics rather than thermodynamics

4. the crystal nucleus is spherical
   - lowest surface area
   - number of particles in sphere scales as $N \sim r^3$

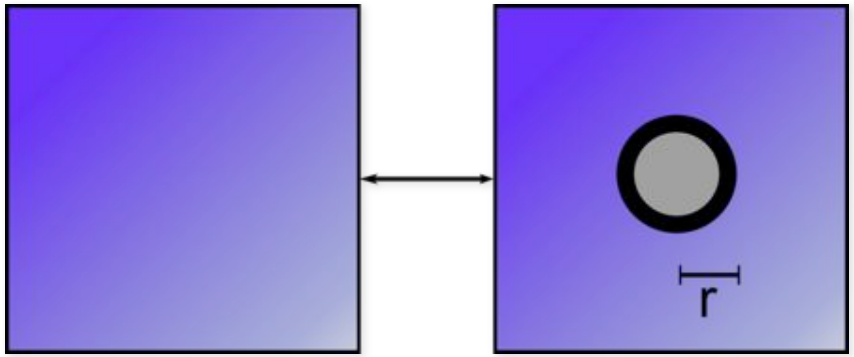
Here we investigate assumptions 3 and 4

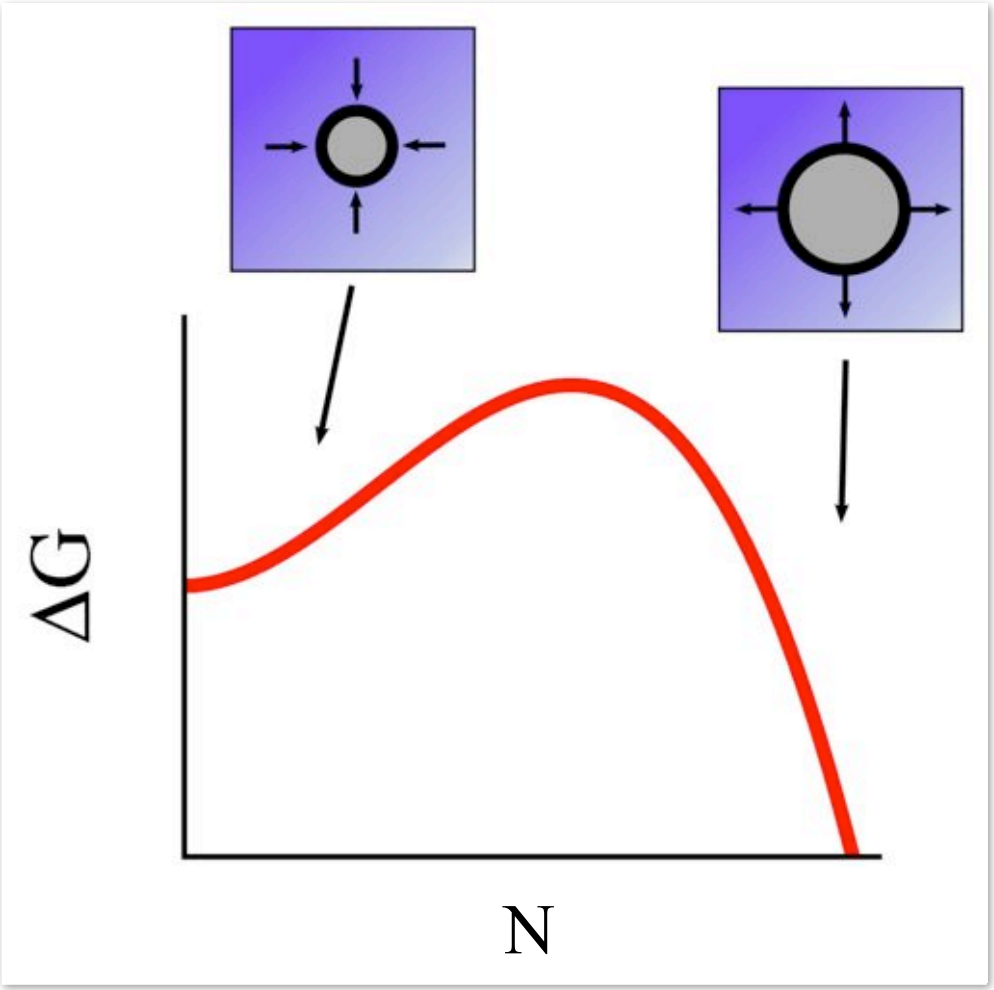
Lechner, Dellago, Bolhuis, PRL 2011, JCP 135 154110 (2011)

Why do crystals nucleate?

Volume ↓ Surface ↓

$$\Delta G = N\Delta\mu + cN^{\frac{2}{3}}\gamma$$

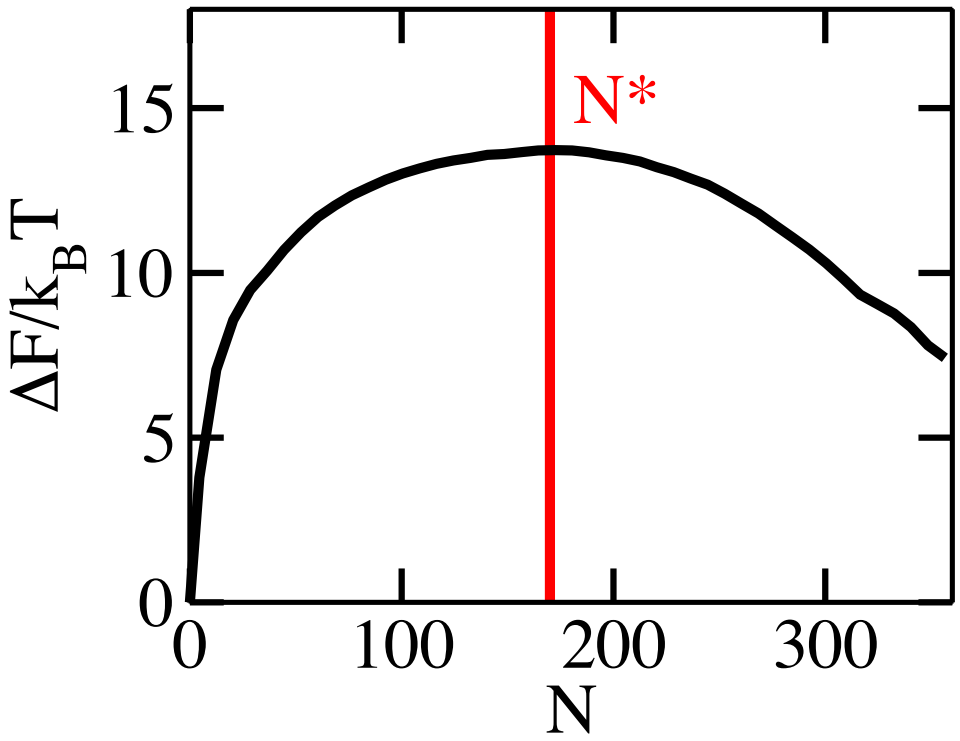


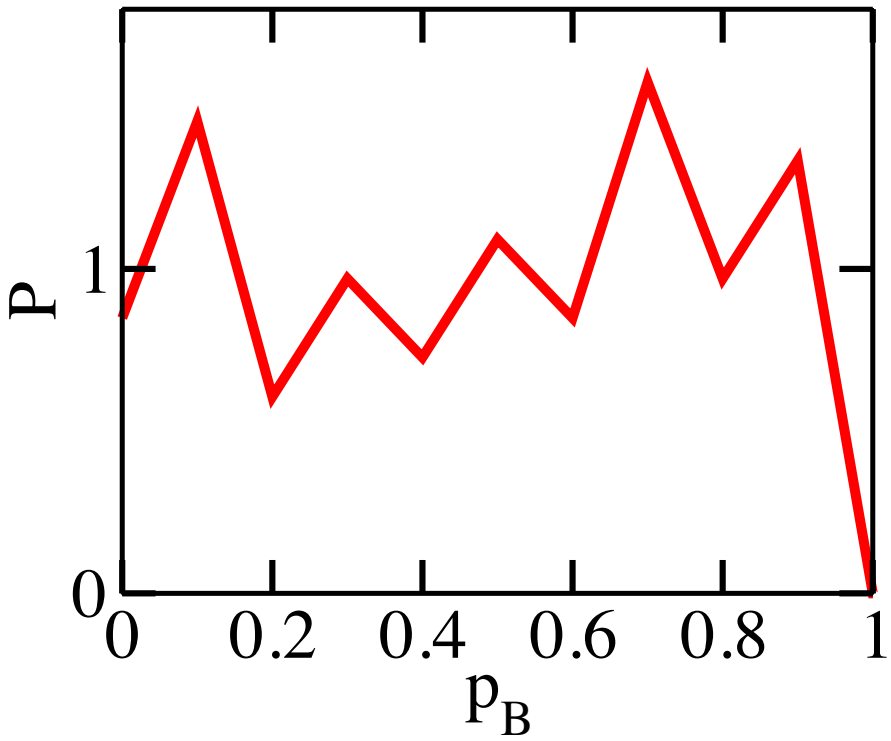


is spherical assumption justified?

Is N really a good reaction coordinate?

Committor analysis of nucleation

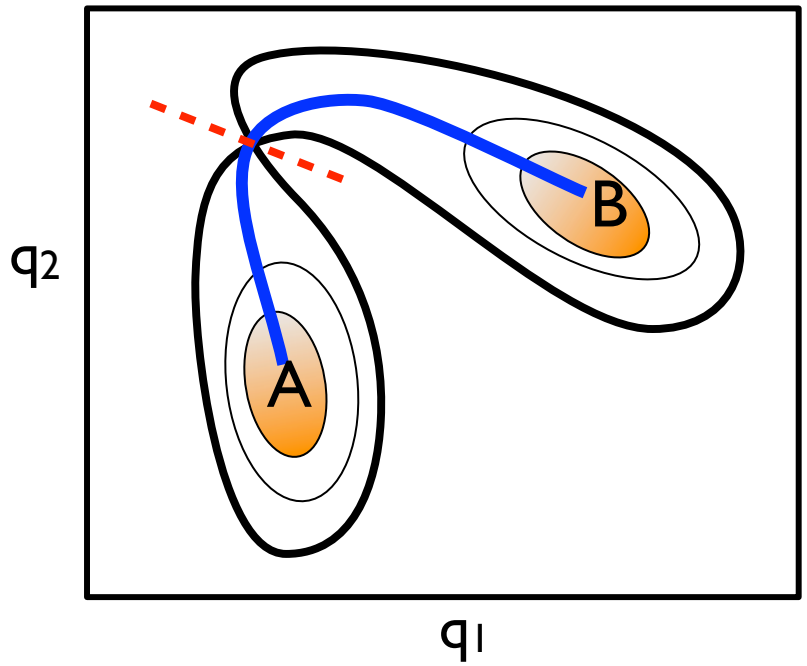




- Size N is not a sufficient ingredient for describing nucleation.
- Can we get THE reaction coordinate?

How do crystals nucleate?

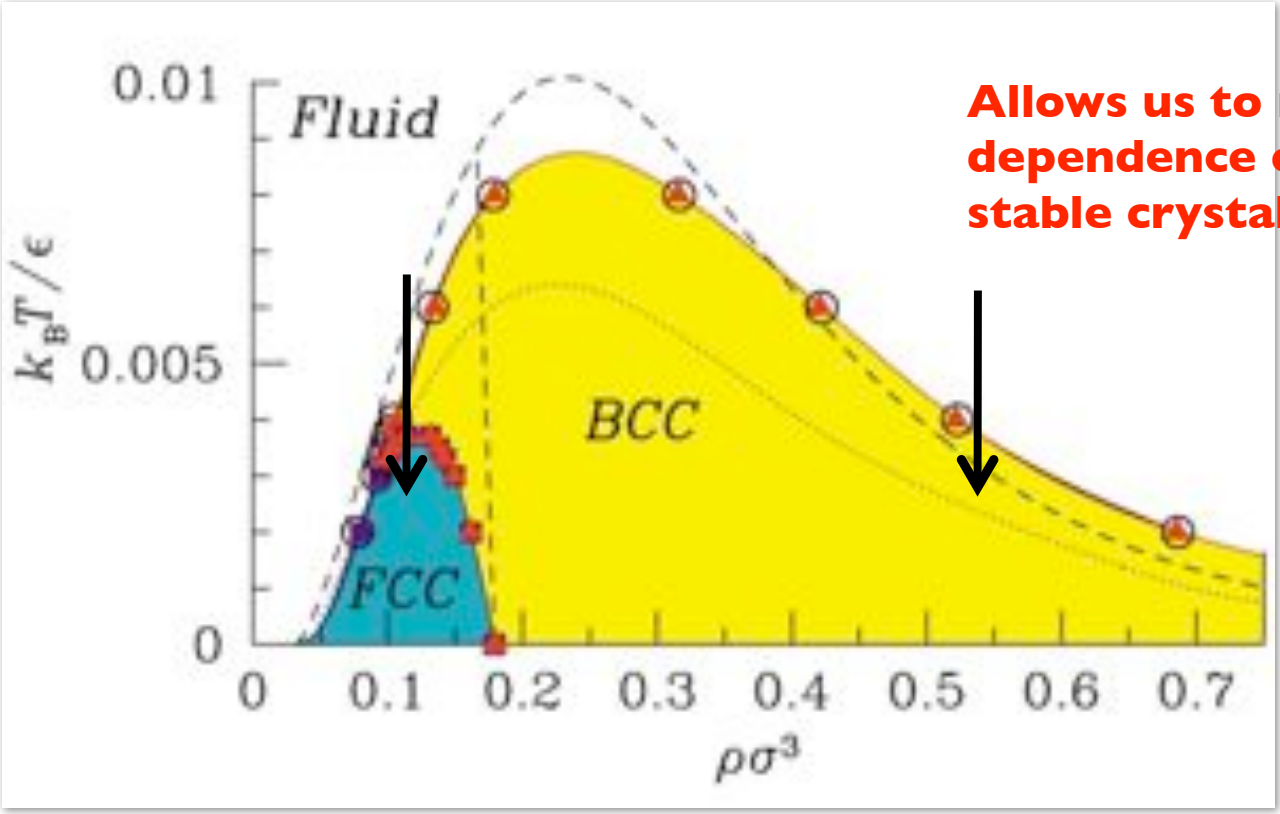
- Aim: best low dimensional parameterization of $p_B(x) = f[q_1(x), q_2(x) \ldots]$



- **Strategy**:
  - define order parameters
  - sample complete nucleation process in reweighted path ensemble
  - find best reaction coordinate by likelihood maximization
  - obtain insight from reaction coordinate

Gaussian Core Model

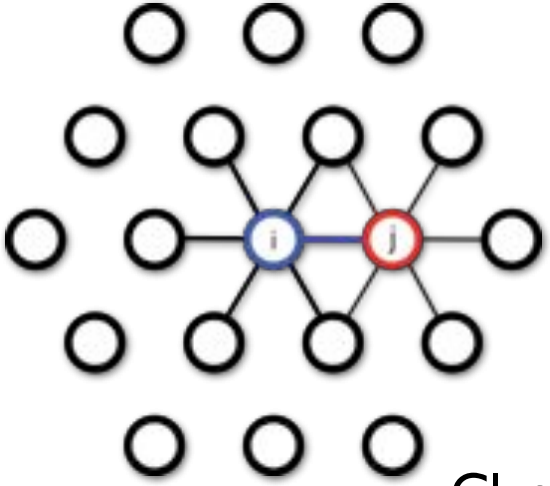
$$v(r) = \varepsilon\, e^{-\frac{x^2}{\sigma^2}}$$



Allows us to investigate dependence on underlying stable crystal phase

Reference: S. Prestipino, F. Saija, P.V. Giaquinta ,Phys. Rev. E **71**, 050102 (2005)

Solid-Fluid Distinction

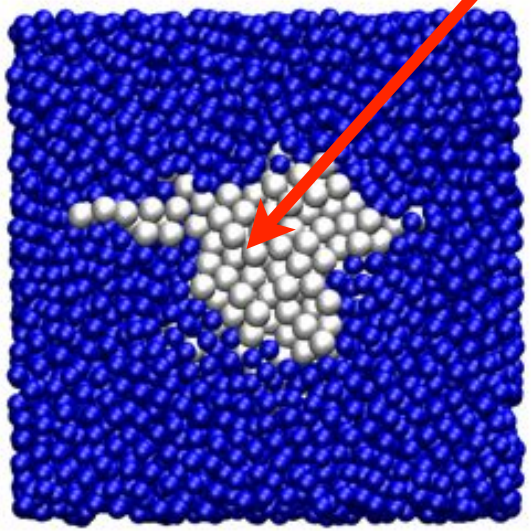


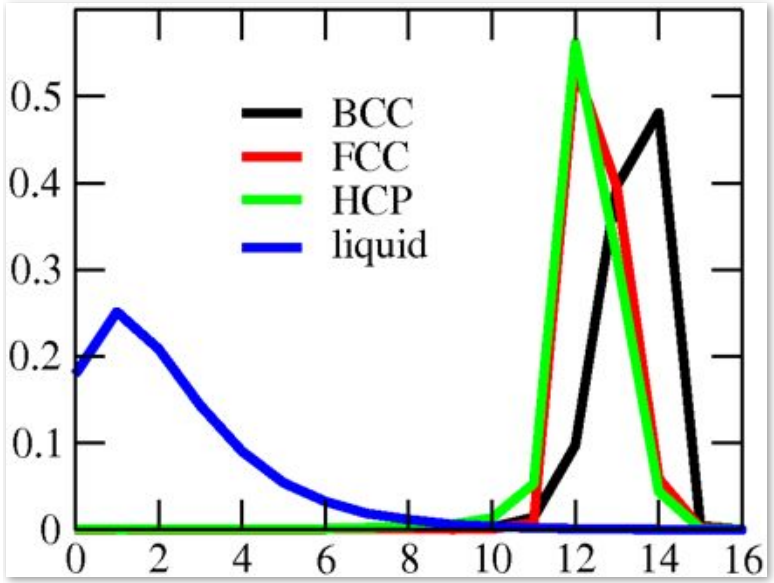
$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(r_{ij})$$

$$s_{ij} = \left\langle q_{lm}(i) q_{lm}(j) \right\rangle$$

$$N_{bonds} = \sum_{N_b} \Theta(s_{ij} - 0.5)$$

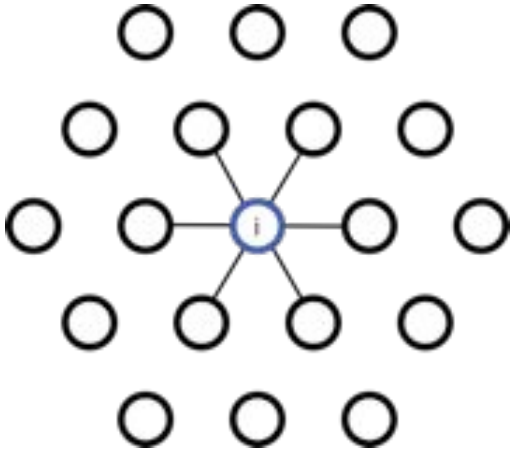
Cluster size N



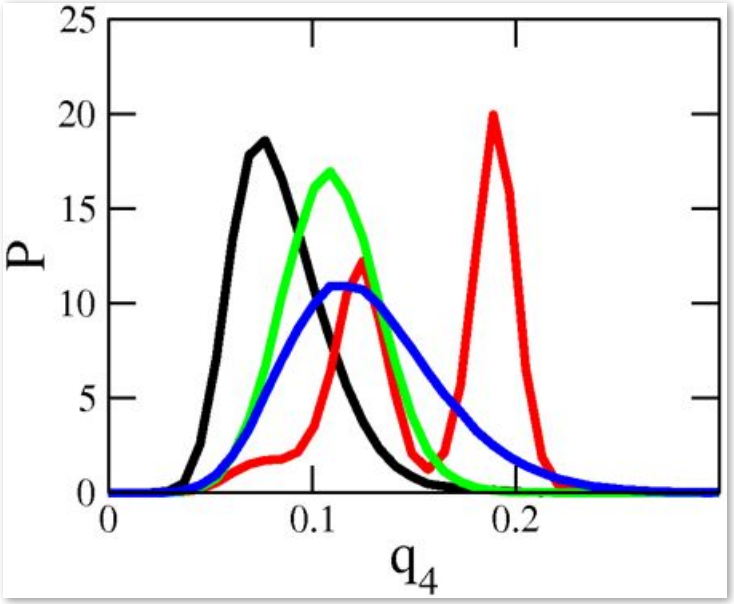


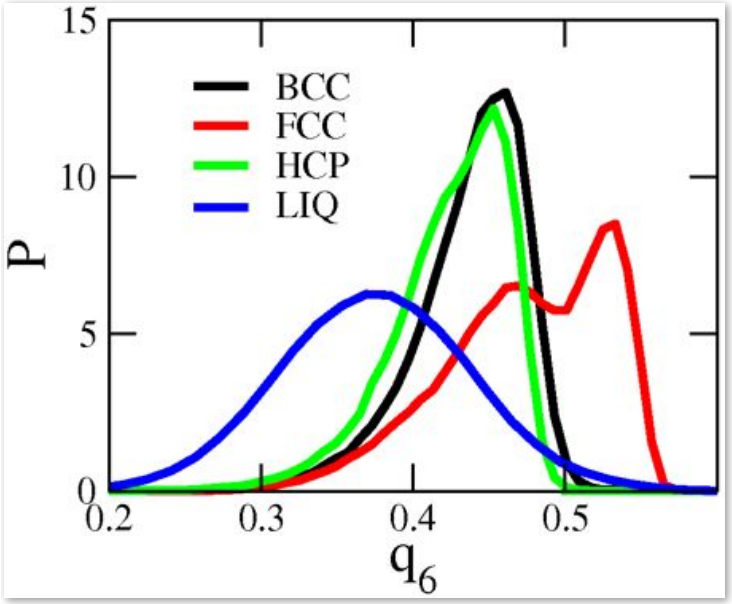
ten Wolde, Ruiz-Montero Frenkel, *Phys. Rev. Lett.* **75,** 2714 (1995)

Structure Analysis

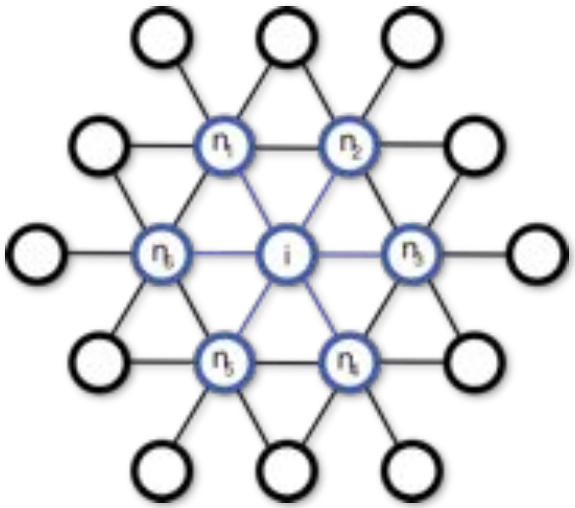


$$q_l(i) = \sqrt{\frac{4\pi}{2l+1}\sum_{m=-l}^{l}\left|q_{lm}(i)\right|^2}$$



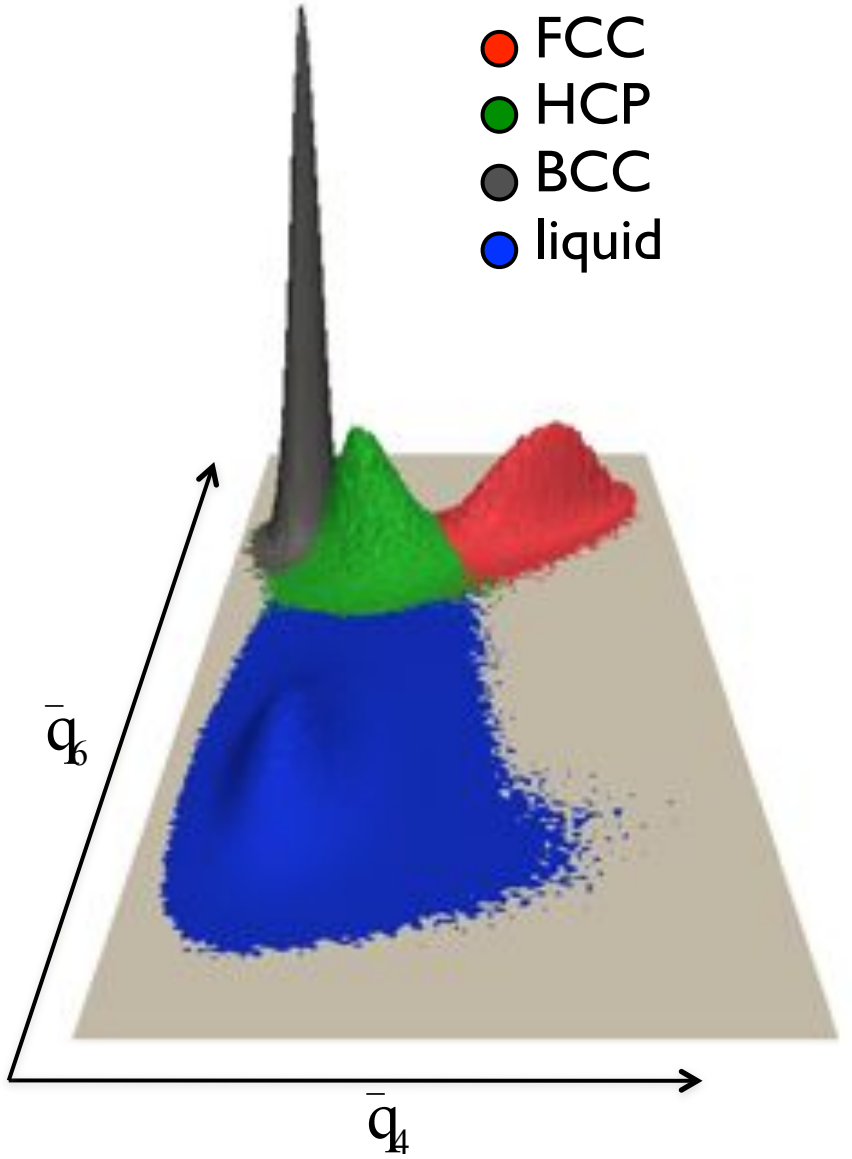


Averaged Bond Order Parameters



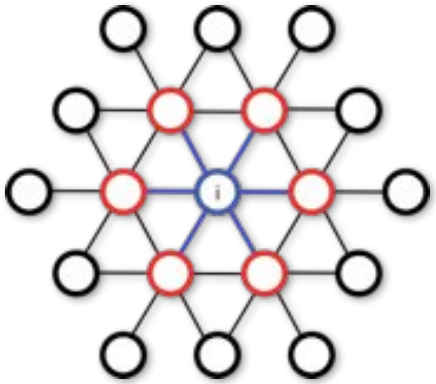
$$\bar{q}_{lm}(i) = \frac{\sum_{k=0}^{N_b} q_{lm}(k) + q_{lm}(i)}{N_b + 1}$$

$$\bar{q}_l(i) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} \left|\bar{q}_{lm}(i)\right|^2}$$

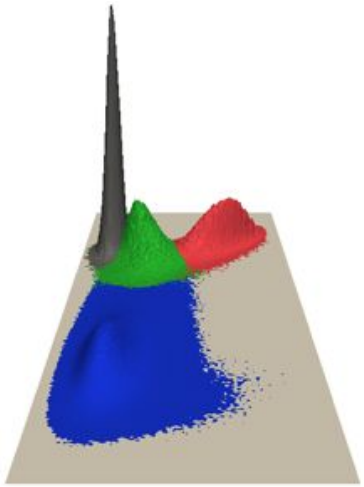


W. Lechner and C. Dellago, *J. Chem. Phys.* **129**,114707 (2008)

Novel Definition of Local Solidity

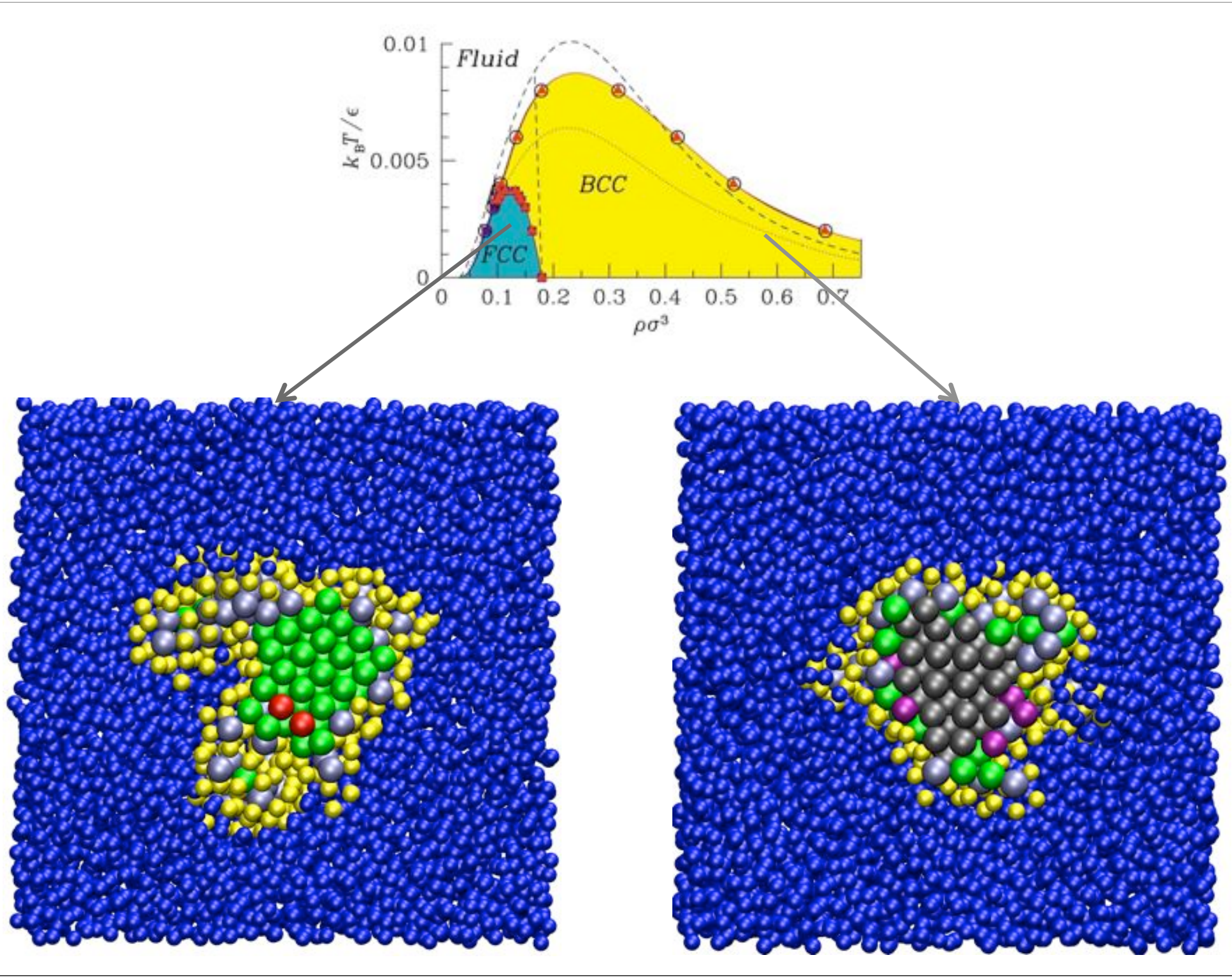


Local structure is correlated with that of neighboring particles

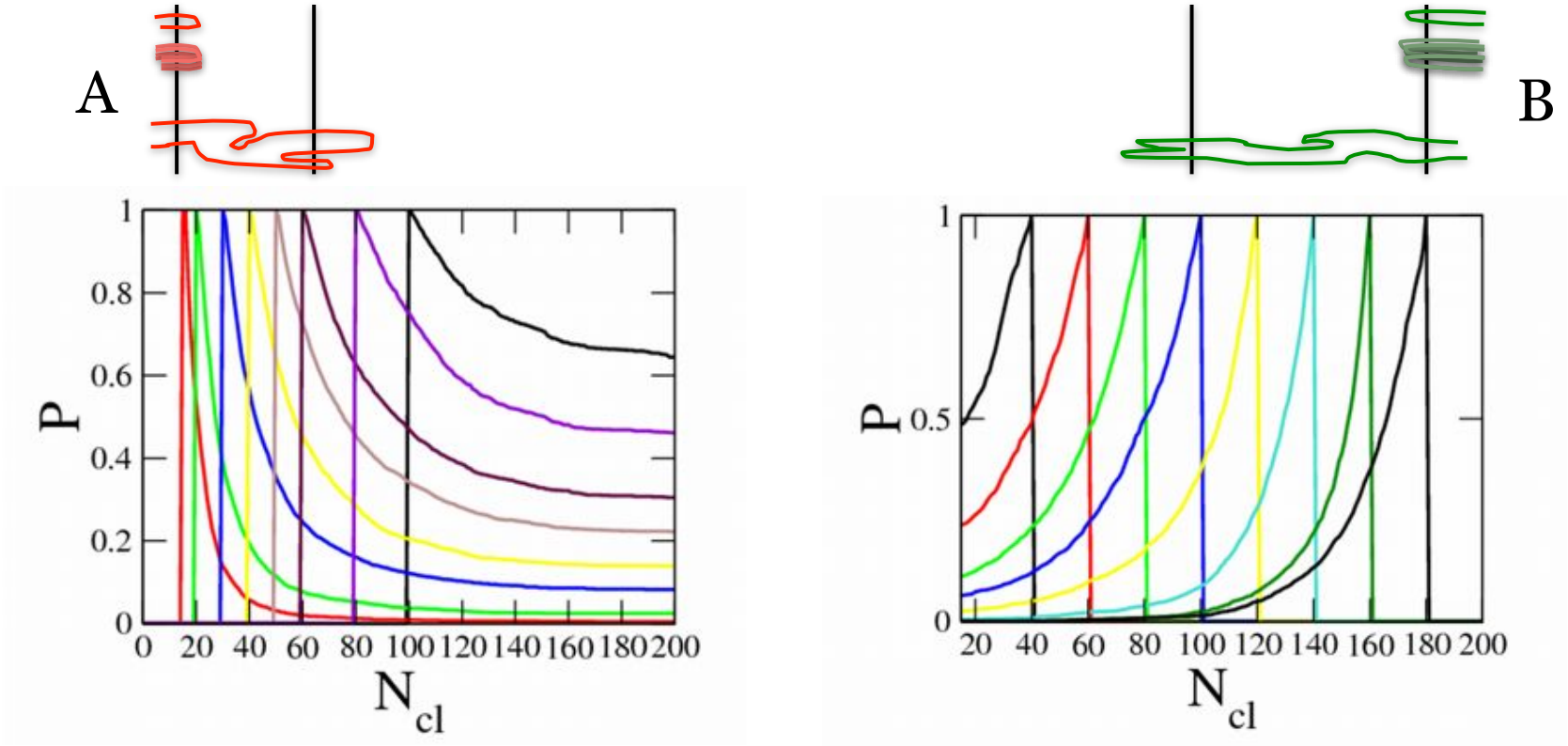


Probability of finding the local structure in the liquid vanishes

W. Lechner, C. Dellago and P.G. Bolhuis, *Phys. Rev. Lett.* **106**, 085701(2011)


0.01
Fluid
k_BT/ϵ
0.005
BCC
FCC
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
$\rho\sigma^3$

The Reweighted Path Ensemble

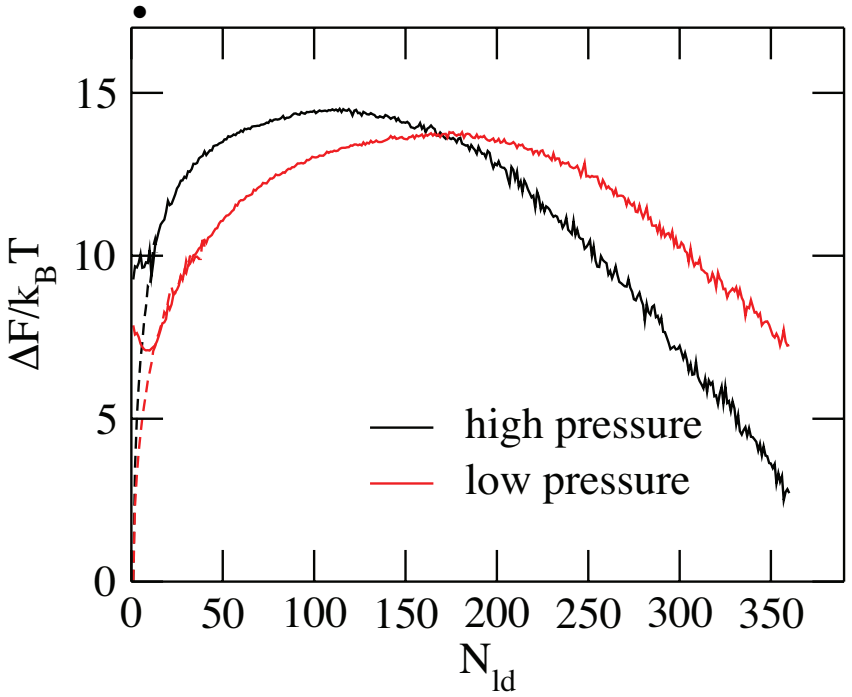


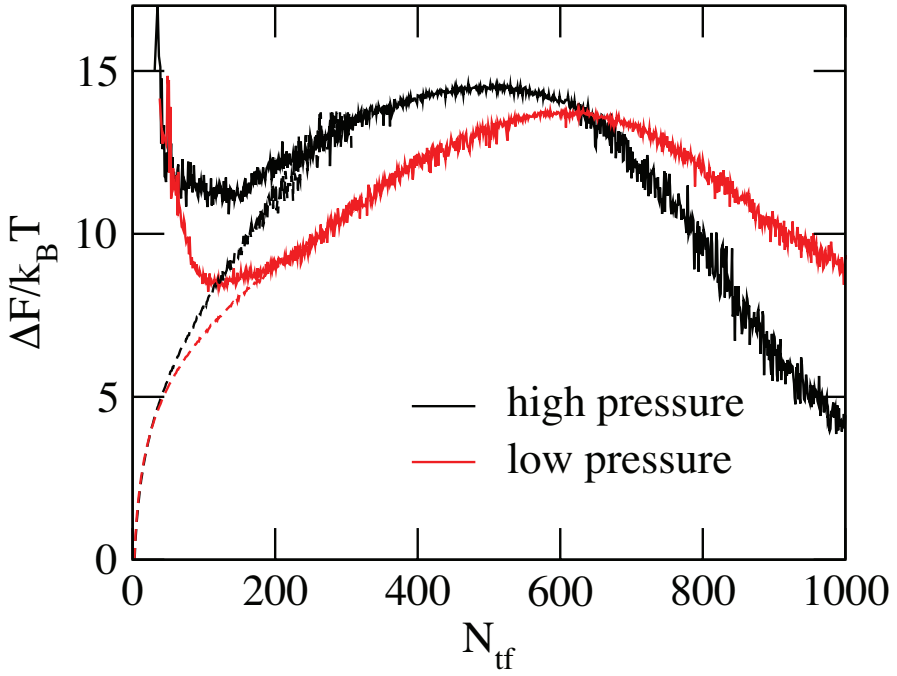
$$\mathcal{P}[\mathbf{x}^L] = c_A \sum_{j=1}^{n-1} \mathcal{P}_{A\Lambda_j}[\mathbf{x}^L] W^A[\mathbf{x}^L] + c_B \sum_{j=1}^{n-1} \mathcal{P}_{B\Lambda_j}[\mathbf{x}^L] W^B[\mathbf{x}^L]$$

$$W^A[\mathbf{x}^L] = \sum_{i=1}^{n-1} \bar{w}_i^A \theta(\lambda_{max}[\mathbf{x}^L] - \lambda_i)\theta(\lambda_{i+1} - \lambda_{max}[\mathbf{x}^L])$$

Rogal et al , J. Chem. Phys. 133, 174109 (2010)
Lechner et al .J. Chem. Phys. 133 174110 (2010).

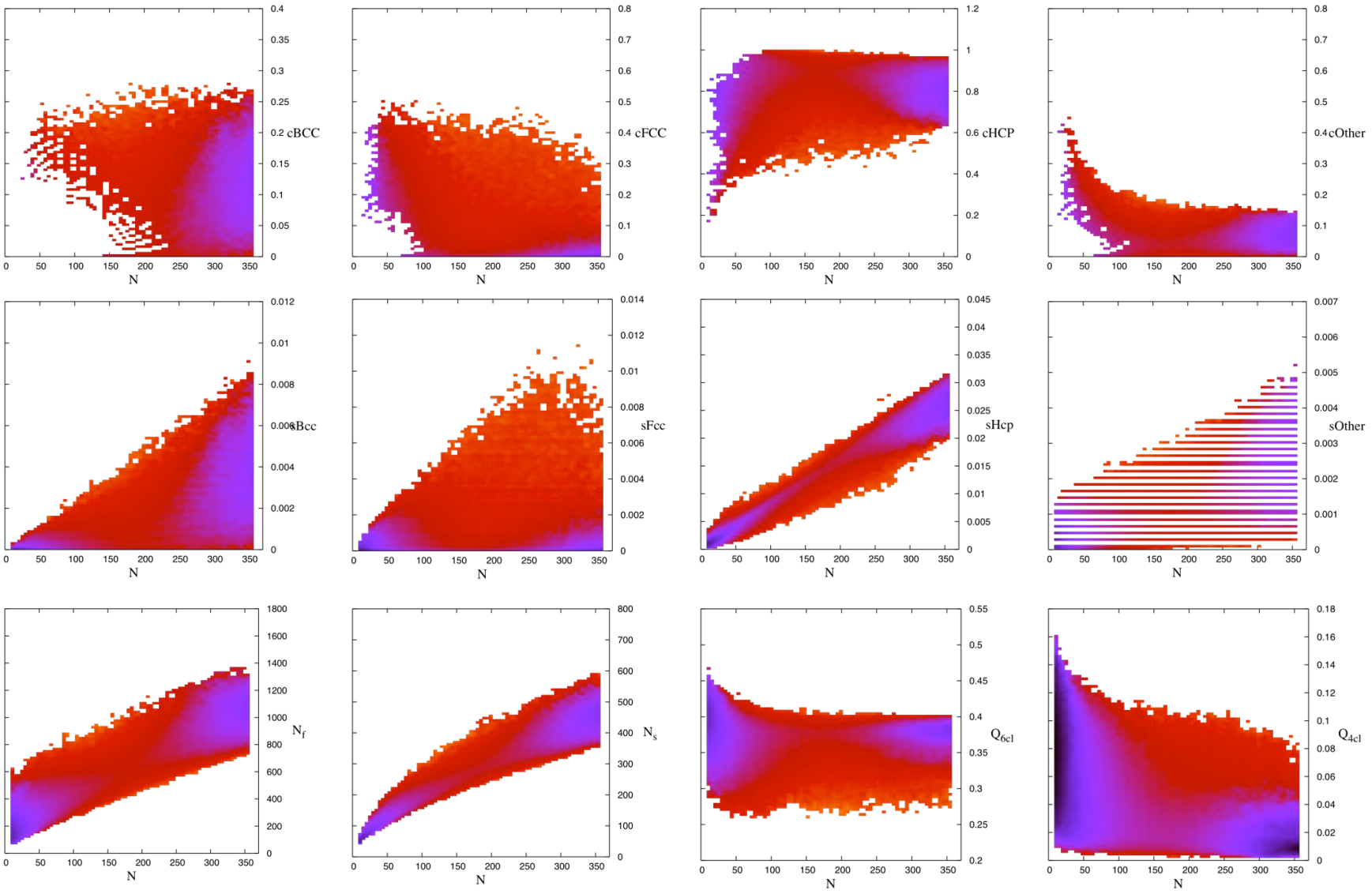
Free energies and rates



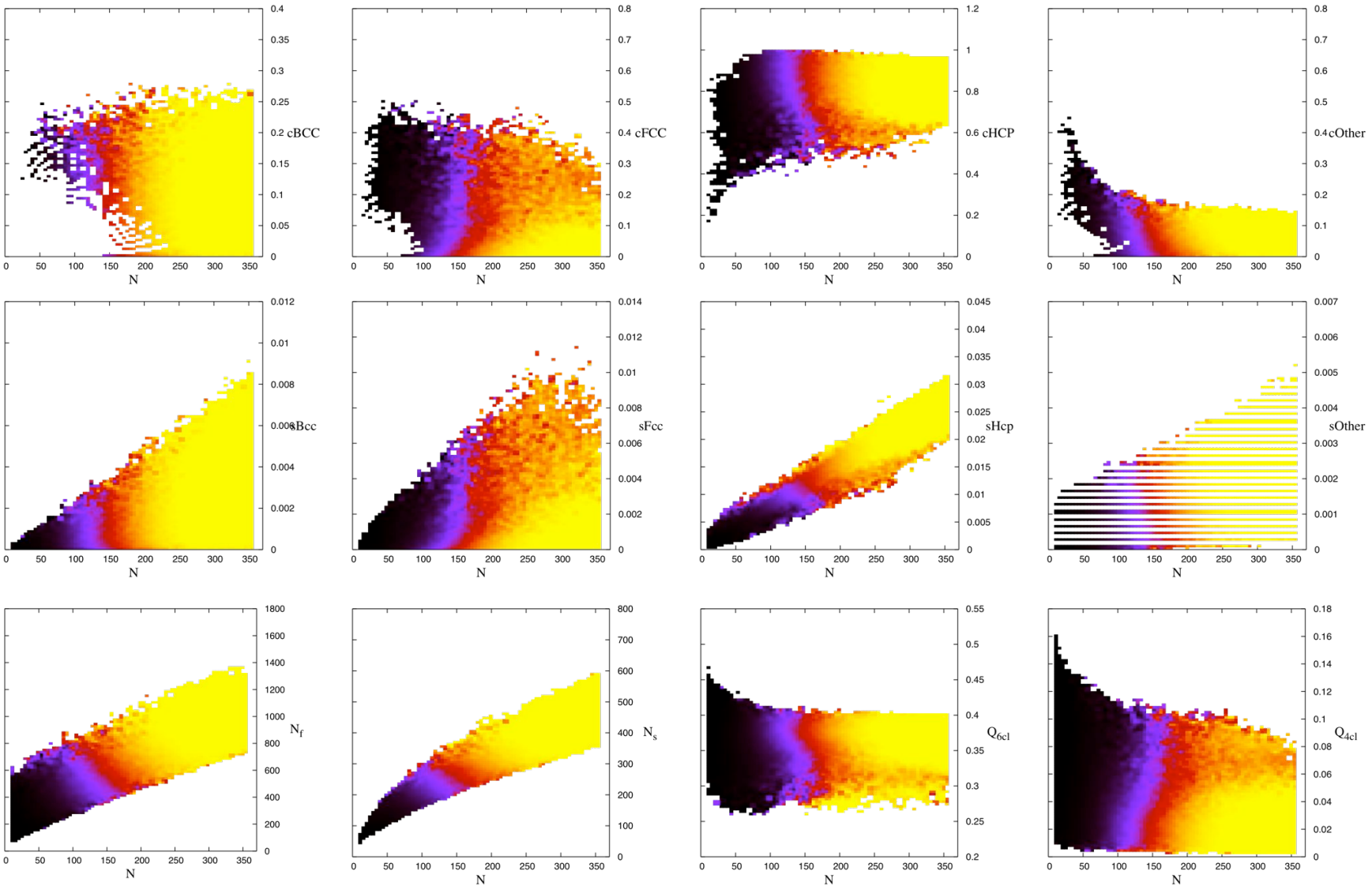


- rates :
  - $k_{high} = 6.32 \times 10{-}7$ ($\Phi_{A,\lambda 0} = 0.018$)
  - $k_{low} = 1.53 \times 10{-}6$ ($\Phi_{A,\lambda 0} = 0.092$)

Free Energy Projection



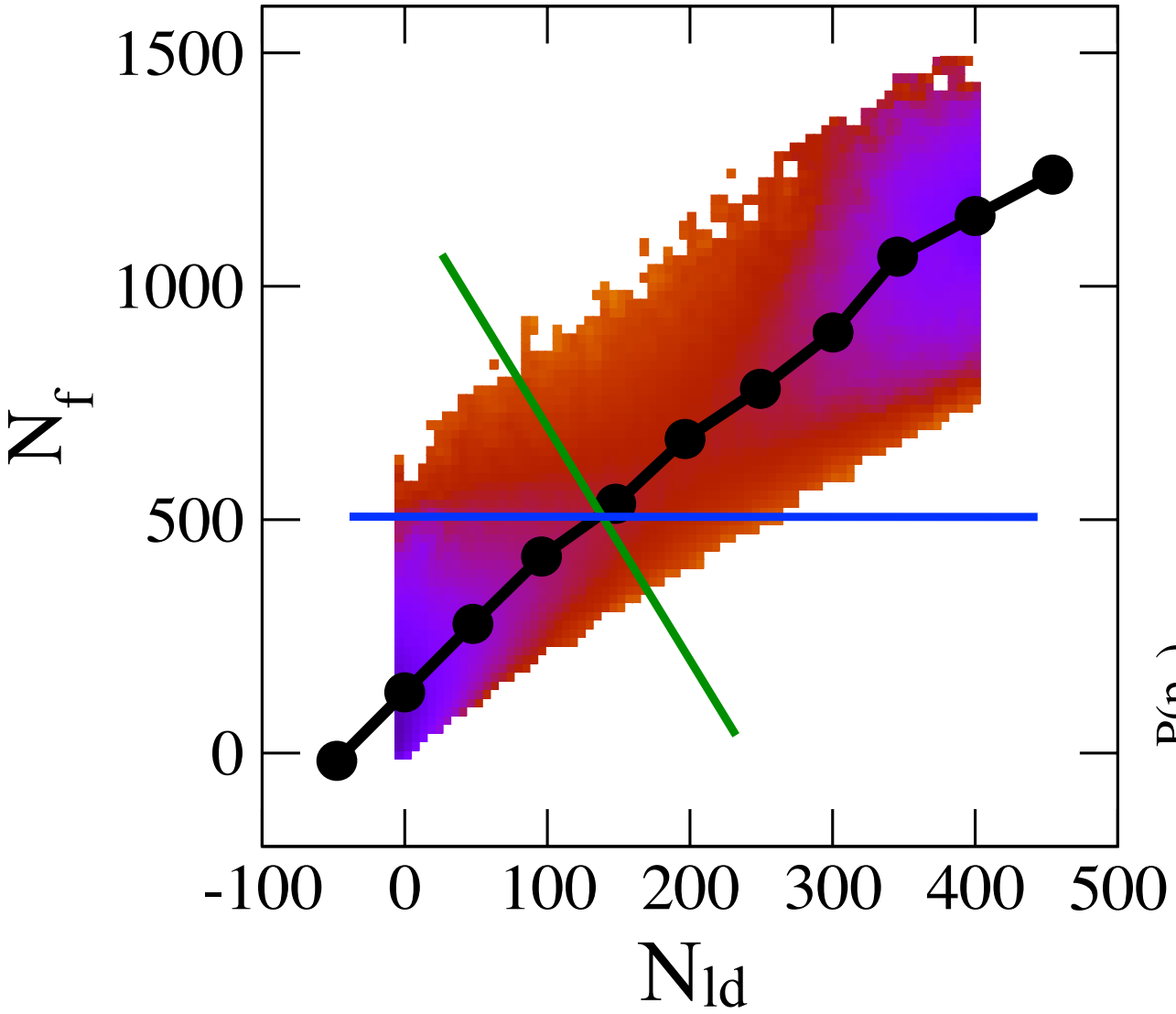
Committor Surface

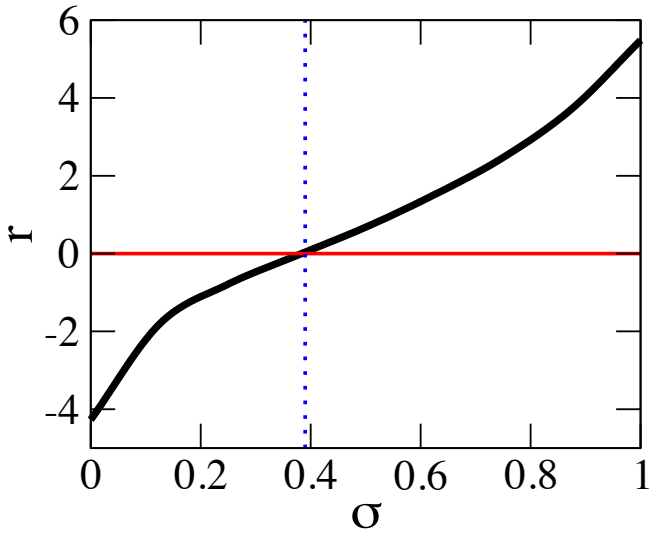


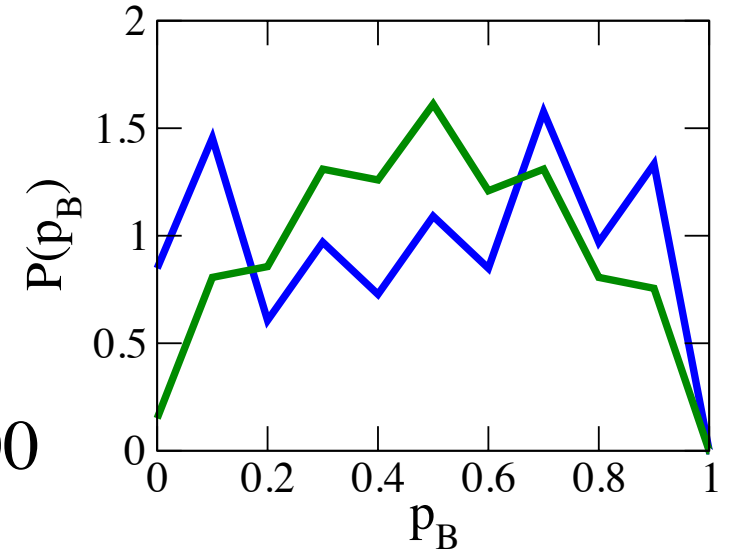
Order parameters

N_{cl}	average bond order parameter
N_f	standard ten Wolde Frenkel order parameter
n_{bcc}	fraction of bcc particles in cluster
n_{fcc}	fraction of fcc particles in cluster
n_{hcp}	fraction of hcc particles in cluster
n_{und}	fraction of undefined particles in cluster
s_{bcc}	fraction of bcc particles in system
s_{fcc}	fraction of fcc particles in system
s_{hcp}	fraction of hcp particles in system
s_{und}	fraction of undefined particles in system
s_{sol}	fraction of solid particles in system
s_{liq}	fraction of liquid particles in system
Q_{6cl}	local q6 of the cluster
Q_{4cl}	local q4 of cluster
Ns	number of liquid particles next to cluster
Nl	number of links to liquid particles

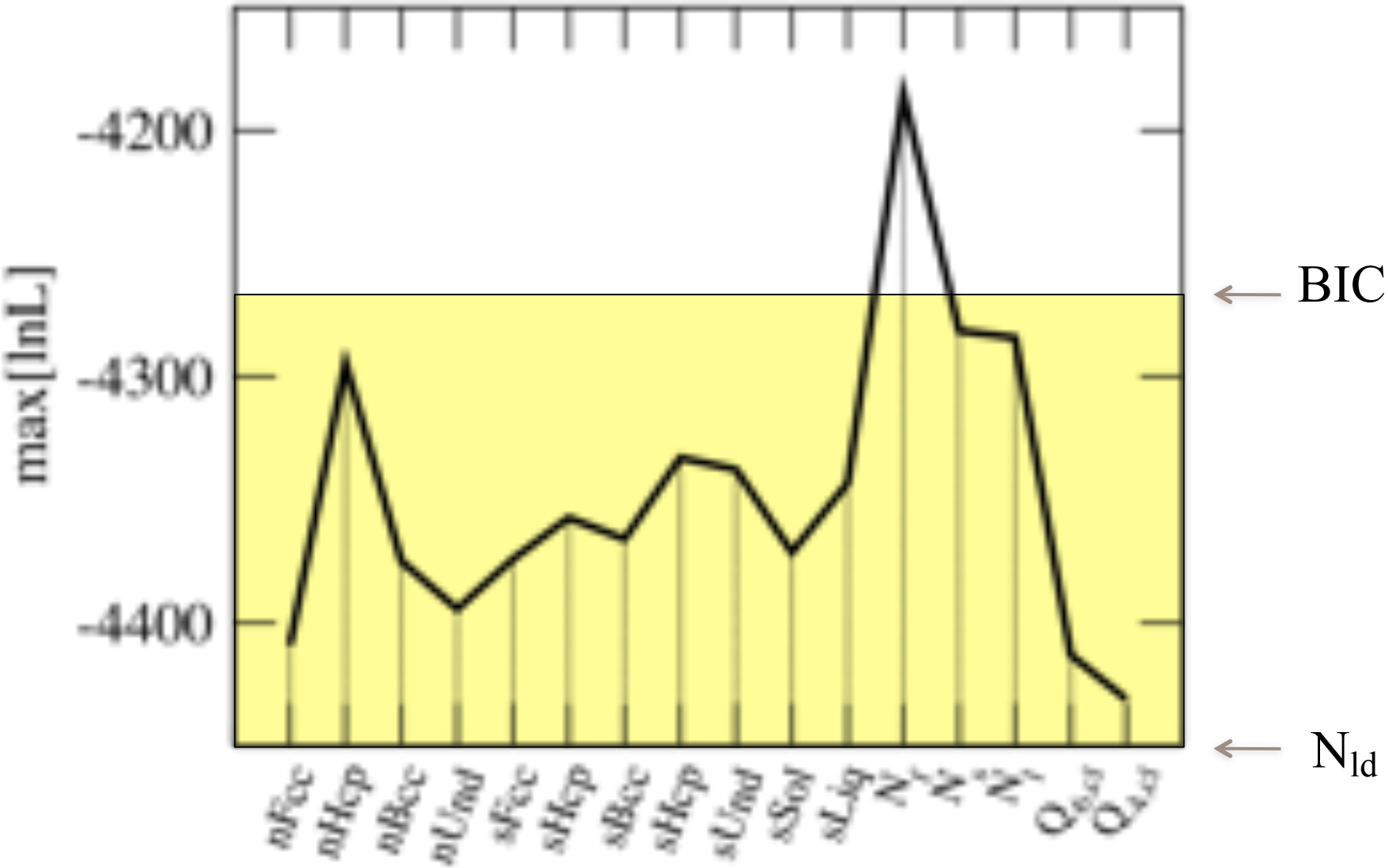
Optimal reaction coordinate



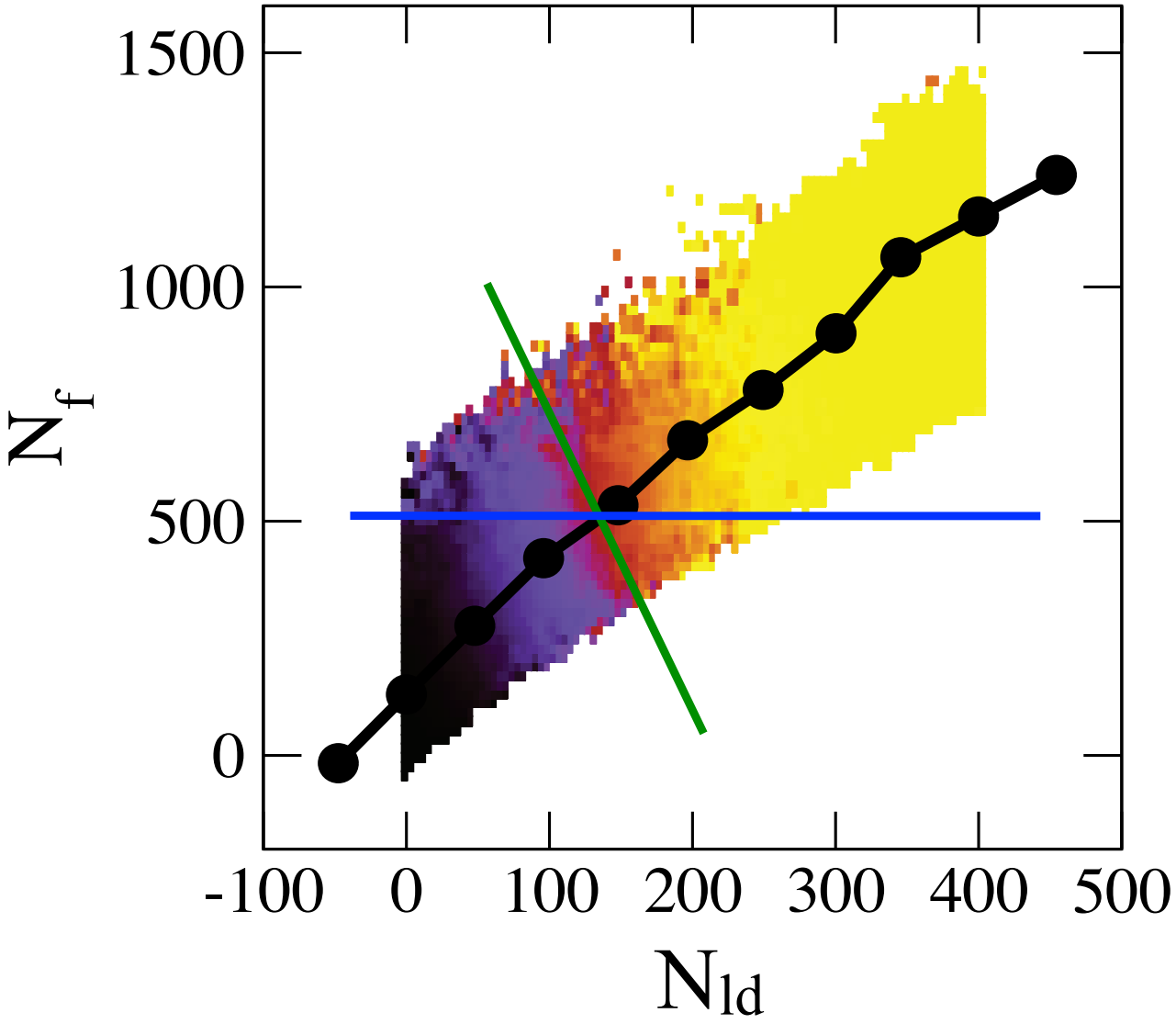


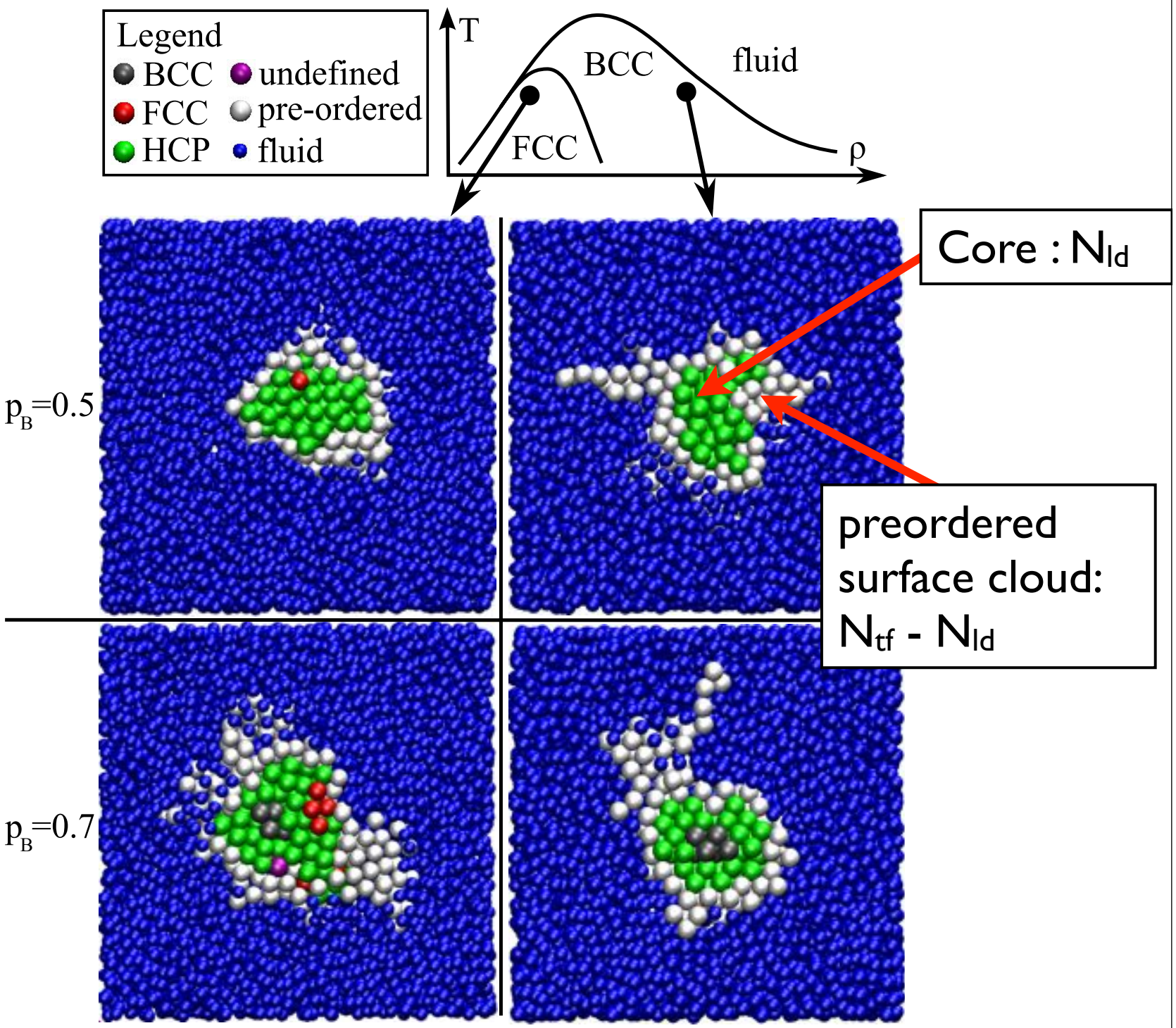


Optimal Reaction Coordinate

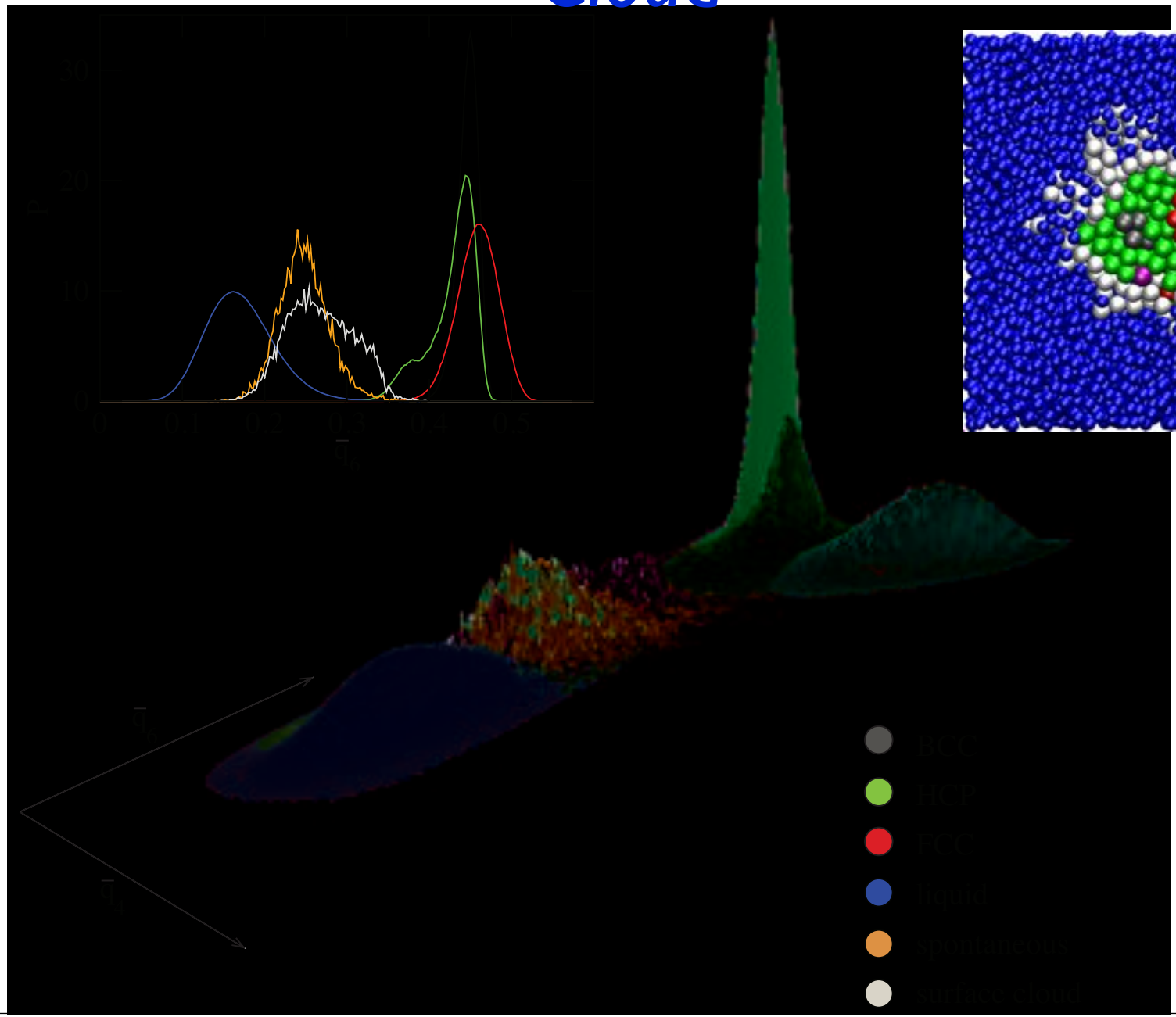


Transition State




Legend
BCC
undefined
FCC
pre-ordered
HCP
fluid
T
BCC
fluid
FCC
ρ
Core : N_{ld}
preordered surface cloud: N_{tf} - N_{ld}
p_B=0.5
p_B=0.7

Structural Properties of the Surface Cloud

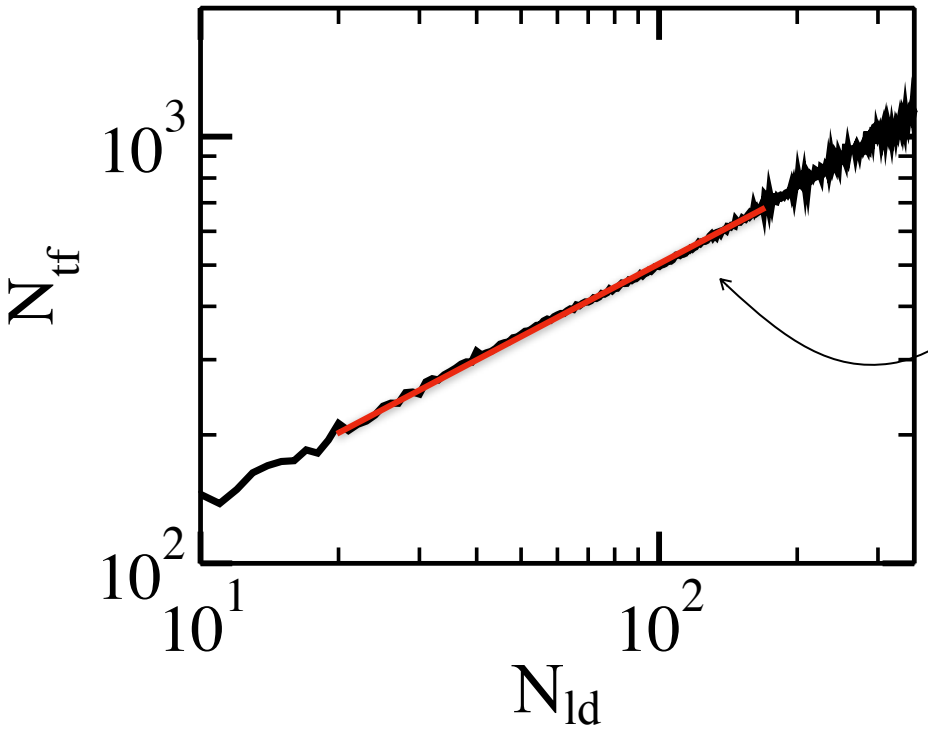


Classical nucleation theory revisited

$$\Delta G = \Delta\mu N_{ld} + \gamma(N_{tf} - N_{ld}) = c_1 + c_2 N_{ld} + c_3 N_{tf}$$

Core

preordered surface cloud

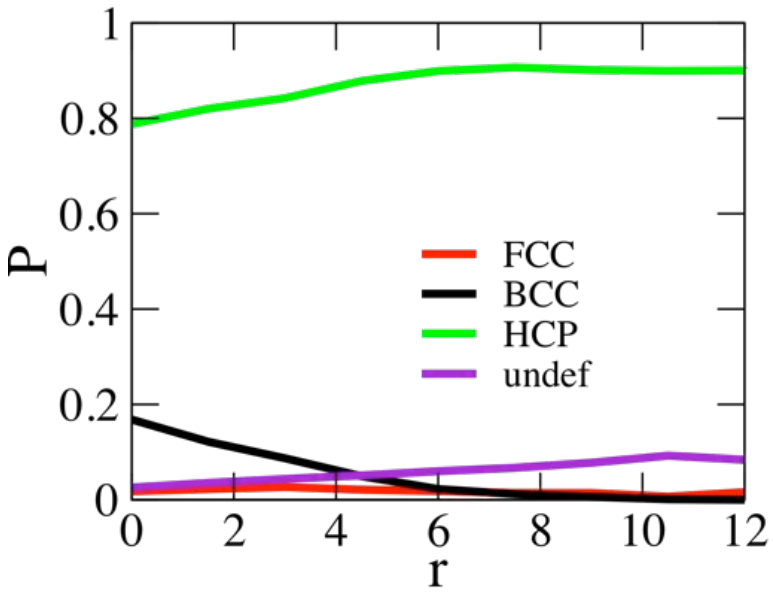
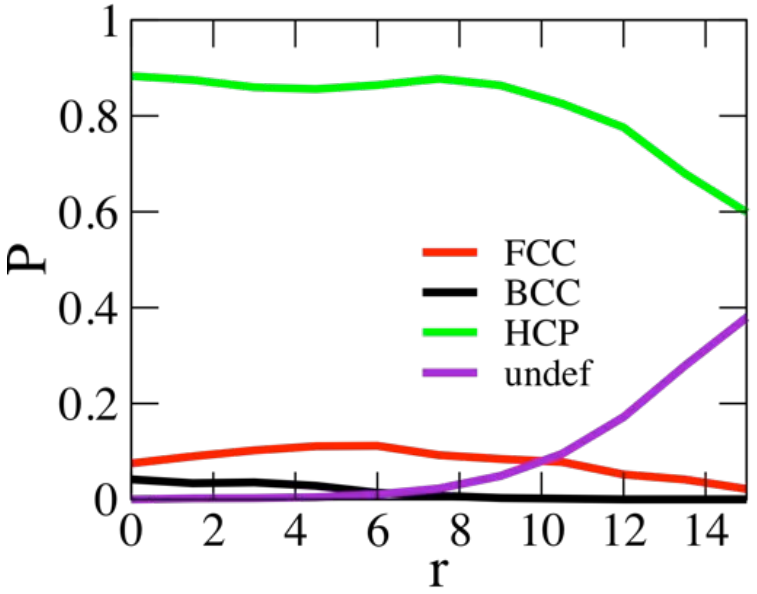
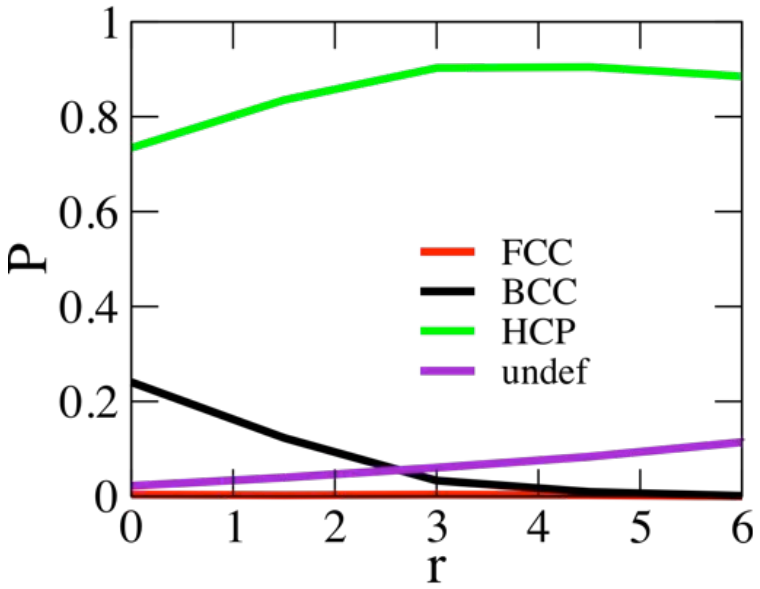
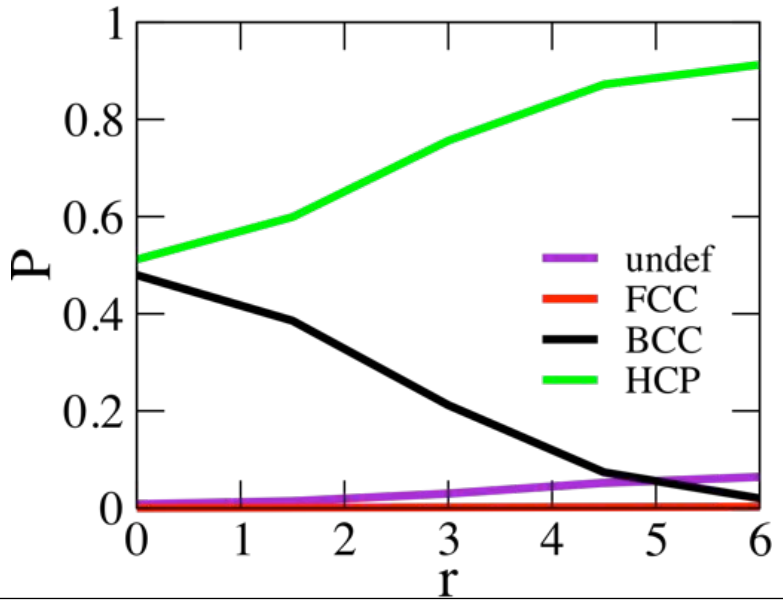


$$N_{tf} \sim N_{ld}^{2/3}$$

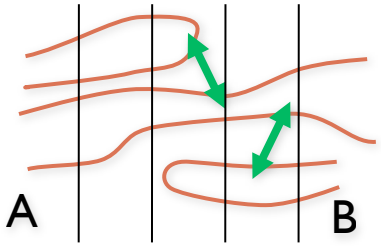
$$\Delta G = c_1 + c_2 N_{ld} + c_3 N_{ld}^{2/3}$$

No spherical assumption needed!

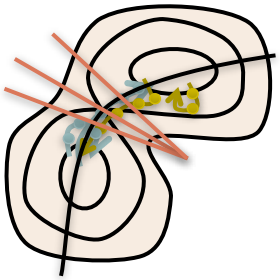
Ostwald Step-Rule

	Critical Size	Post-Critical Size
low pressure FCC		
high pressure BCC		

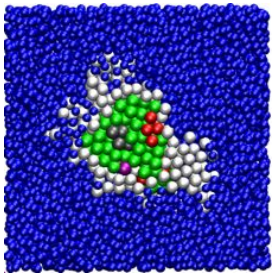
Take home messages



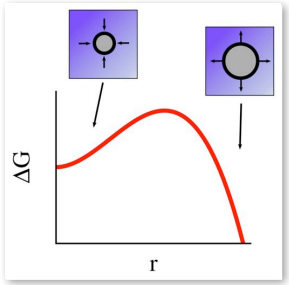
Replica exchange TIS can sample the unbiased complete path ensemble by reweighting for any complex rare event



Non-linear committor analysis of the complete path ensemble can find optimal reaction coordinates for any complex rare event



The pre-ordered surface cloud and crystalline core as revealed by novel order parameters are important reaction coordinates for nucleation



Free energy profile scales like CNT by using surface cloud and core: removing the assumption of spherical clusters.

explains why CNT works even if nuclei are clearly aspherical