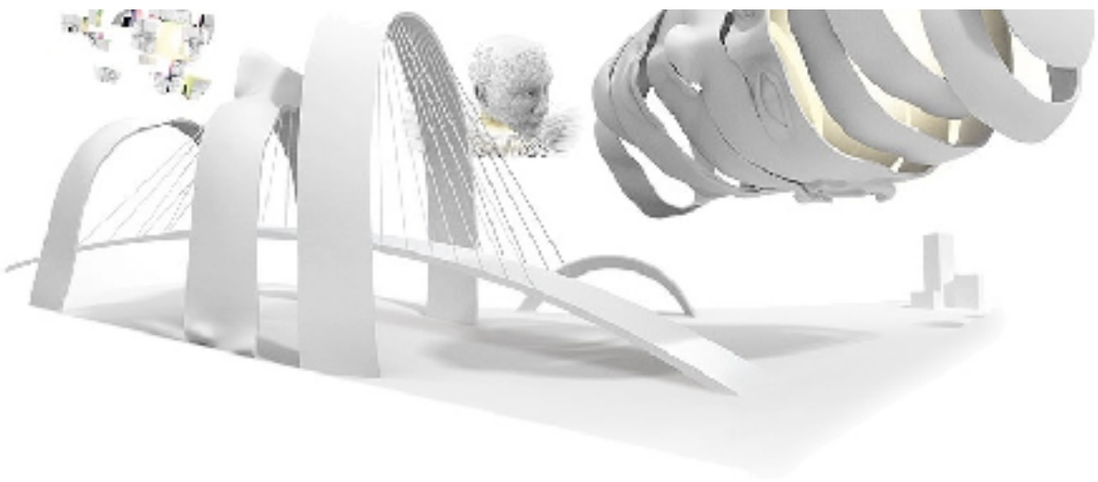
NTNU

Norwegian University of
Science and Technology



Advances in Computational Statistical Physics

Perspectives en physique statistique computationnelle

17 - 21 September 2018

CIRM-Marseille (France) 2018

Titus van Erp

NTNU, Trondheim, Norway
University of Ghent, Belgium

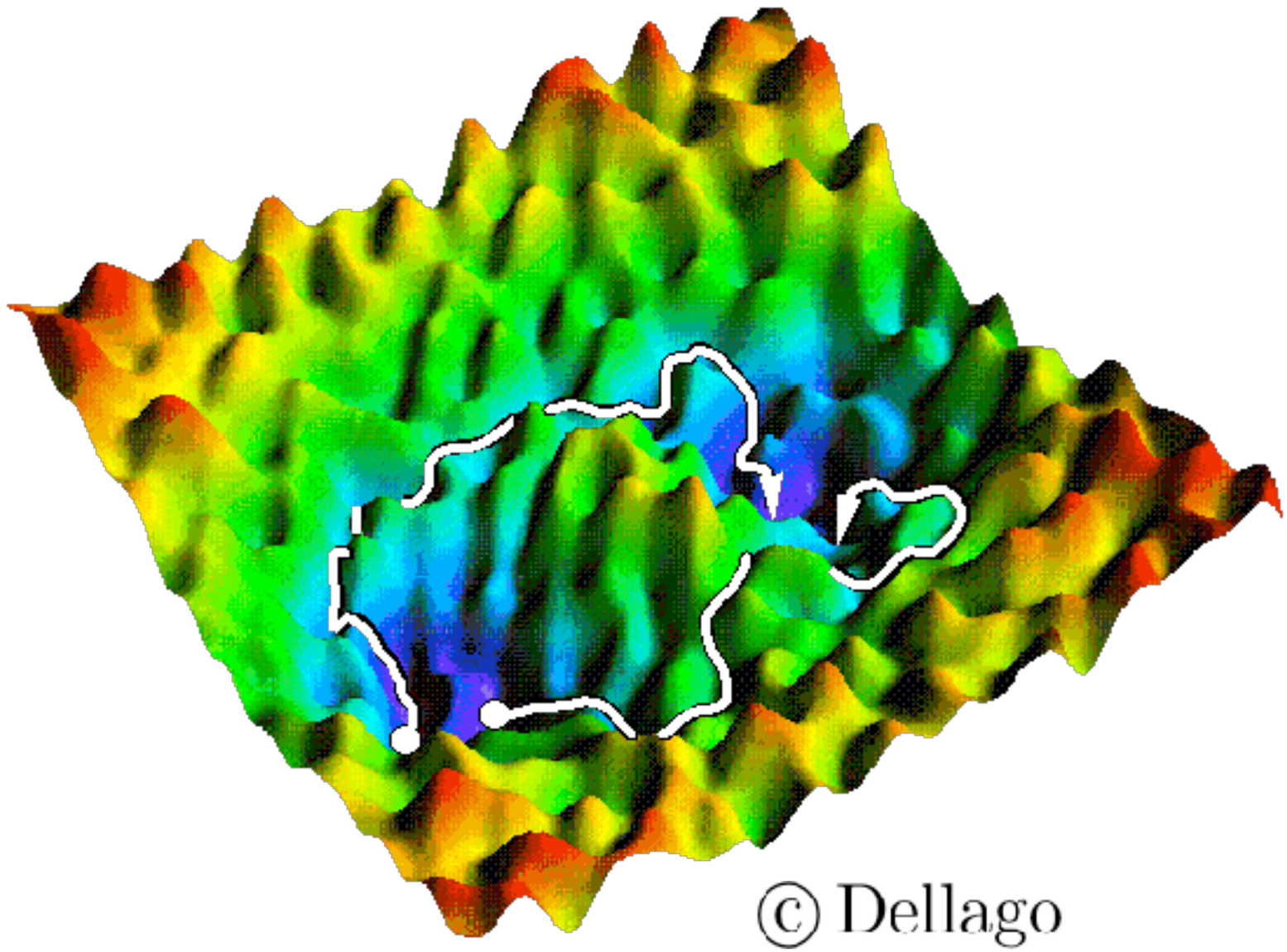
Transition path sampling approaches for the study of rare events

Outline

- General introduction on path sampling: TPS -> TIS -> RETIS
- A demonstration (life simulation using PyRETIS!)
- Some more path sampling approaches: PPTIS/Milestoning, TS-PPTIS, FFS
- New MC moves: Stone Skipping and Web Throwing
- Machine Learning for identifying reaction triggers using TIS/RETIS/FFS data
- Conclusions, prospectives, and challenges

Titus S. van Erp

Transition Path Sampling (TPS): Sampling of unbiased dynamical trajectories using a Monte Carlo approach (Dellago, Bolhuis, Chandler 1998)



2003: Transition Interface Sampling (TIS):

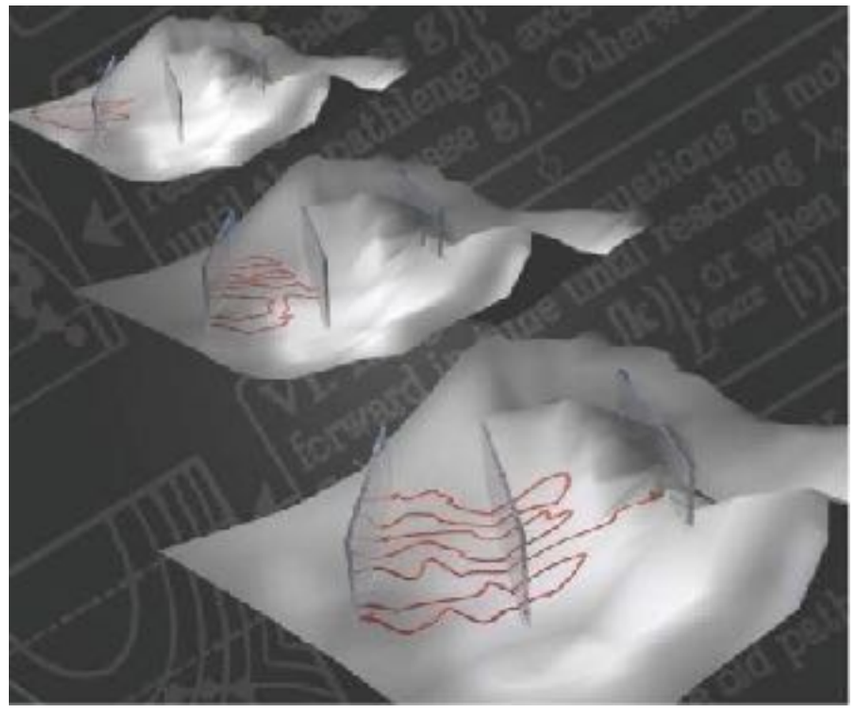
Improves the original TPS approach by:

Different theoretical rate equations based on *overall states*

Path ensembles based on interfaces

Flexible path length ensemble

Faster convergence due to *effective positive flux* expression



2007: Replica Exchange Transition Interface Sampling (RETIS):

Improvement of TIS (but more complex to implement) by:

Replica Exchange (RE) moves between path ensembles

Fully based on path ensembles, no initial MD run: increases flexibility of RE moves.

Titus S. van Erp

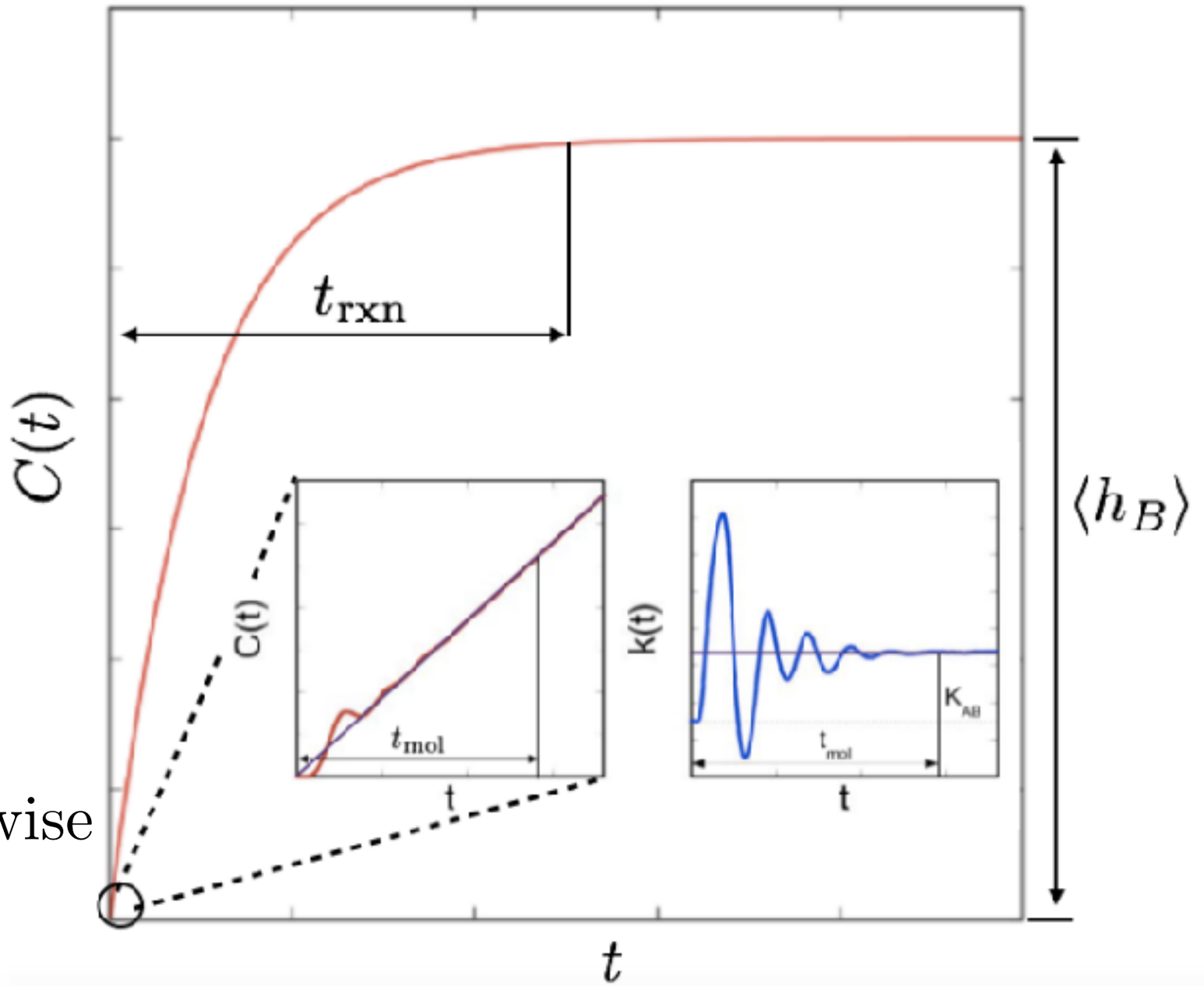
Rate constant theory in Transition Path Sampling vs (RE)TIS

$$k_{AB} = k(t') \text{ for } t_{\text{mol}} < t' \ll t_{\text{rxn}}$$

$$k(t) = \frac{dC(t)}{dt},$$

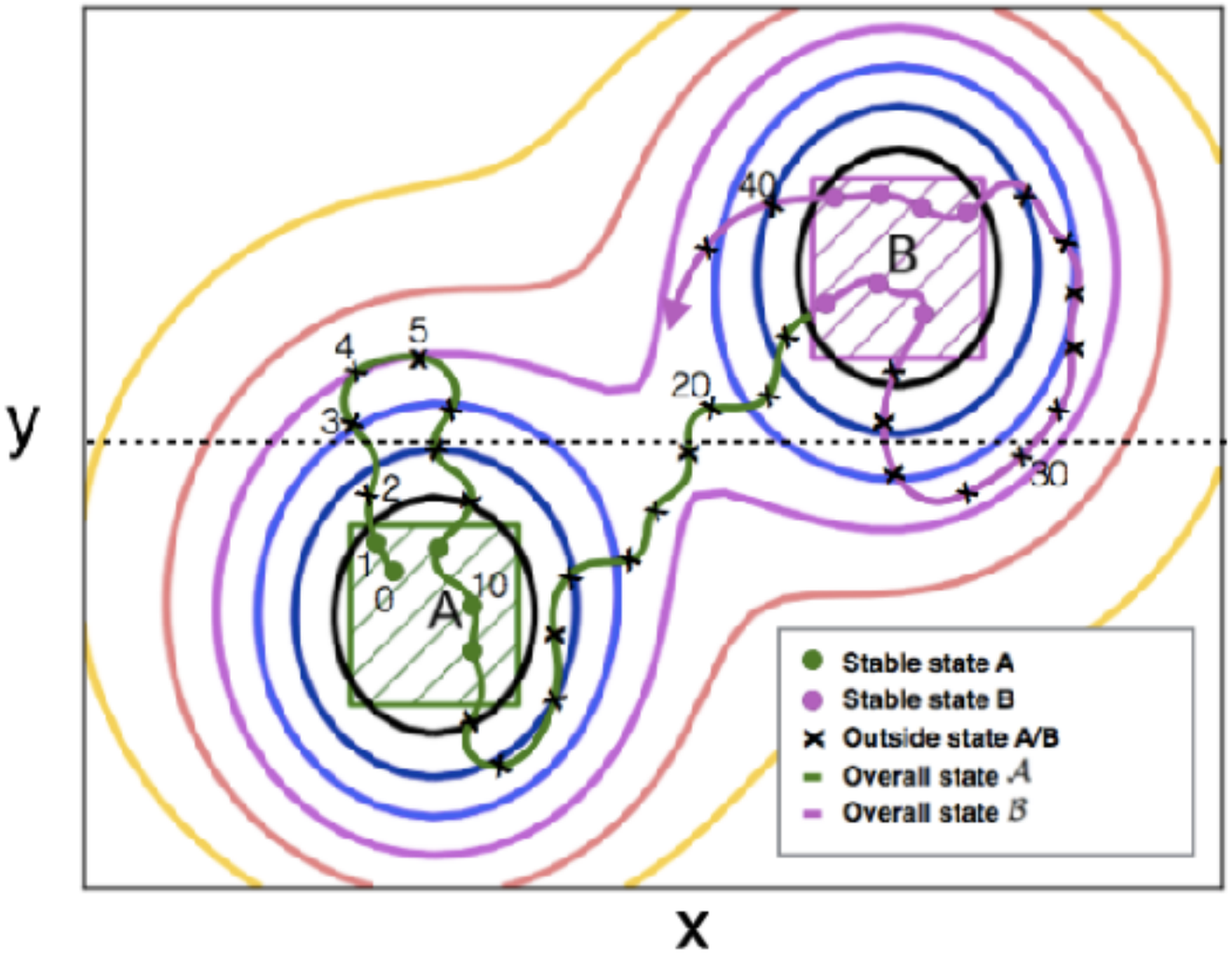
$$C(t) = \frac{\langle h_A(0) h_B(t) \rangle}{\langle h_A \rangle}$$

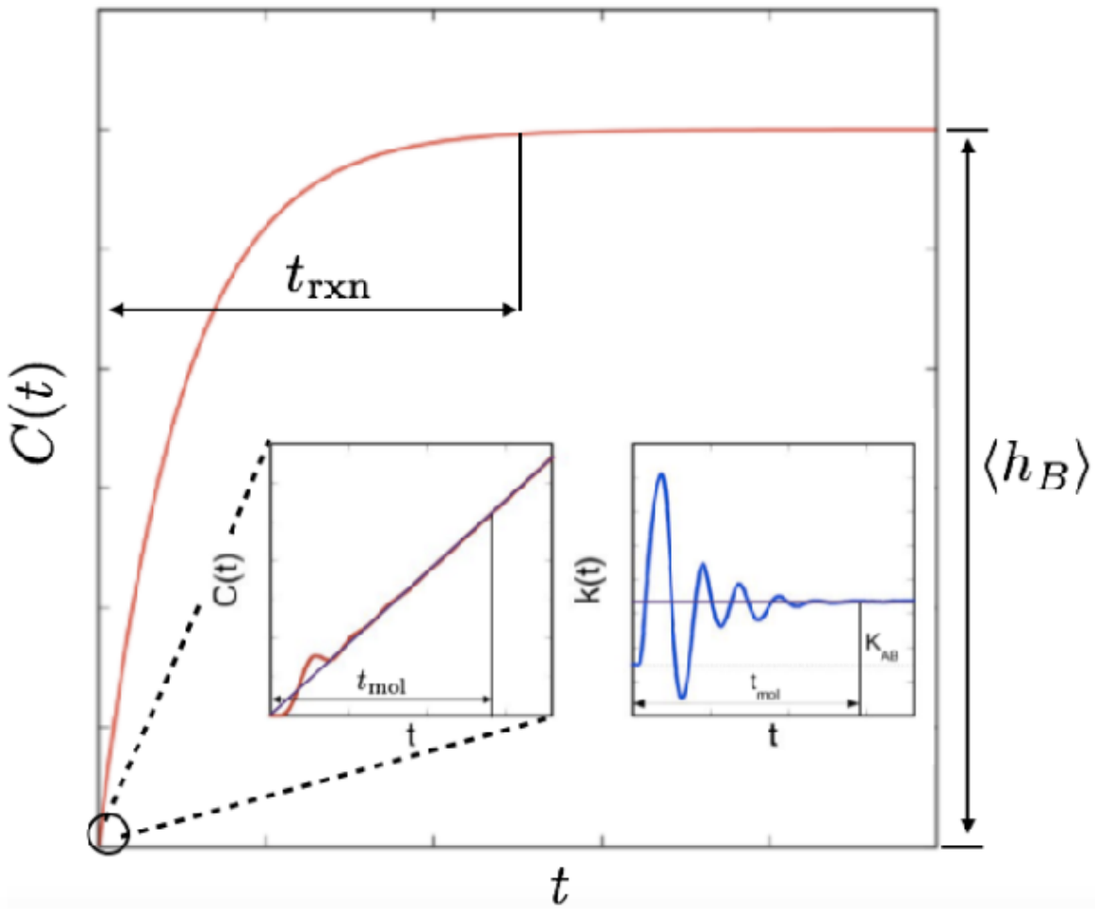
$$h_X = 1 \text{ if } \in X \text{ and 0 otherwise}$$



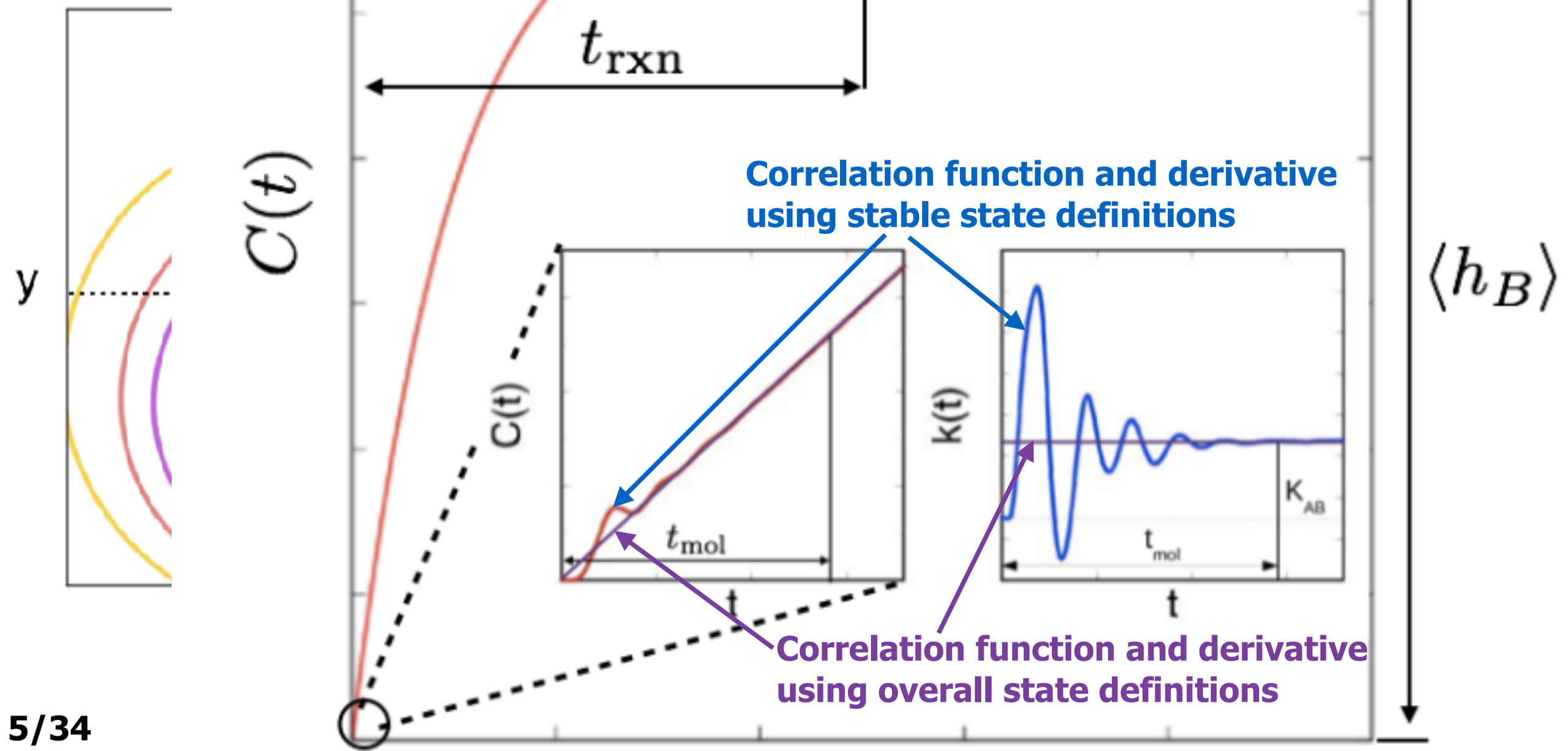
Titus S. van Erp

Rate constant theory in Transition Path Sampling vs (RE)TIS

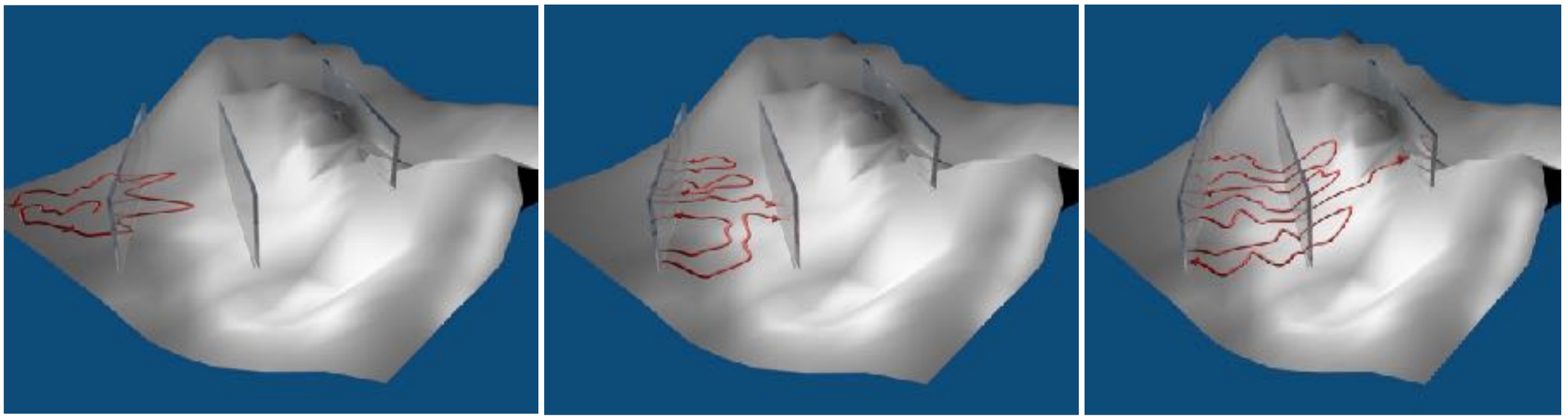




Rate constant theory in Transition Path Sampling vs (RE)TIS



The TIS Algorithm



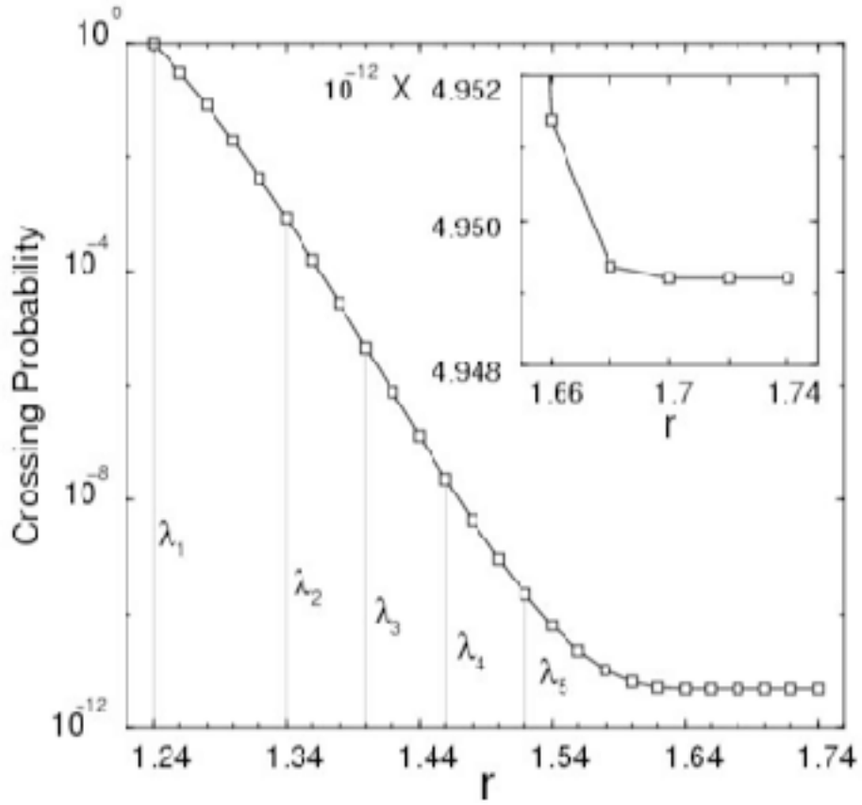
MD **[0⁺] path ensemble** **[1⁺] path ensemble**

Transition Interface Sampling, van Erp, Moroni, and Bolhuis, J. Chem. Phys. 118, 7762 (2003)

The TIS Algorithm

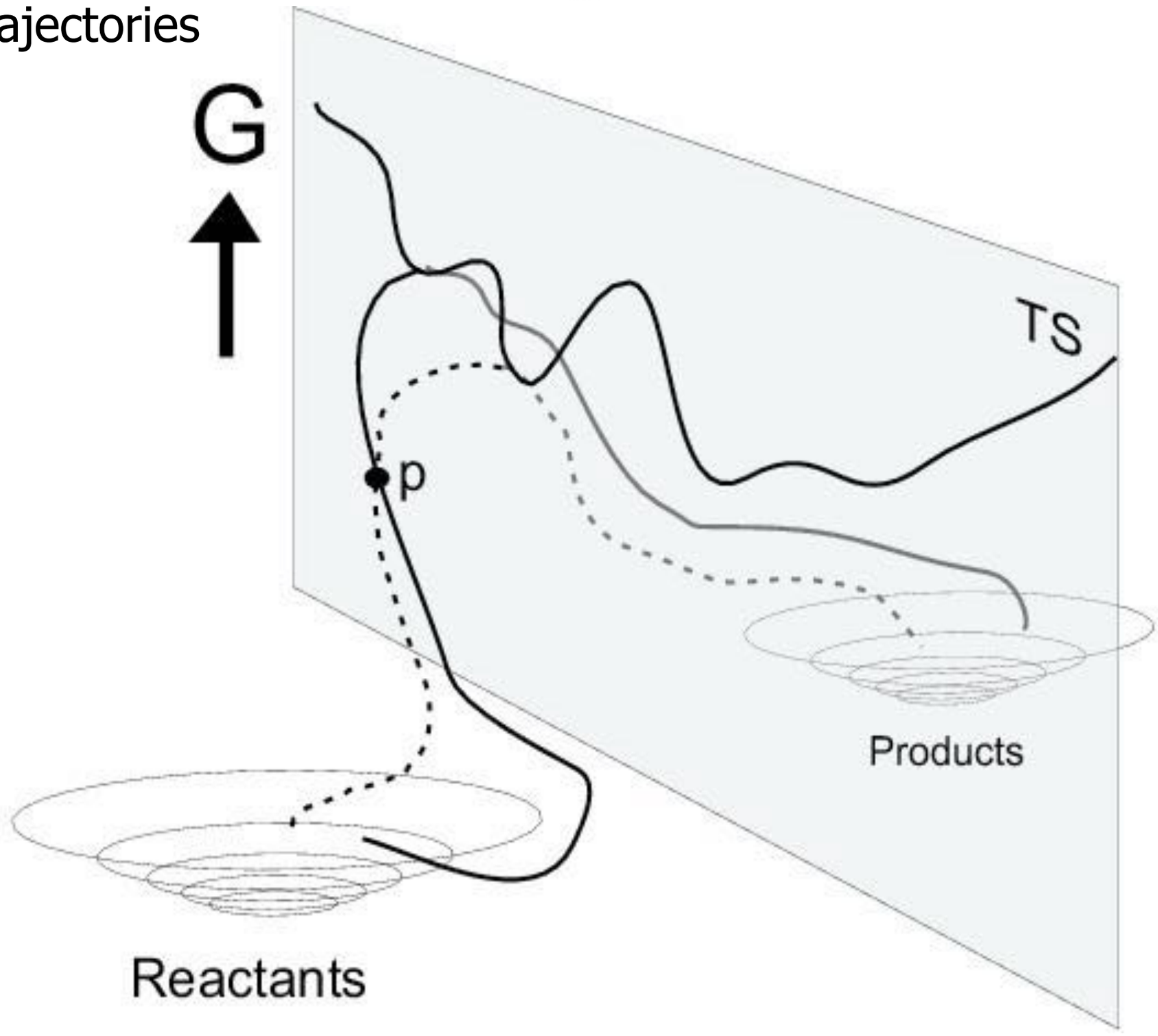
$$k_{AB} = f_A \mathcal{P}_A(\lambda_B|\lambda_A)$$

$$\mathcal{P}_A(\lambda_B|\lambda_A) = \mathcal{P}_A(\lambda_n|\lambda_0) = \prod_{i=0}^{n-1} \mathcal{P}_A(\lambda_{i+1}|\lambda_i)$$



Transition Interface Sampling, van Erp, Moroni, and Bolhuis, J. Chem. Phys. 118, 7762 (2003)

MC sampling of MD trajectories



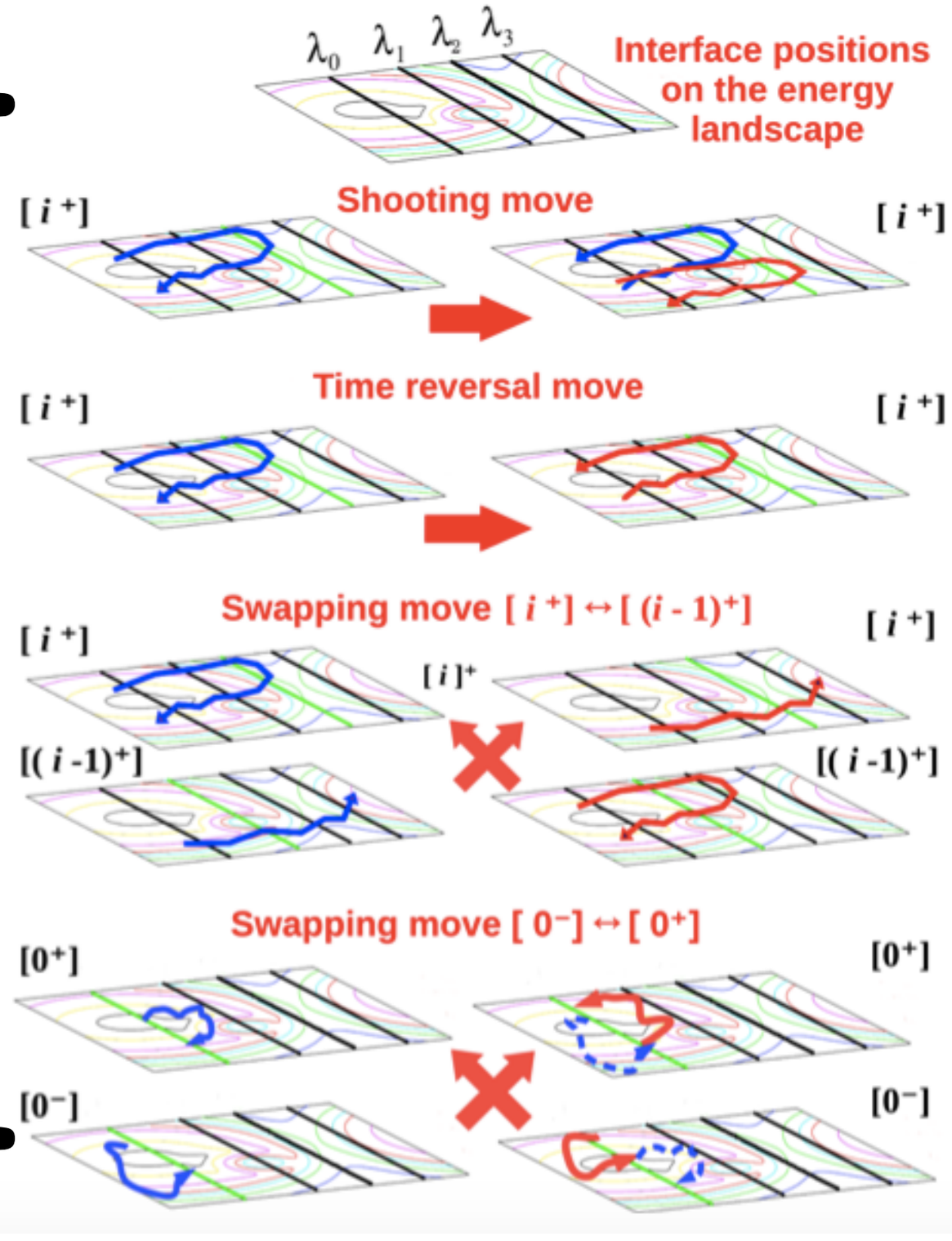
picture by B. Ensing

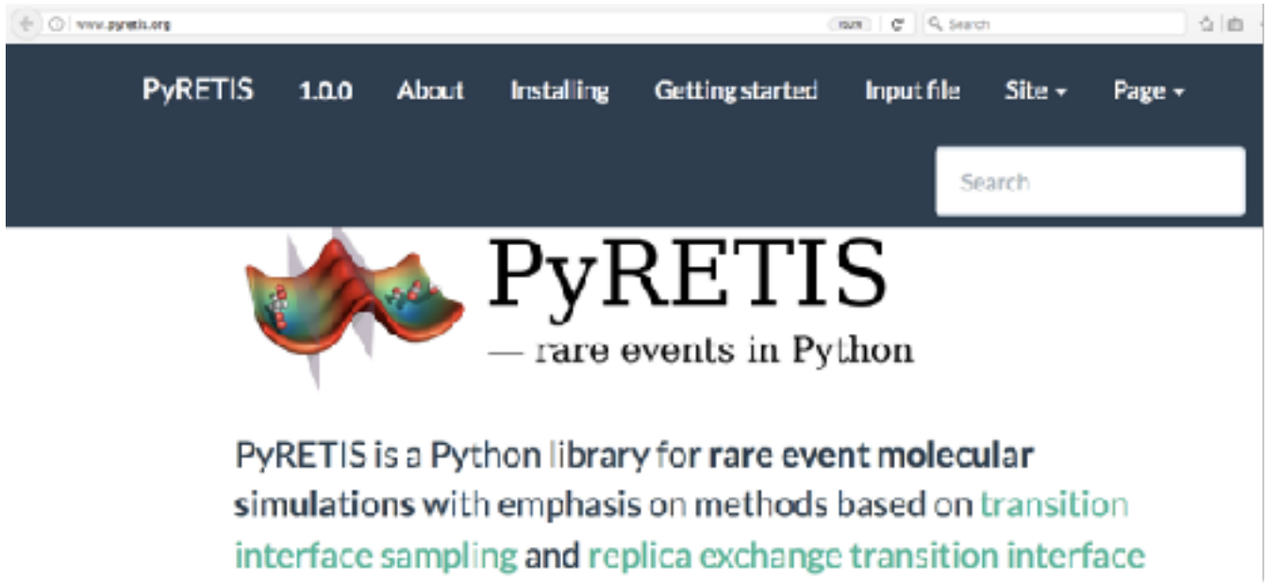
Titus van Erp

Replica Exchange TIS:
T. S. Van Erp, Phys. Rev. Lett. 2007

No MD
Swap-Moves

MC moves:



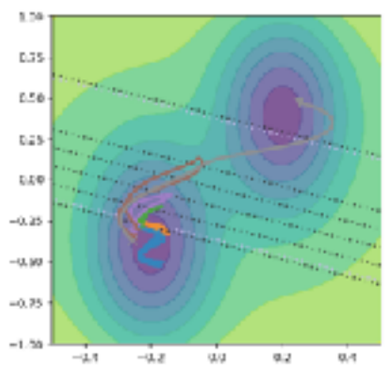


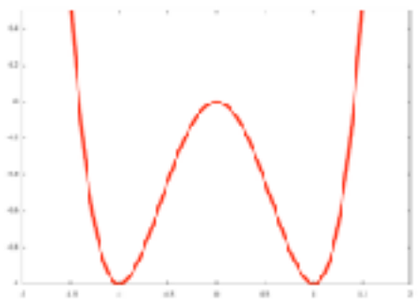
PyRETIS: A well-done, medium-sized python library for rare events

Anders Lervik, Enrico Riccardi, Titus van Erp, J. Comput. Chem. 2017

And now a demonstration!

(Sponsored by the Olav Thon foundation for combined educational/research projects)





Conclusions part I:

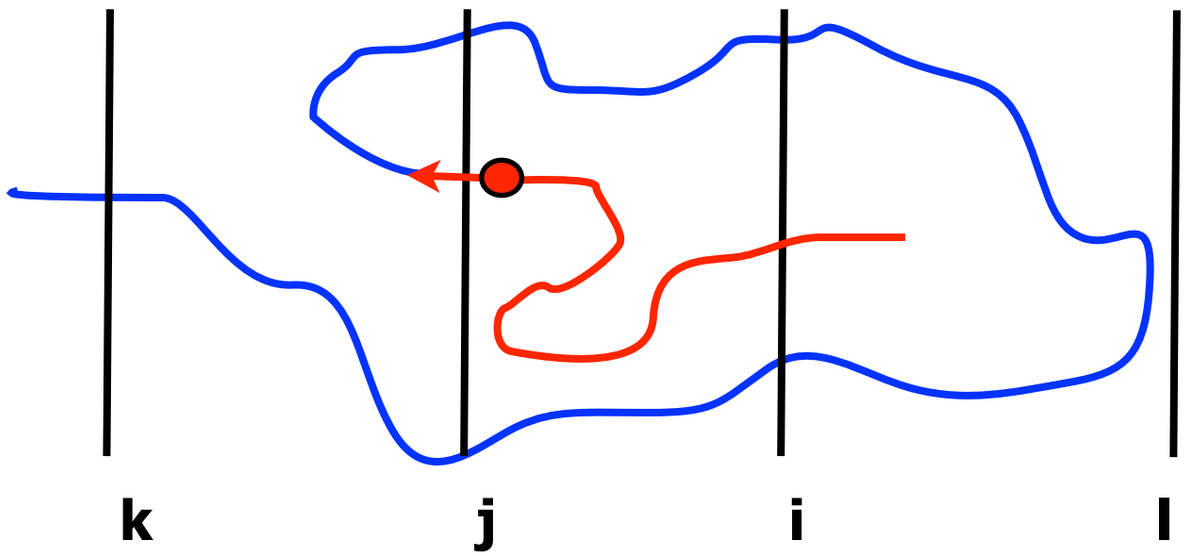
- Path sampling (TPS) is a MC sampling of short MD trajectories
- Can be applied for all kind of dynamics: Newtonian, Langevin, Brownian, kinetic MC
- TIS and later RETIS have improved the efficiency of the original TPS approach without the need of approximations.
- Implementation of RETIS is non-trivial but user-friendly open-source programs have now been developed (PyRETIS, OPS)

More about path algorithms

- Approximative local path methods: PPTIS/Milestoning, TS-PPTIS
- Splitting based path method: FFS
- Efficiency and reaction coordinate

The conditional crossing probability

$$P\left(\begin{matrix}k\\l\end{matrix}\middle|\begin{matrix}j\\i\end{matrix}\right)$$



The conditional crossing probability

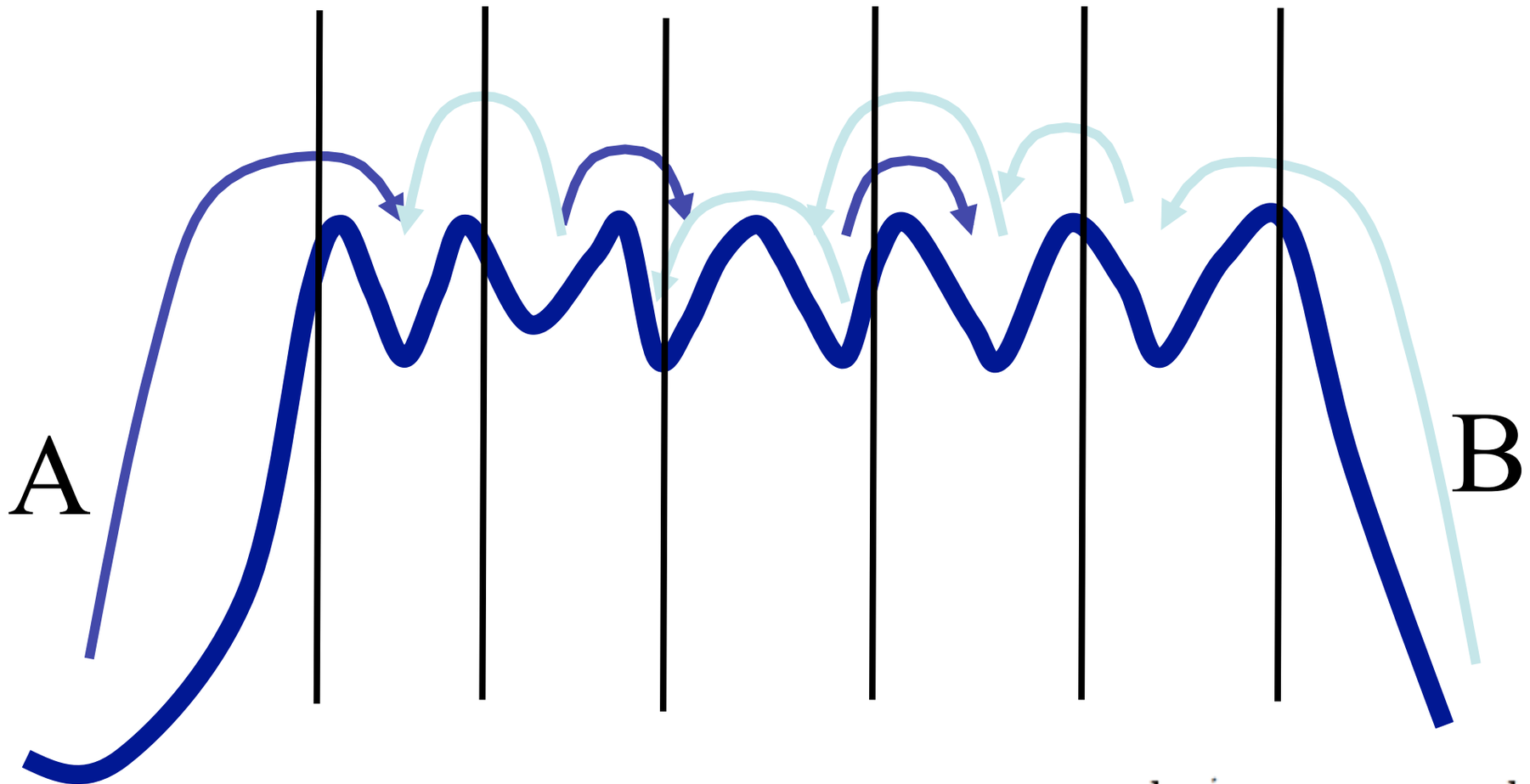
$$P\binom{k}{l}\Big|\binom{j}{i}$$

$$\mathcal{P}_A(\lambda_{i+1}|\lambda_i) = P\binom{i+1}{0}\Big|\binom{i}{0}$$

$$\begin{aligned} k_{AB} &= f_A \mathcal{P}_A(\lambda_B|\lambda_A) \\ \mathcal{P}_A(\lambda_B|\lambda_A) &= \mathcal{P}_A(\lambda_n|\lambda_0) = \prod_{i=0}^{n-1} \mathcal{P}_A(\lambda_{i+1}|\lambda_i) \end{aligned}$$

Partial Path TIS

D. Moroni, P. G. Bolhuis, and T. S. van Erp, JCP **120**, 4055 (2004)

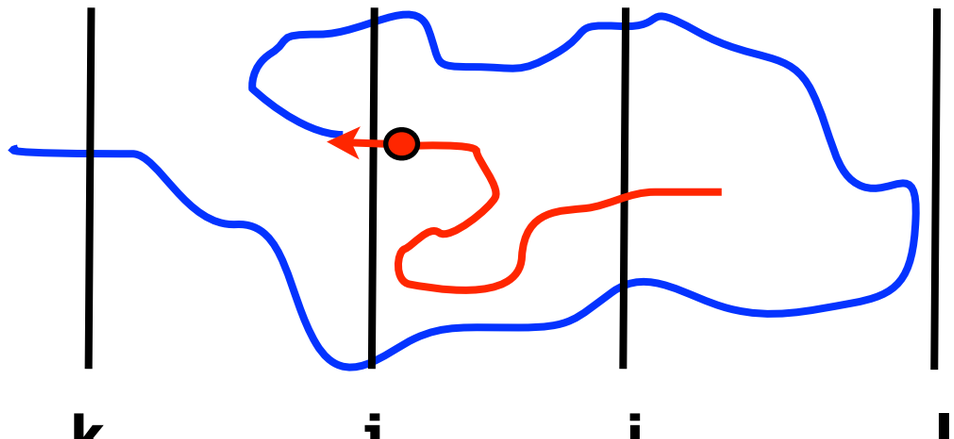


PPTIS: time-independent. Memory-loss assumption: $P(^{l}_{k}|^{j}_{j\pm m}) \approx P(^{l}_{k}|^{j}_{j\pm 1})$

Milestoning: time-dependent hopping sequence. Assumes an equilibrium at each Milestone (stronger assumption)

Partial Path TIS

$$P\binom{k}{l}\Big|\binom{j}{i}$$



$$p_i^{\pm} \equiv P\binom{i+1}{i-1}\Big|\binom{i}{i-1}, \quad p_i^{\mp} \equiv P\binom{i-1}{i+1}\Big|\binom{i}{i+1},$$

$$p_i^{=} \equiv P\binom{i-1}{i+1}\Big|\binom{i}{i-1}, \quad p_i^{\ddagger} \equiv P\binom{i+1}{i-1}\Big|\binom{i}{i+1},$$

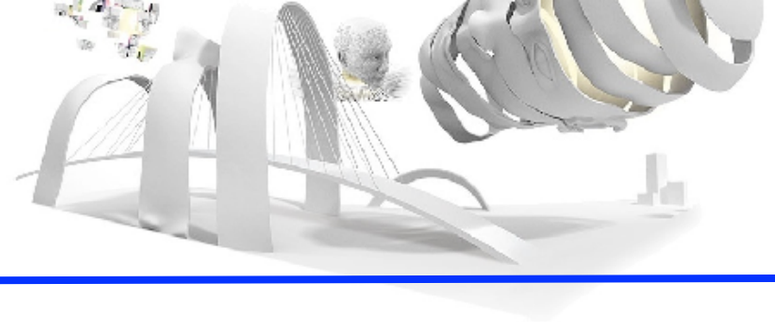
which fulfill the following relations:

$$p_i^{\pm} + p_i^{=} = p_i^{\mp} + p_i^{\ddagger} = 1.$$

Short-distance hopping probabilities

$$P_i^{+} \equiv P\binom{i}{0}\Big|\binom{1}{0}, \quad P_i^{-} \equiv P\binom{0}{i}\Big|\binom{i-1}{i}.$$

Long-distance hopping probabilities



Partial Path TIS

$$p_i^{\pm} \equiv P(^{i+1}_{i-1}|^{i}_{i-1}), \quad p_i^{\mp} \equiv P(^{i-1}_{i+1}|^{i}_{i+1}),$$

$$p_i^{=} \equiv P(^{i-1}_{i+1}|^{i}_{i-1}), \quad p_i^{\ddagger} \equiv P(^{i+1}_{i-1}|^{i}_{i+1}),$$

which fulfill the following relations:

$$p_i^{\pm} + p_i^{=} = p_i^{\mp} + p_i^{\ddagger} = 1.$$

$$P_i^{+} \equiv P(^{i}_{0}|^{1}_{0}), \quad P_i^{-} \equiv P(^{0}_{i}|^{i-1}_{i}).$$

$$P_j^{+} = \frac{p_{j-1}^{\pm} P_{j-1}^{+}}{p_{j-1}^{\pm} + p_{j-1}^{=} P_{j-1}^{-}},$$

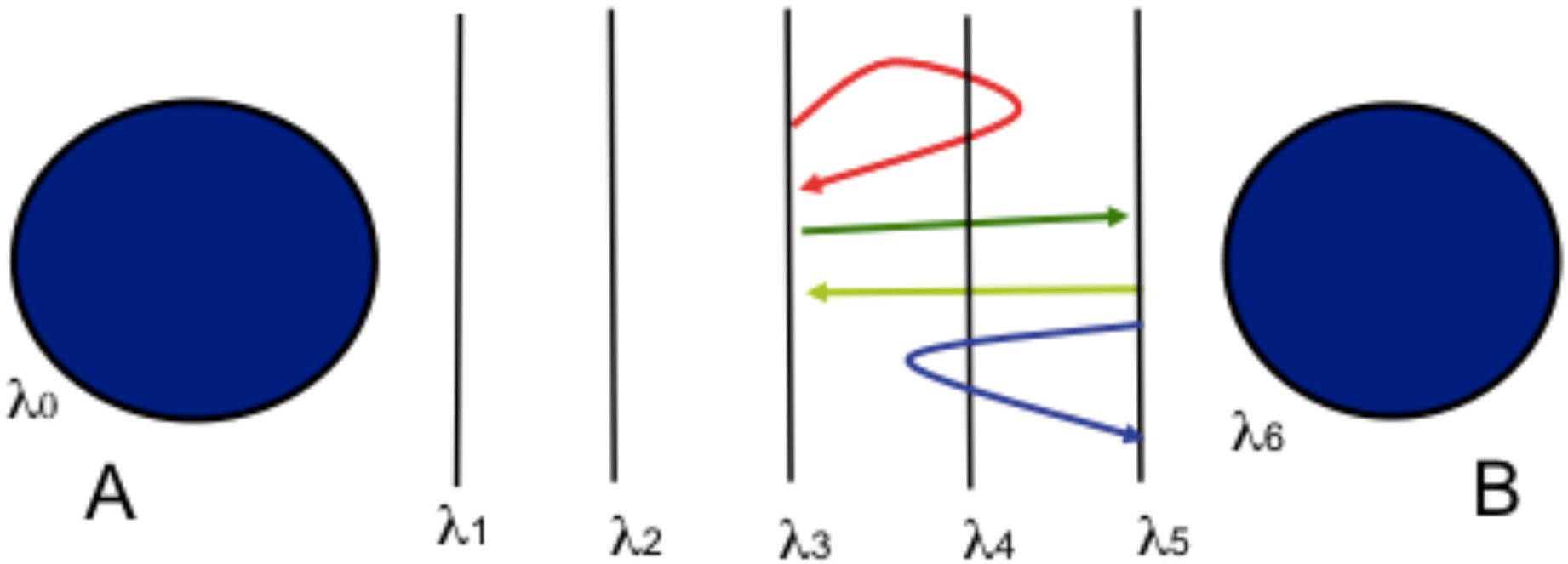
$$P_j^{-} = \frac{p_{j-1}^{\mp} P_{j-1}^{-}}{p_{j-1}^{\pm} + p_{j-1}^{=} P_{j-1}^{-}}.$$

$$P_1^{+} = P_1^{-} = 1$$

$$\mathcal{P}_A(\lambda_B|\lambda_A) = \mathcal{P}_A(\lambda_n|\lambda_0) = P_n^{+}$$

if we choose: $\lambda_1 = \lambda_0 + \epsilon$

Partial Path sampling



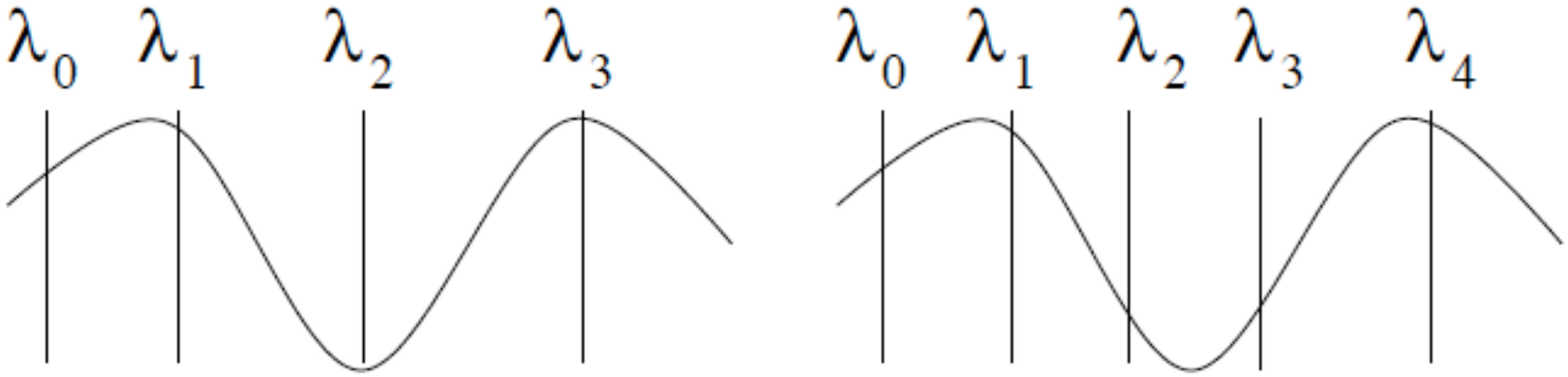
$$p_4^{\pm} = \frac{\#p_{35}}{\#p_{33}+\#p_{35}}$$

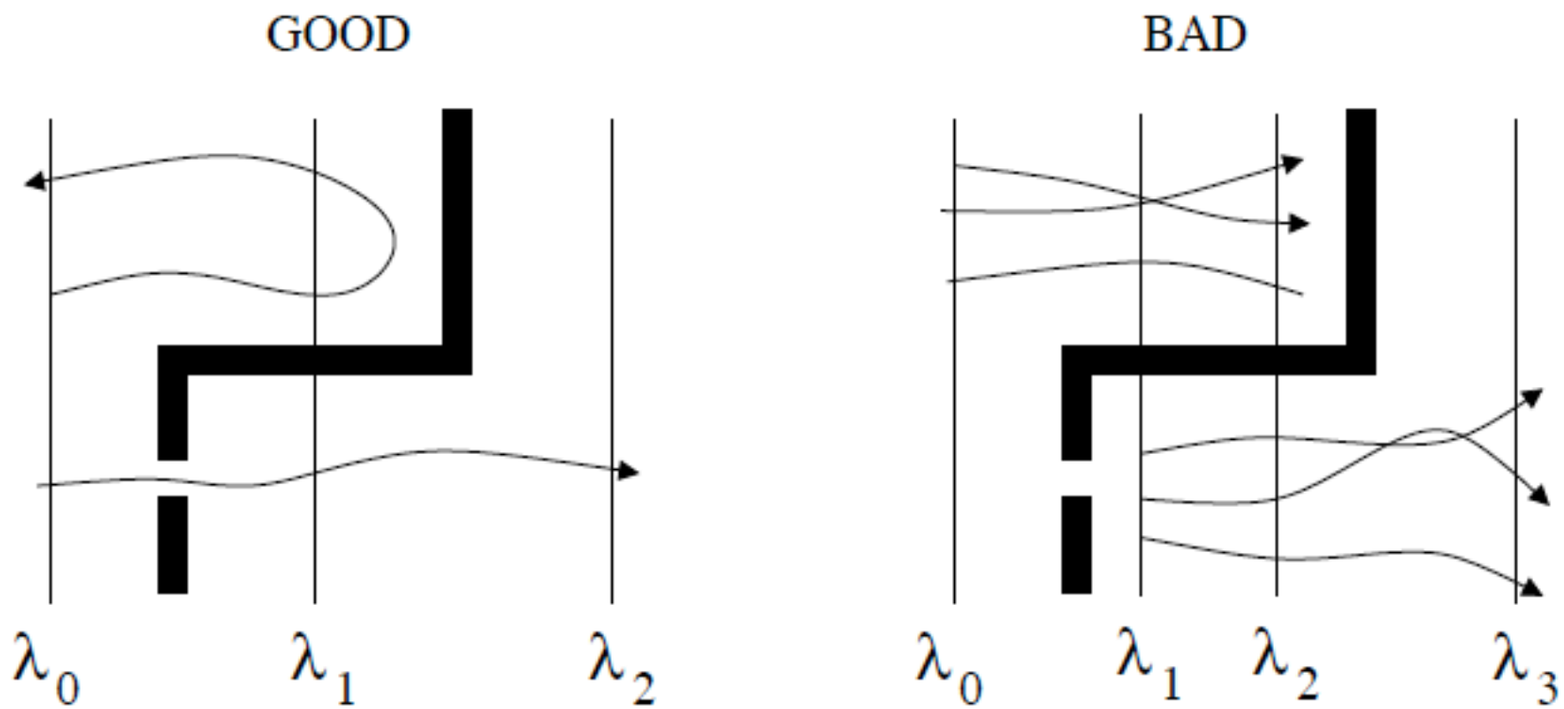
$$p_4^{\mp} = \frac{\#p_{53}}{\#p_{55}+\#p_{53}}$$

$$p_4^{=} = 1 - p_4^{\pm}$$

$$p_4^{=} = 1 - p_4^{\pm}$$

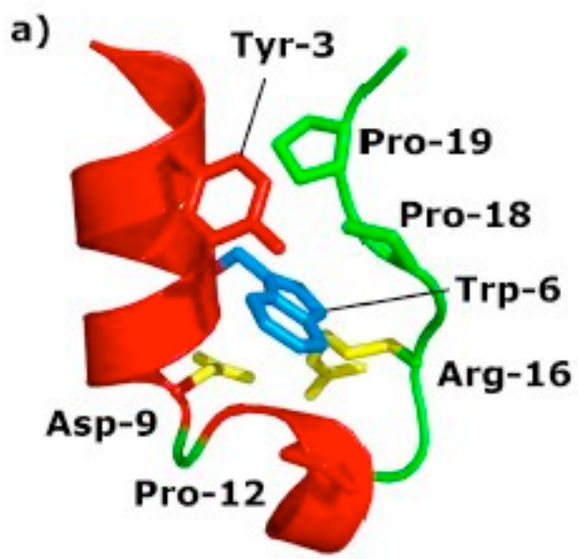
On the memory loss assumption in PPTIS

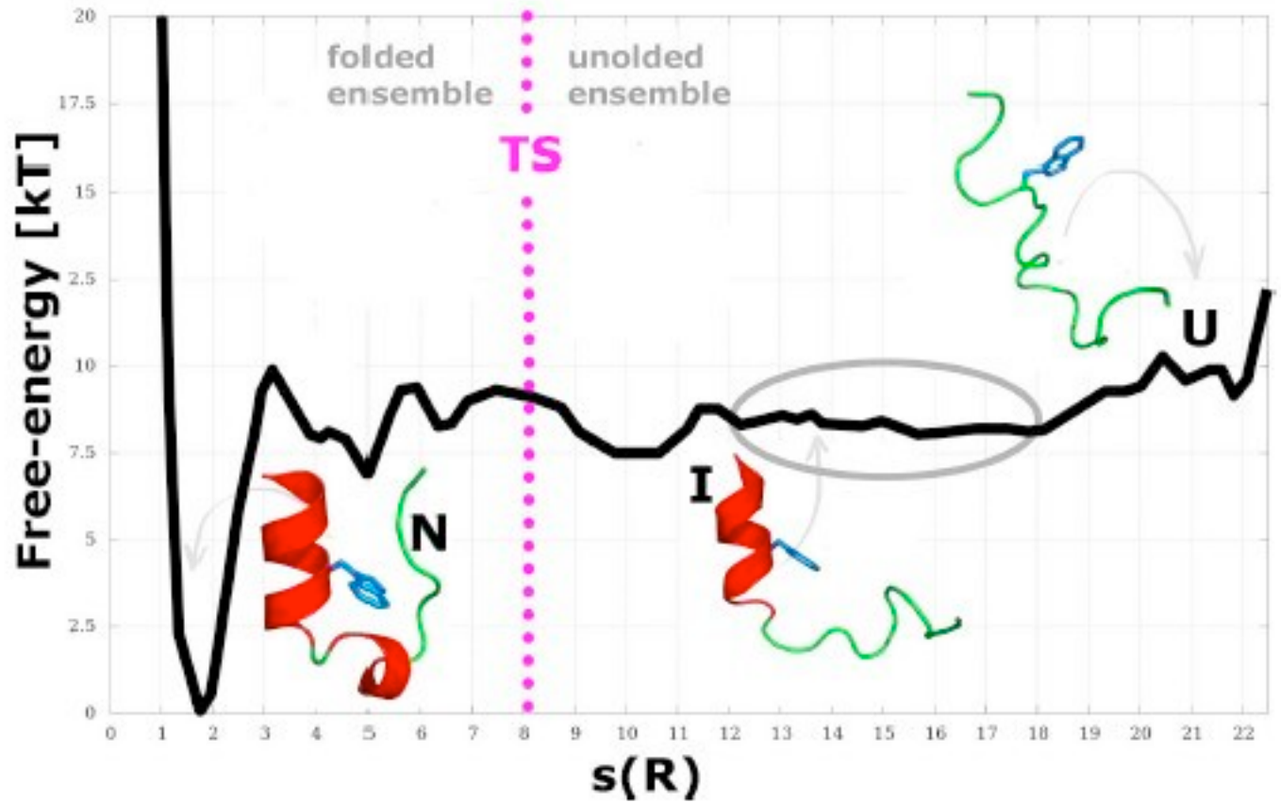




Transition State PPTIS

Juraszek,Saladino, van Erp,Gervasio, Phys. Rev. Lett., **110**, 108106, (2013).





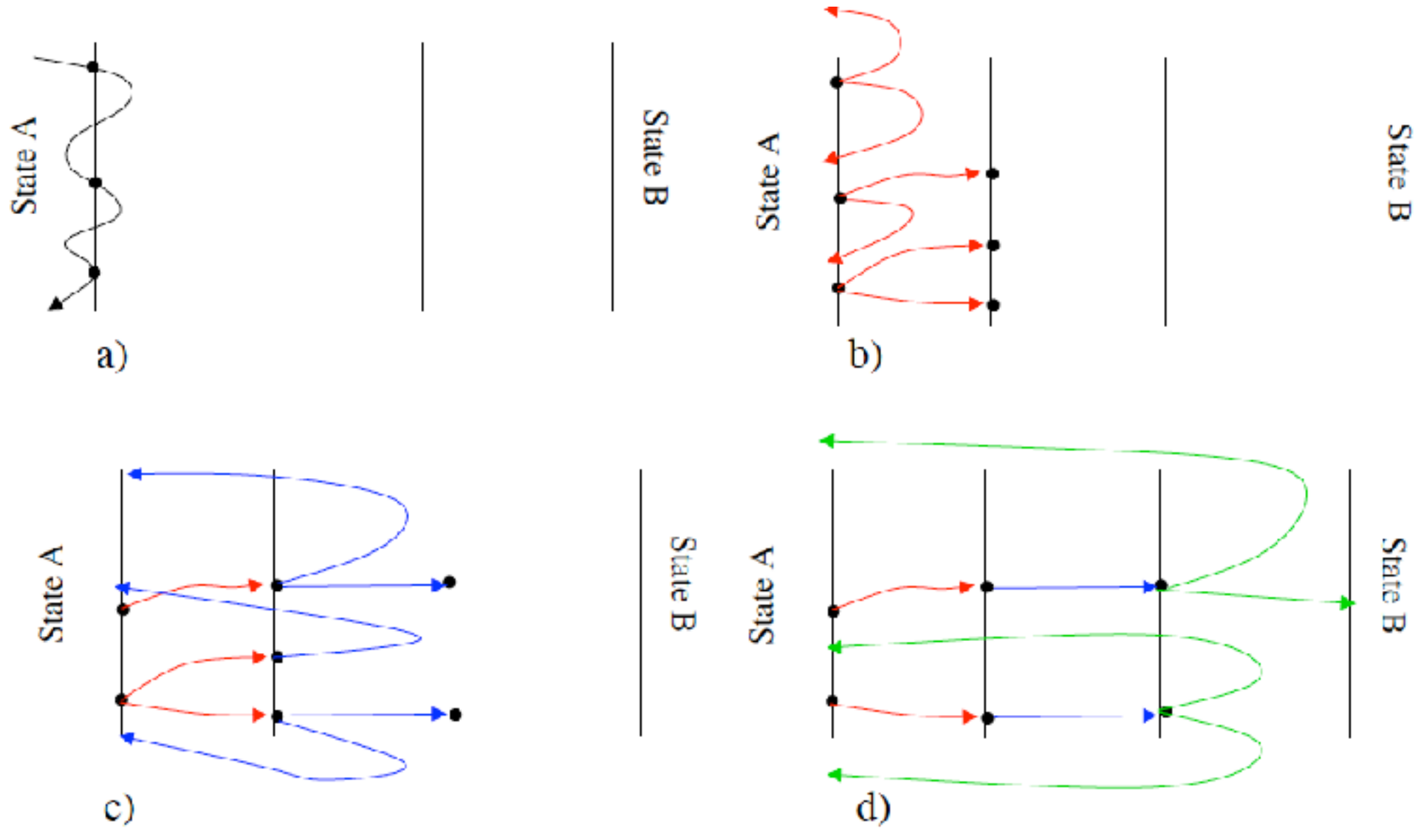
Applied to the Trp-cage miniprotein folding

Basically a Reactive Flux (RF) method using a PPTIS approach to compute small transmission coefficients

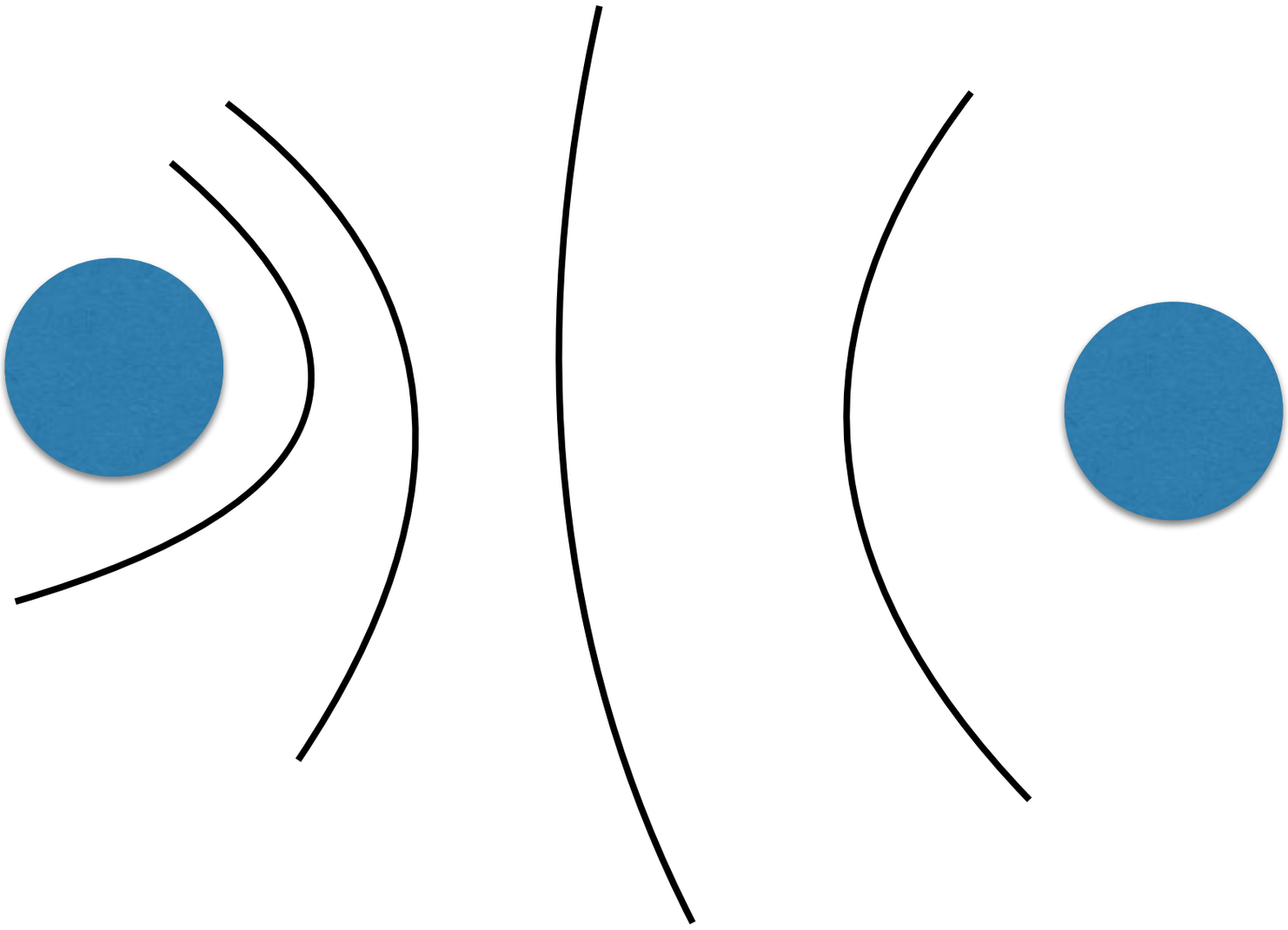
Forward Flux Sampling

R. J. Allen, P. B. Warren, and P. R. ten Wolde. Phys. Rev. Lett., **94**, 018104, (2005).

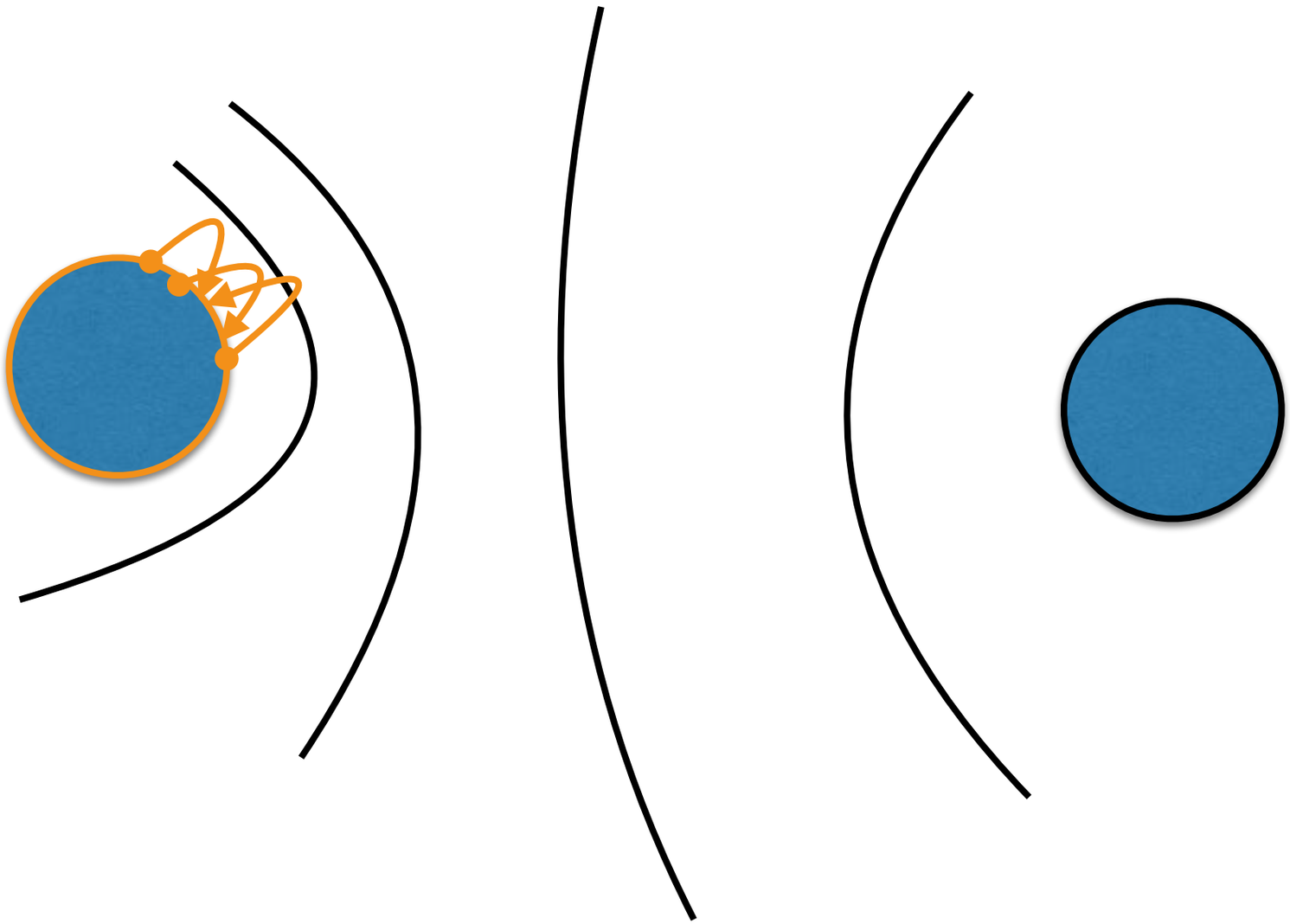
Based on TIS rate constant expression, but uses <u>splitting</u> instead of <u>shooting</u>



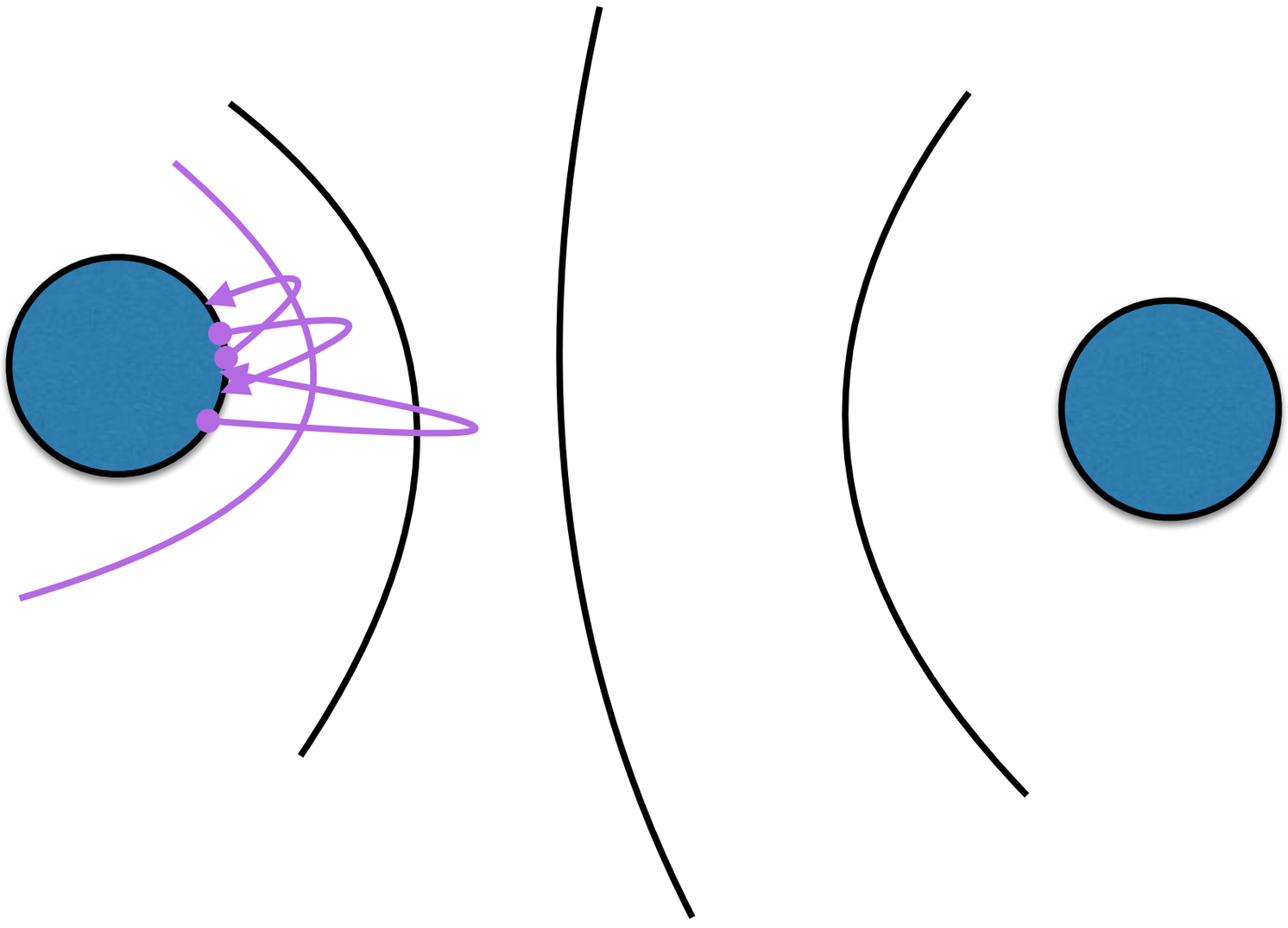
Metropolis-Hastings MC allows you to update the initial distribution.
Big advantage compared to splitting-type methods!



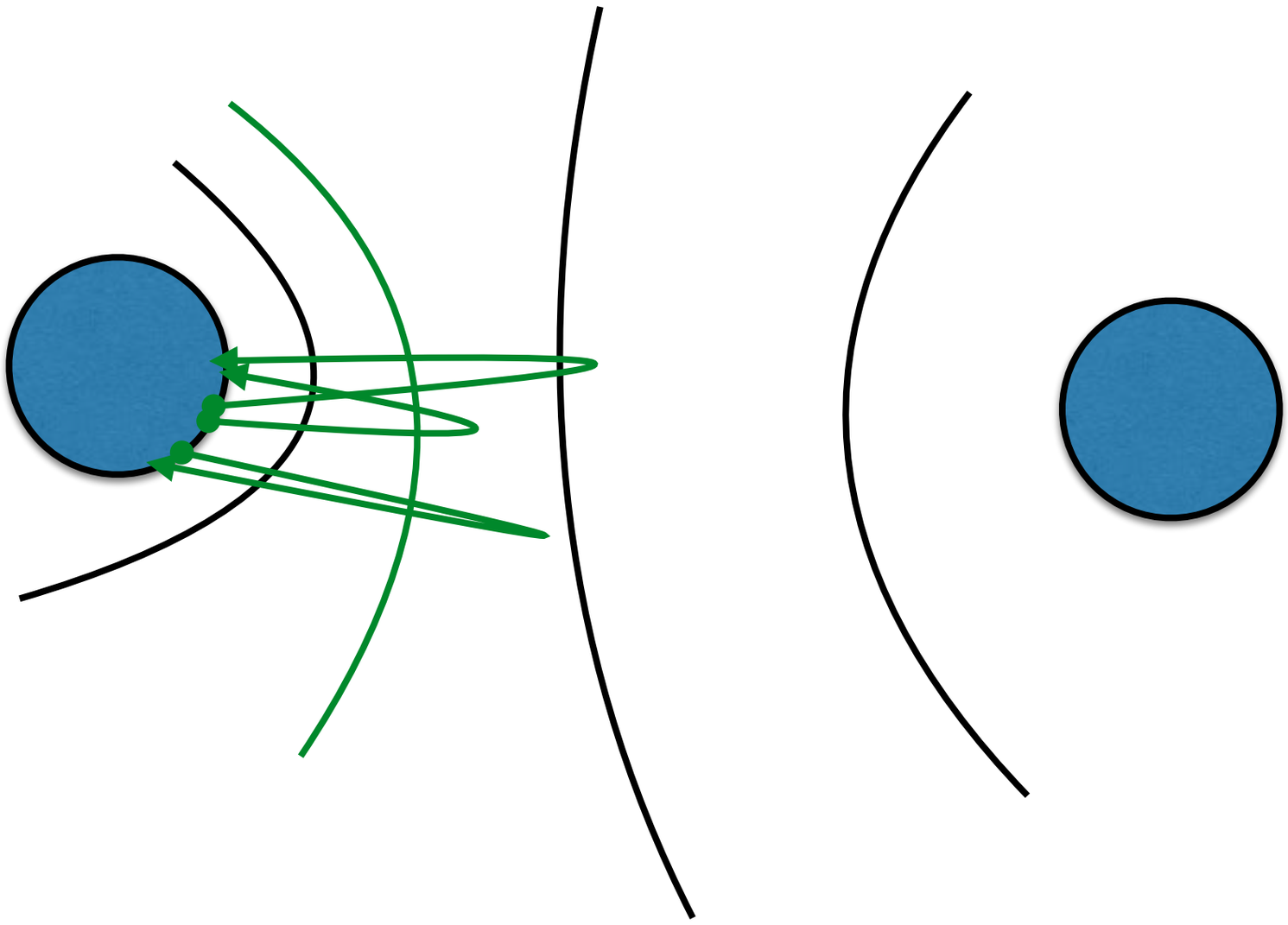
Metropolis-Hastings MC allows you to update the initial distribution.
Big advantage compared to splitting-type methods!



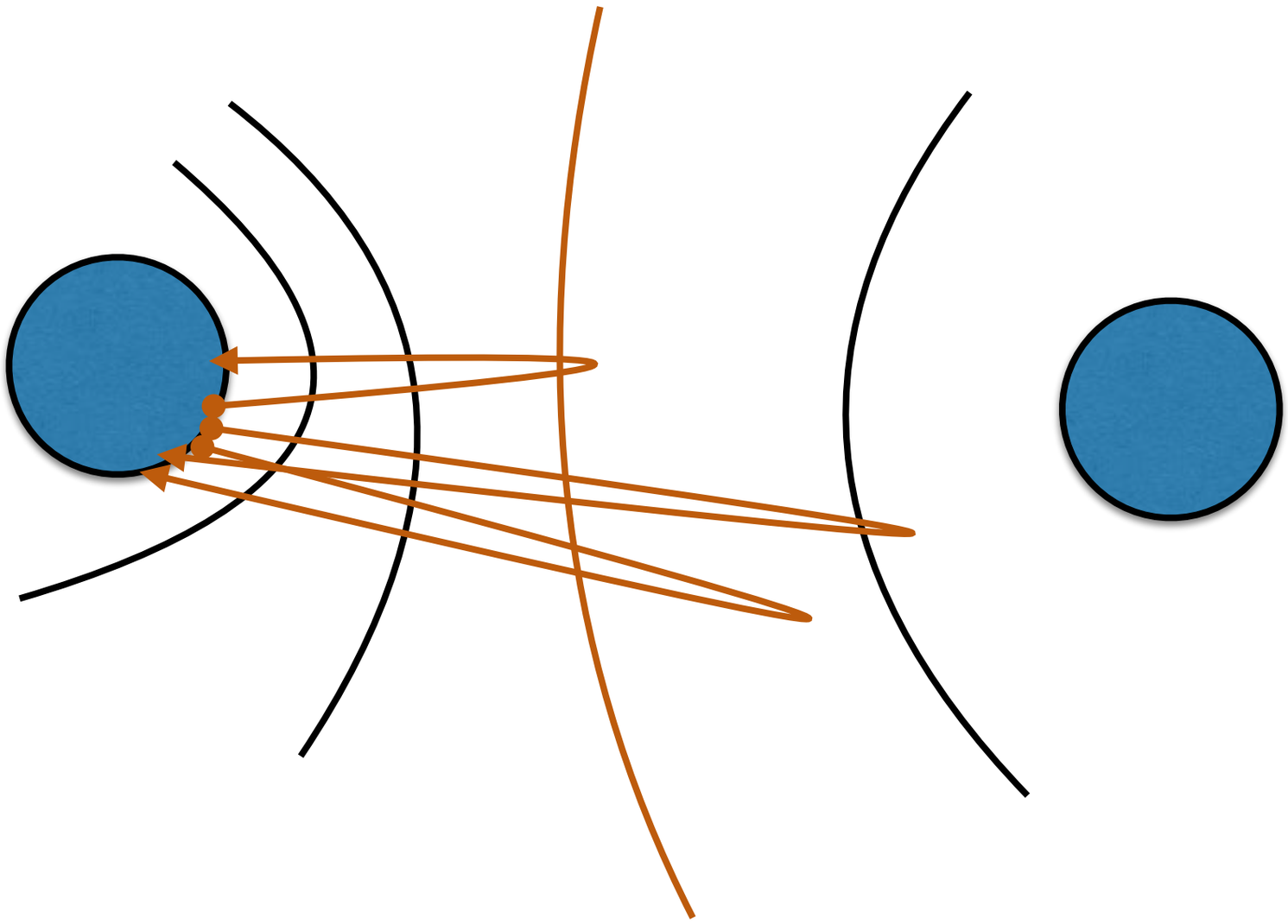
Metropolis-Hastings MC allows you to update the initial distribution.
Big advantage compared to splitting-type methods!



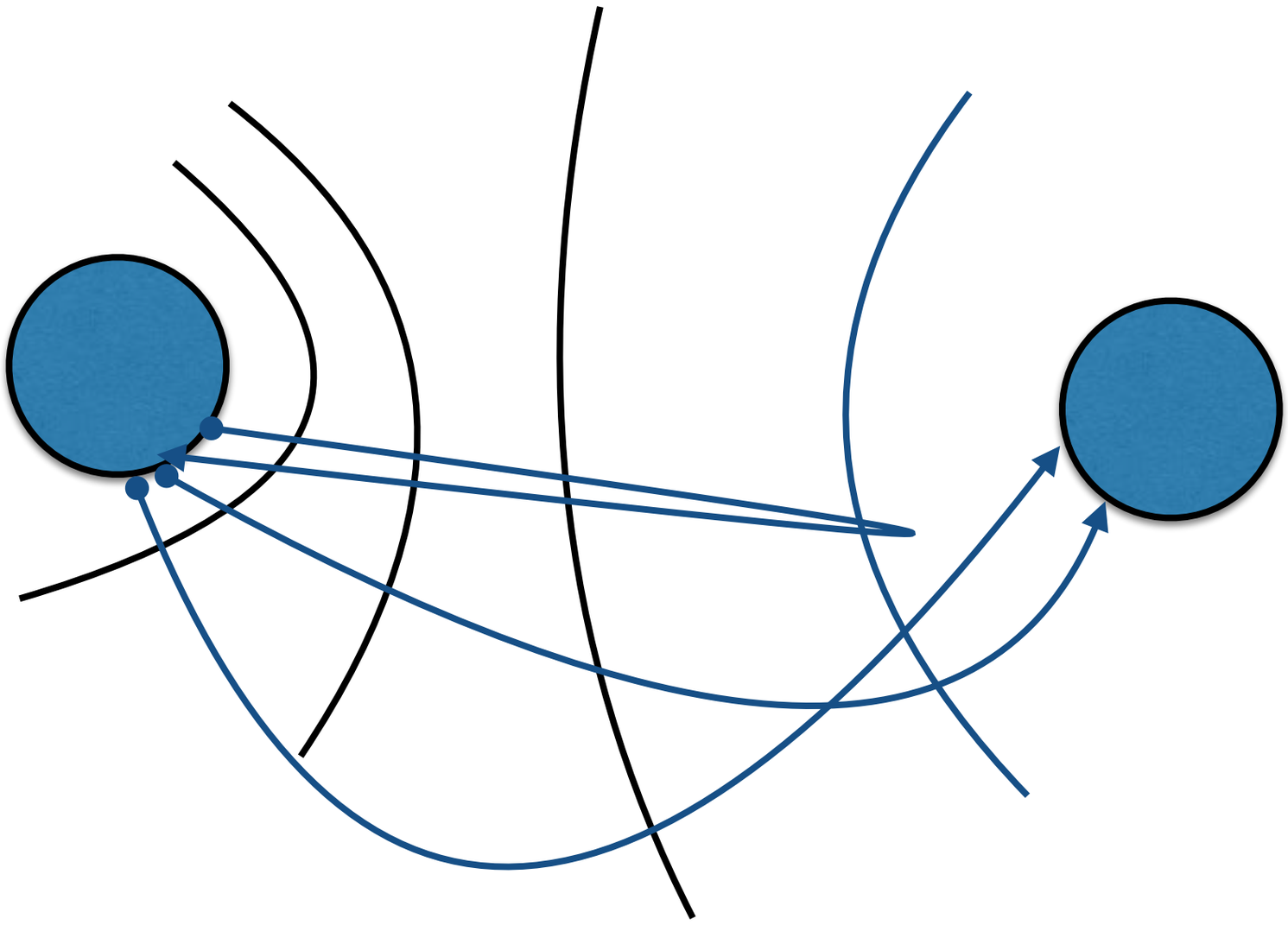
Metropolis-Hastings MC allows you to update the initial distribution.
Big advantage compared to splitting-type methods!



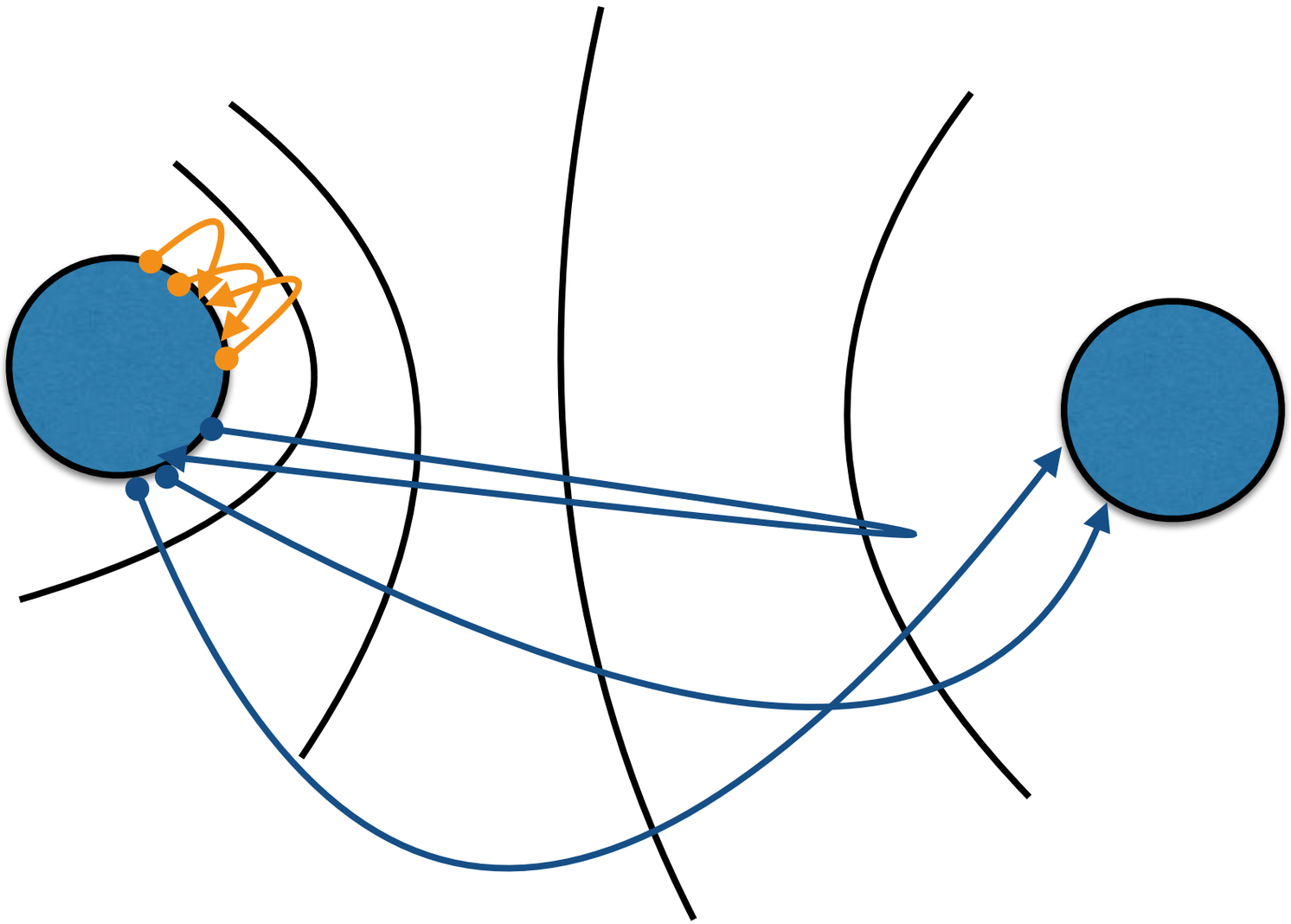
Metropolis-Hastings MC allows you to update the initial distribution.
Big advantage compared to splitting-type methods!



Metropolis-Hastings MC allows you to update the initial distribution.
Big advantage compared to splitting-type methods!



Metropolis-Hastings MC allows you to update the initial distribution.
Big advantage compared to splitting-type methods!

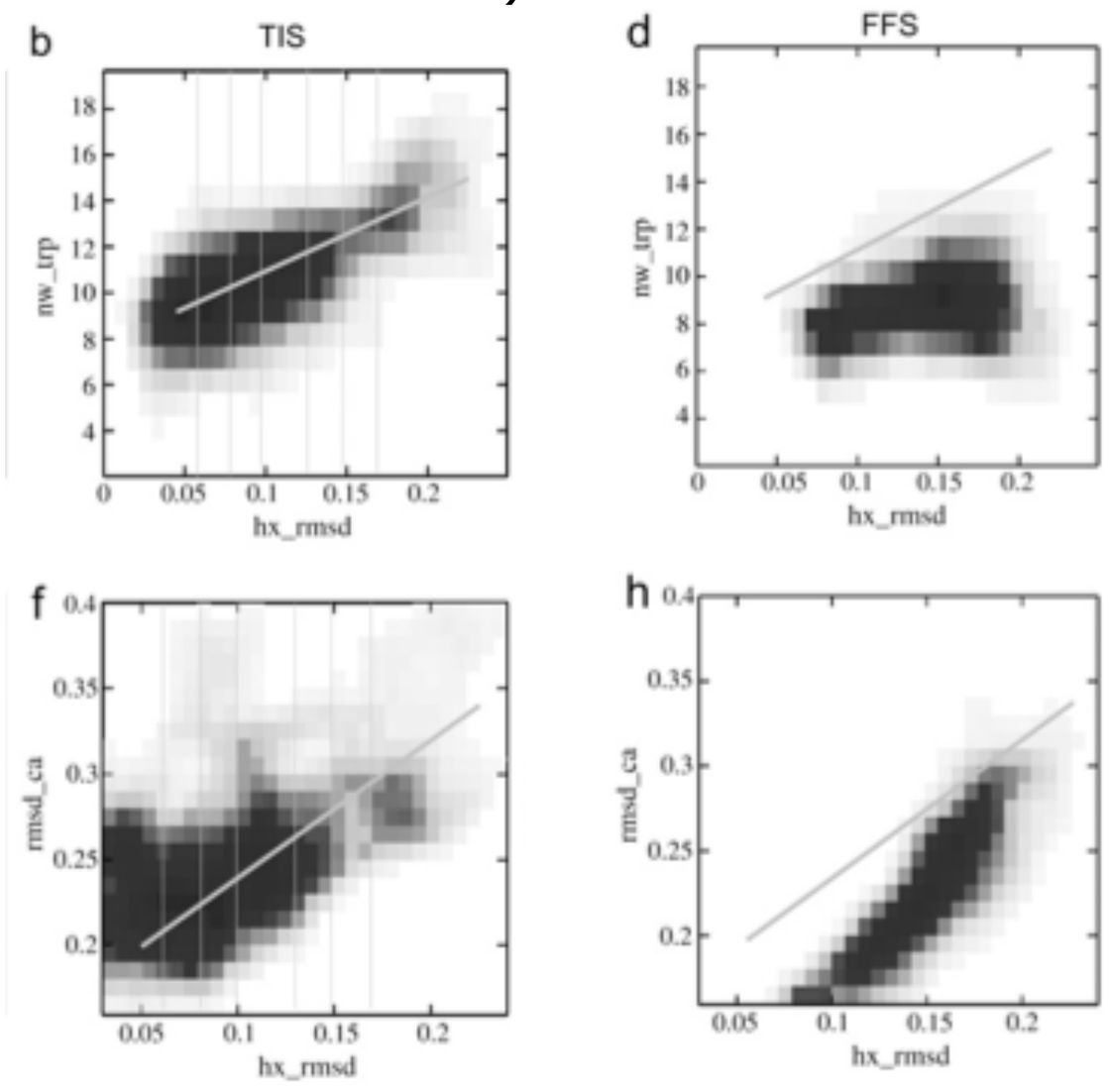


An example of FFS giving a too low reaction rate:

Rate Constant and Reaction Coordinate of Trp-Cage Folding in Explicit Water
Juraszek and Bolhuis, Biophys. J. **95**, 4246–4257 (2008)

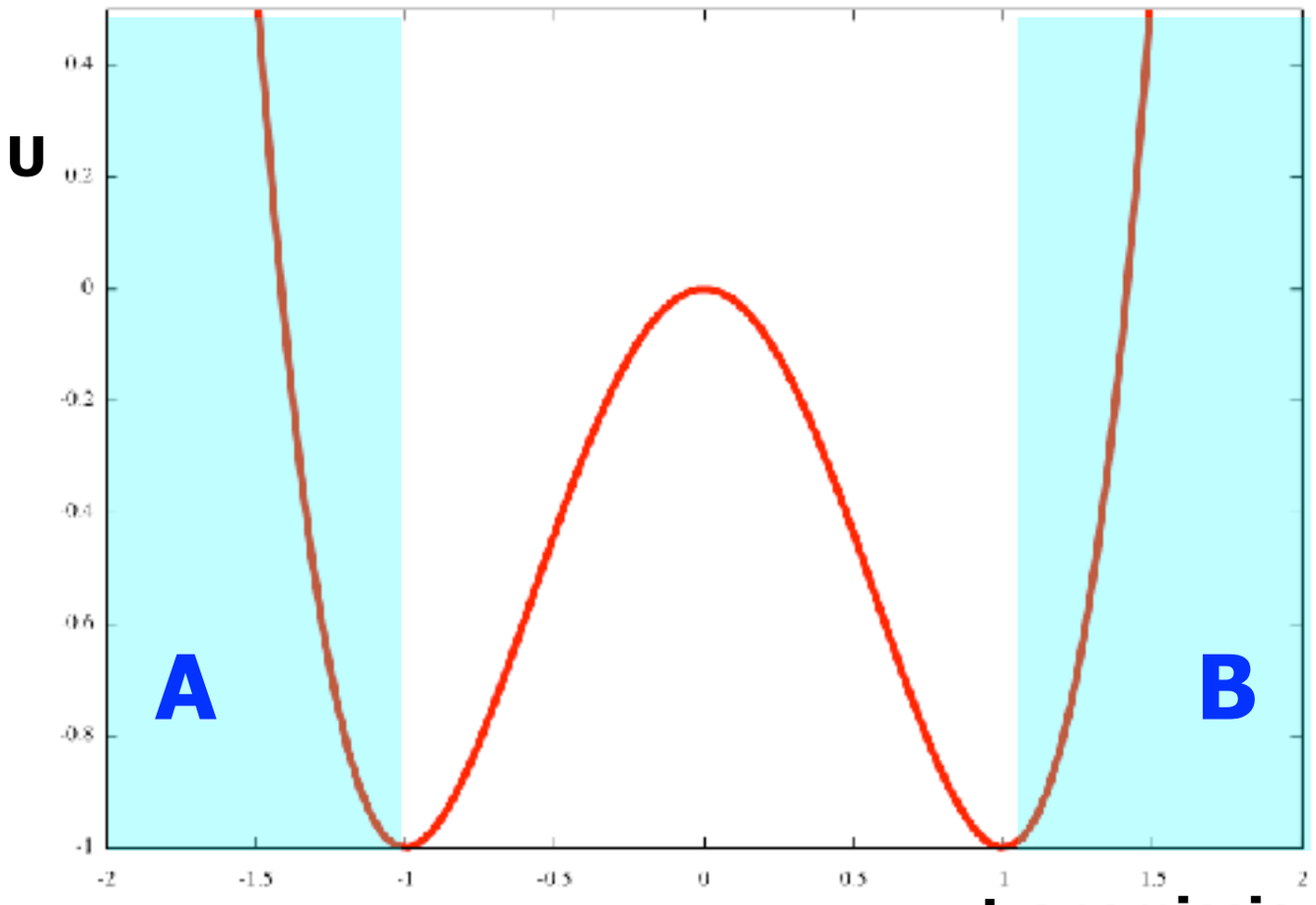
TIS: $k_{\mathrm{NL}} = (1.2\,\mu s)^{-1}$

FSS: $k_{\mathrm{NL}} = (100\,\mu s)^{-1}$



Comparison of RF, TIS, RETIS, and FFS on a 1D ! example

T. S. Van Erp Adv. Chem. Phys. 2012



$$U(r) = r^4 - r^2$$

Langevin Dynamics

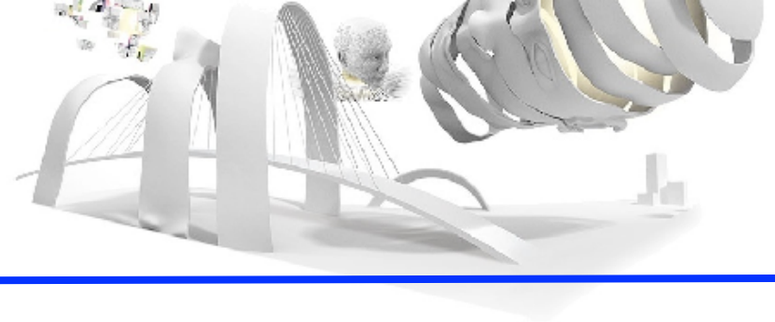
$$k_B = m = 1$$

$$\gamma = 0.3,\ T = 0.7$$

8 interfaces

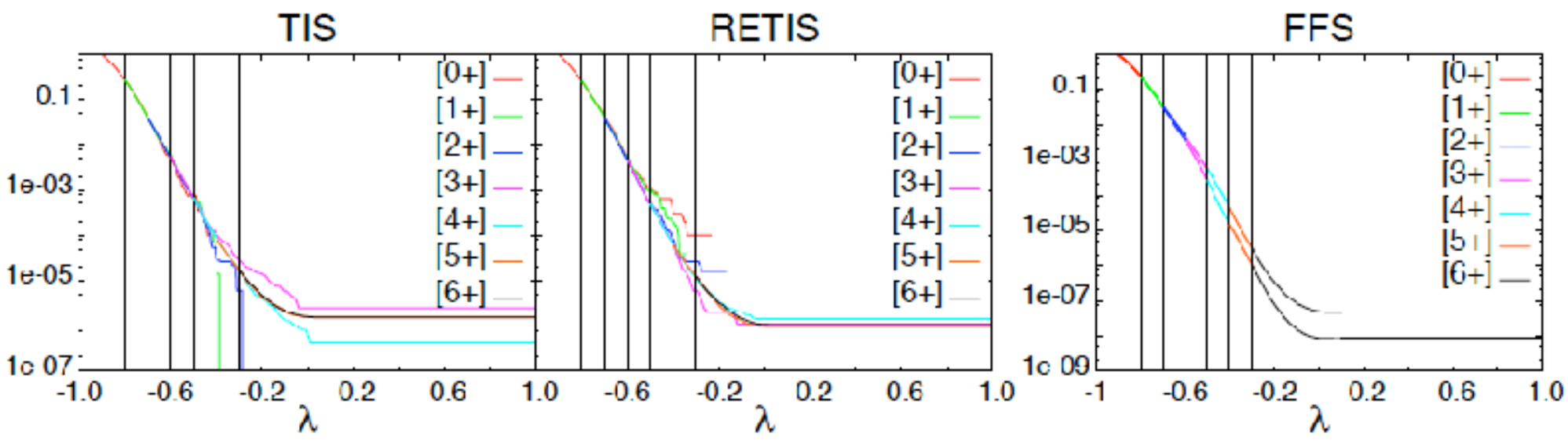
$$\lambda_0 = -0.9, \lambda_1 = -0.8, \lambda_2 = -0.7, \lambda_3 = -0.6,$$
$$\lambda_4 = -0.5, \lambda_5 = -0.4, \lambda_6 = -0.3, \lambda_7 = 1.0$$

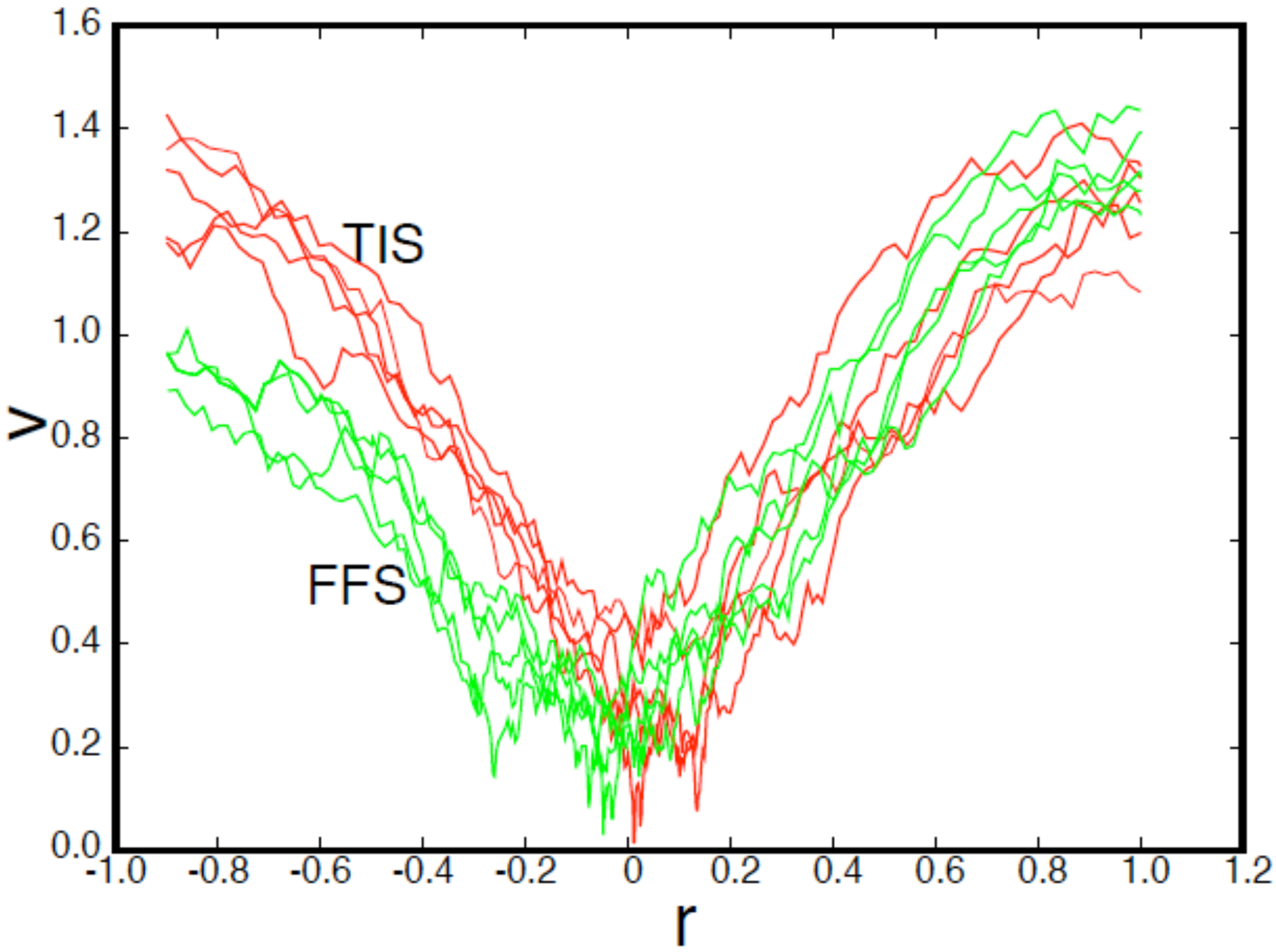
transmission coefficient (RF): 100,000 trajectories
flux (TIS, PPTIS, FFS): 10,000,000 MD-steps run
path=ensembles (TIS, PPTIS, FFS, RETIS):
20,000 trajectories each



reactive flux method	$\frac{1}{\sqrt{2\pi\beta m}}$	$\frac{e^{-\beta F(0)}}{\int_{-\infty}^{0} d\lambda\, e^{-\beta F(\lambda)}}$	κ	$k - \kappa \times \frac{1}{\sqrt{2\pi\beta m}} \times \frac{e^{-\beta F(0)}}{\int_{-\infty}^{0} d\lambda\, e^{-\beta F(\lambda)}}$
EPF algorithm	0.106	$2.63 \cdot 10^{-6}$	$0.874 \perp 4\%$	$2.42 \cdot 10^{-7} \perp 4\ \%$

path sampling	f_A	$\mathcal{P}_A(\lambda_n\|\lambda_0)$	k– $f_A \times \mathcal{P}_A(\lambda_n\|\lambda_0)$
TIS	$0.263 \pm 1\%$	$1.52 \cdot 10^{-6} \pm 20\%$	$4.02 \cdot 10^{-7} \pm 20\%$
PPTIS	$0.263 + 1\%$	$1.04 \cdot 10^{-6} + 19\%$	$2.73 \cdot 10^{-7} + 19\%$
RETIS	$0.265 \pm 1\%^*$	$1.05 \cdot 10^{-6} \pm 25\%^*$	$2.79 \cdot 10^{-7} \pm 25\%^*$
FFS (long MD run)	$0.263 \perp 1\%$	$4.69 \cdot 10^{-8} \perp 6\%^*$	$1.23 \cdot 10^{-8} \perp 6\%^*$
FFS (short MD run)	$0.259 \pm 2\%$	$8.45 \cdot 10^{-9} \pm 9\%^*$	$2.18 \cdot 10^{-9} \pm 9\%^*$



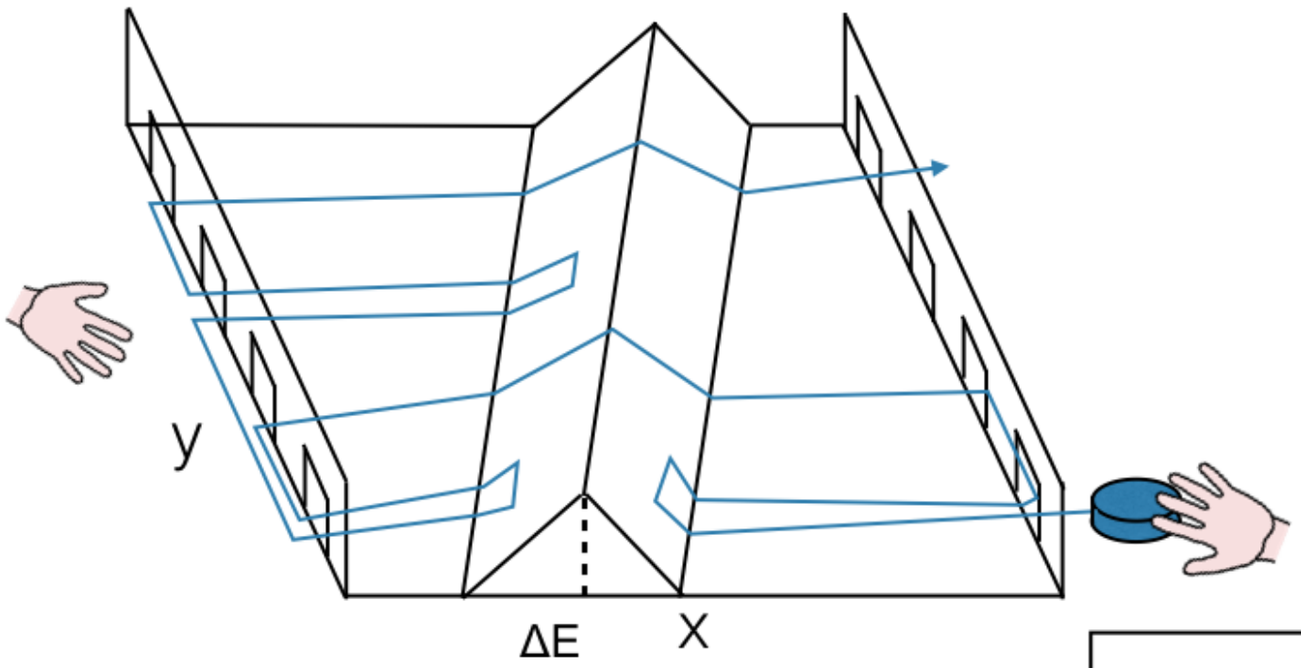


TIS: symmetric, good!

FFS: asymmetric, not good!

Efficiency of TIS/RETIS is relatively insensitive to choice of RC

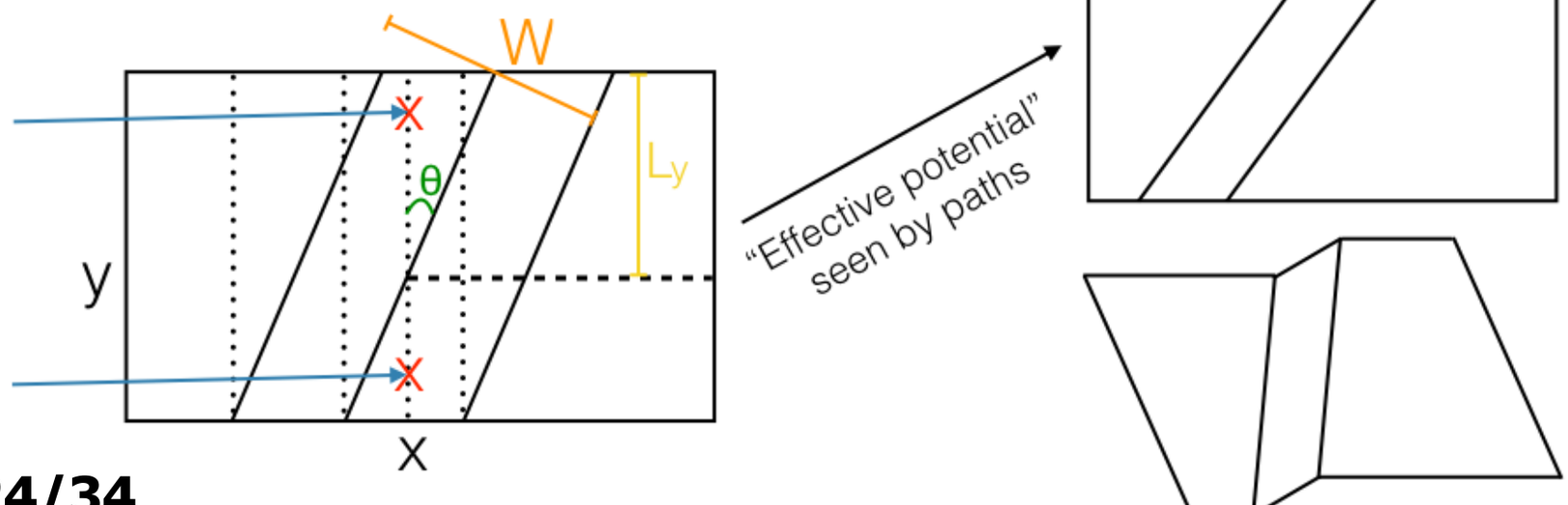
RF: Reactive Flux method=dynamically corrected Transition State Theory



$$\theta = 0$$

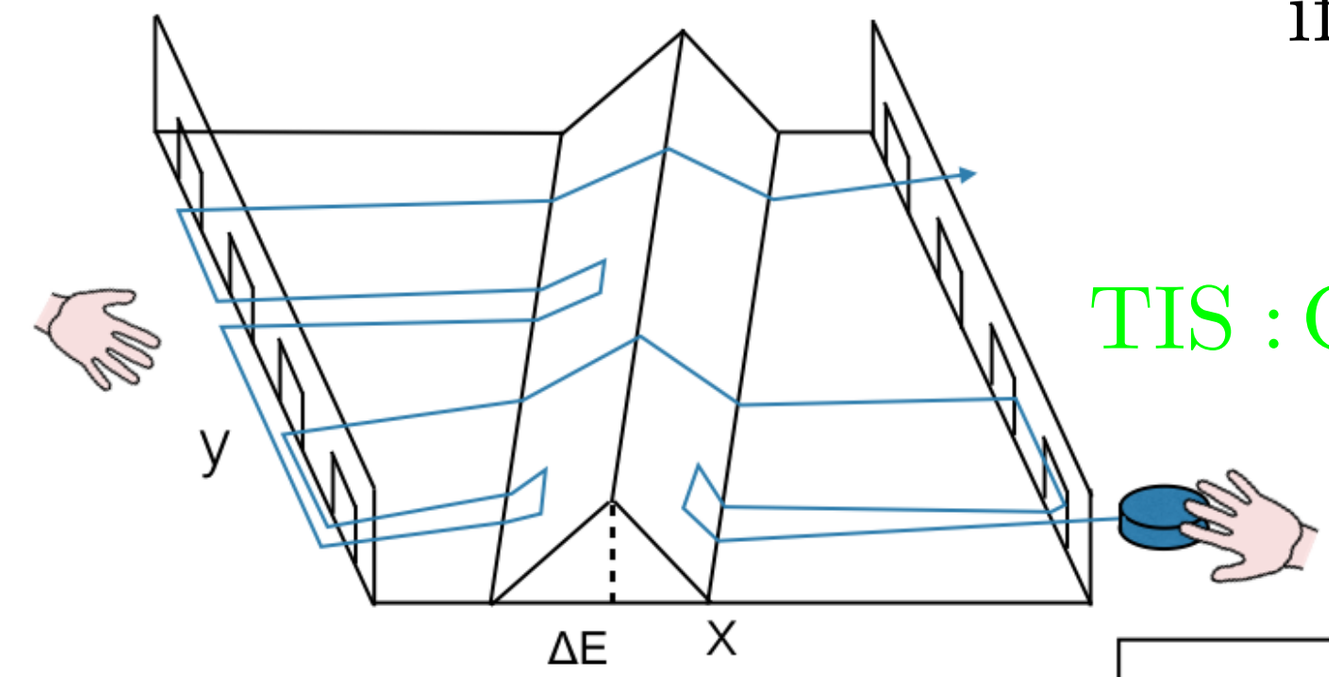
$$\text{MD: CPU} \propto \exp(\beta \Delta E)$$

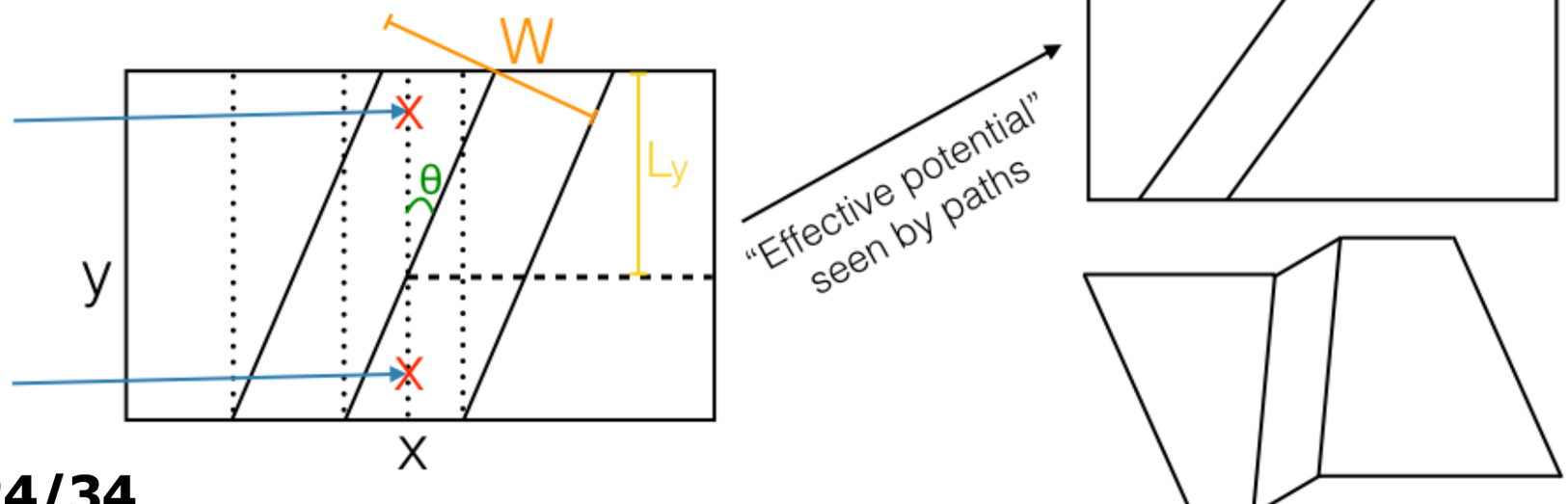
$$\text{RF/TIS : CPU} \propto (\beta \Delta E)^2$$



Efficiency of TIS/RETIS is relatively insensitive to choice of RC

RF: Reactive Flux method=dynamically corrected Transition State Theory





$$\alpha = 2\beta\Delta E \sin\theta L_y/W$$

if $\alpha \gg 0$

$$\mathrm{RF:CPU} \propto \frac{\exp(\alpha)}{\sqrt{\alpha}}$$

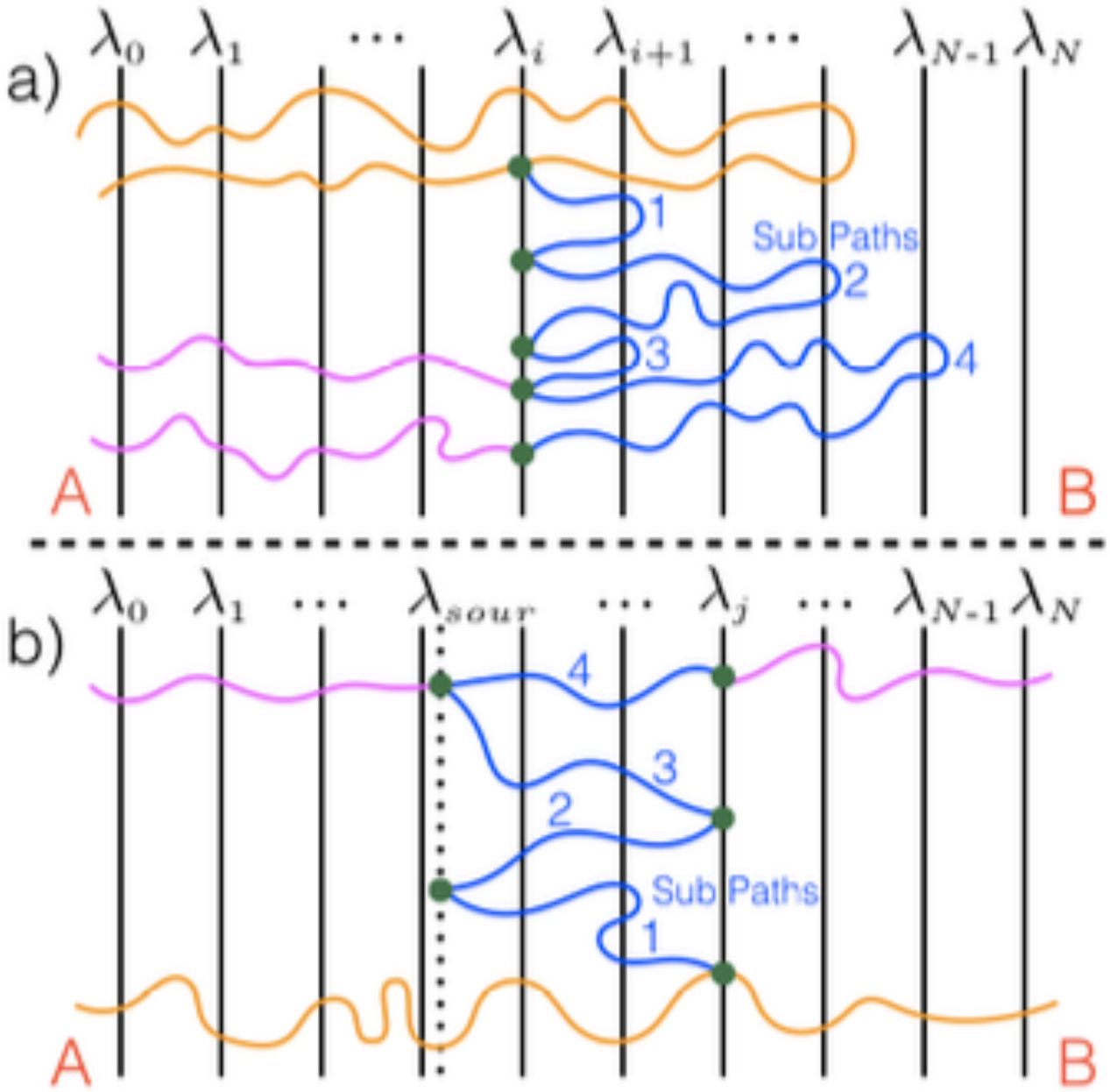
$$\mathrm{TIS:CPU} \propto (\beta\Delta E(1 + 2\alpha/\beta\Delta E))^2$$

FFS: no analytical result available but must be exponential as well

NTNU
Norwegian University of Science and Technology

How to improve further the RETIS Efficiency?

E. Riccardi, O. Dahlen, and T. S. Van Erp, J. Phys. Chem. Lett. 2017



Source path

Shooting Point

New path

Stone-Skipping

New path

Shooting Point

Web-Throwing

Source path

NTNU
Norwegian University of Science and Technology

Acceptance rule based on Super-detailed balance

$$P_{\text{acc}} = \min\left[1, \frac{P(p^{(n)})P_{\text{gen}}(p^{(n)} \to p^{(o)} \text{via } \overline{\chi})}{P(p^{(o)})P_{\text{gen}}(p^{(o)} \to p^{(n)} \text{via } \chi)}\right]$$

which leads to

$$\text{SS}: P_{\text{acc}} = \min\left[1, \frac{n_c^{(o)}}{n_c^{(n)}}\right], \text{WT}: P_{\text{acc}} = \min\left[1, \frac{n_s^{(o)}}{n_s^{(n)}}\right]$$

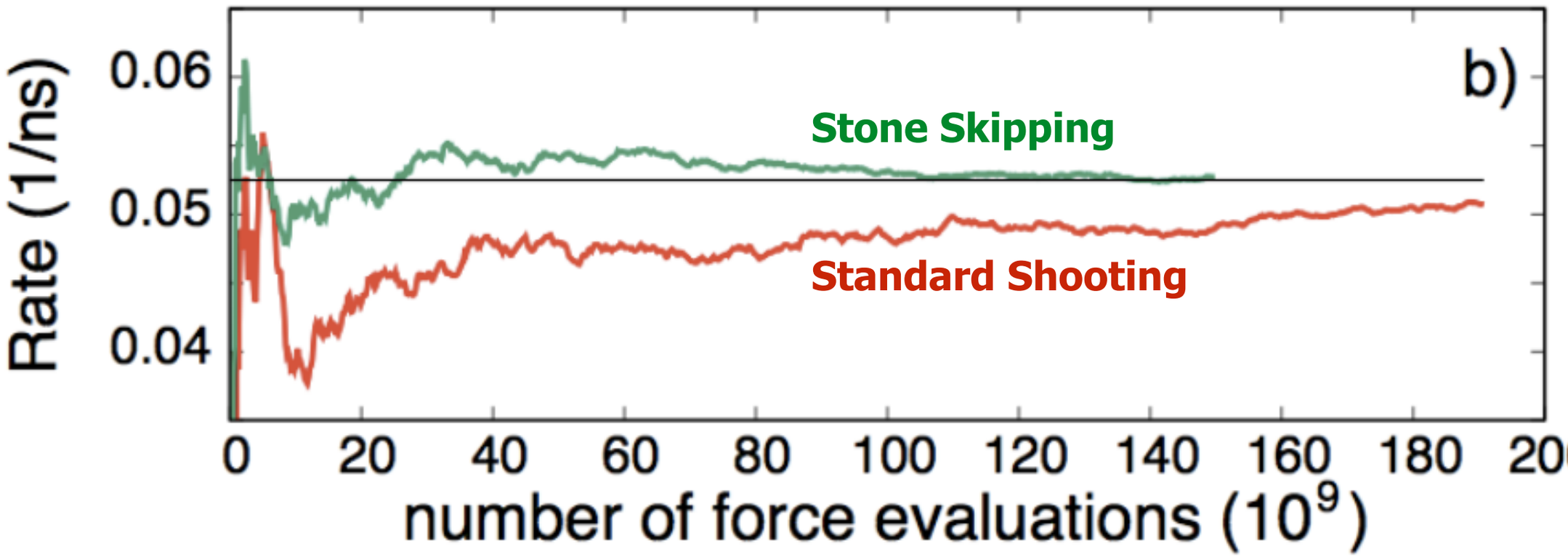
but with alternative weights

$$\tilde{P}(p) = w_i(p)P(p) \equiv n_{c,i}(p)q(p)P(p) \text{ for } p \in [i^+]$$

$$q(p) \text{ equals } 2 \text{ for } A \to B \text{ and } 1 \text{ for } A \to A \text{ paths.}$$

all paths can be accepted except if completion of last sub-path ends in B in both time directions (B->A paths are time-inverted and then accepted)

In the end, correct statistics is obtained by reweighting each path with $w_i(p)^{-1}$



Rate of DNA denaturation using the mesoscopic PBD model.
Horizontal line is result based on partition function integration (nearly exact result).

Stone skipping was found 12 times faster

Conclusions part II:

- Partial Path TIS (PPTIS/Partial Path Sampling) uses a Markovian approximation (though softer than the one used in Milestoning) to reduce average path length.

- RETIS with Stone-Skipping and Web-Throwing reduces the correlation between paths using short sub-trajectories -> can become nearly as efficient as PPTIS or Milestoning but exact!

- FFS allows the study non-equilibrium dynamics but has a serious (underestimated) sampling problem. Solutions?

- Further improvement of RETIS: how to parallelize efficiently path-ensembles which have different path lengths?

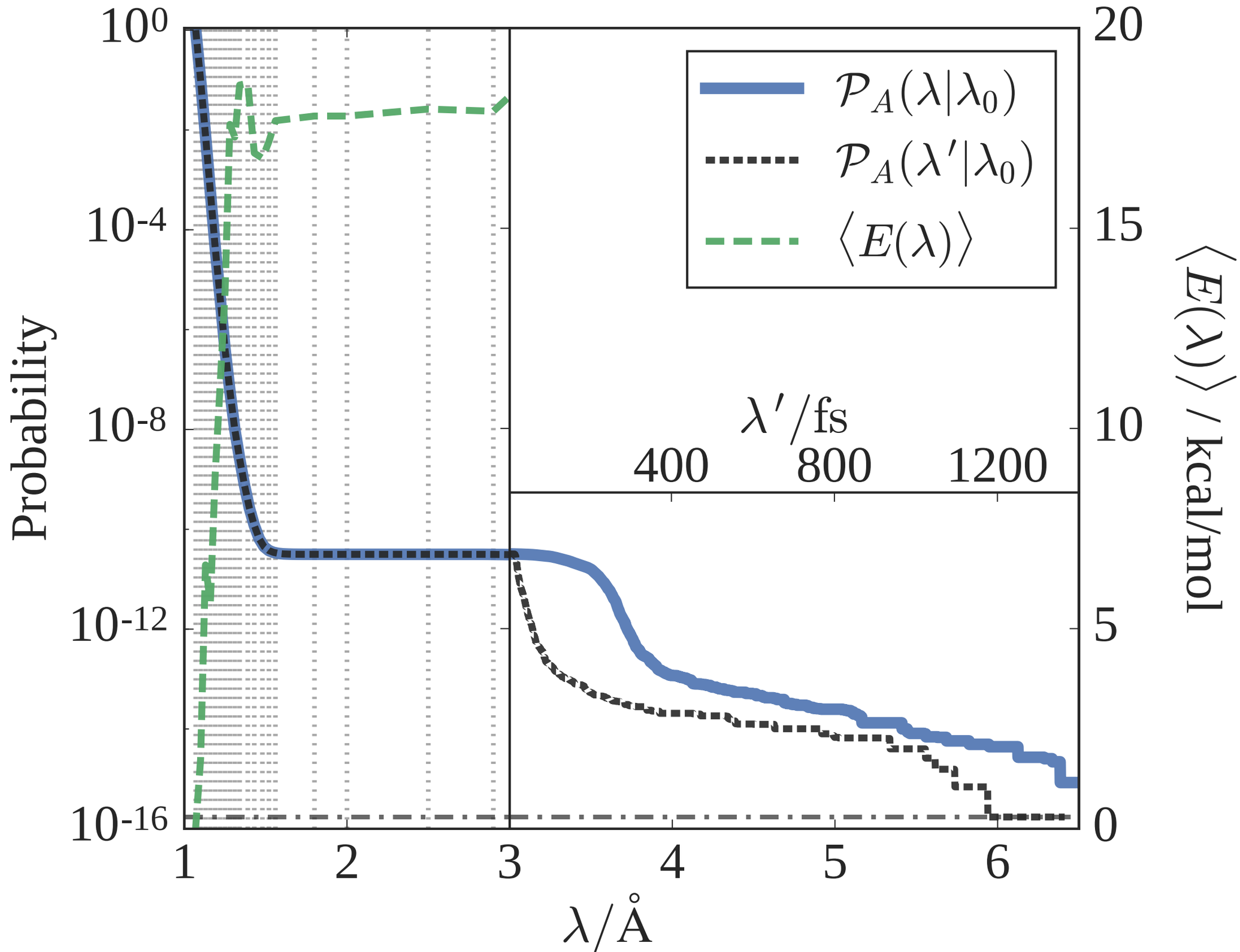
NTNU

Norwegian University of Science and Technology

Auto-ionization of water: a RETIS/CP2K study

NTNU

Norwegian University of Science and Technology



NTNU

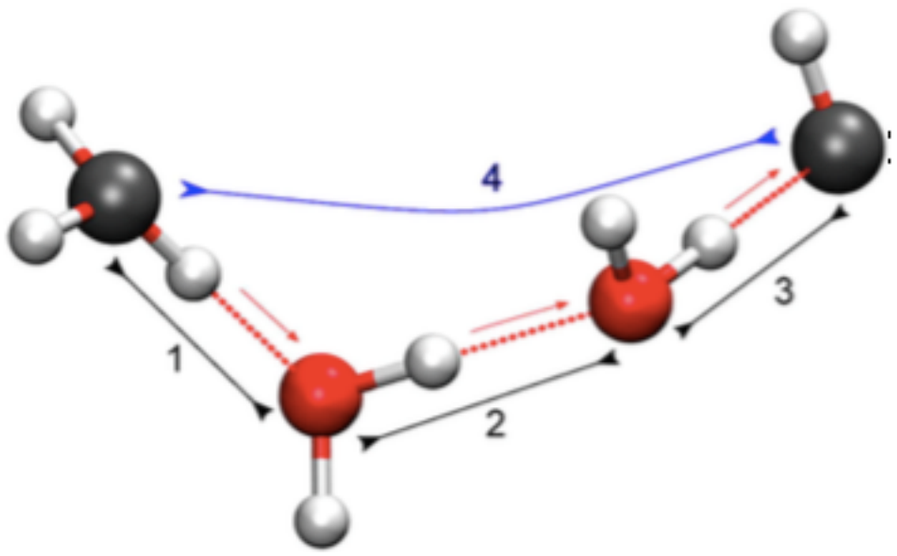
Norwegian University of Science and Technology

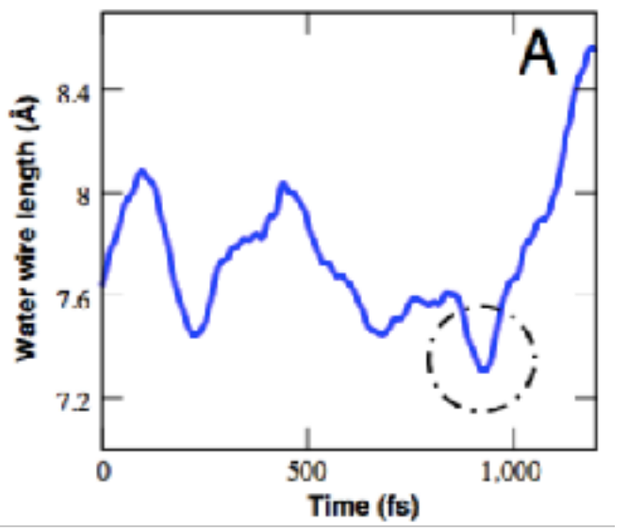
On the recombination of hydronium and hydroxide ions in water

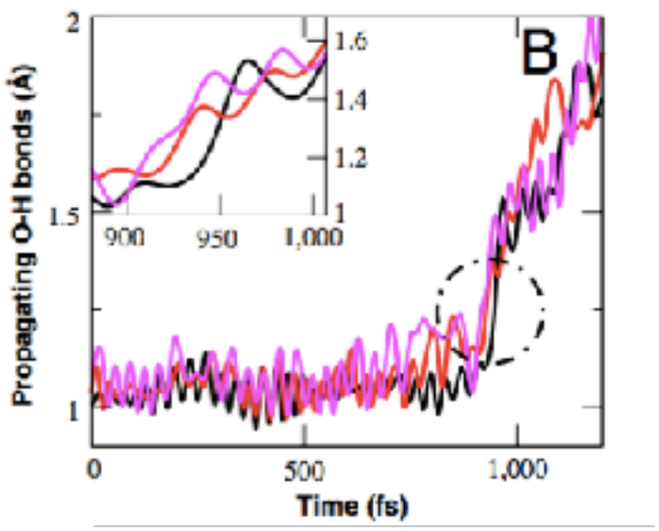
PNAS

Ali Hassanali[1], Meher K. Prakash, Hagai Eshet, and Michele Parrinello

The neutralization event involves a collective compression of the water-wire bridging the ions, which occurs in approximately 0.5 ps, triggering a concerted triple jump of the protons. This process leaves the neutralized hydroxide in a hypercoordinated state, with the implications that enhanced **collective compressions of several water molecules around similarly hypercoordinated states** are likely to serve as nucleation events for the autoionization of liquid water.







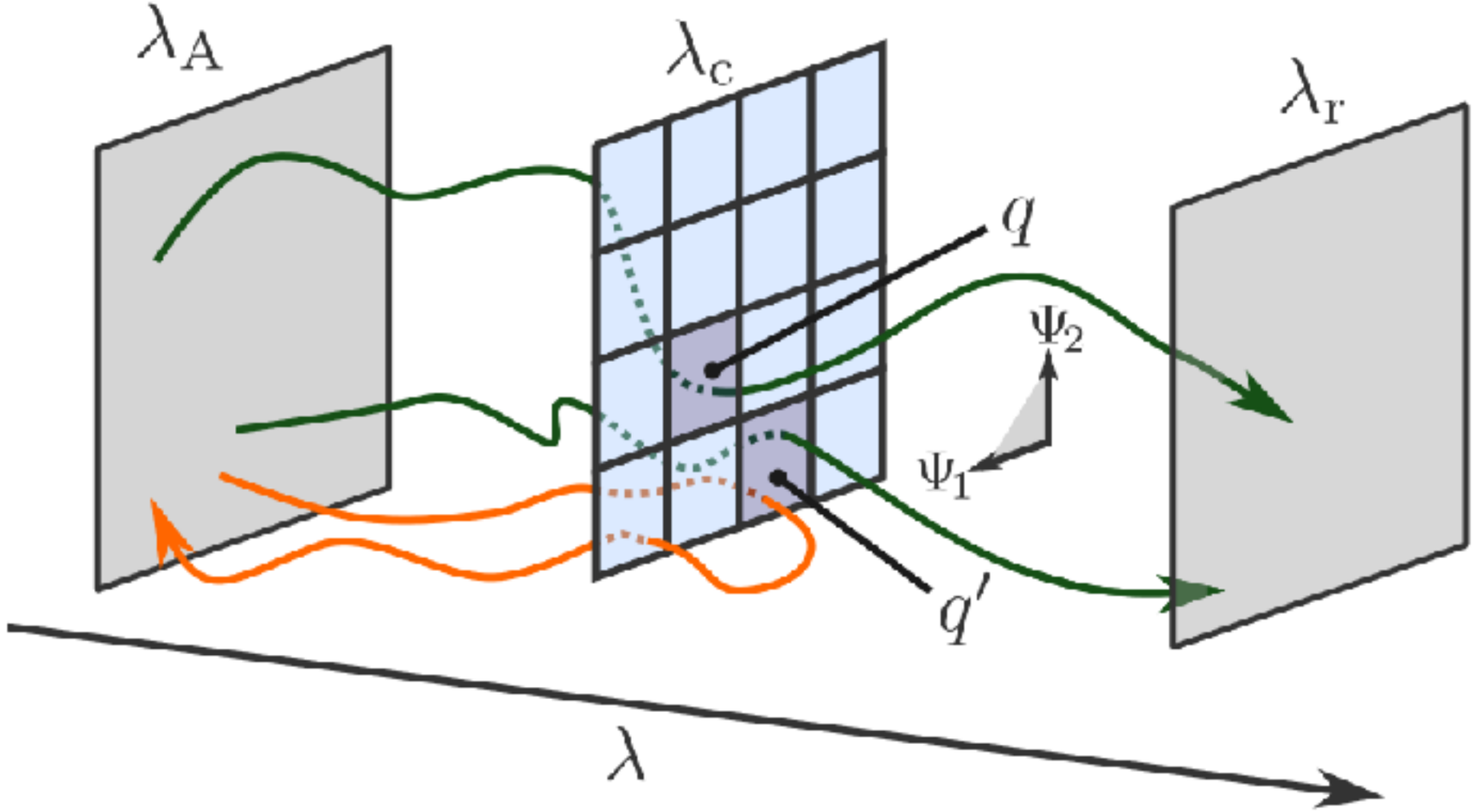
NTNU

Norwegian University of Science and Technology

Analyzing complex reaction mechanisms using path sampling

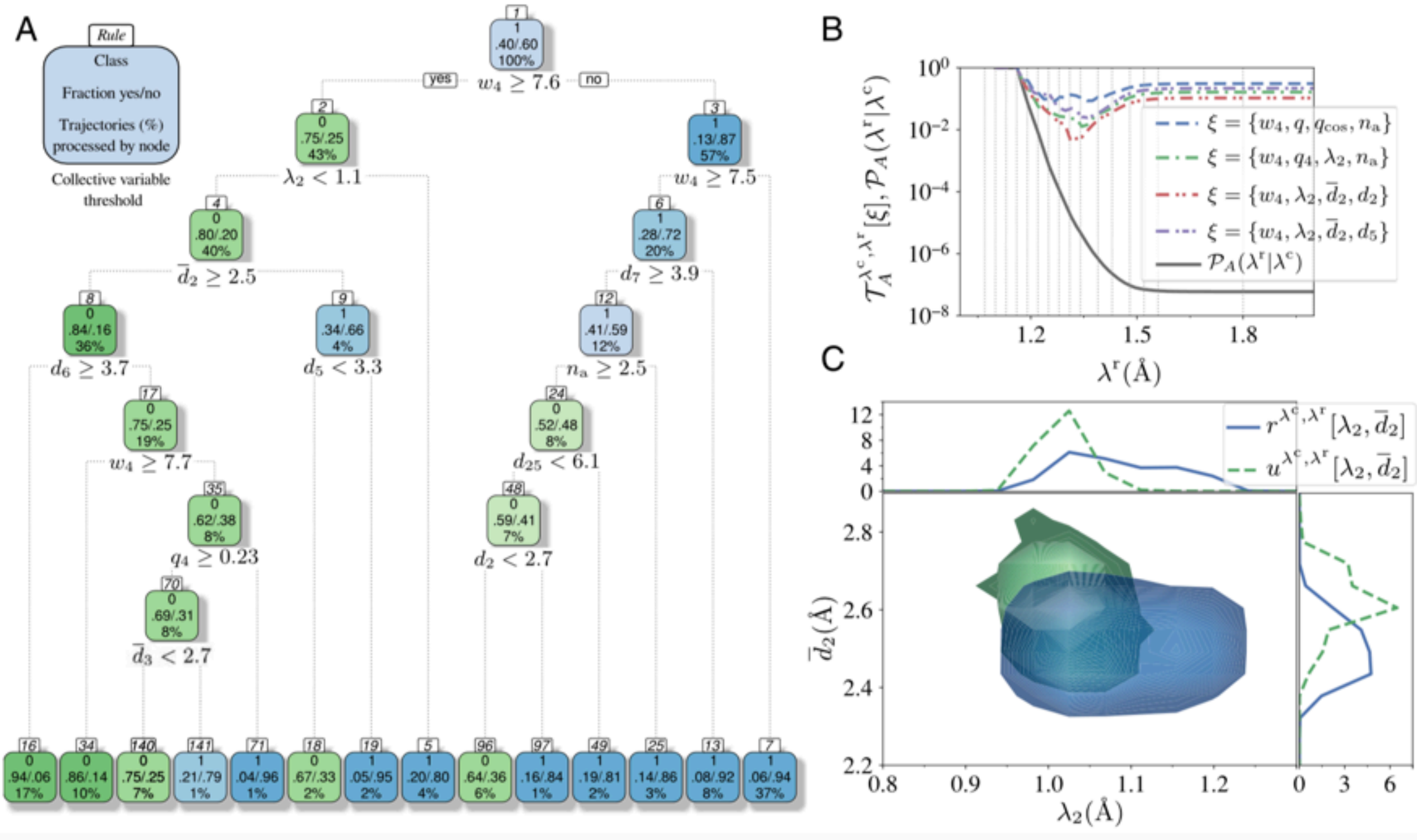
Titus S. van Erp, Mahmoud Moqadam, Enrico Riccardi, and Anders Lervik

J. Chem. Theory Comput. 2016



NTNU

Norwegian University of Science and Technology



Parrinello's w4 parameter pops-up as the most important variable of 138 collective variables that were considered in classification and regression decision trees

Conclusions part III:

- Using RETIS we can reach the minute timescale using ab-initio MD and still get exact rates and unbiased dynamics.

- We designed a new approach to analyze the data of path sampling simulation methods like **TIS**, **RETIS**, or **FFS** which can be used to test hypotheses on the reaction mechanism.

- In water auto-ionization, the compressions of several water molecules around a hypercoordinated state seems to be a **necessary** (the most predictive parameter) but by far **not a sufficient** condition for the initiation of water splitting.

- Also other parameters (distortion from tetrahedral ordering, the length of the stretched hydrogen bond of a nearby water molecule) need to be in the right range to let a dissociation event happen.

Thanks! Questions?