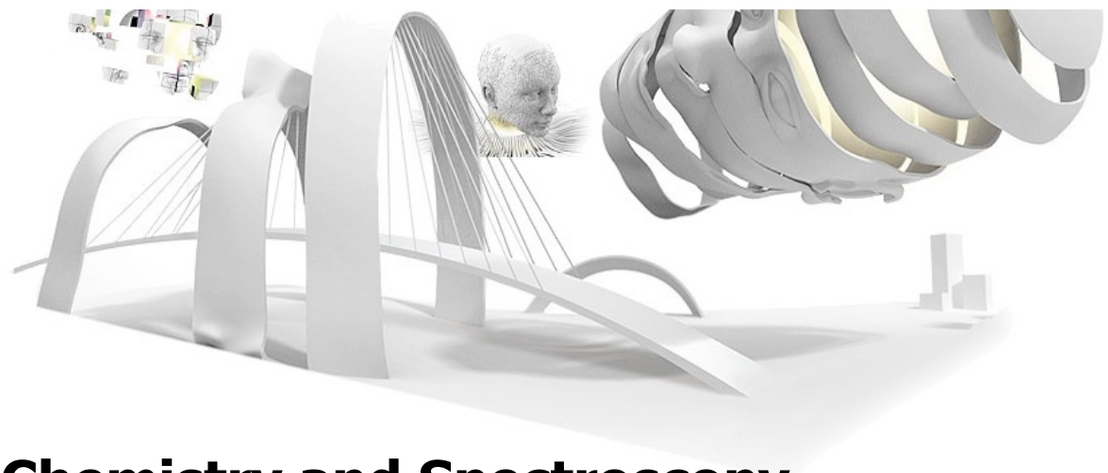
NTNU

Norwegian University of
Science and Technology



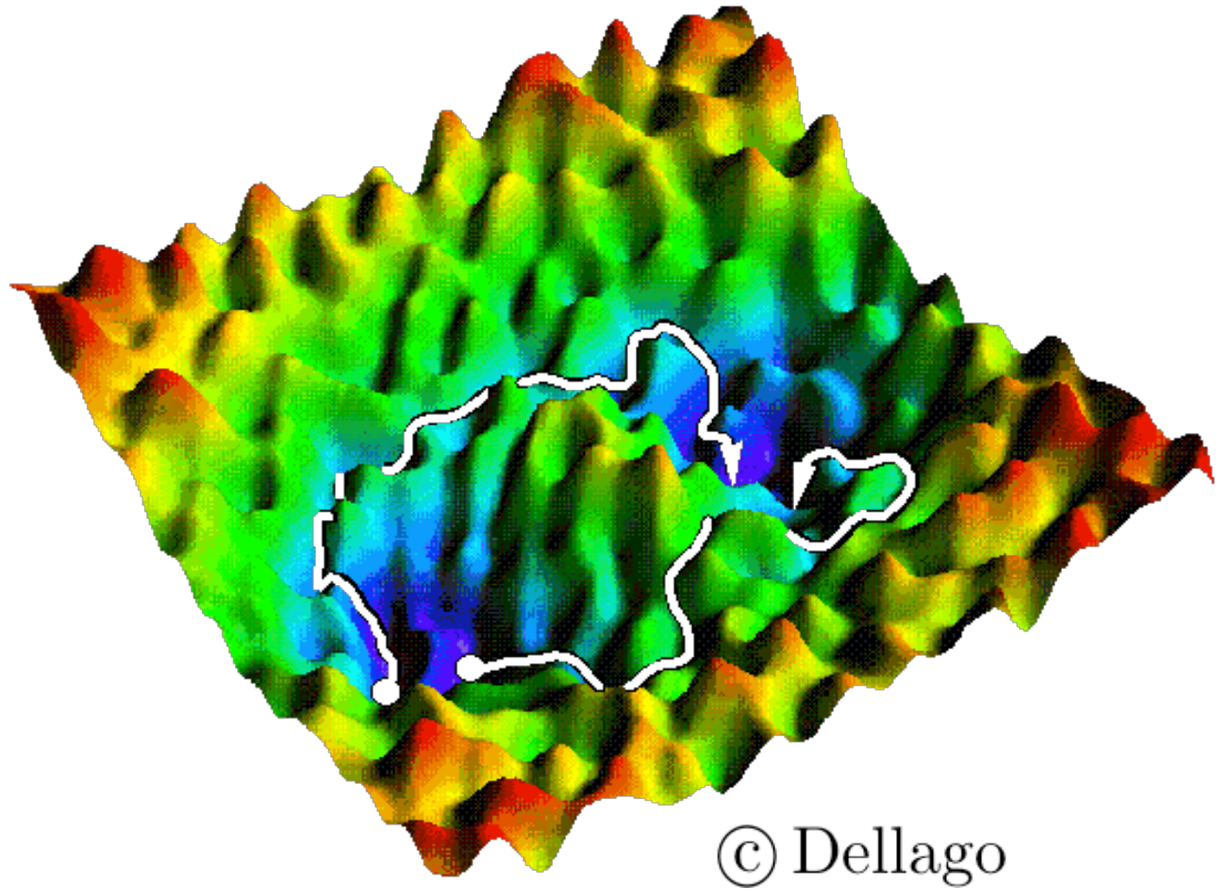
Theoretical Chemistry and Spectroscopy
Han-sur-Lesse, Belgium

Titus S. van Erp, Dep. Chem. NTNU, Norway

Transition Path Sampling methods

Complex energy landscapes
(e.g. reaction is a solvent)

- saddle points uncountable
- reaction coordinates unknown
- many pathways possible



Transition Path Sampling

Chandler, Dellago, Bolhuis,

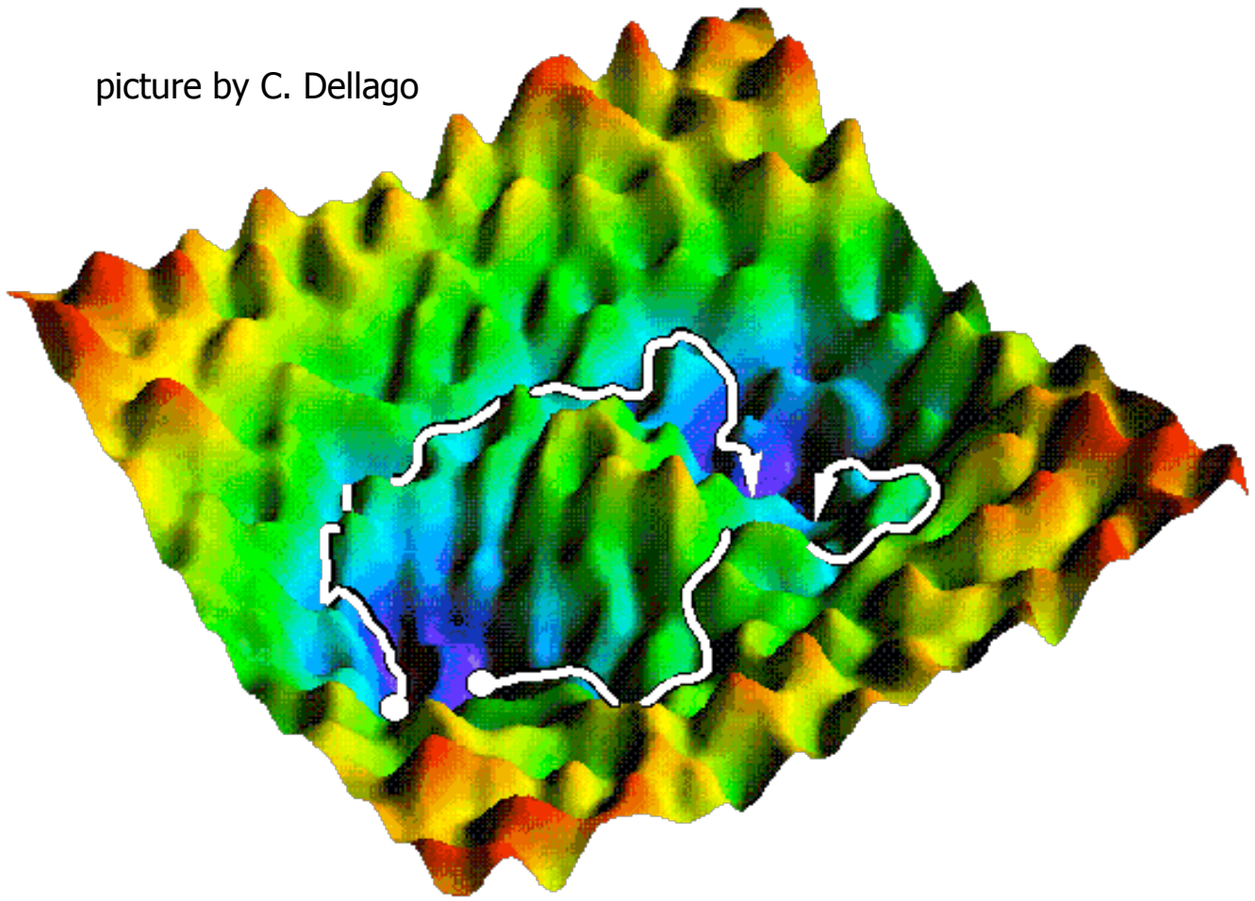
1998

NTNU
Norwegian University of
Science and Technology

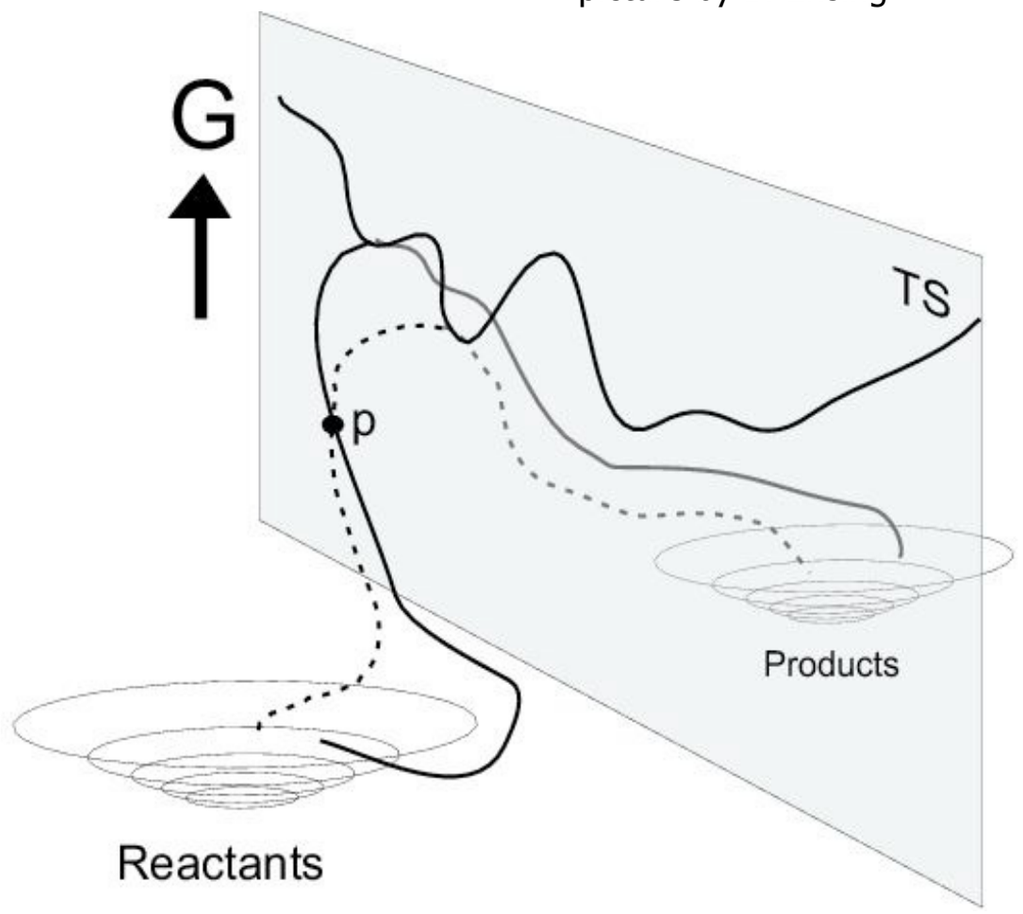
Titus S. van Erp

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

picture by C. Dellago



picture by B. Ensing



(Transition) Path Sampling

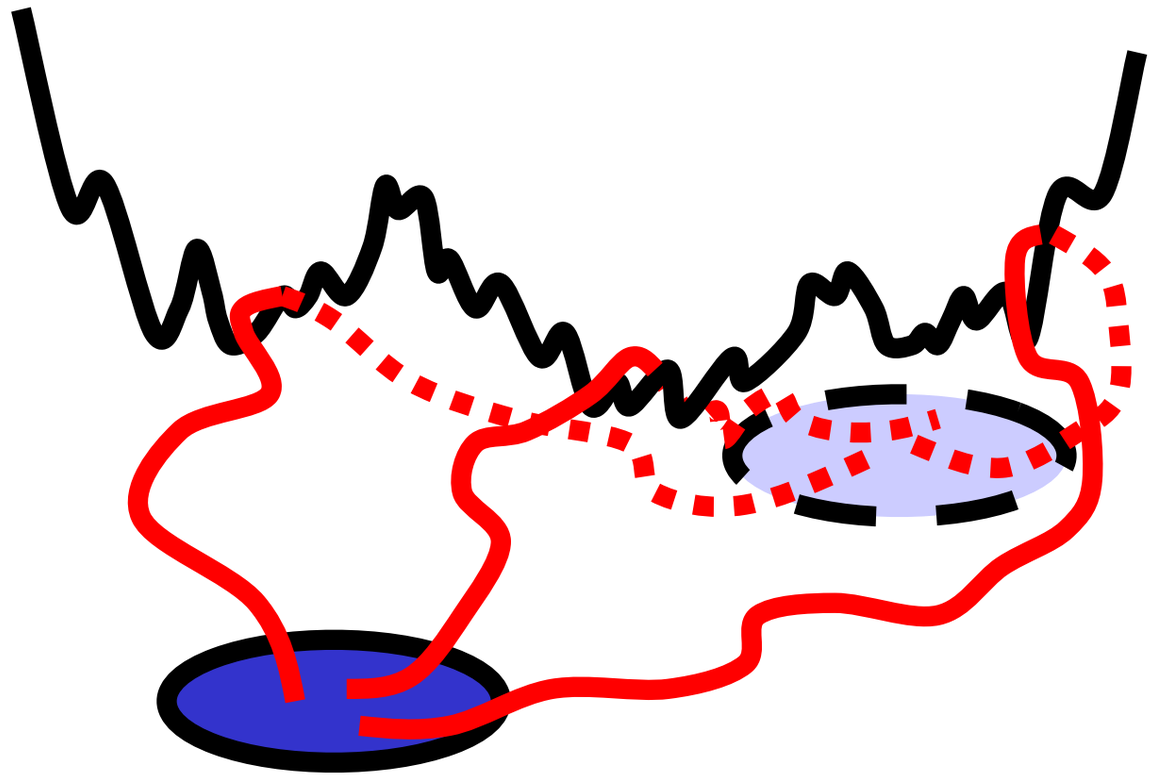
There is no dictionary where these concepts are defined in a manner everyone agrees on. So depending to who you talk, you might get slightly different definitions

1. A robust method to generate an ensemble of reactive trajectories or trajectories for a well-defined path ensemble (containing not necessarily transition paths alone) with their correct statistical weight. This approach does not require the definition of a Reaction Coordinate (RC).

2. A method to calculate reaction rates. Computational expensive and still requires an order parameter (~Reaction Coordinate). This approach has become redundant as TIS and RETIS are faster and more accurate (no fixed path length).

Transition path sampling

Reactive path ensemble:
all trajectories that lead over barrier

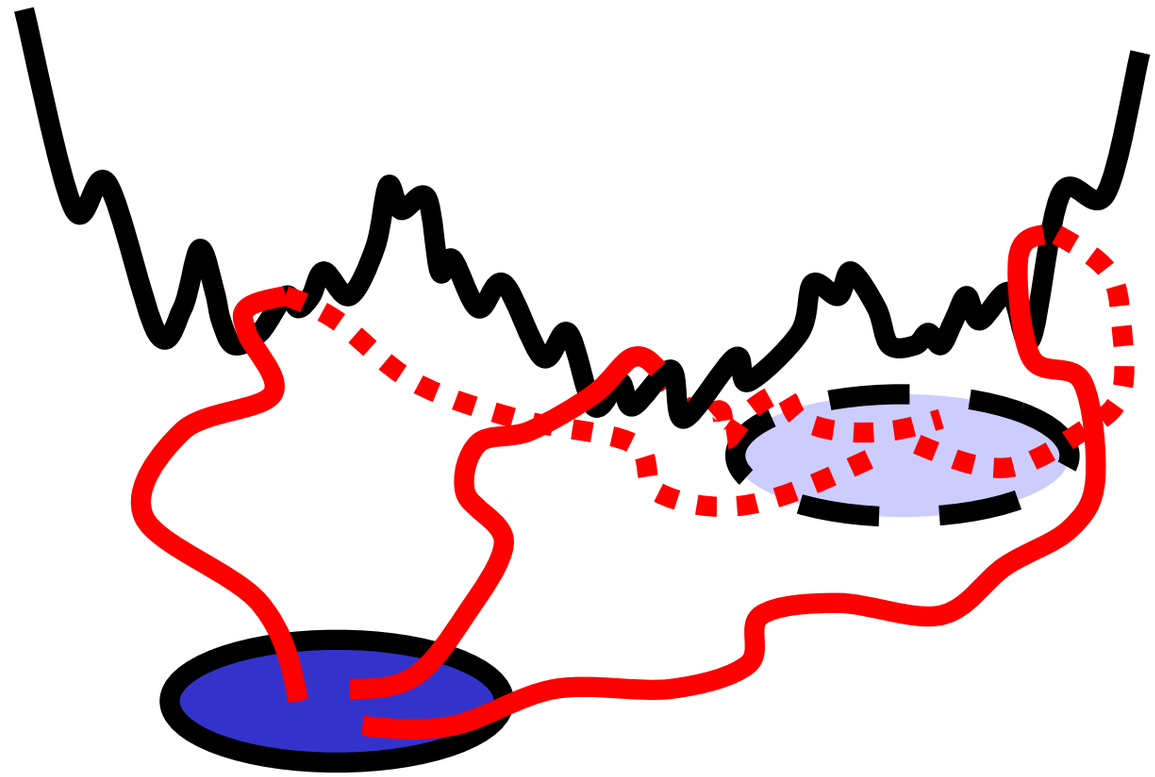


- Sampling by Monte Carlo
- Results in ensemble of pathways
- Requires definition of stable states A,B only

Apply when process of interest
- is a rare event
- is complex and “reaction coordinate” is not known
- If the reaction happens (which is rare), it should go fast

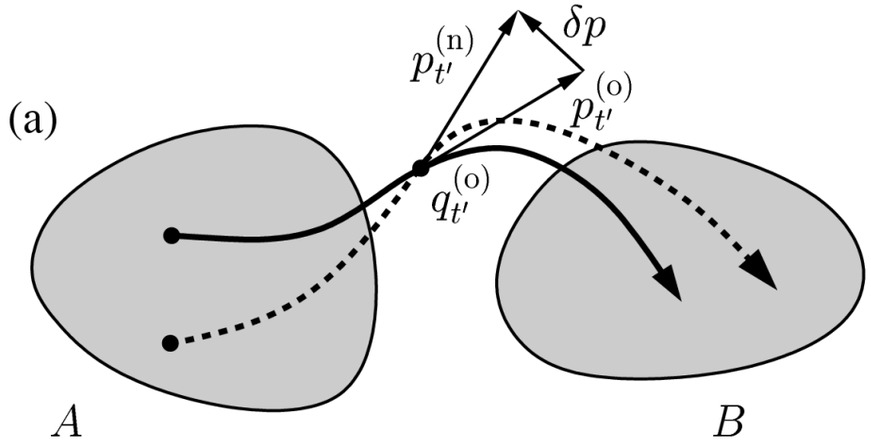
Examples: autodissociation of water, organic reactions in solution, protein folding

Transition path sampling

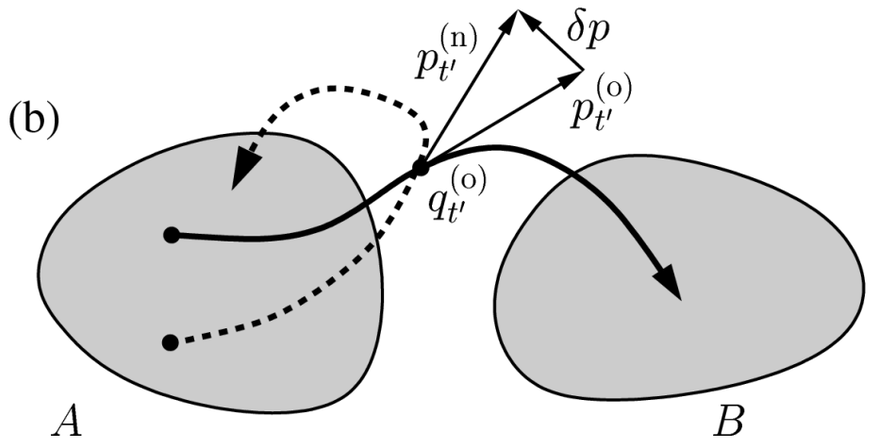


The MC sampling is based on detailed-balance using a Metropolis-Hastings type of algorithm

Shooting moves

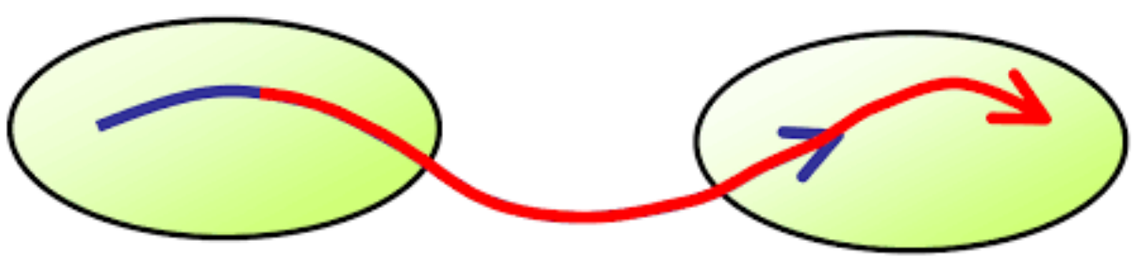


accept



reject

Shifting moves



Detailed balance

$$\rho(o)\pi(o \to n) = \rho(n)\pi(n \to o)$$ **o: old state, n: new state**

Boltzmann statistics: $$\rho(x) = \frac{e^{-\beta E(x)}}{Z}$$

Decomposition into generation and acceptance probability:

$$\pi(x \to x') = P_{\text{gen}}(x \to x')P_{\text{acc}}(x \to x')$$

Metropolis method: use MC moves with symmetric generation probability

$$P_{\text{gen}}(x \to x') = P_{\text{gen}}(x' \to x)$$

and use acceptance probability:

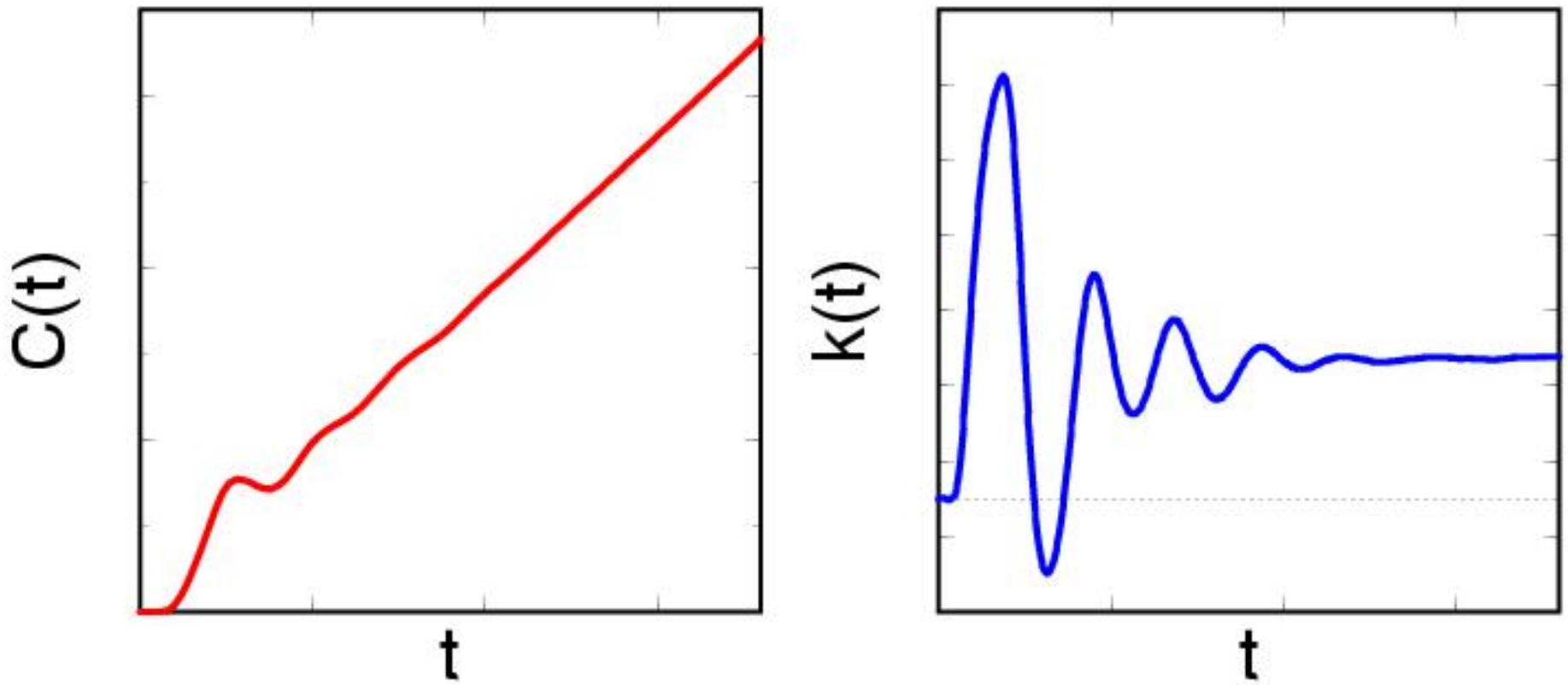
$$P_{\text{acc}}(x \to x') = \min[1, \frac{\rho(x')}{\rho(x)}] = \min[1, \frac{e^{-\beta E(x')}}{e^{-\beta E(x)}}]$$

Metropolis-Hastings method: generalisation for non-symmetric generation probability

$$P_{\text{acc}}(x \to x') = \min[1, \frac{\rho(x')P_{\text{gen}}(x' \to x)}{\rho(x)P_{\text{gen}}(x \to x')}]$$

How this is translated into a practical path sampling algorithm we will see in a few moments .....

Rate constants in TPS

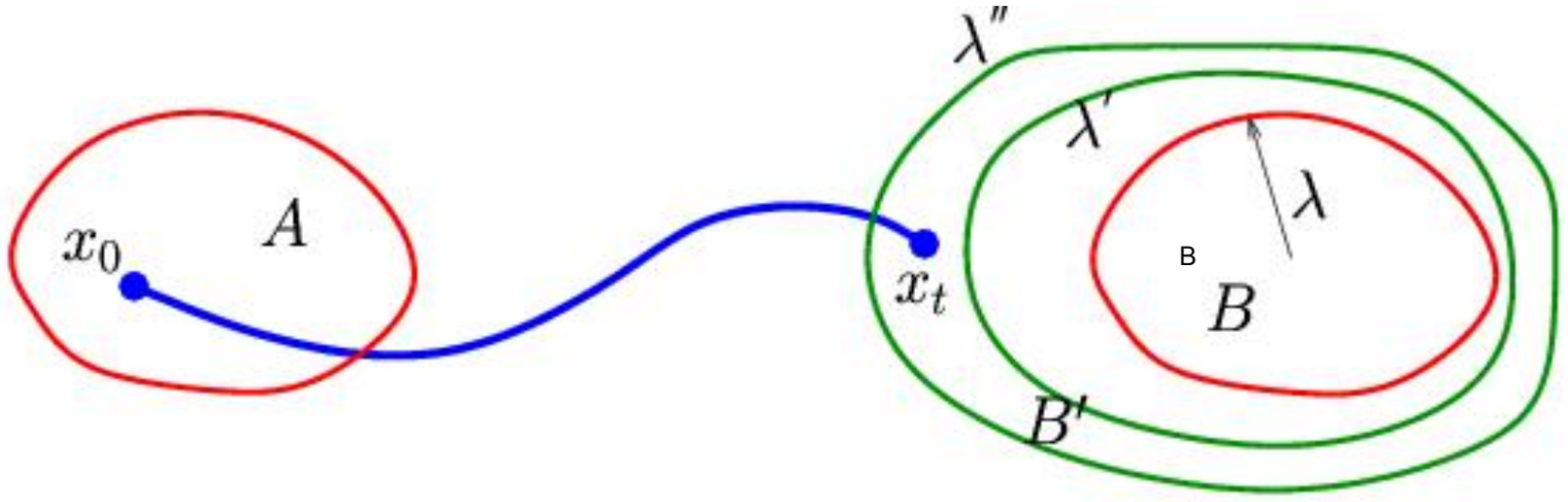


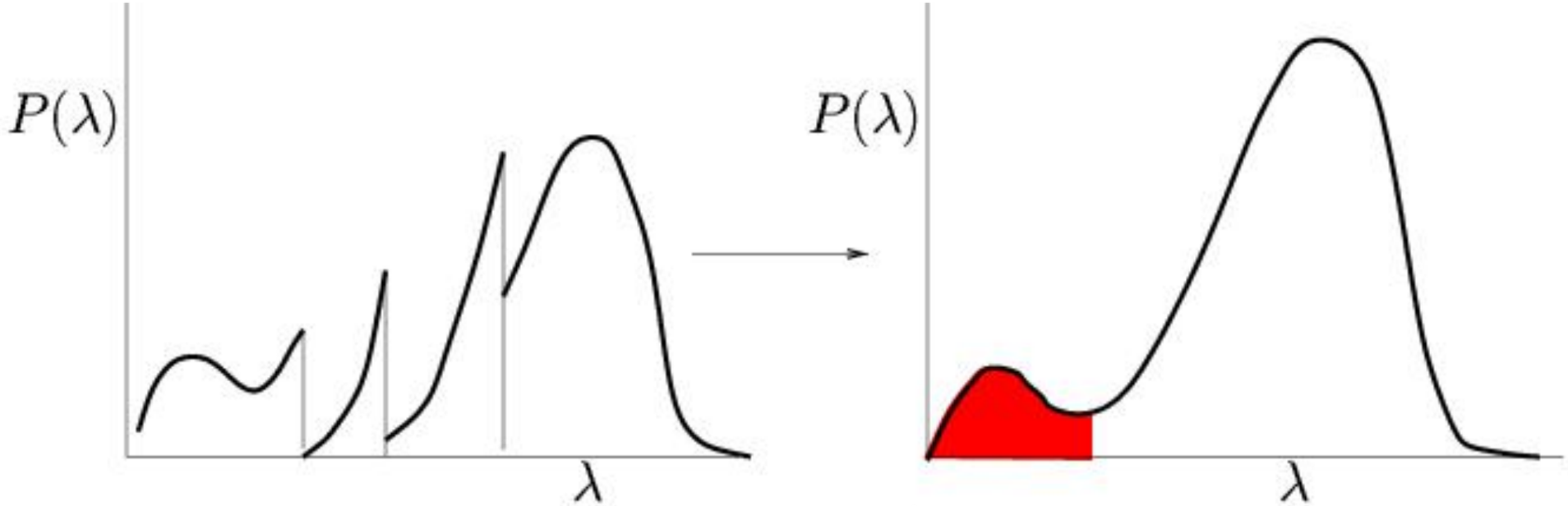
$$C(t) = \frac{\langle h_A(0) h_B(t) \rangle}{\langle h_A(0) \rangle}$$

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

$h_X = 1$ if $\in X$ and 0 otherwise

C(t') *by umbrella sampling*





$$C(t) = \frac{\int_0^{\lambda_B} P(\lambda, t) d\lambda}{\int_0^{\infty} P(\lambda, t) d\lambda}$$

NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

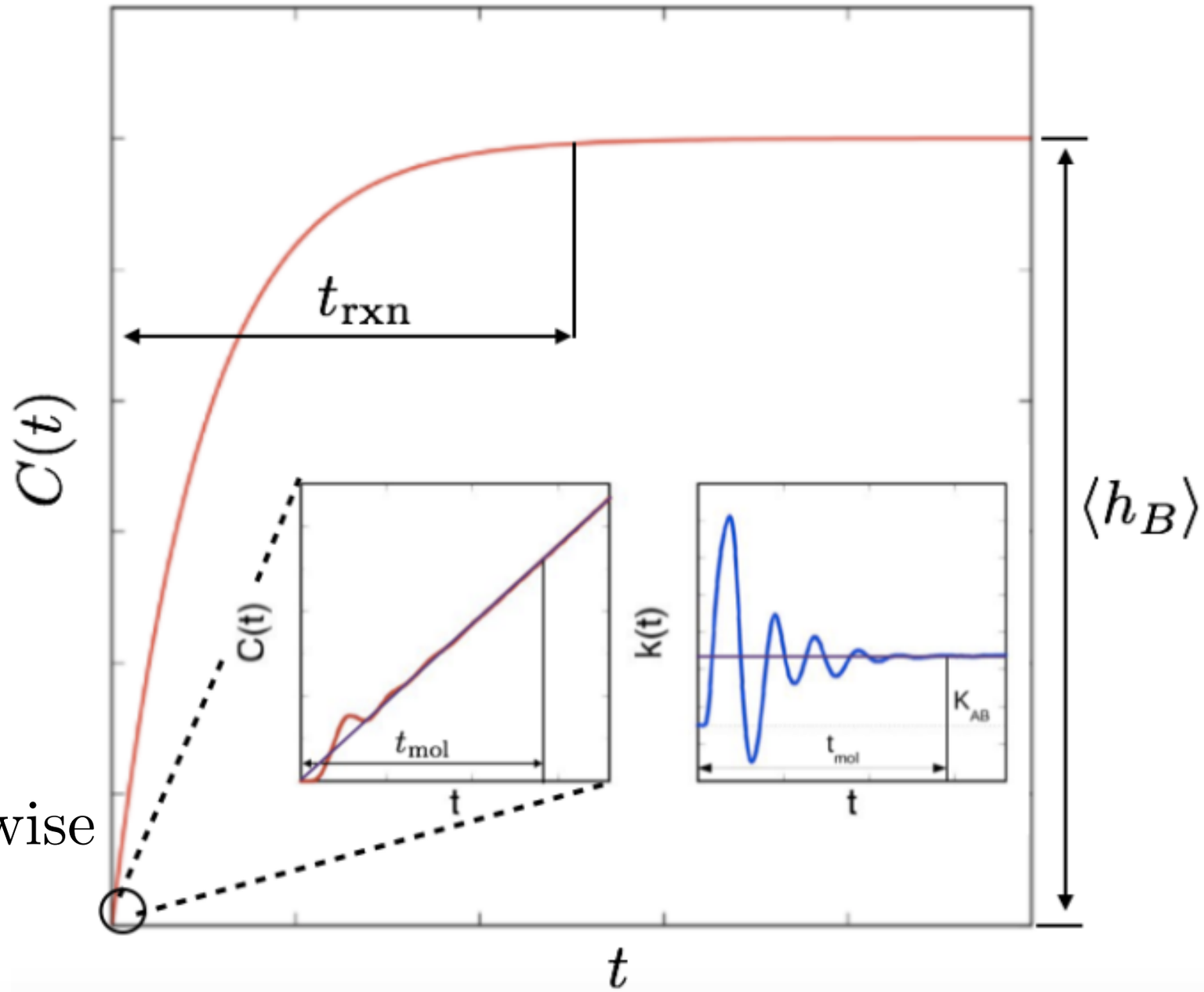
Rate constant theory in Transition Path Sampling vs (RE)TIS

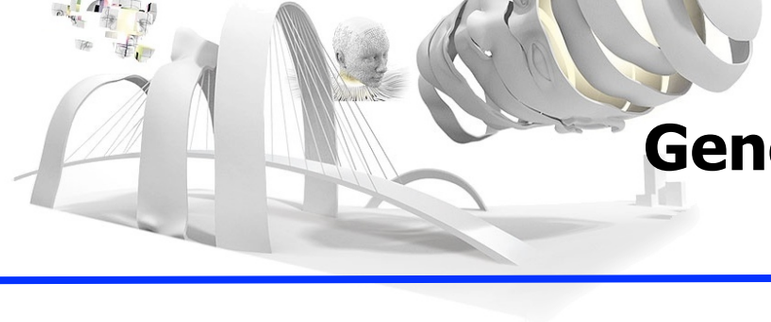
$$k_{AB} = k(t') \text{ for } t_{\text{mol}} < t' \ll t_{\text{rxn}}$$

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

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

$h_X = 1$ if $\in X$ and 0 otherwise

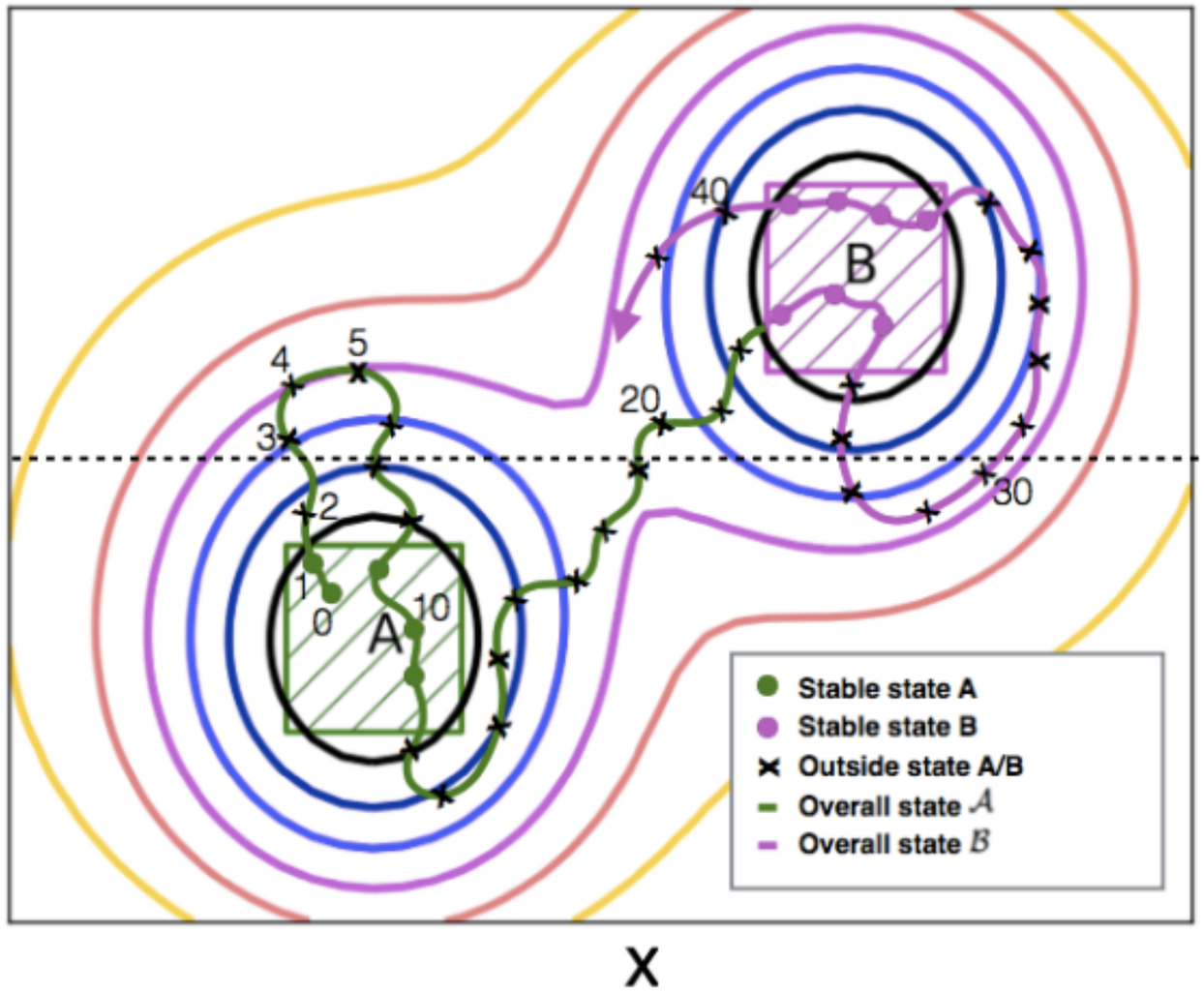


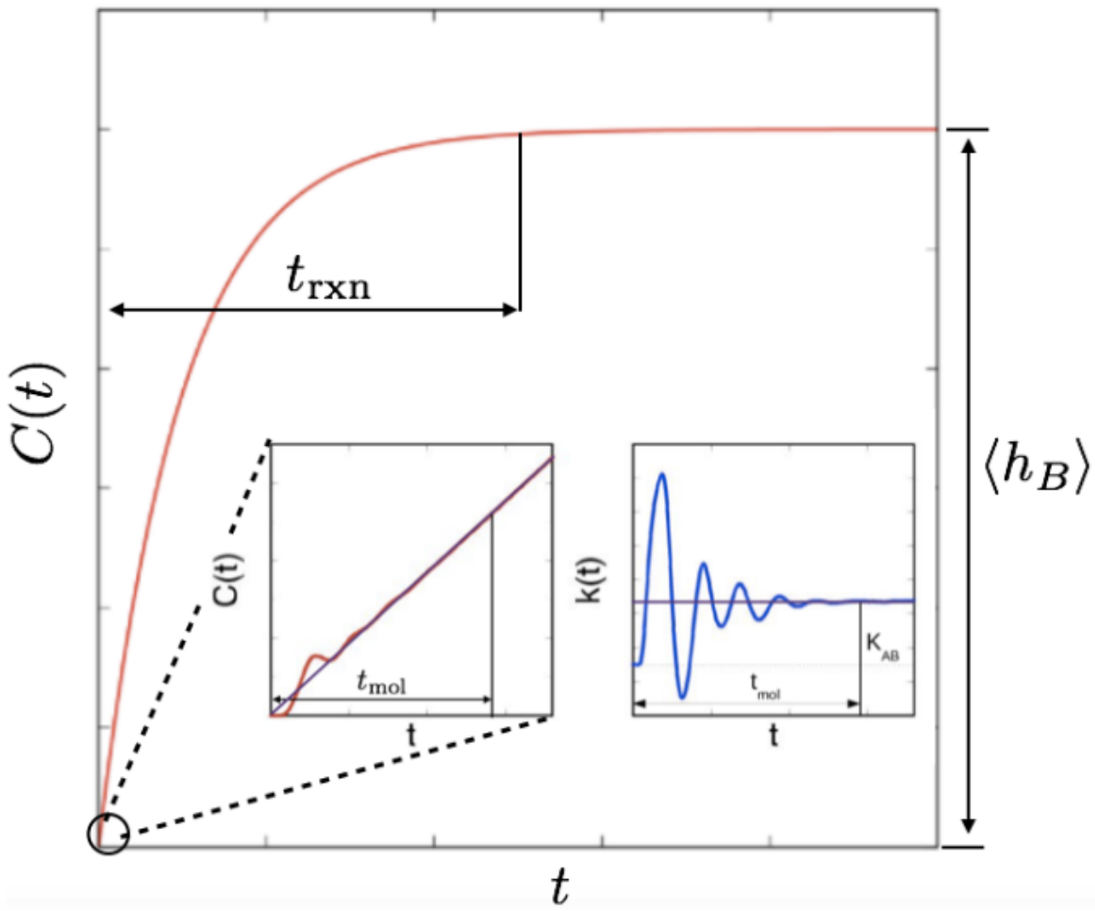


Titus S. van Erp

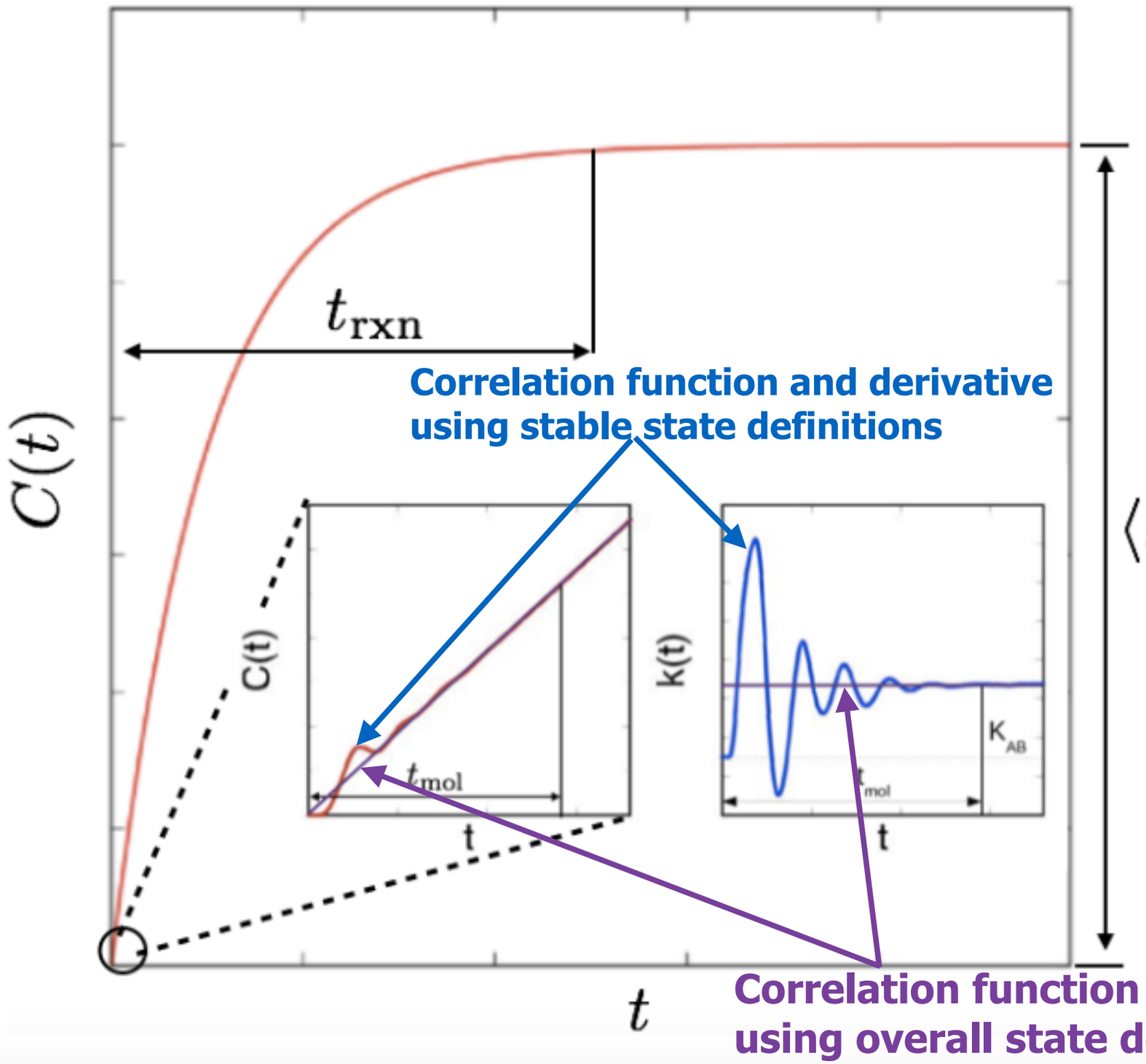
Generation of paths: RETIS

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





Titus S. van Erp



Correlation function and derivative using stable state definitions

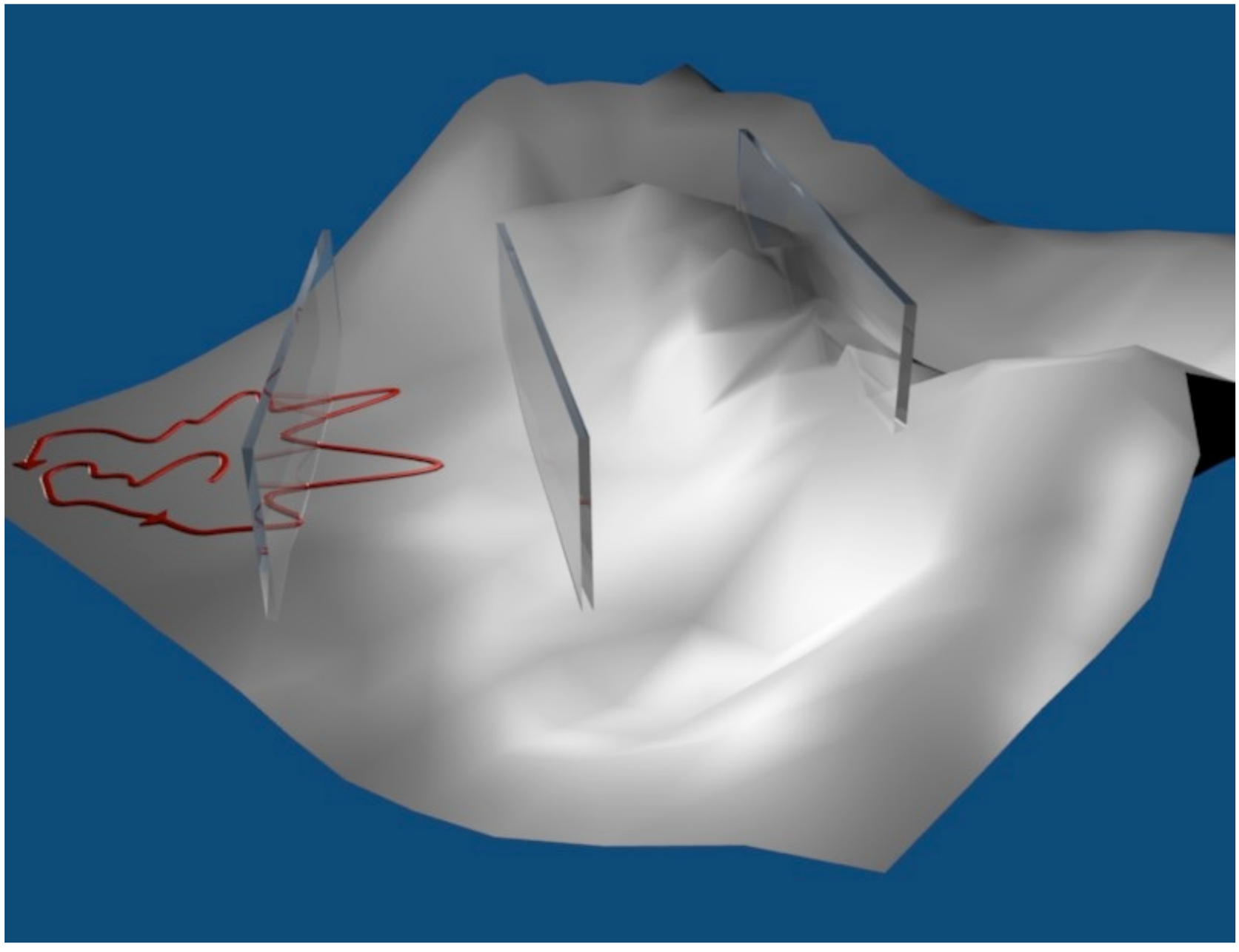
Correlation function and derivative using overall state definitions

NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

Generation of paths: RETIS

The TIS Algorithm



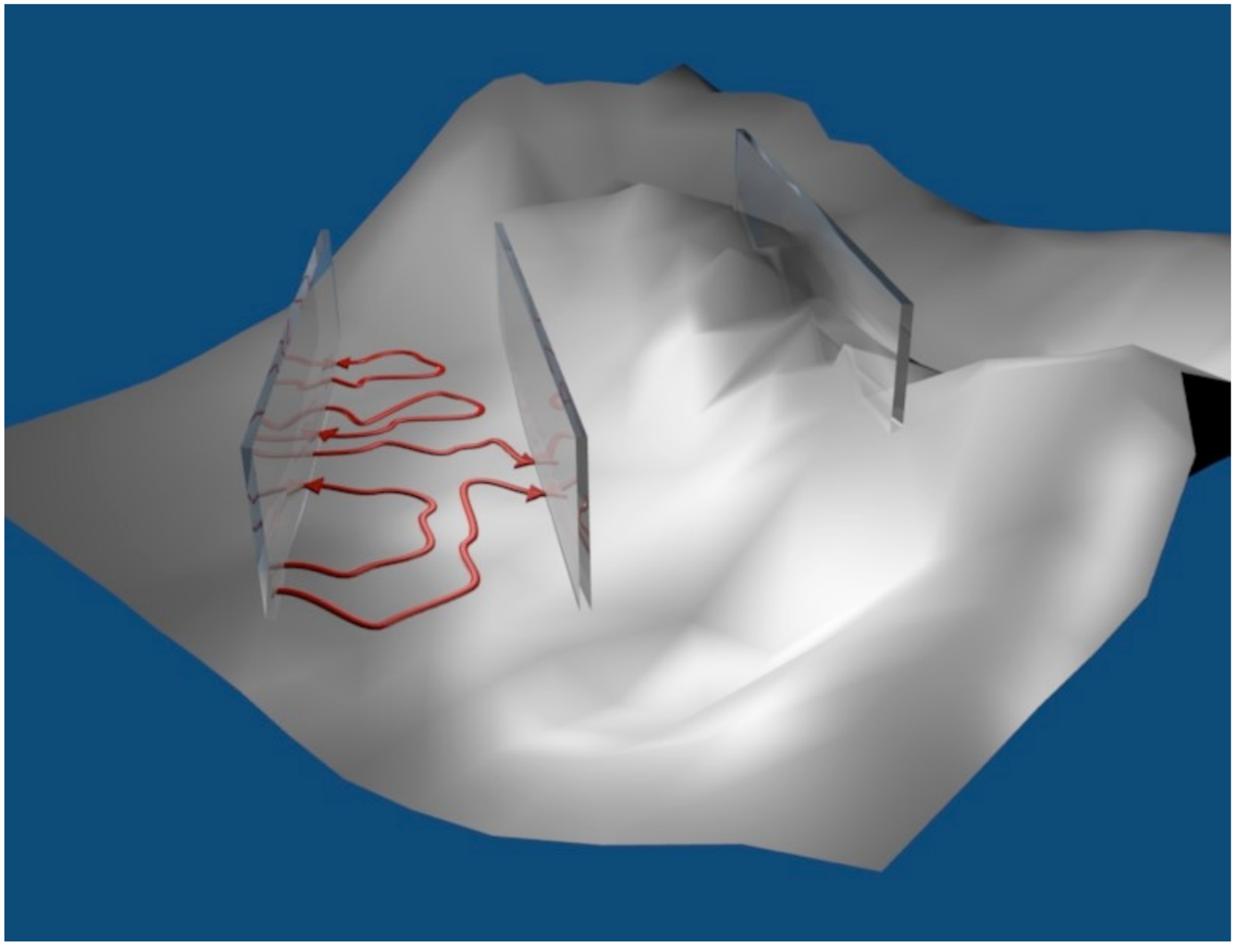
Transition Interface Sampling, van Erp, Moroni, and Bolhuis, JCP 118, 7762 (2003)

NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

Generation of paths: RETIS

The TIS Algorithm



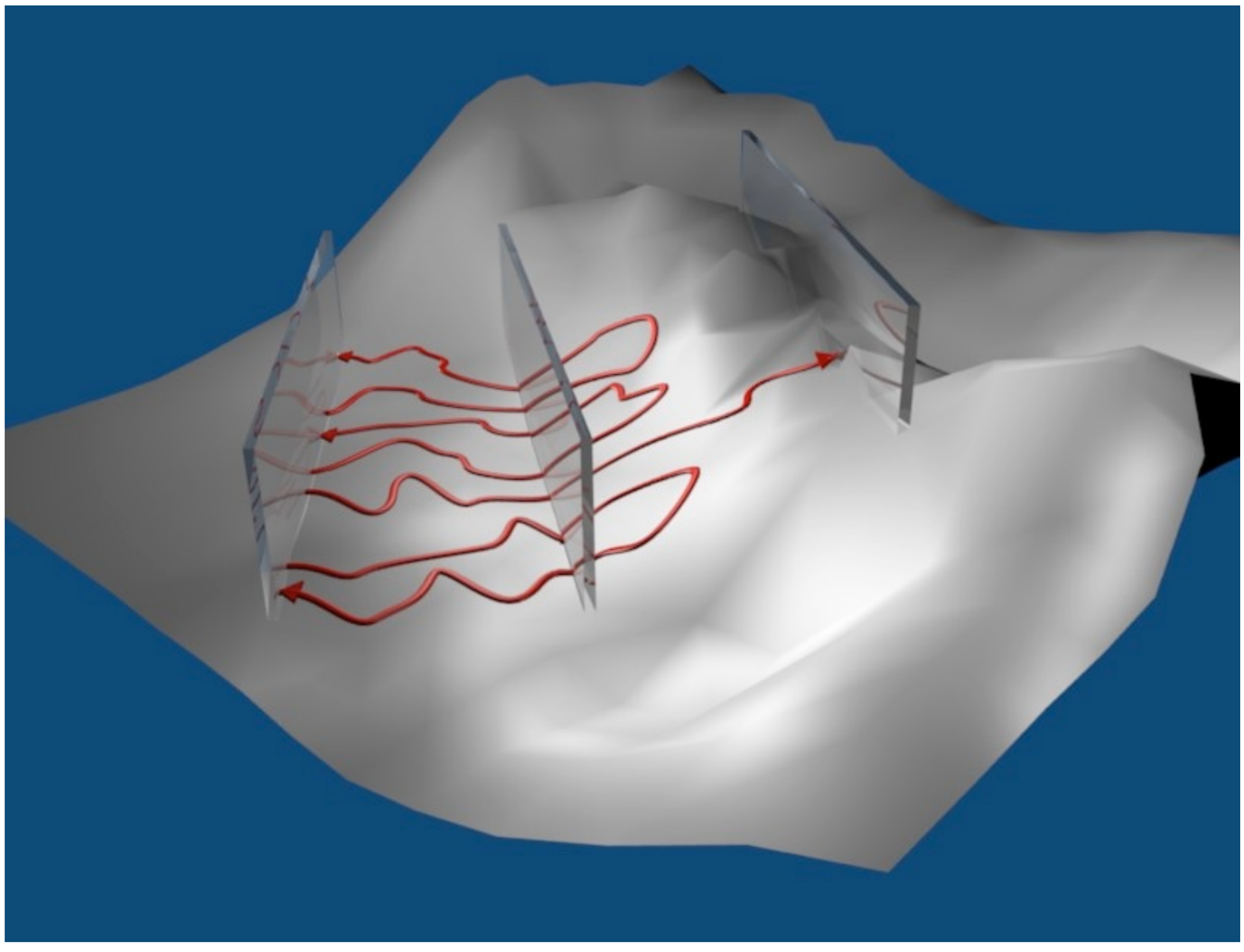
Transition Interface Sampling, van Erp, Moroni, and Bolhuis, JCP 118, 7762 (2003)

NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

Generation of paths: RETIS

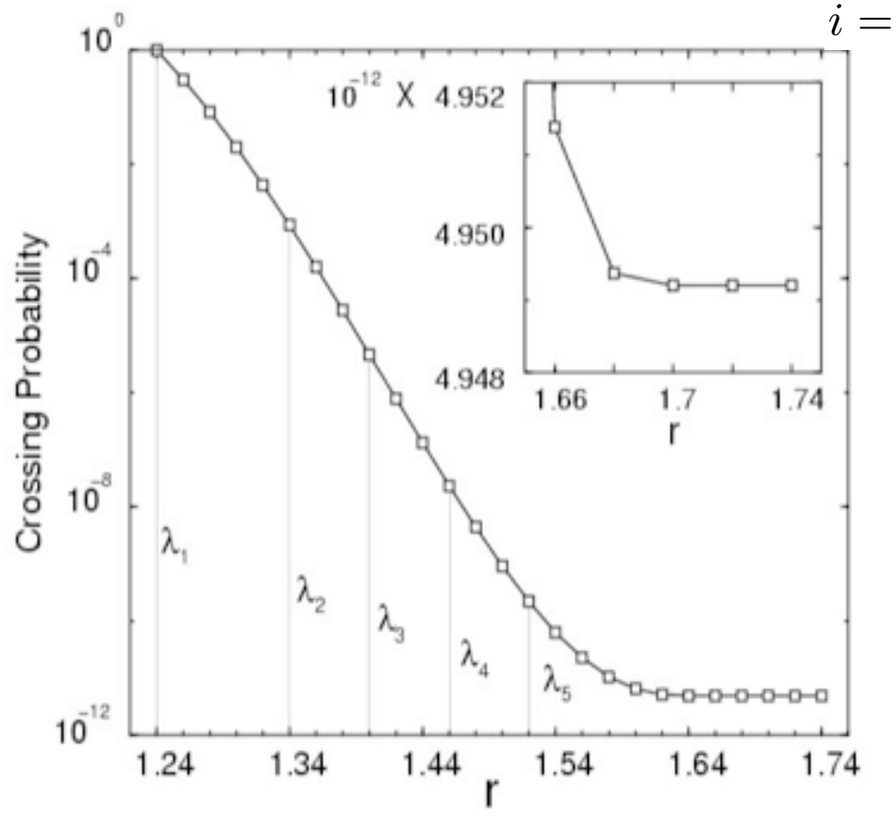
The TIS Algorithm



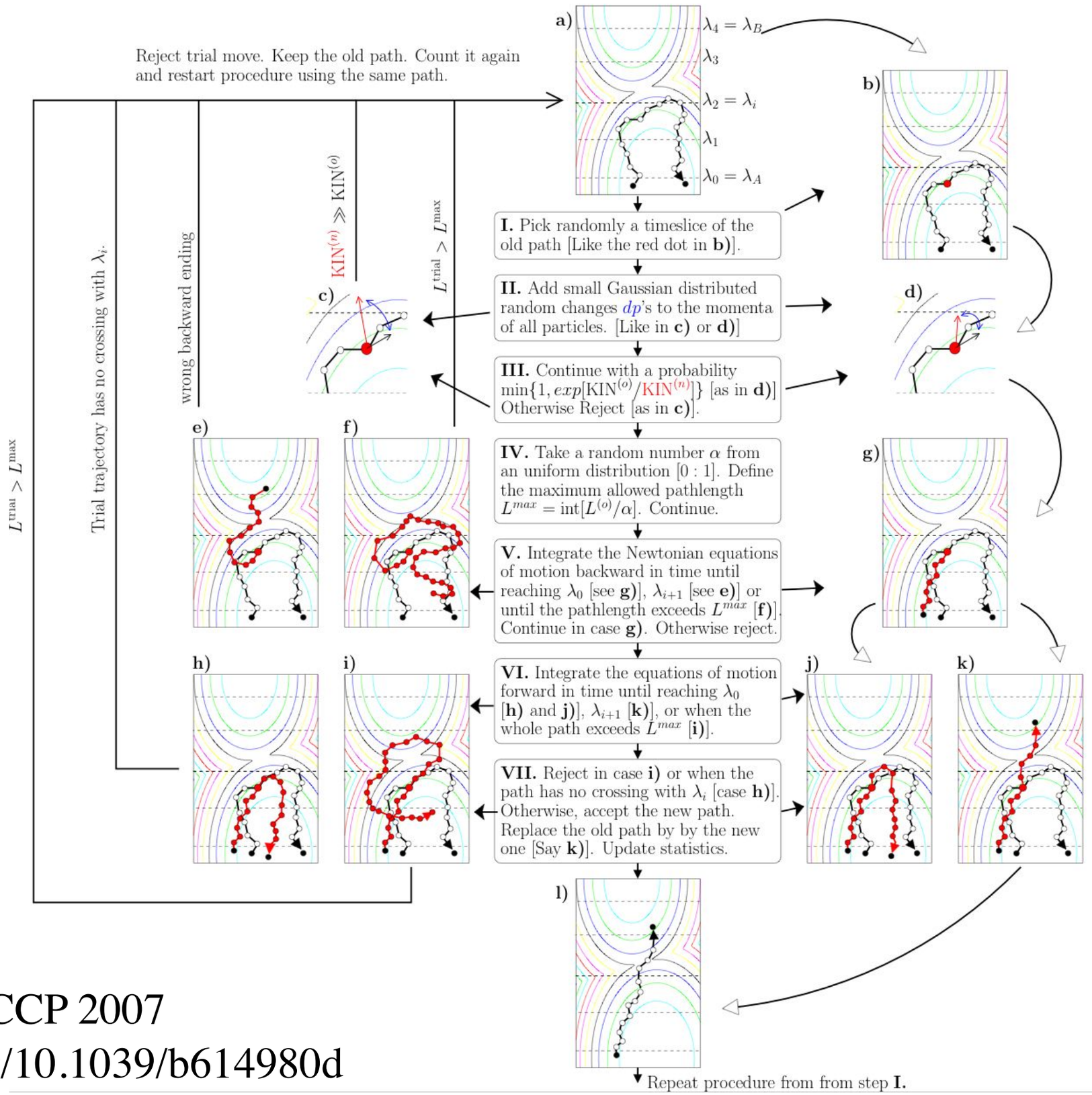
Transition Interface Sampling, van Erp, Moroni, and Bolhuis, JCP 118, 7762 (2003)

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

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



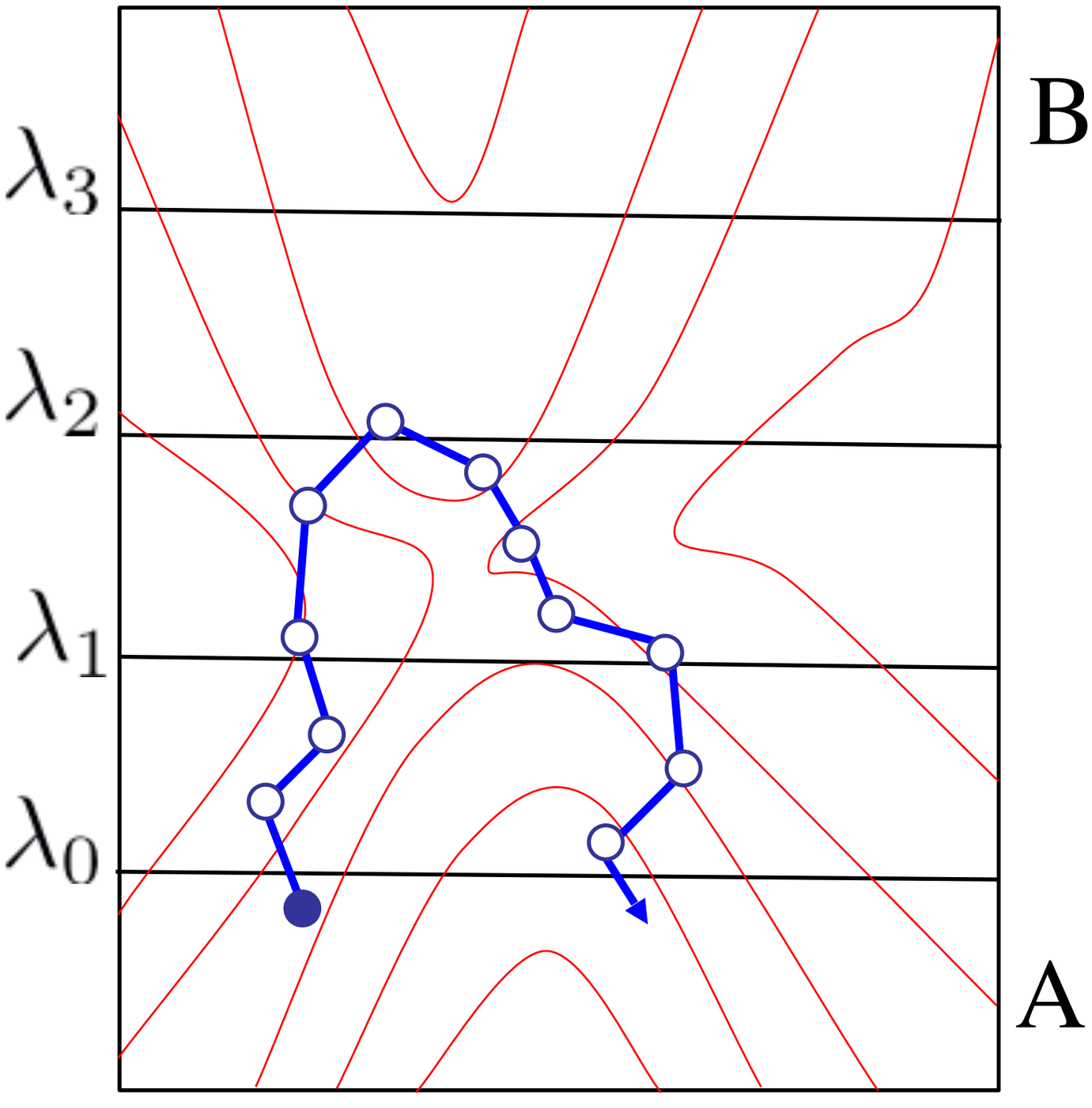
Transition Interface Sampling, van Erp, Moroni, and Bolhuis, J. Chem. Phys. 118, 7762 (2003)

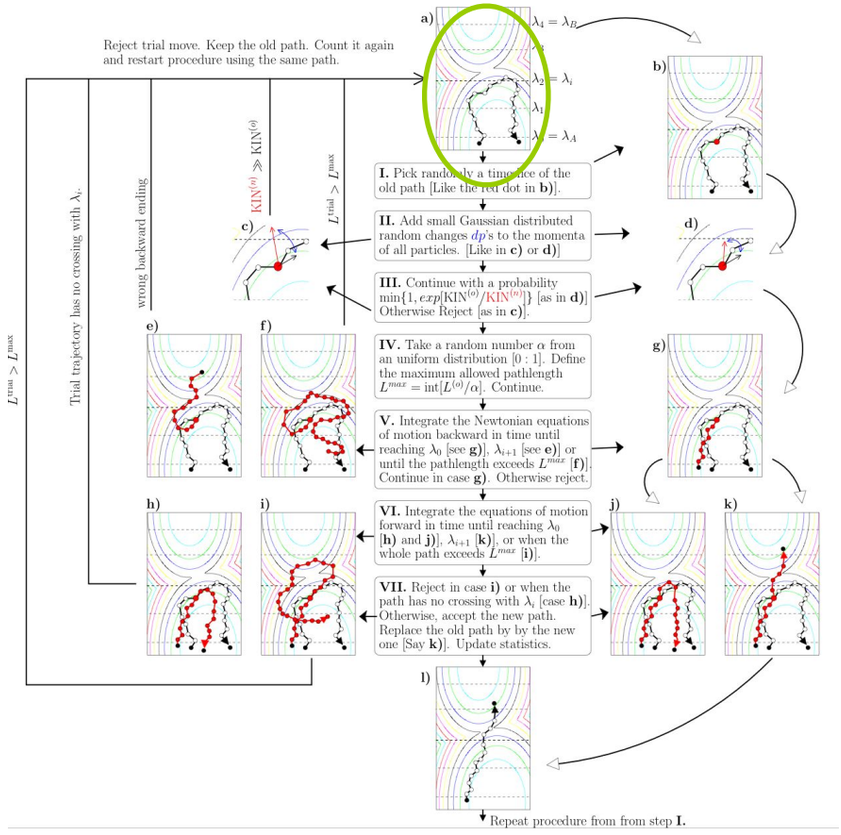


van Erp et al, PCCP 2007
http://dx.doi.org/10.1039/b614980d

Simulating the $[2^+]$ ensemble

The algorithm requires to have
an initial path that fulfils the condition



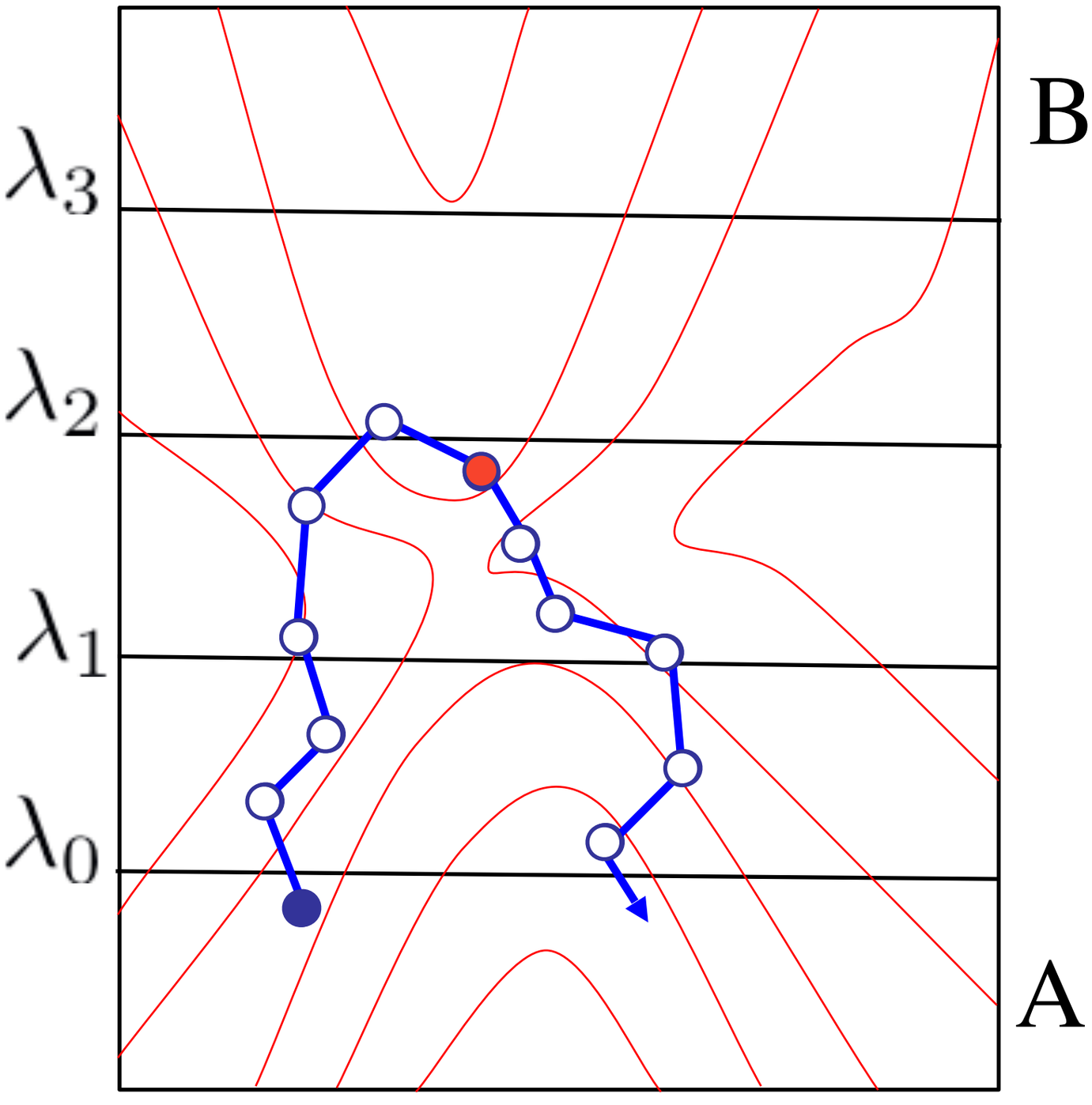


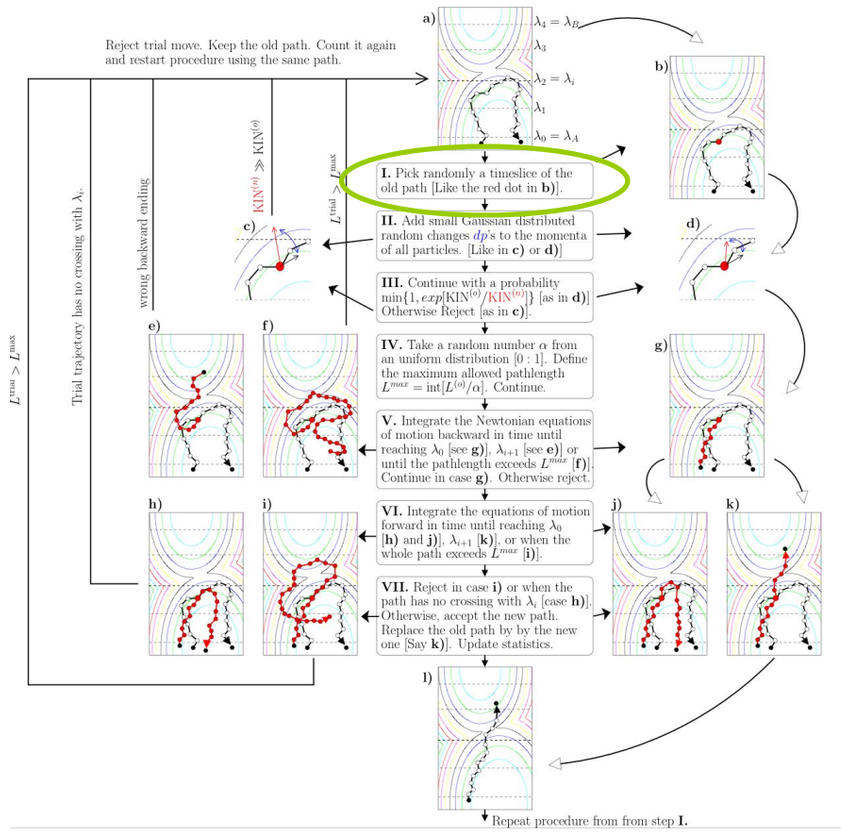
Length of the old path:

$$L^{(o)} = 11$$

Simulating the $[2^+]$ ensemble

I. Pick a random timeslice [1:11] of the old path. Say timeslice nr 6.



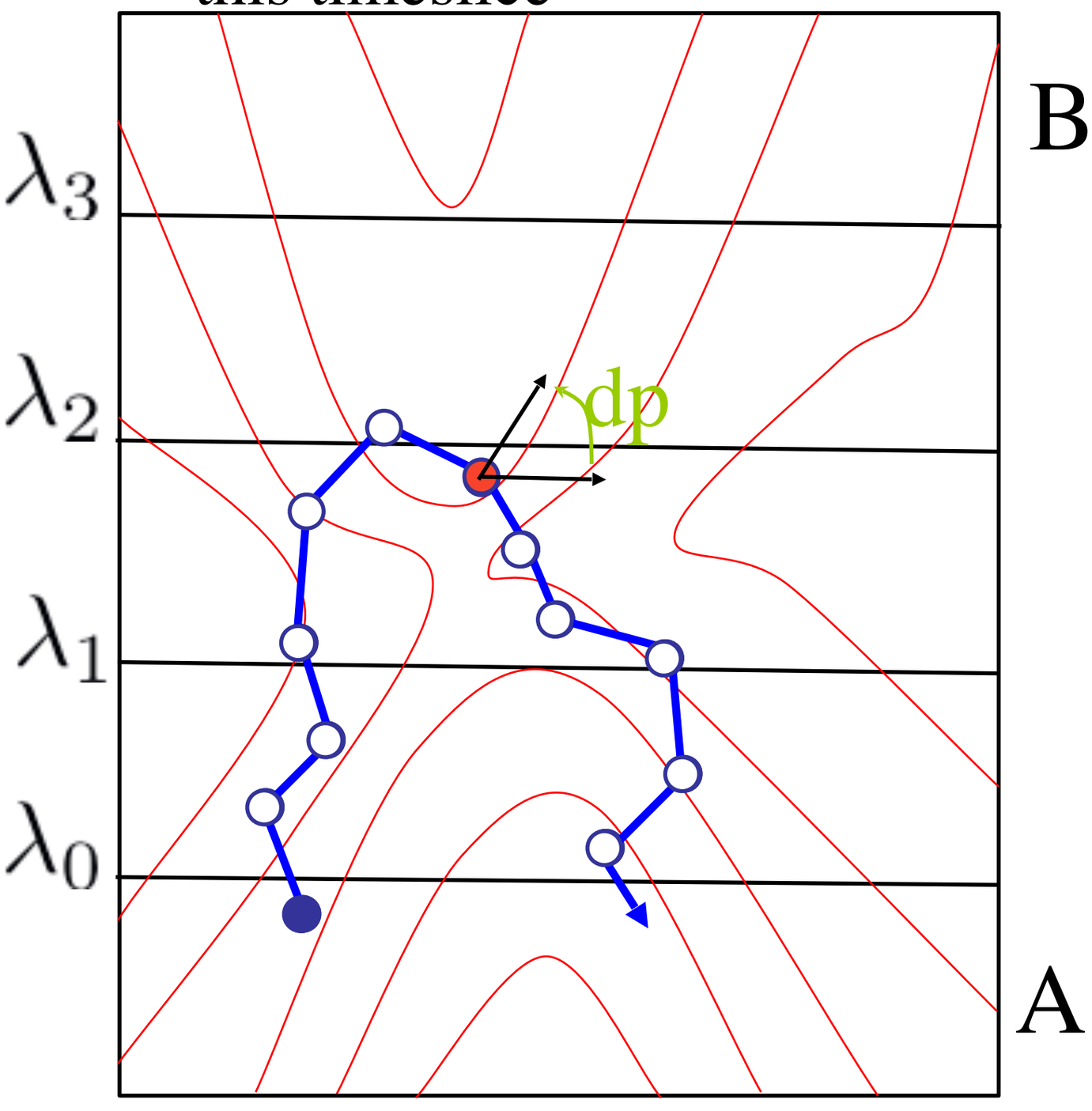


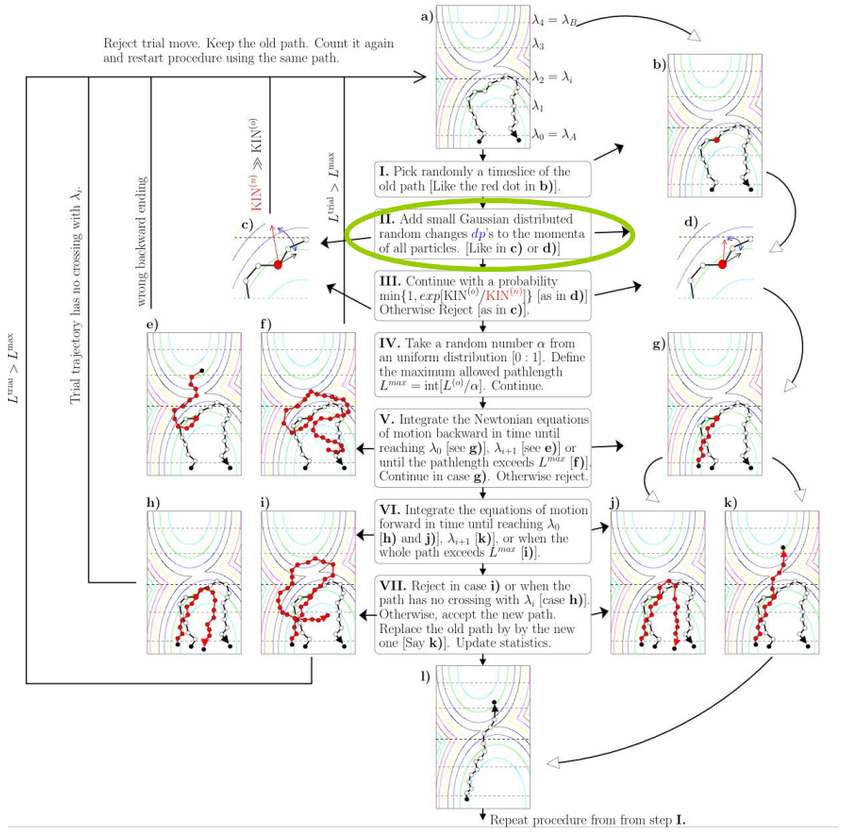
Length of the old path:

$$L^{(o)} = 11$$

Simulating the $[2^+]$ ensemble

II. Add small Gaussian distributed random changes dp to all momenta of this timeslice





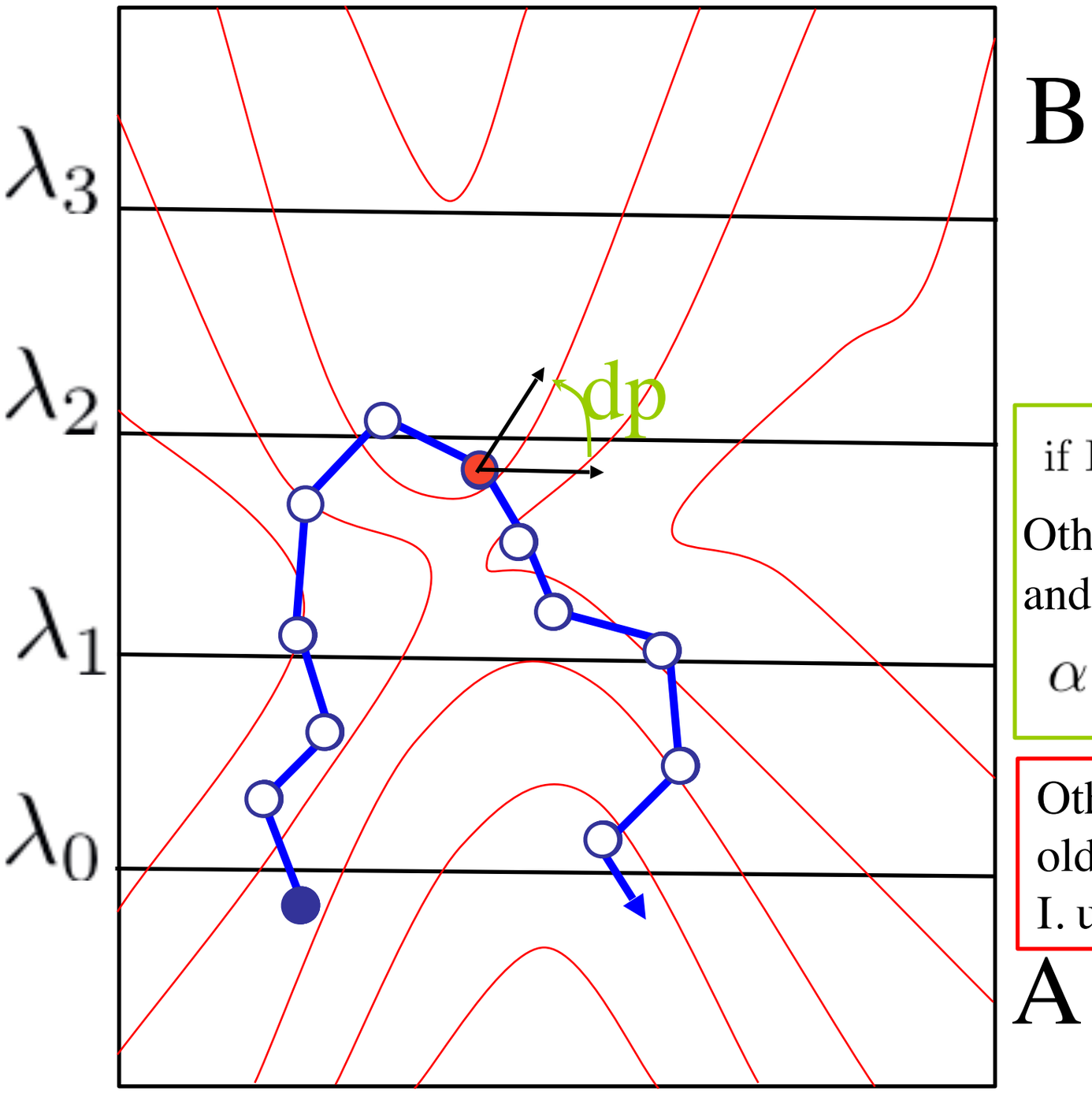
Length of the old path:

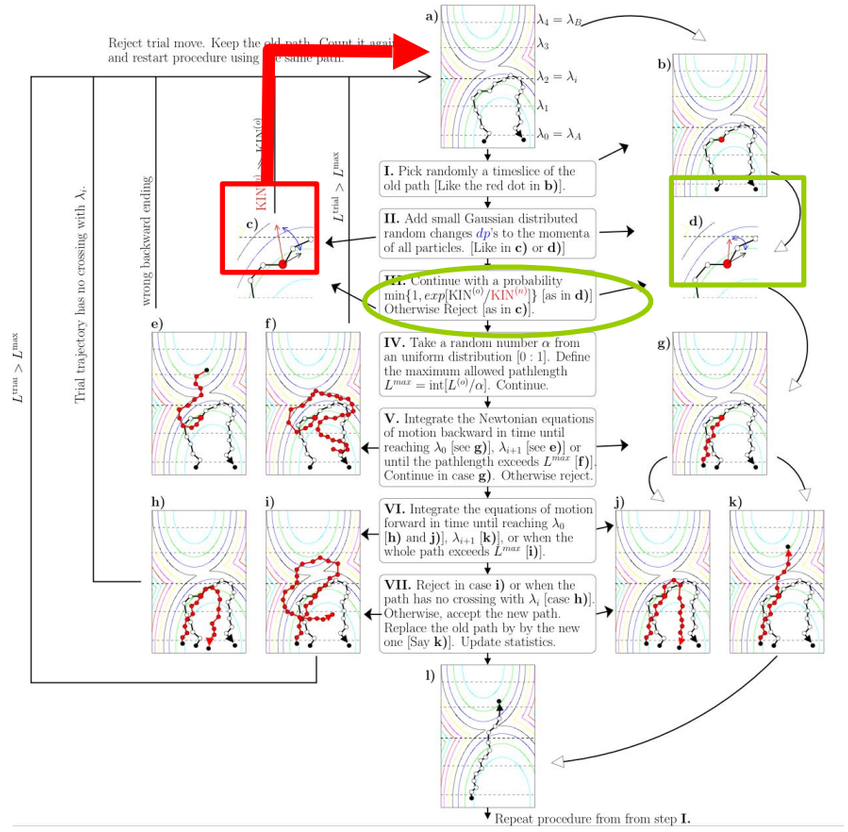
$$L^{(o)} = 11$$

Simulating the $[2^+]$ ensemble

III. Continue with a probability

$$\mathrm{MIN}[1, \exp(\beta[\mathrm{KIN}^{(o)} - \mathrm{KIN}^{(n)}])]$$





if $\mathrm{KIN}^{(n)} < \mathrm{KIN}^{(o)}$ always accept new momenta.

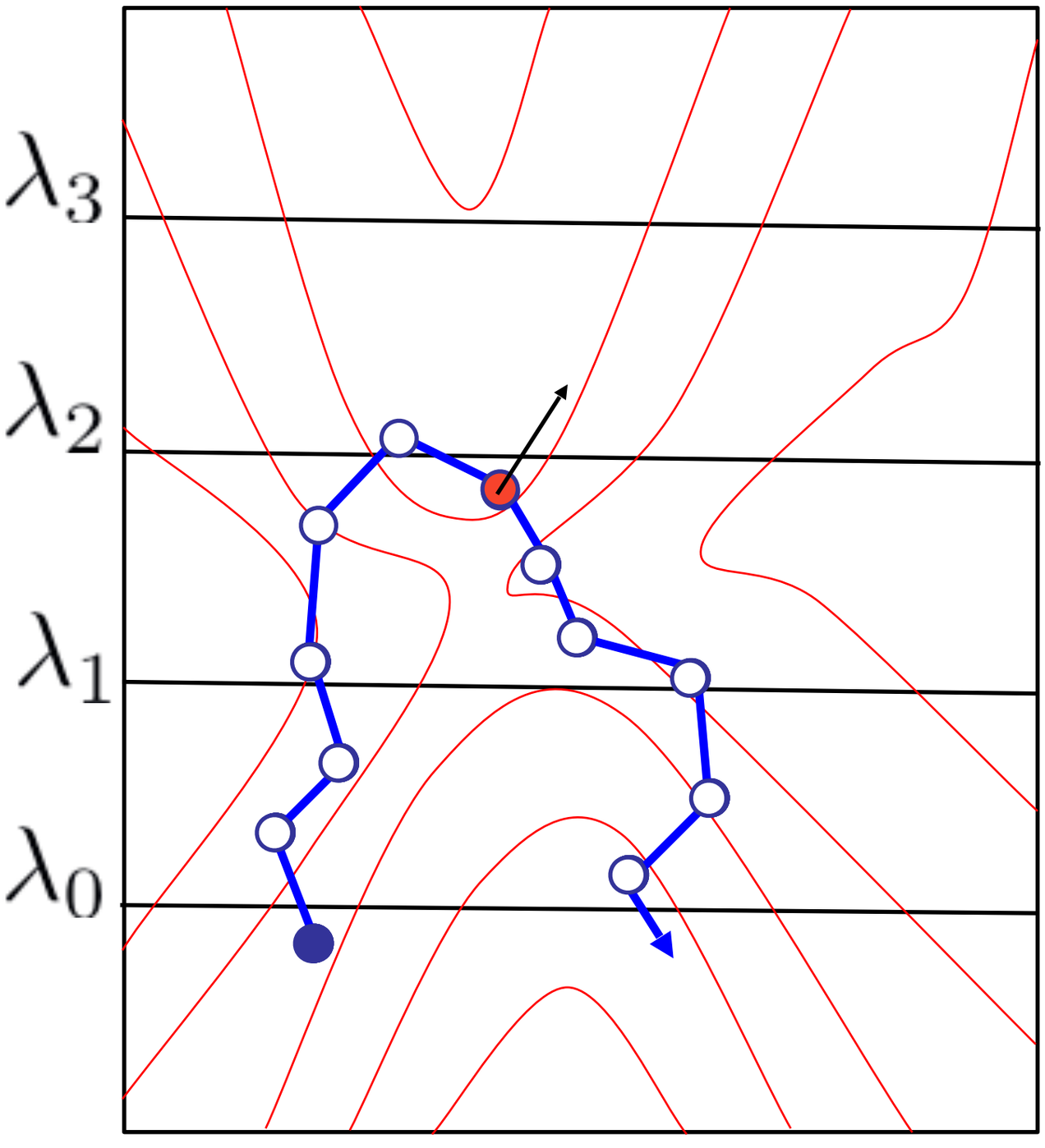
Otherwise take a random number $\alpha \in [0:1]$ and accept if

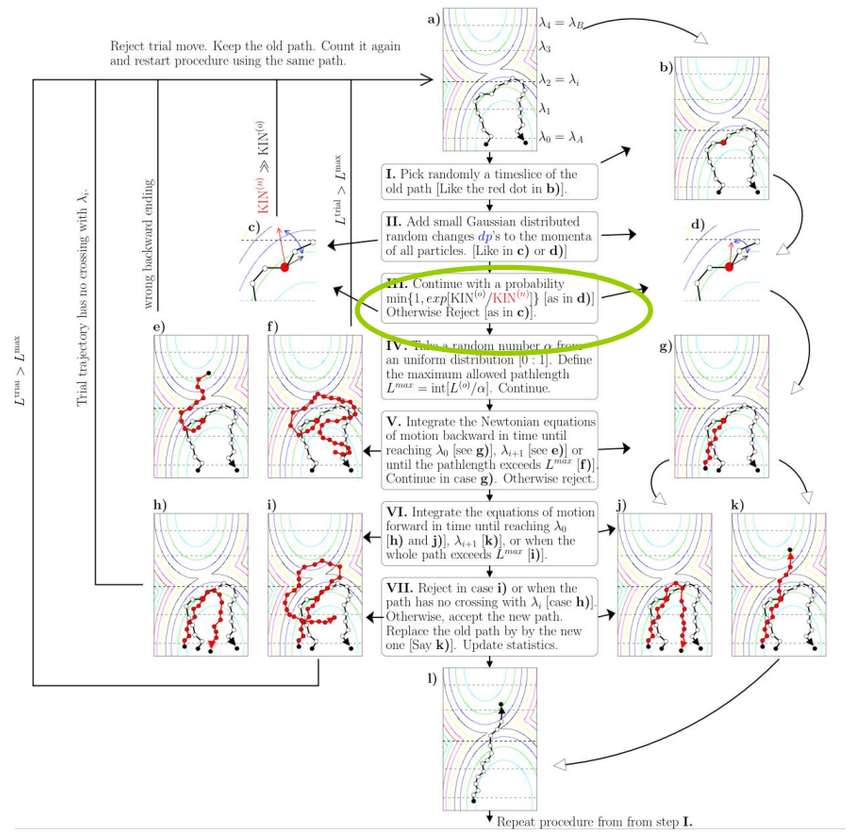
$$\alpha < \exp(-\beta[\mathrm{KIN}^{(n)} - \mathrm{KIN}^{(0)}])$$

Otherwise reject the whole move, keep the old path, count this path again and start from I. using the old path

Simulating the $[2^+]$ ensemble

IV. Define a maximum allowed path length:





Take a new random number $\alpha \in [0:1]$.
The maximum allowed path length $L^{\max}$ is then defined as

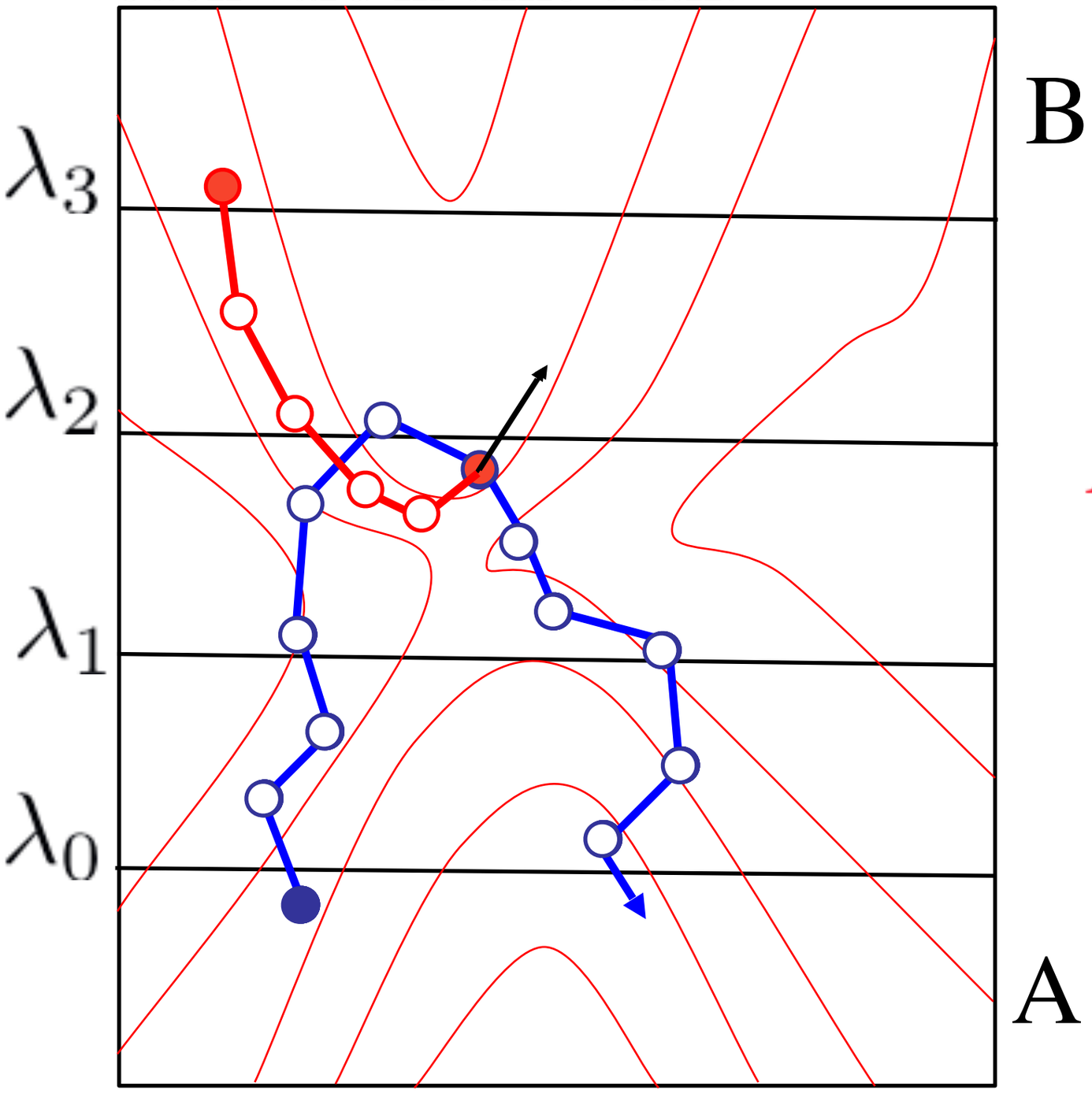
$$L^{\max} = \text{int}[L^{(o)}/\alpha]$$

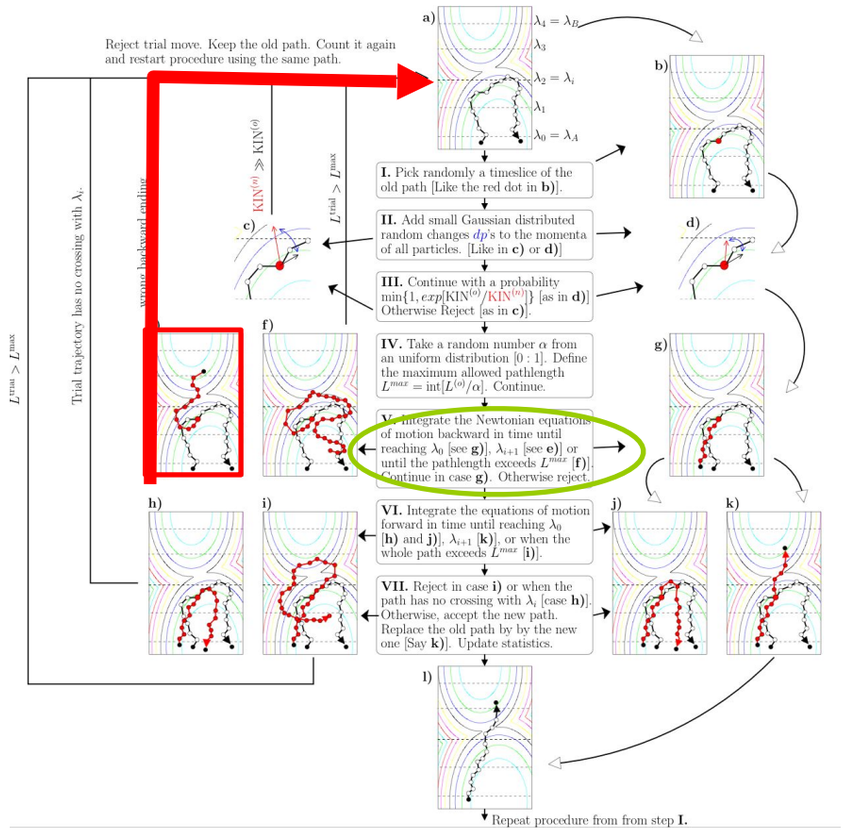
Say $\alpha = 0.888$ then $L^{\max} = 12$

$L^{(o)}$=11. (excluding end-points)
int[11/.888]=int[12.387]=12

Simulating the $[2^+]$ ensemble

V. Integrate the equations of motion backward in time until:

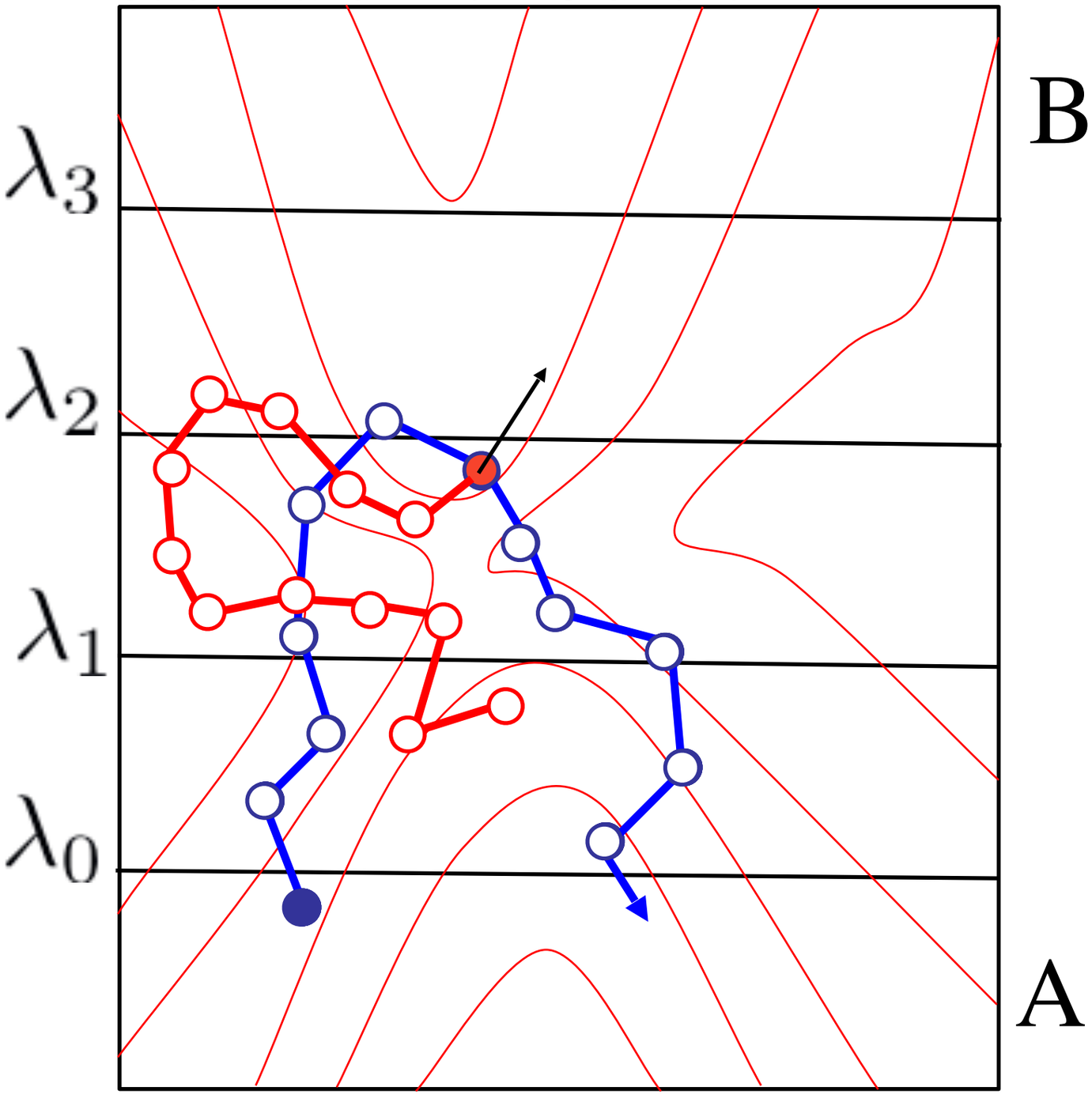


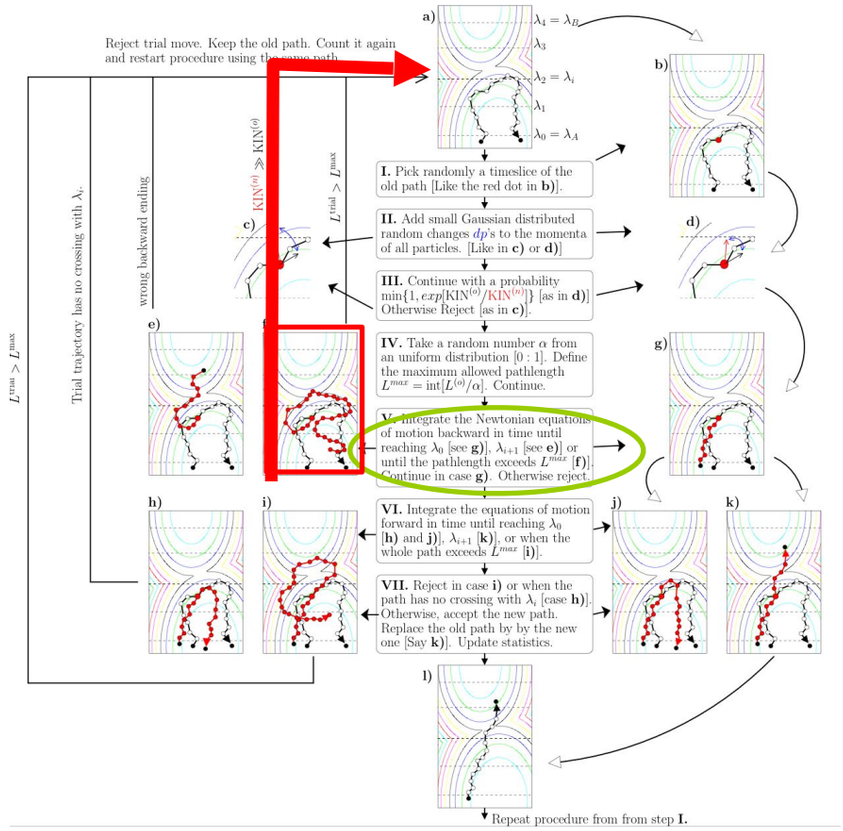


A) You reach $\lambda_3 \Rightarrow$ reject

Simulating the $[2^+]$ ensemble

V. Integrate the equations of motion backward in time until:

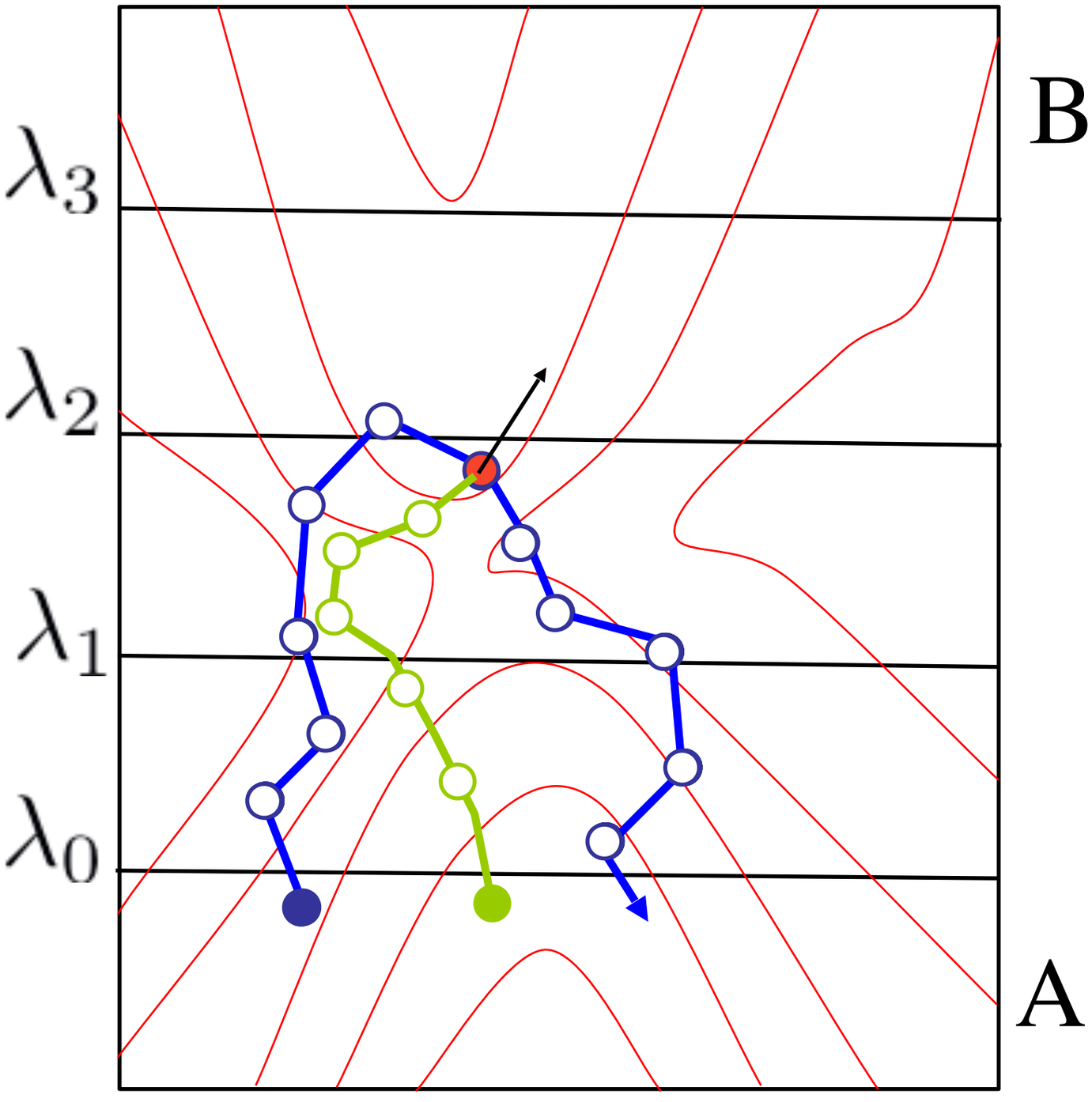


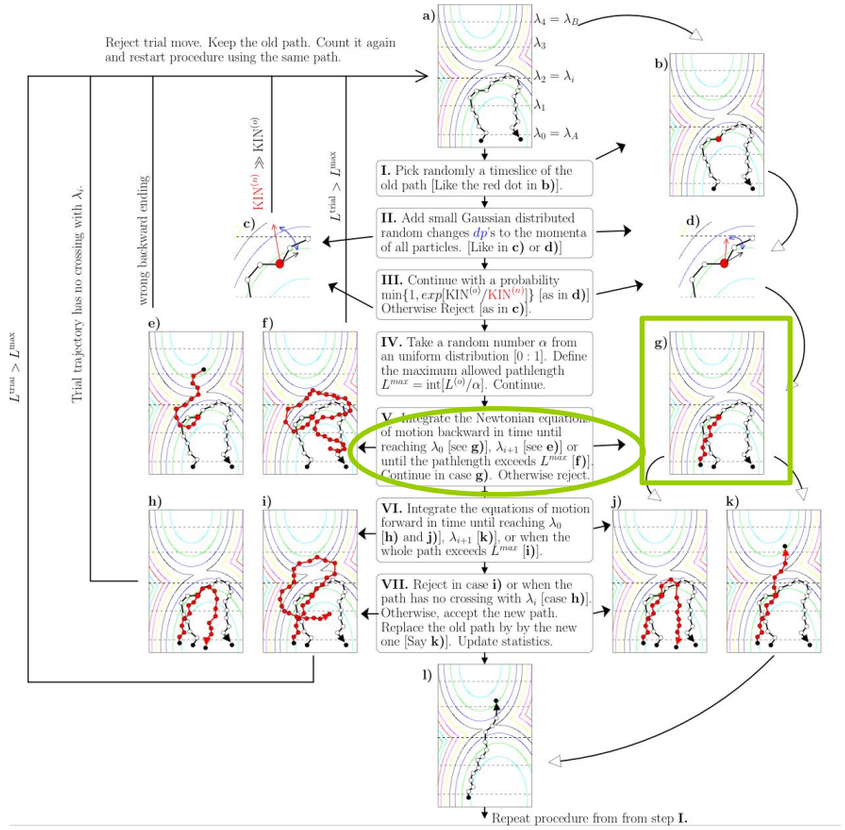


B) Backward trajectory exceeds $L^{\max}$

Simulating the $[2^+]$ ensemble

V. Integrate the equations of motion backward in time until:

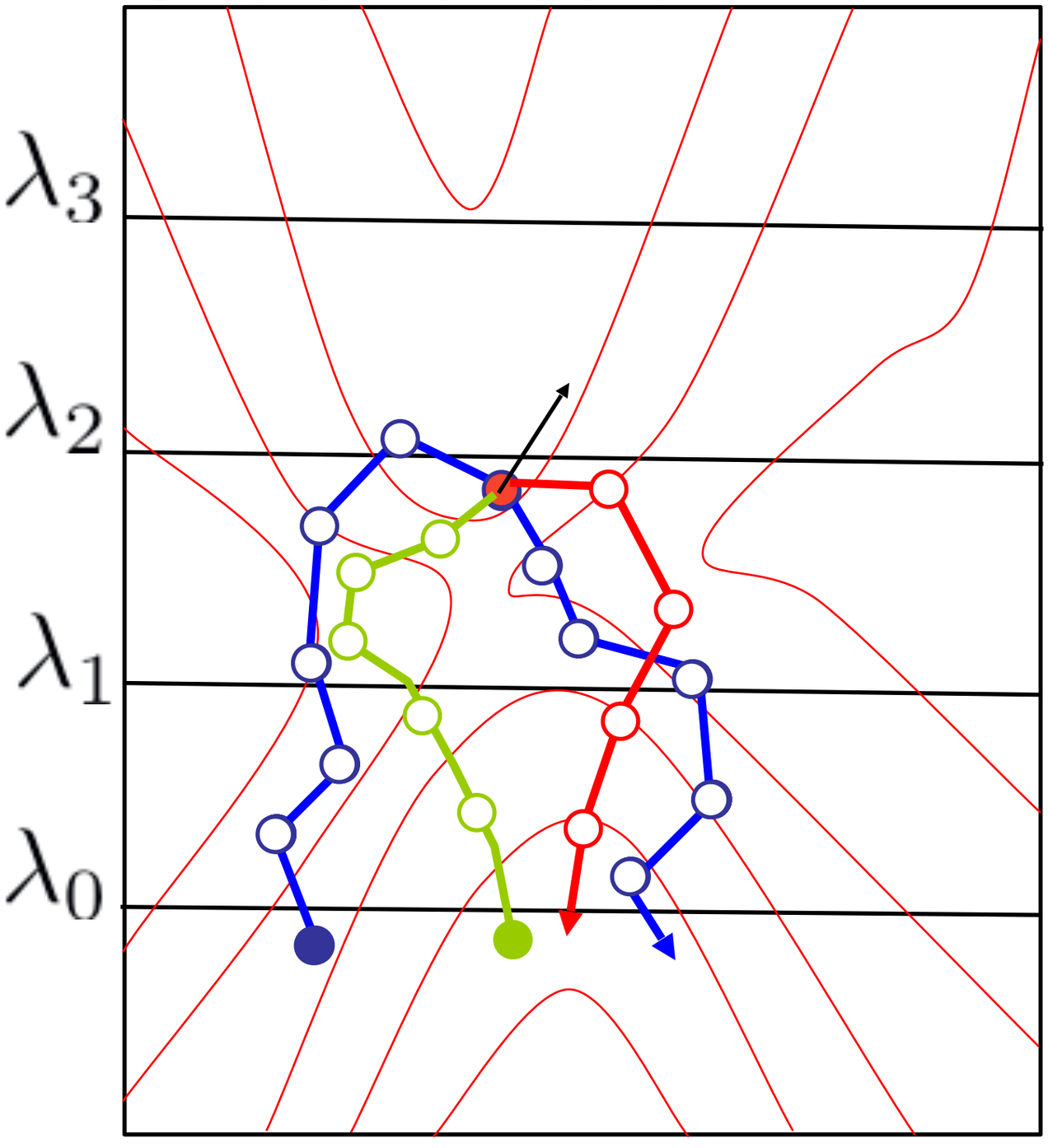


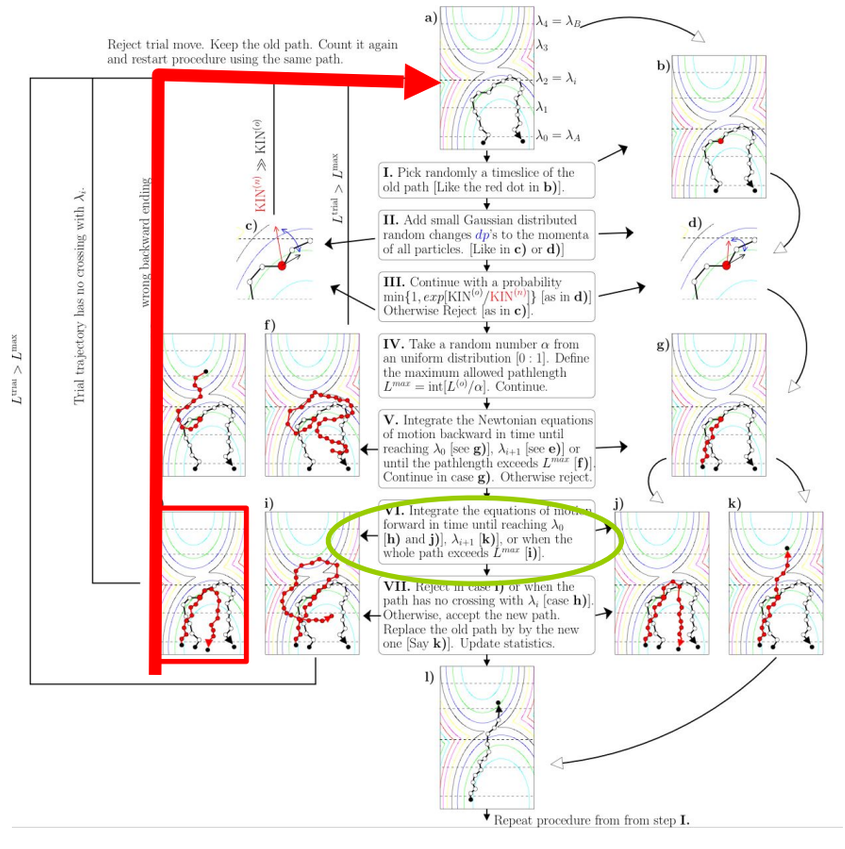


C) You end up at λ_0

Simulating the $[2^+]$ ensemble

VI. Integrate the equations of motion forward in time until:



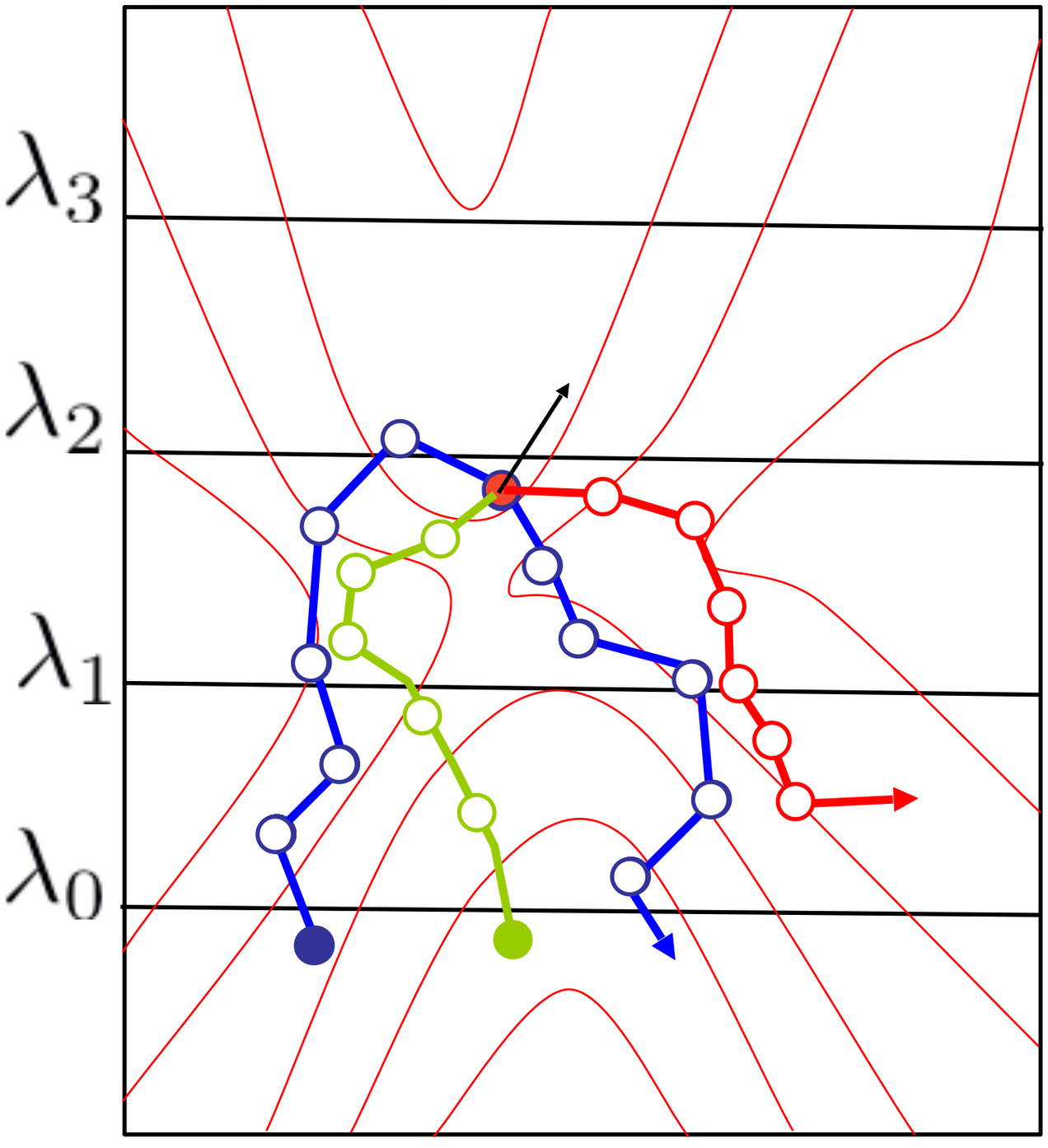


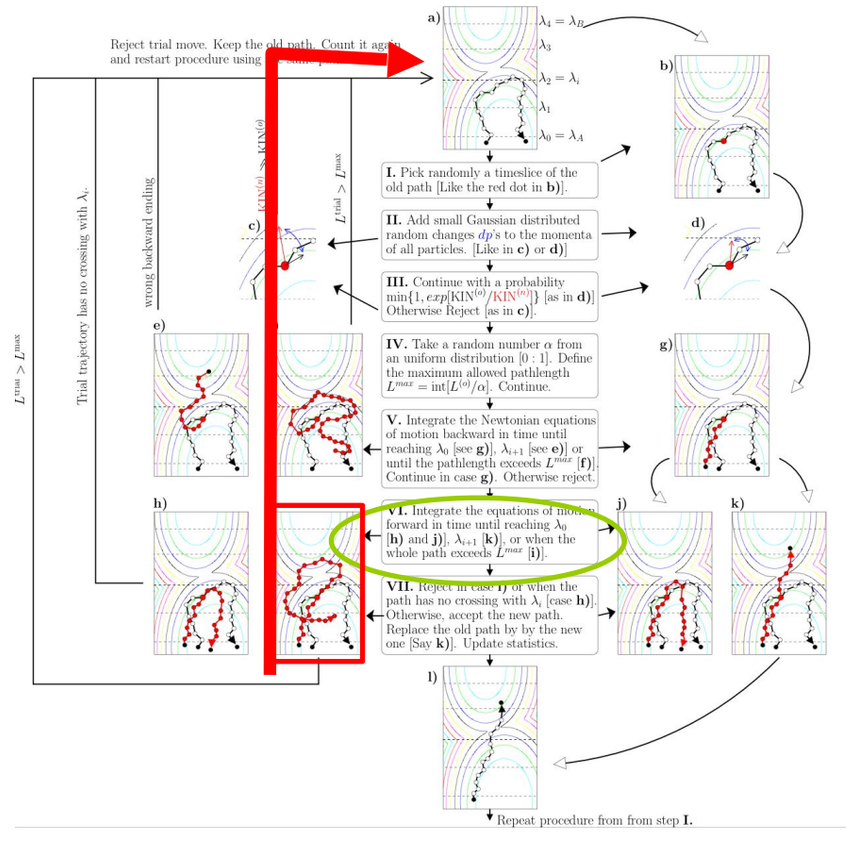
You reach λ_0 or λ_3 or $L^{\max}$

A) Total trajectory fails to cross λ_2

Simulating the $[2^+]$ ensemble

VI. Integrate the equations of motion forward in time until:





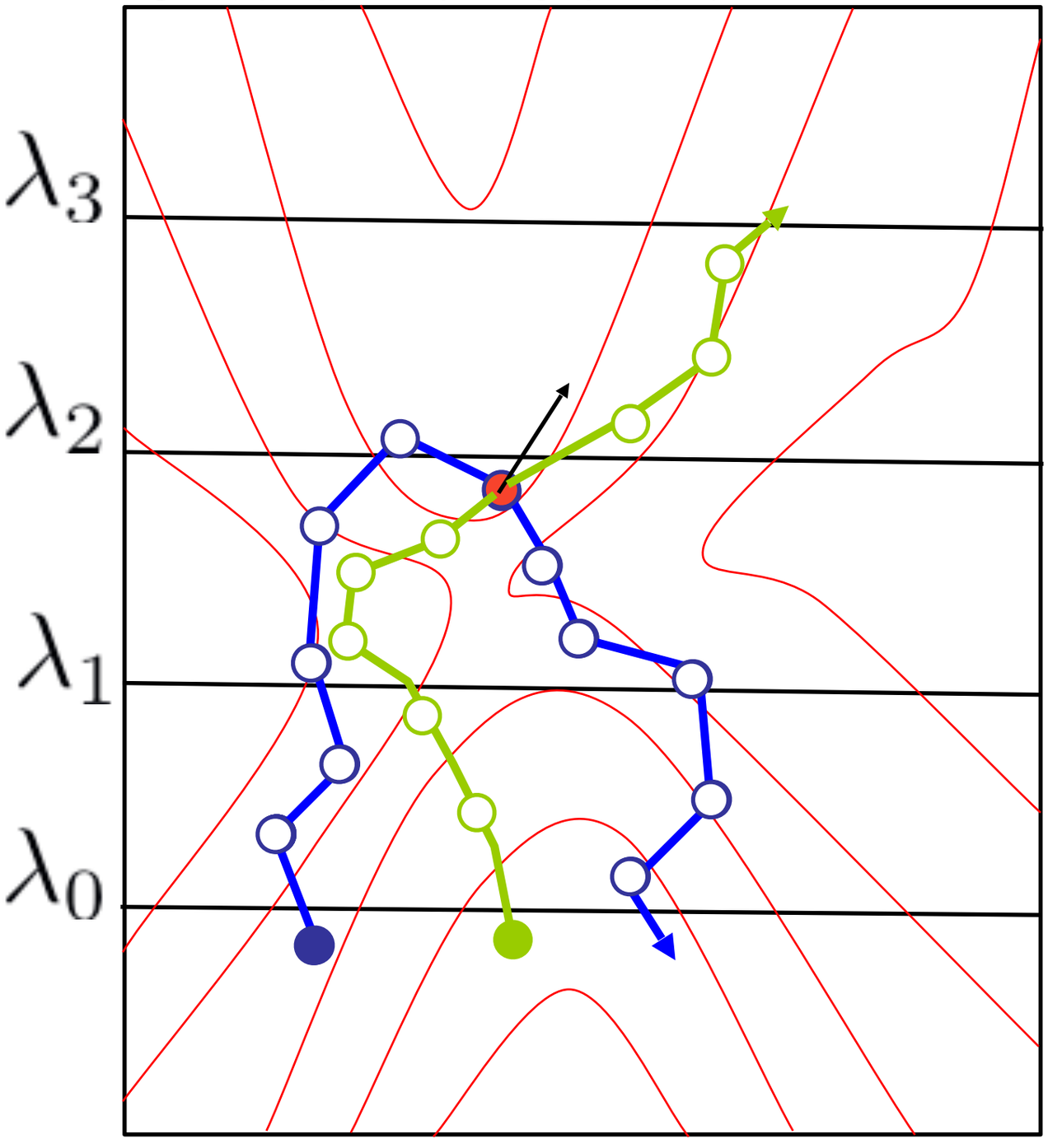
You reach λ_0 or λ_3 or $L^{\max}$

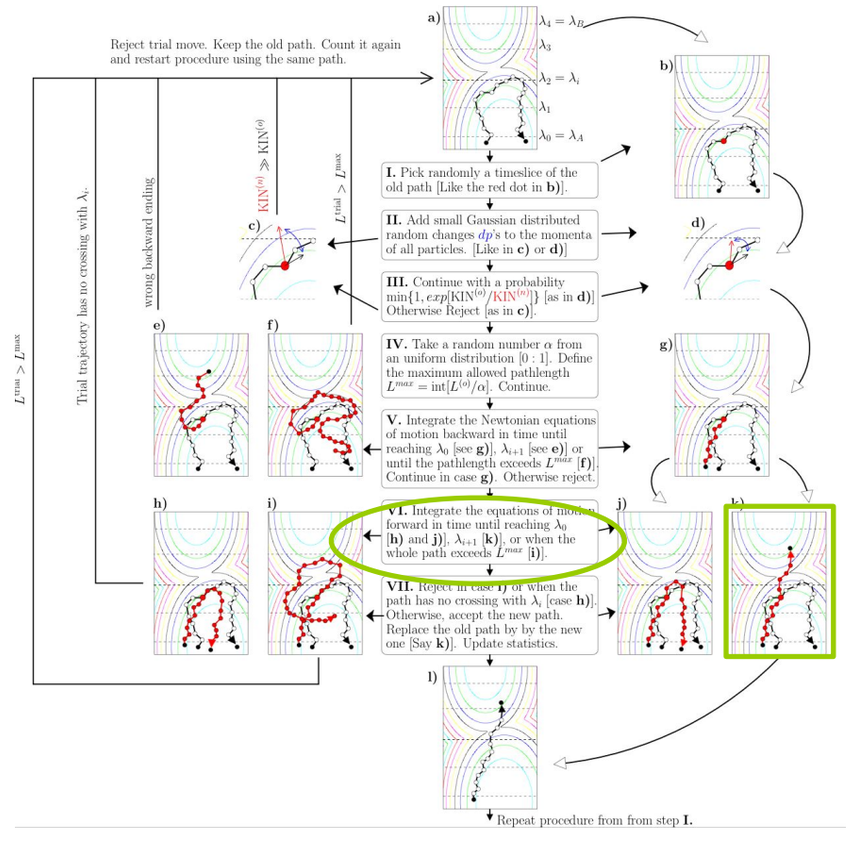
A) Total trajectory fails to cross λ_2

B) $L^{\max}$ is exceeded

Simulating the $[2^+]$ ensemble

VI. Integrate the equations of motion forward in time until:





You reach λ_0 or λ_3 or $L^{\max}$

A) Total trajectory fails to cross λ_2

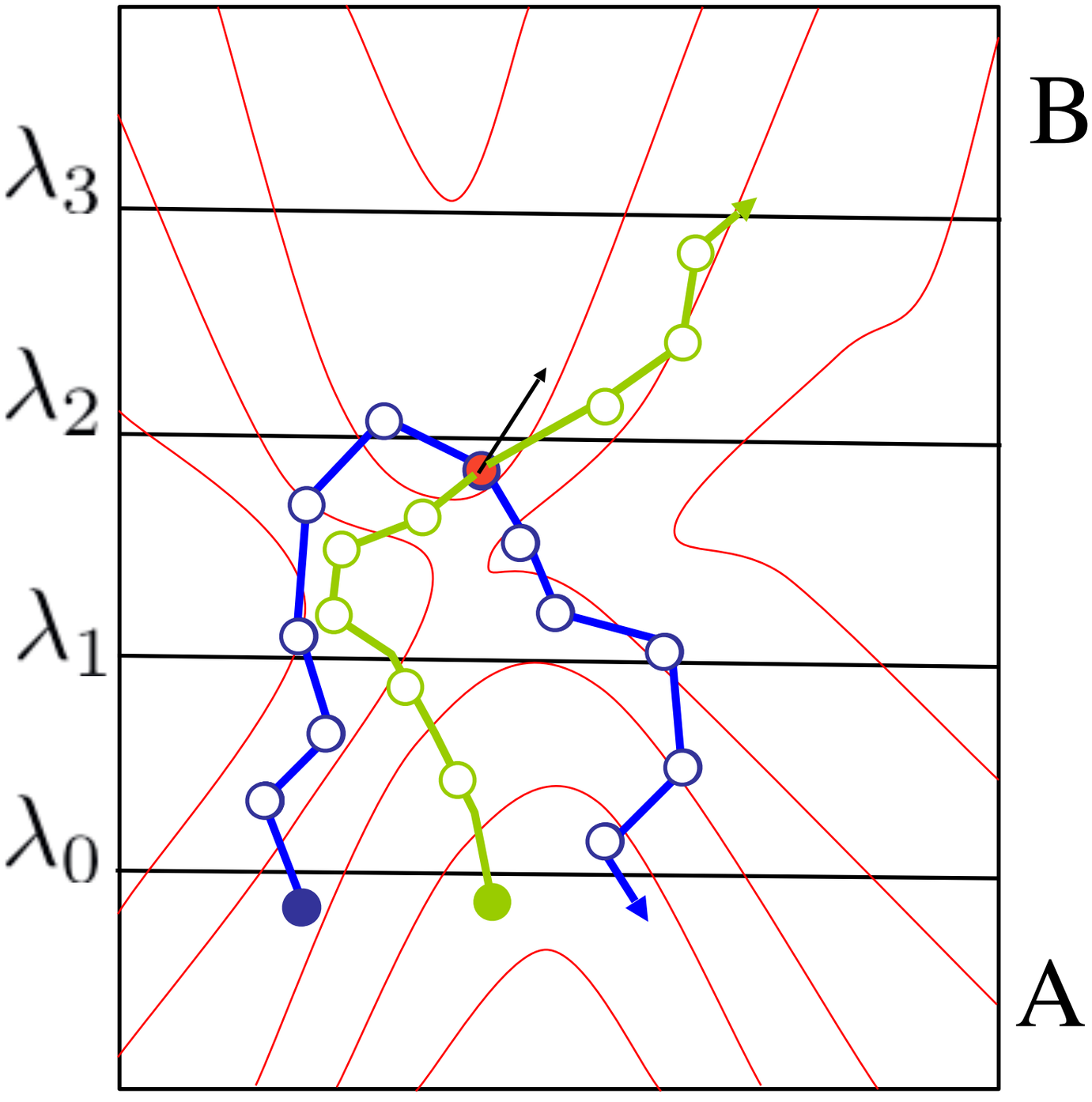
B) $L^{\max}$ is exceeded

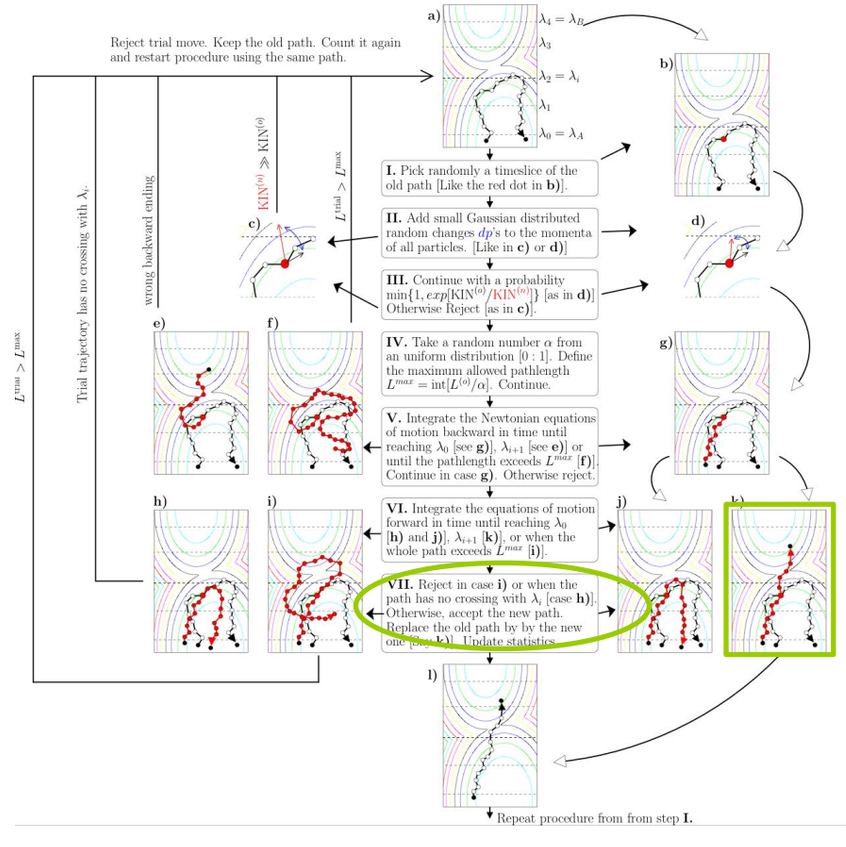
C) Valid trajectory that crosses λ_2

D) Valid trajectory that crosses λ_2 and λ_3

Simulating the $[2^+]$ ensemble

VII. If accepted …

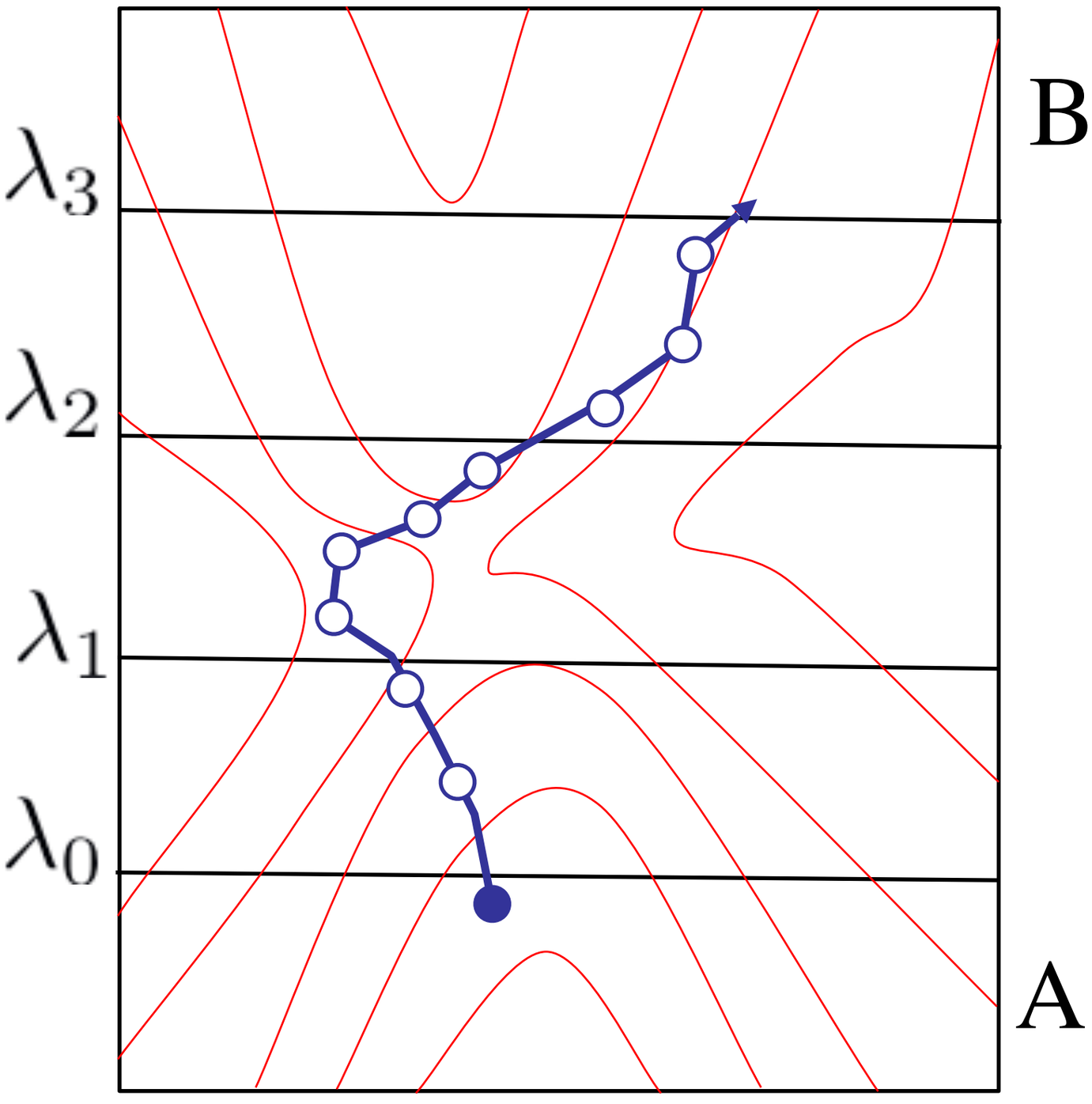


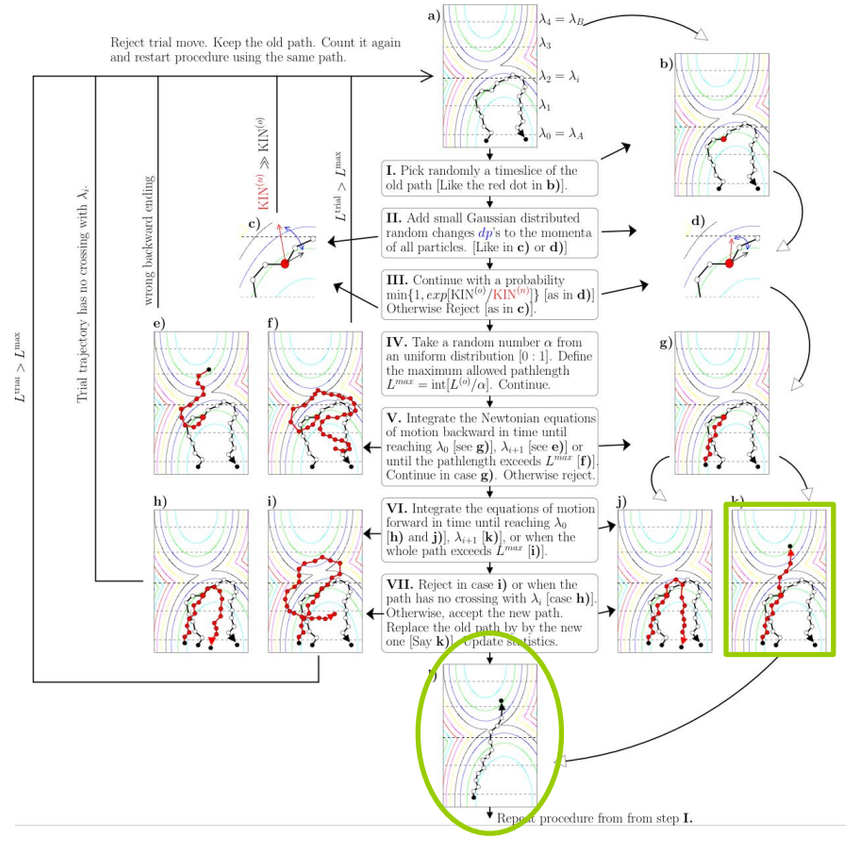


Replace the old path by the new one, update statistics and restart procedure from I. using this new path.

Simulating the $[2^+]$ ensemble

VII. If accepted …

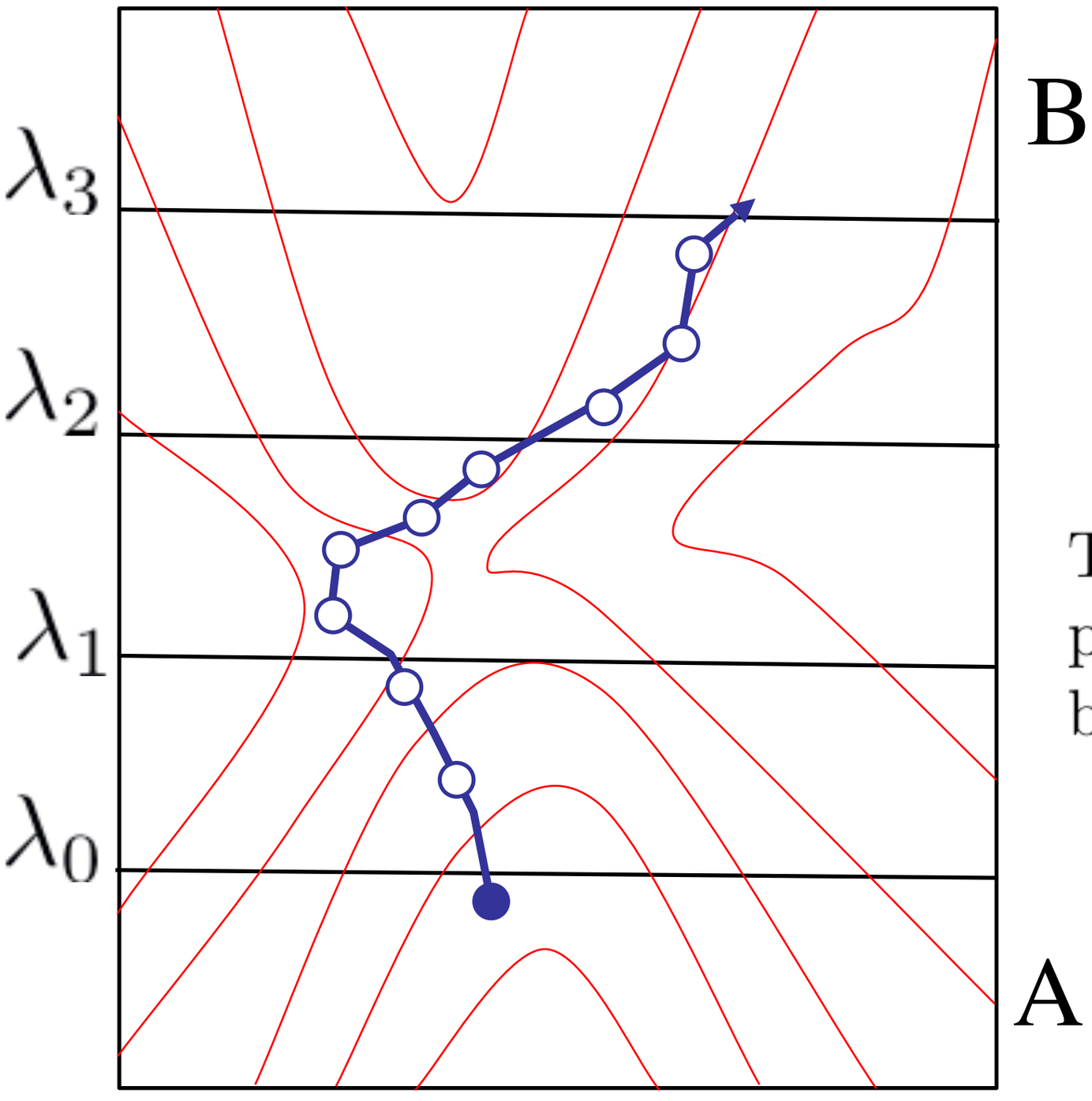


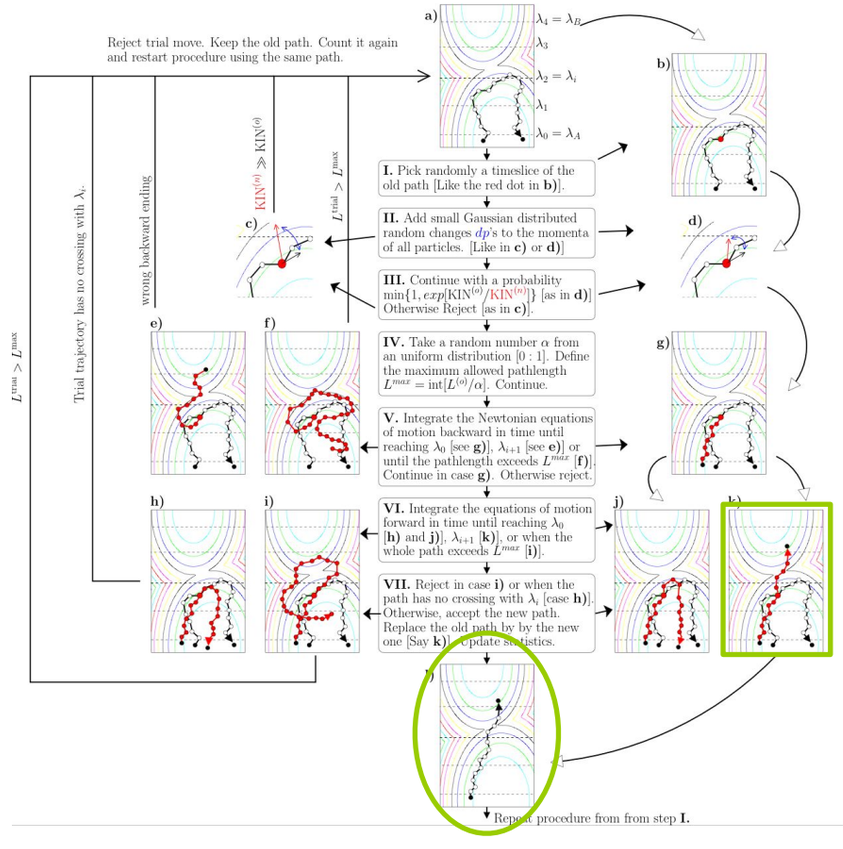


Replace the old path by the new one, update statistics and restart procedure from I. using this new path.

Simulating the $[2^+]$ ensemble

VII. If accepted …





The fraction of accepted paths that crosses λ_3 besides λ_2 equals $\mathcal{P}_A(\lambda_3|\lambda_2)$

Detailed balance continued …..

$$\rho(x_{\text{path}}^{L}) = \frac{e^{-\beta E(x_0)}}{Z} p^{\text{MD}}(x_0 \to x_1) p^{\text{MD}}(x_1 \to x_2) \ldots p^{\text{MD}}(x_{L-1} \to x_L)$$

x_i : phase point (r_i,v_i) of trajectory at MD step i. (time slice)

$p^{\text{MD}}(x \to x')$: probability that MD integrator generators x' after x.

If the dynamics obeys microscopic reversibility (true for MD, Langevin, Brownian dynamics, …):

$$\rho(x) p^{\text{MD}}(x \to x') = \rho(x') p^{\text{MD}}(\bar{x}' \to \bar{x}) \quad \text{with} \quad \bar{x} = (r, -v) \quad \text{and } \rho(x) = \rho(\bar{x})$$

This implies that we can write the path probability also as:

$$\rho(x_{\text{path}}^{L}) = p^{\text{MD}}(\bar{x}_1 \to \bar{x}_0) \frac{e^{-\beta E(x_1)}}{Z} p^{\text{MD}}(x_1 \to x_2) \ldots p^{\text{MD}}(x_{L-1} \to x_L)$$

$$= p^{\text{MD}}(\bar{x}_1 \to \bar{x}_0) p^{\text{MD}}(\bar{x}_2 \to \bar{x}_1) \frac{e^{-\beta E(x_2)}}{Z} p^{\text{MD}}(x_2 \to x_3) \ldots p^{\text{MD}}(x_{L-1} \to x_L)$$

Etc

Generation probability: (1) selection of shooting point, (2) modification of velocities, (3) going backward and forward in time from modified shooting point

$$P_{\text{gen}}(x^{L^o}_{\text{path(o)}} \to x^{L^n}_{\text{path(n)}}) = P_{\text{sel}}(x^o_s \,|\, x^{L^o}_{\text{path(o)}}) P_{\delta v} P_{\text{MD-steps}}$$

$$P_{\text{sel}}(x^o_s \,|\, x^{L^o}_{\text{path(o)}}) = \frac{1}{L^o}$$

$P_{\delta v} = P_{-\delta v}$ (symmetric velocity change probability)

$$P_{\text{MD-steps}} = p^{\text{MD}}(\overline{x_s}^n \to \overline{x_{s-1}}^n) p^{\text{MD}}(\overline{x_{s-1}}^n \to \overline{x_{s-2}}^n) \ldots p^{\text{MD}}(x^n_s \to x^n_{s+1}) p^{\text{MD}}(x^n_{s+1} \to x^n_{s+2}) \ldots$$

$$P_{\text{MD-steps}} = \frac{\rho(x^{L^n}_{\text{path(n)}})}{\rho(x^n_s)}$$

Metropolis-Hastings:

$$P_{\text{acc}}(x^{L^o}_{\text{path(o)}} \to x^{L^n}_{\text{path(n)}}) = \min[1, \frac{\rho(x^{L^n}_{\text{path(n)}}) P_{\text{gen}}(x^{L^n}_{\text{path(n)}} \to x^{L^o}_{\text{path(o)}})}{\rho(x^{L^o}_{\text{path(o)}}) P_{\text{gen}}(x^{L^o}_{\text{path(o)}} \to x^{L^n}_{\text{path(n)}})}] = \min[1, \frac{\rho(x^n_s) L^o}{\rho(x^o_s) L^n}]$$

$$P_{\text{acc}}(x^{L^o}_{\text{path(o)}} \to x^{L^n}_{\text{path(n)}}) = \min[1, \frac{\rho(x^n_s) L^o}{\rho(x^o_s) L^n}]$$

We will equally well obey detailed balance if

$$P_{\text{acc}}(x^{L^o}_{\text{path(o)}} \to x^{L^n}_{\text{path(n)}}) = \min[1, \frac{\rho(x^n_s)}{\rho(x^o_s)}] \times \min[1, \frac{L^o}{L^n}]$$

$$= \min[1, \frac{e^{-\beta E_{kin}(v^n)}}{e^{-\beta E_{kin}(v^o)}}] \times \min[1, \frac{L^o}{L^n}]$$

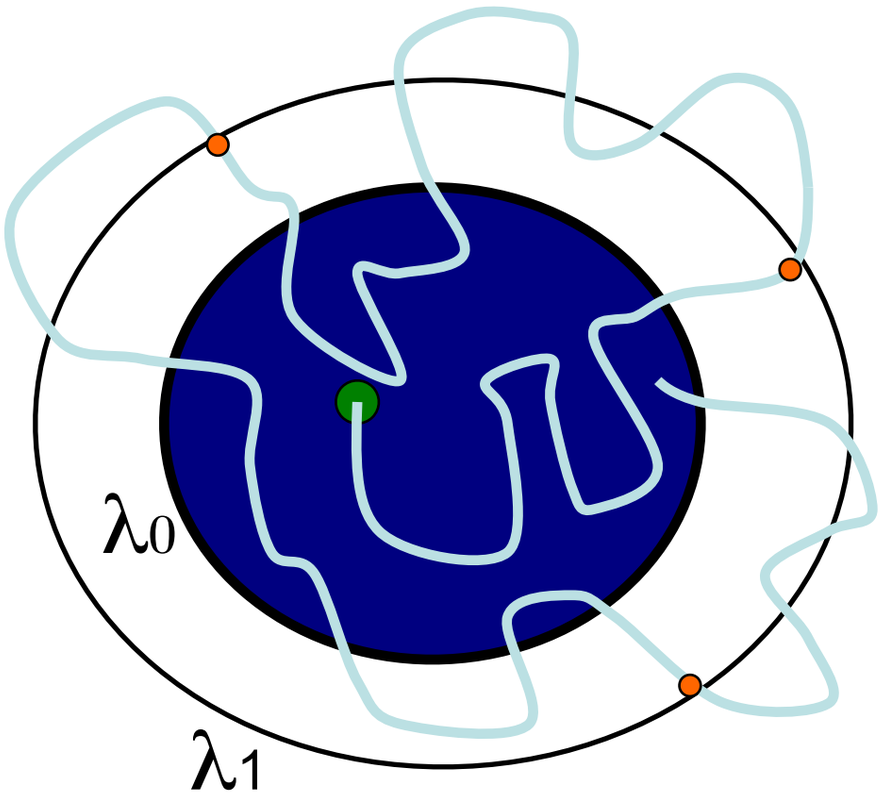
Take a random number $\alpha \in [0:1]$ and accept if $\alpha < e^{-\beta[E_{kin}(v^n) - E_{kin}(v^o)]}$

Take a random number $\alpha' \in [0:1]$ and accept if $\alpha' < L^o/L^n$

$$L^n < L^o/\alpha'$$

Take the random number beforehand and define max. allowed path length: $L^{\max} = \text{int}[L^o/\alpha']$

MD run to determine effective escape flux through interface 1

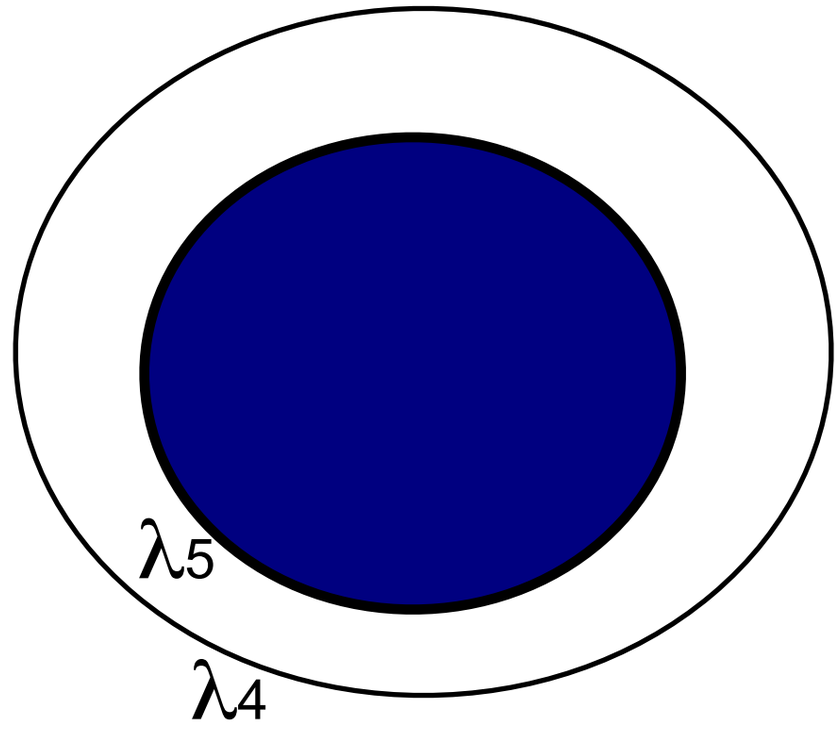


A

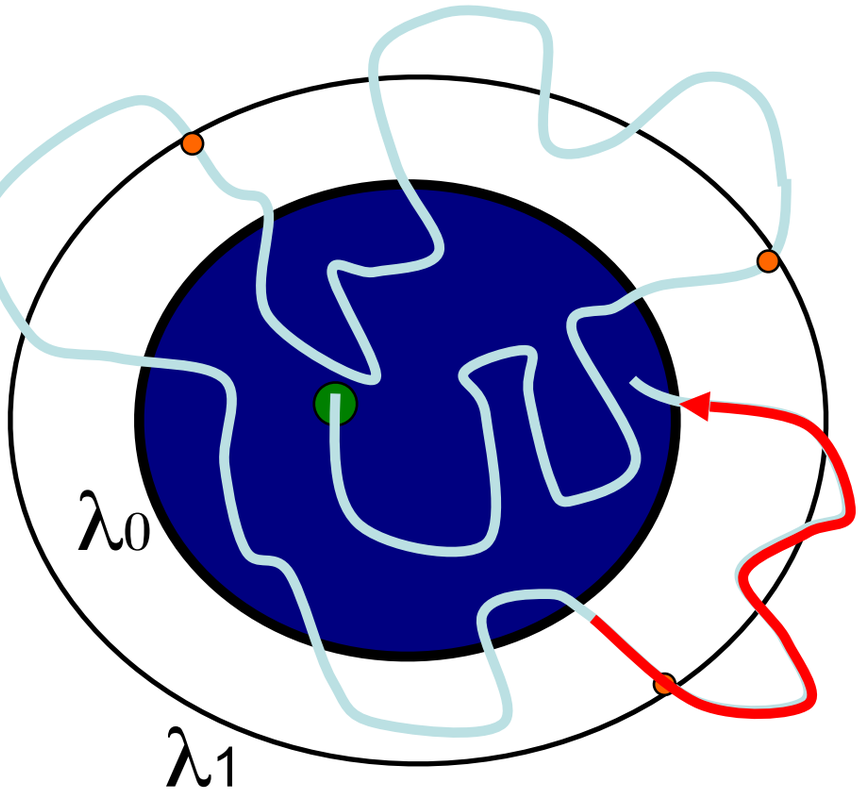
$$\Phi = \frac{1}{\Delta T} \frac{N_C}{N_{MD}}$$

λ2

λ3



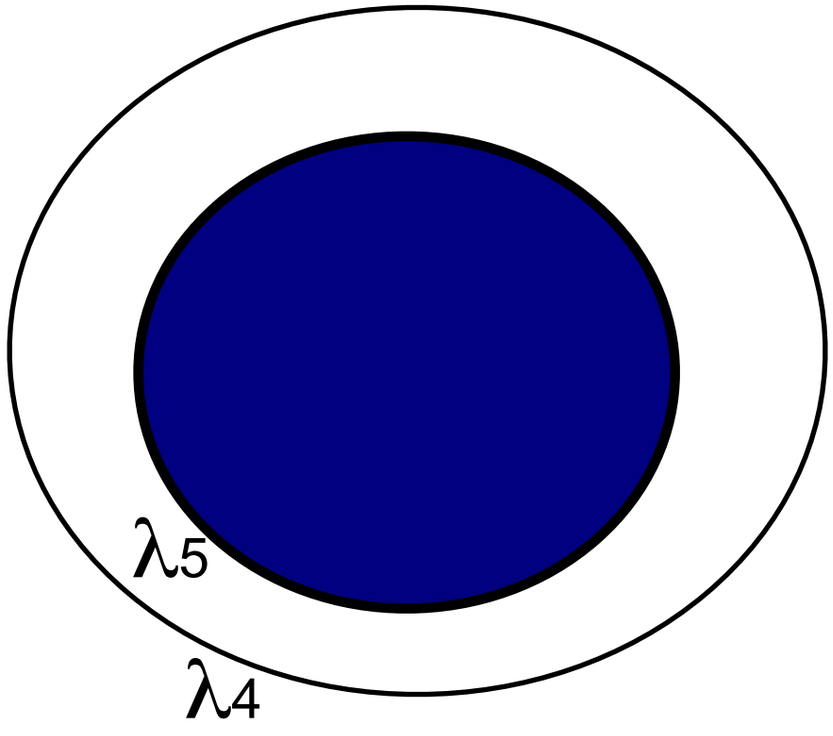
B



A

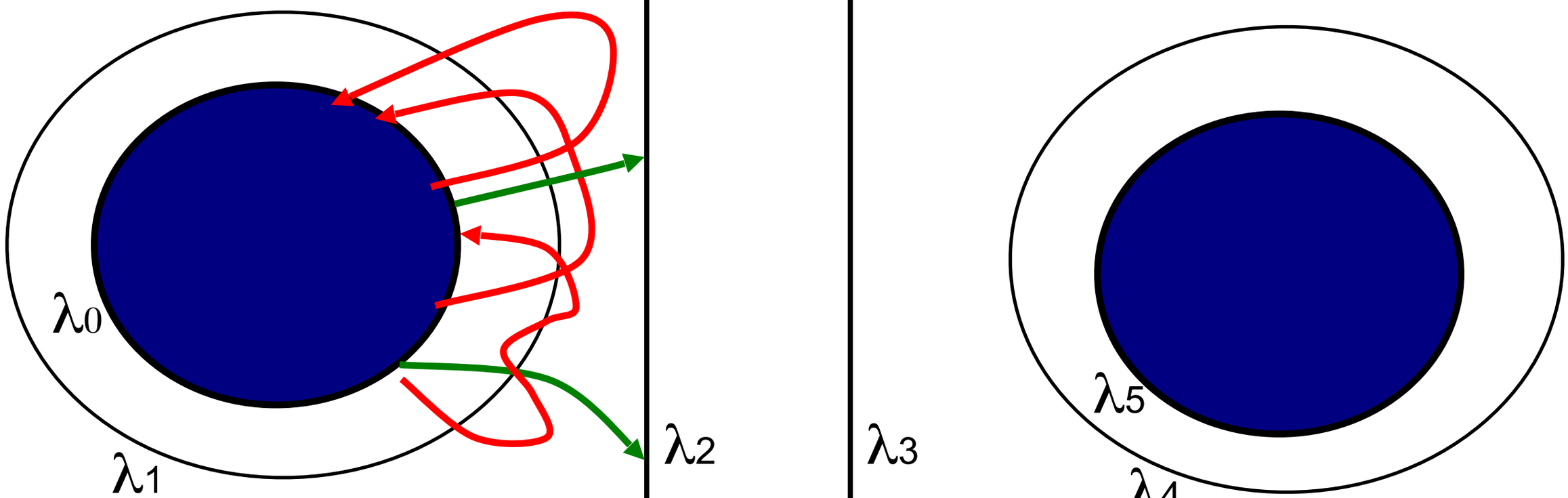
λ2

λ3



B

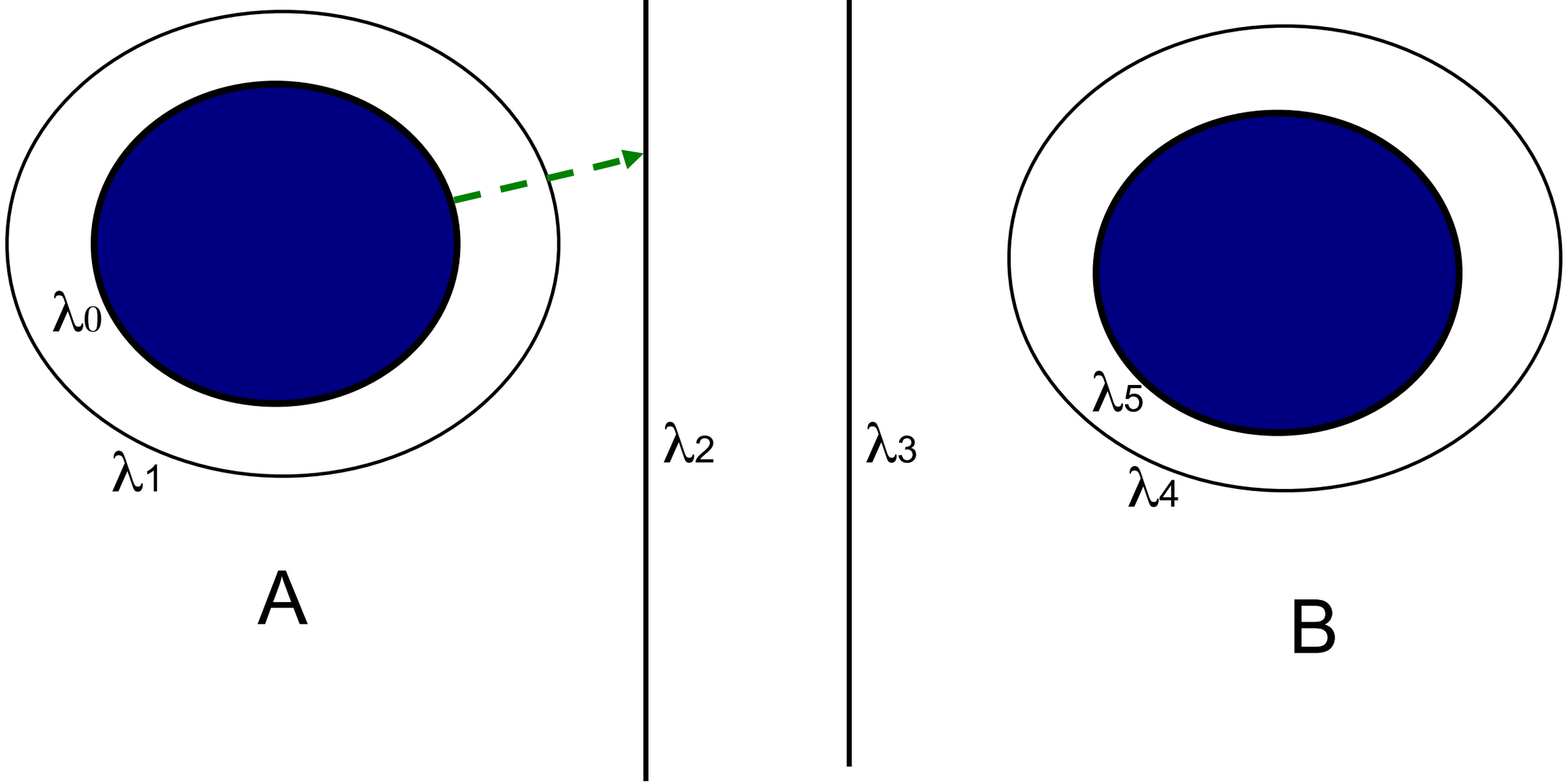
Path sampling
in the
λ_1-interface ensemble



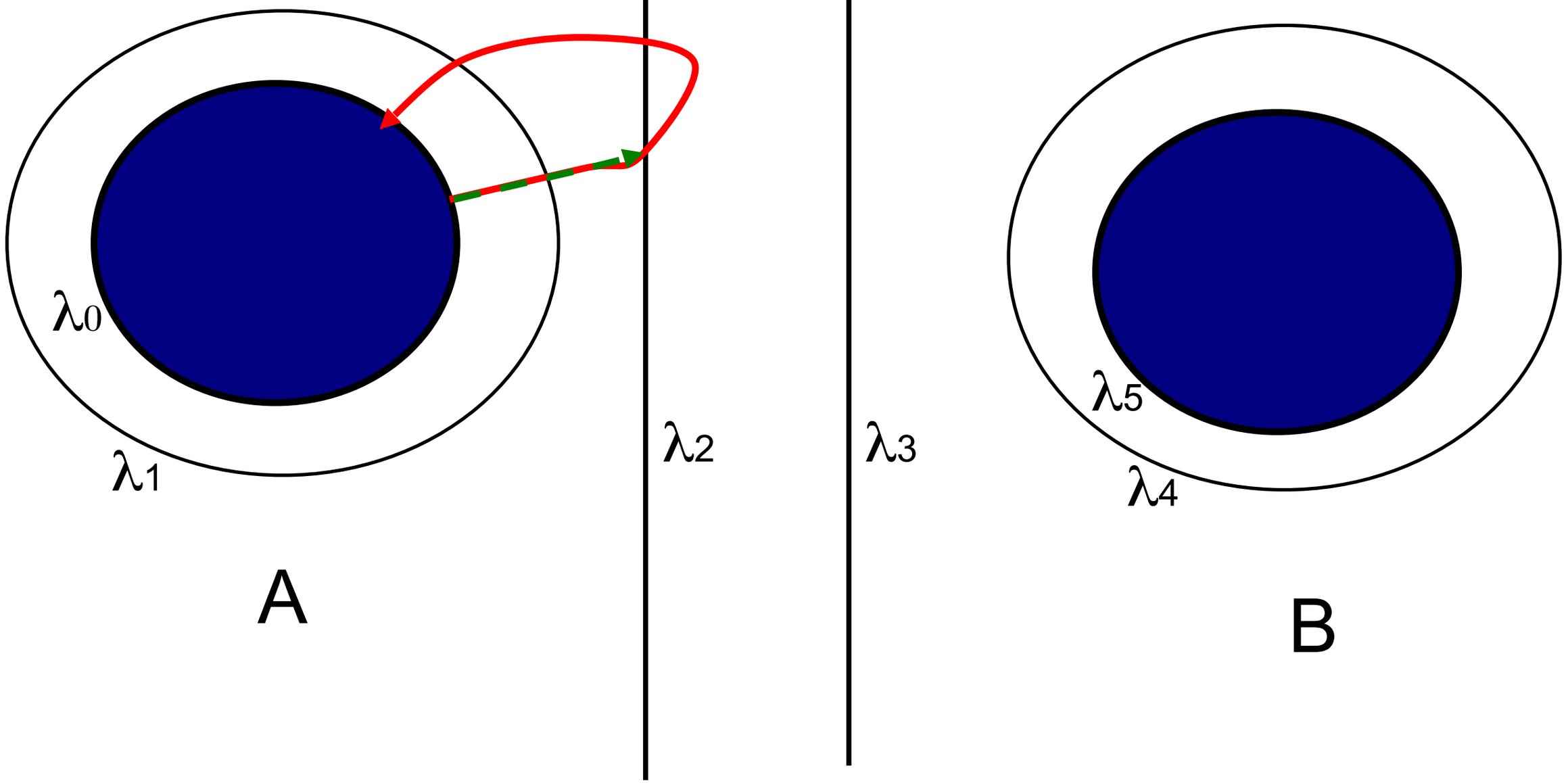
A

B

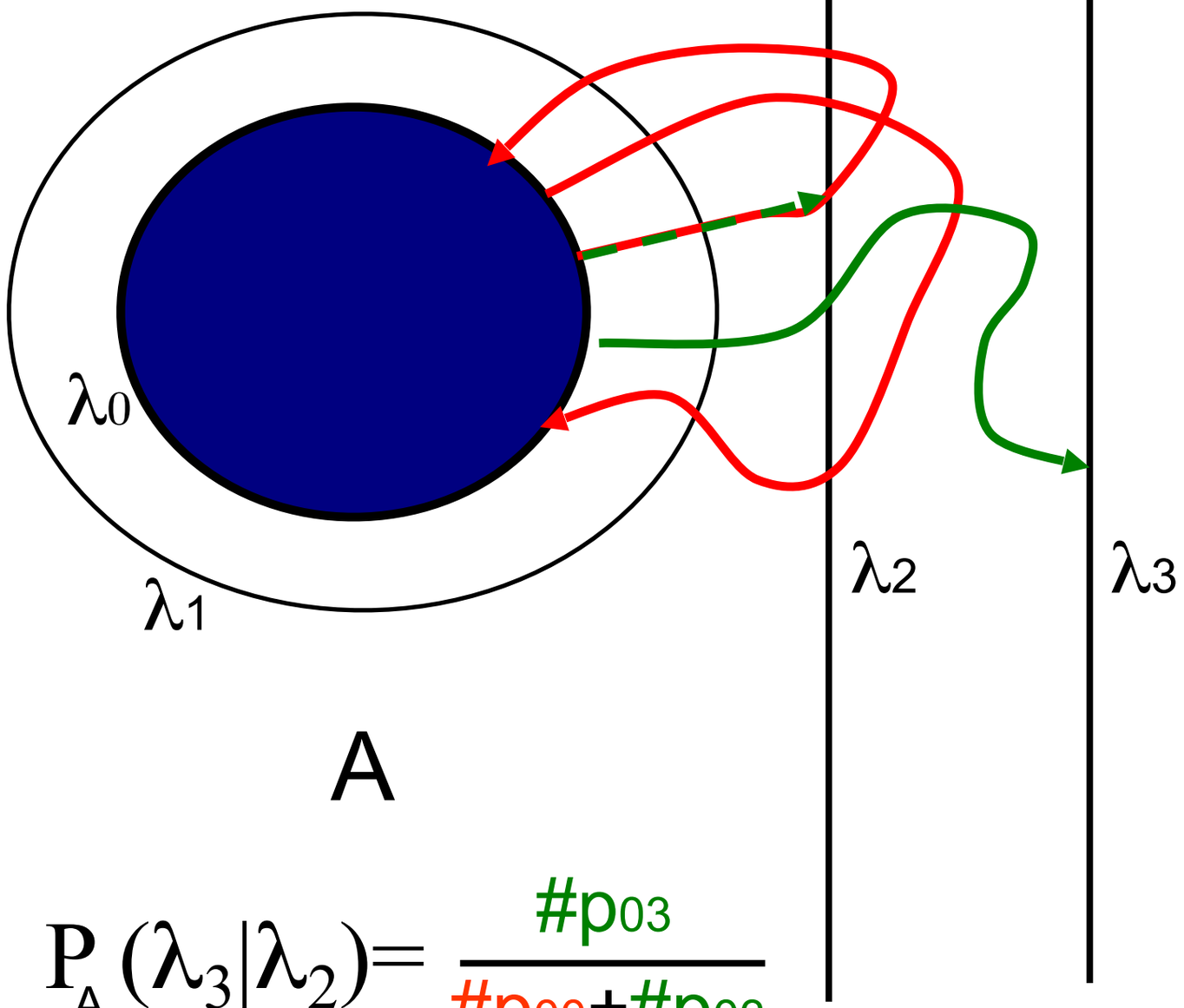
$$P_A(\lambda_2|\lambda_1) = \frac{\#p_{02}}{\#p_{00}+\#p_{02}}$$


λ0
λ1
λ2
λ3
λ4
λ5
A
B

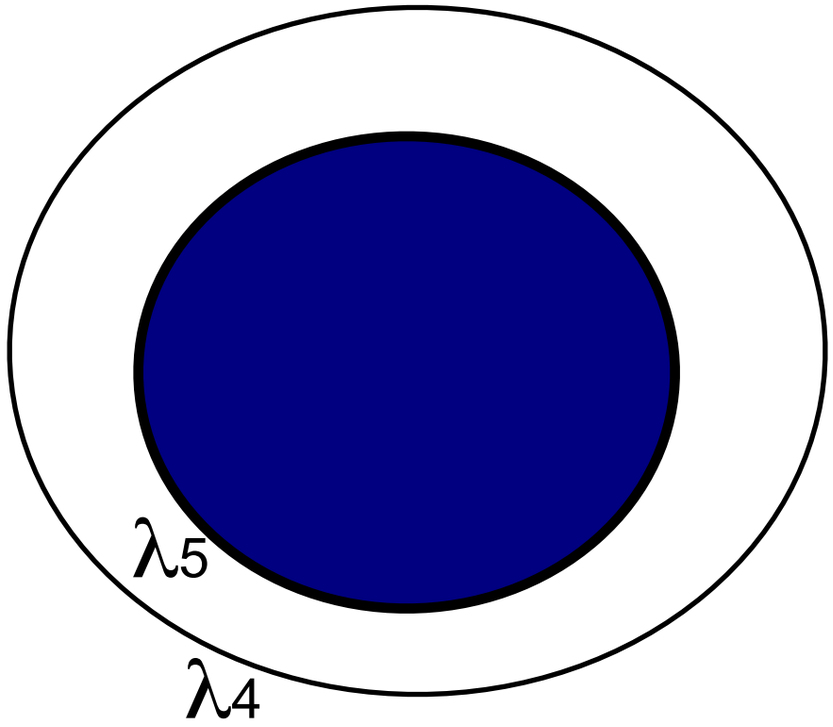
Path sampling
in the
λ_2-interface ensemble



Path sampling
in the
λ_2-interface ensemble

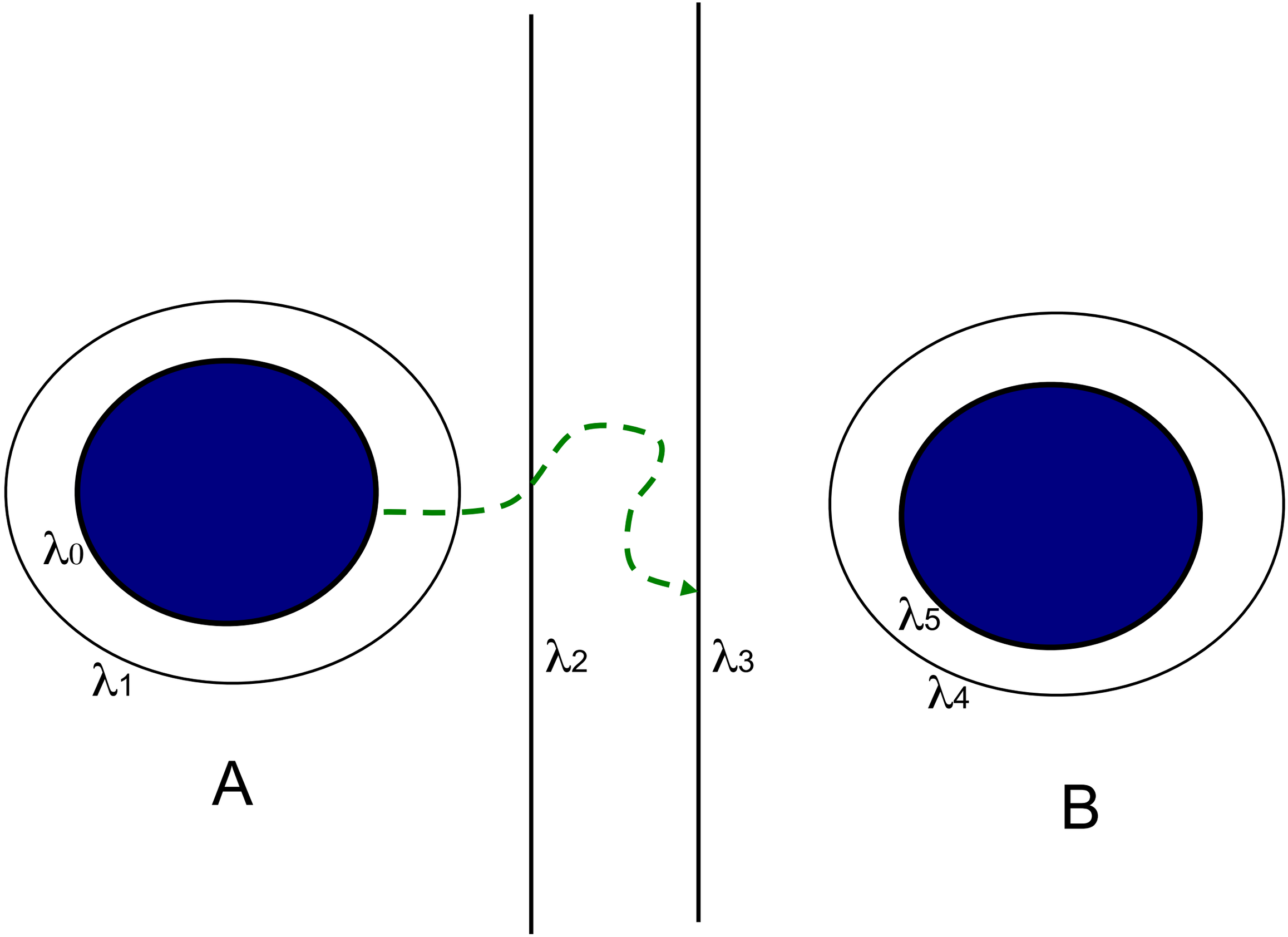


A

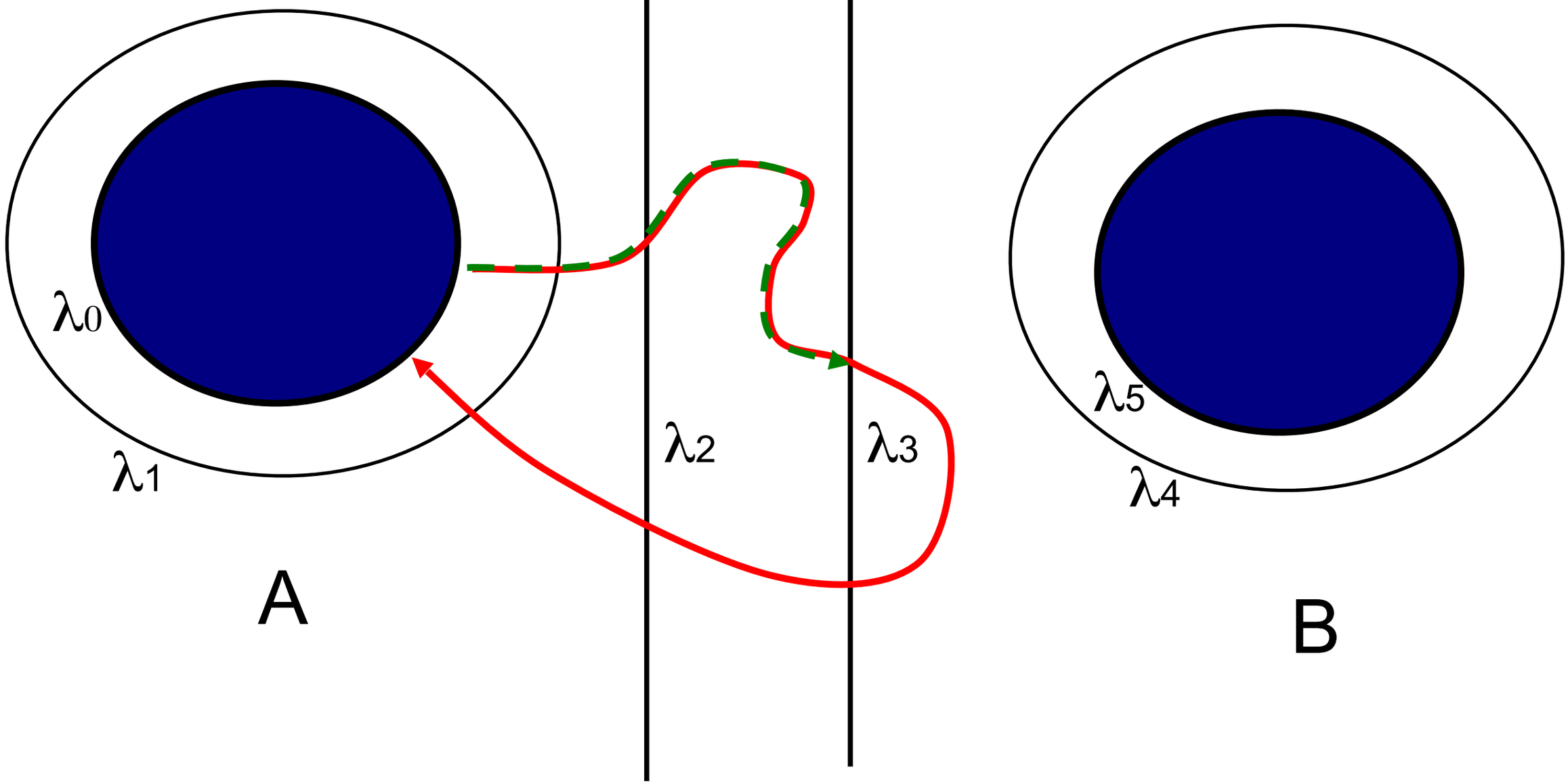


B

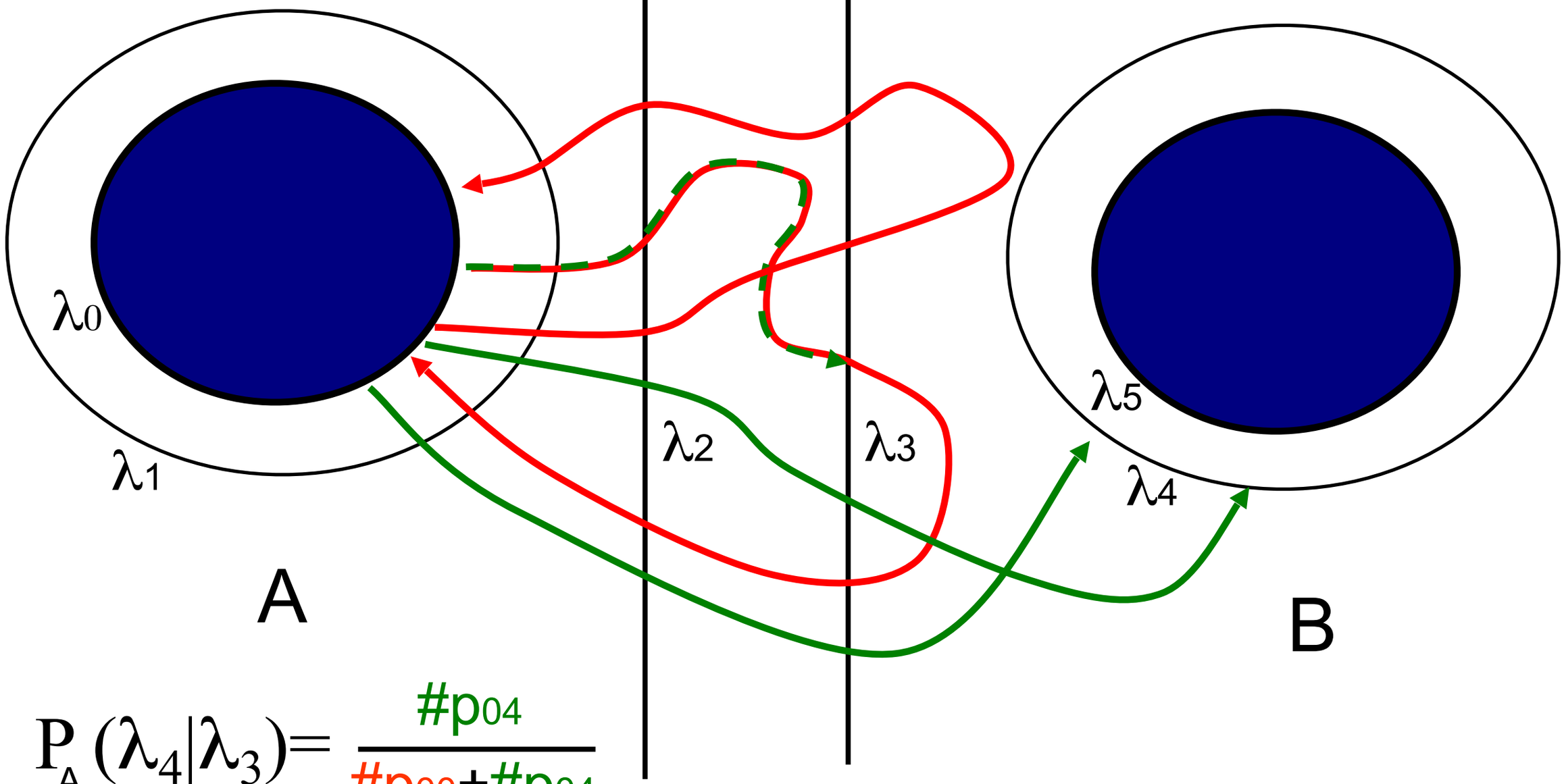
$$P_A(\lambda_3|\lambda_2) = \frac{\#p_{03}}{\#p_{00}+\#p_{03}}$$


λ0
λ1
λ2
λ3
λ4
λ5
A
B

Path sampling
in the
λ_3-interface ensemble

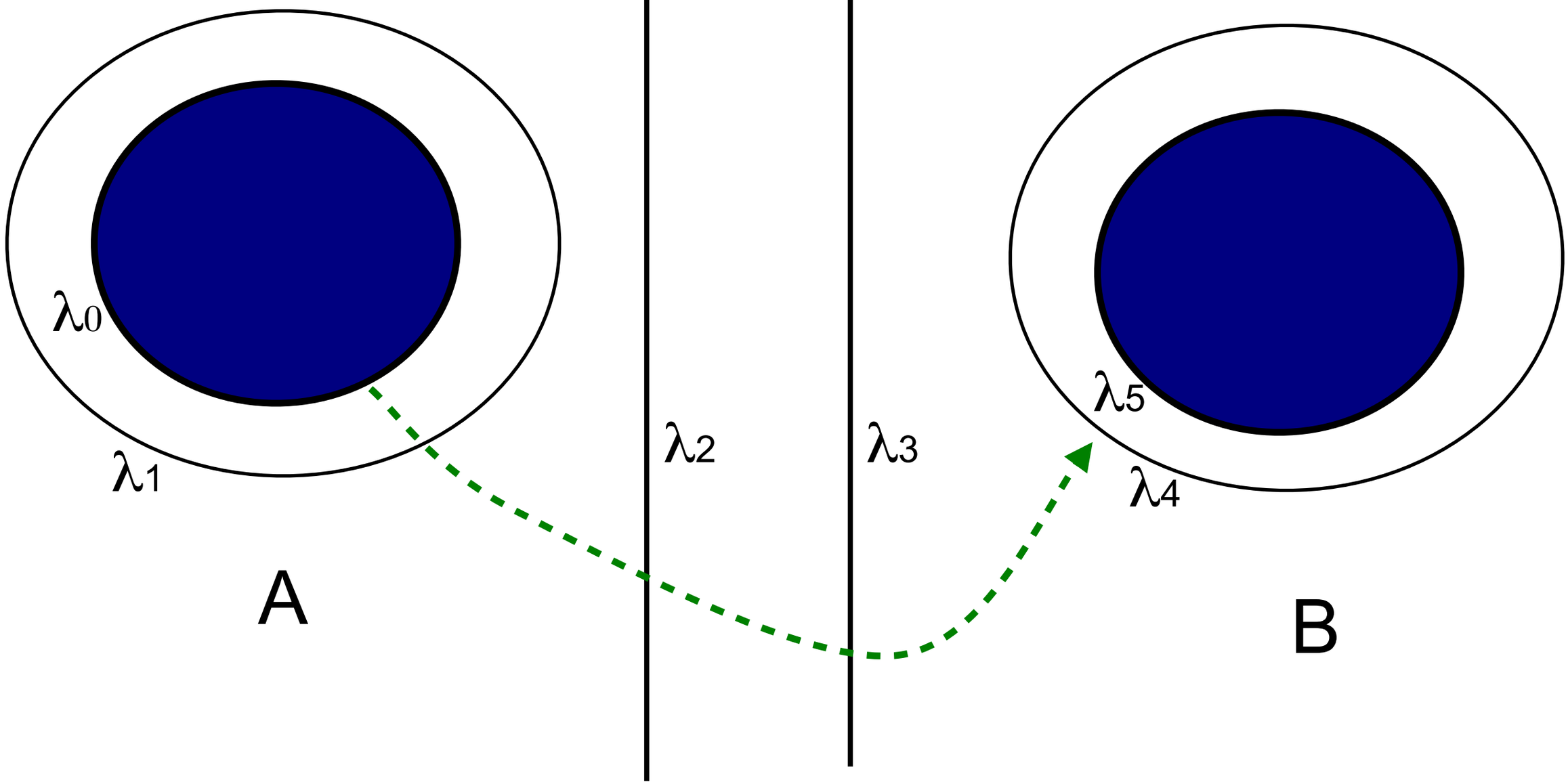


Path sampling in the λ_3-interface ensemble

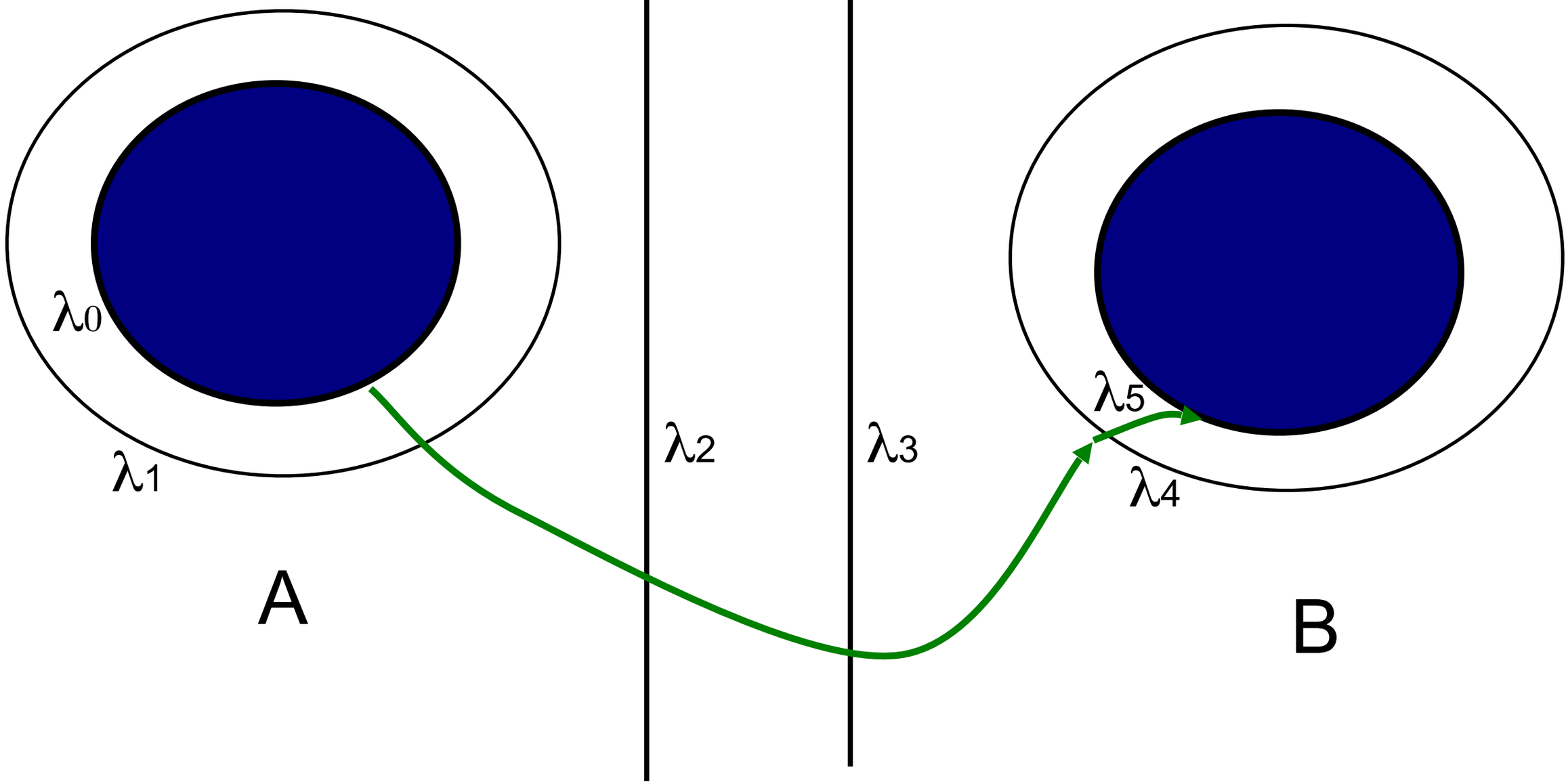


$$P_A(\lambda_4|\lambda_3) = \frac{\#p_{04}}{\#p_{00}+\#p_{04}}$$

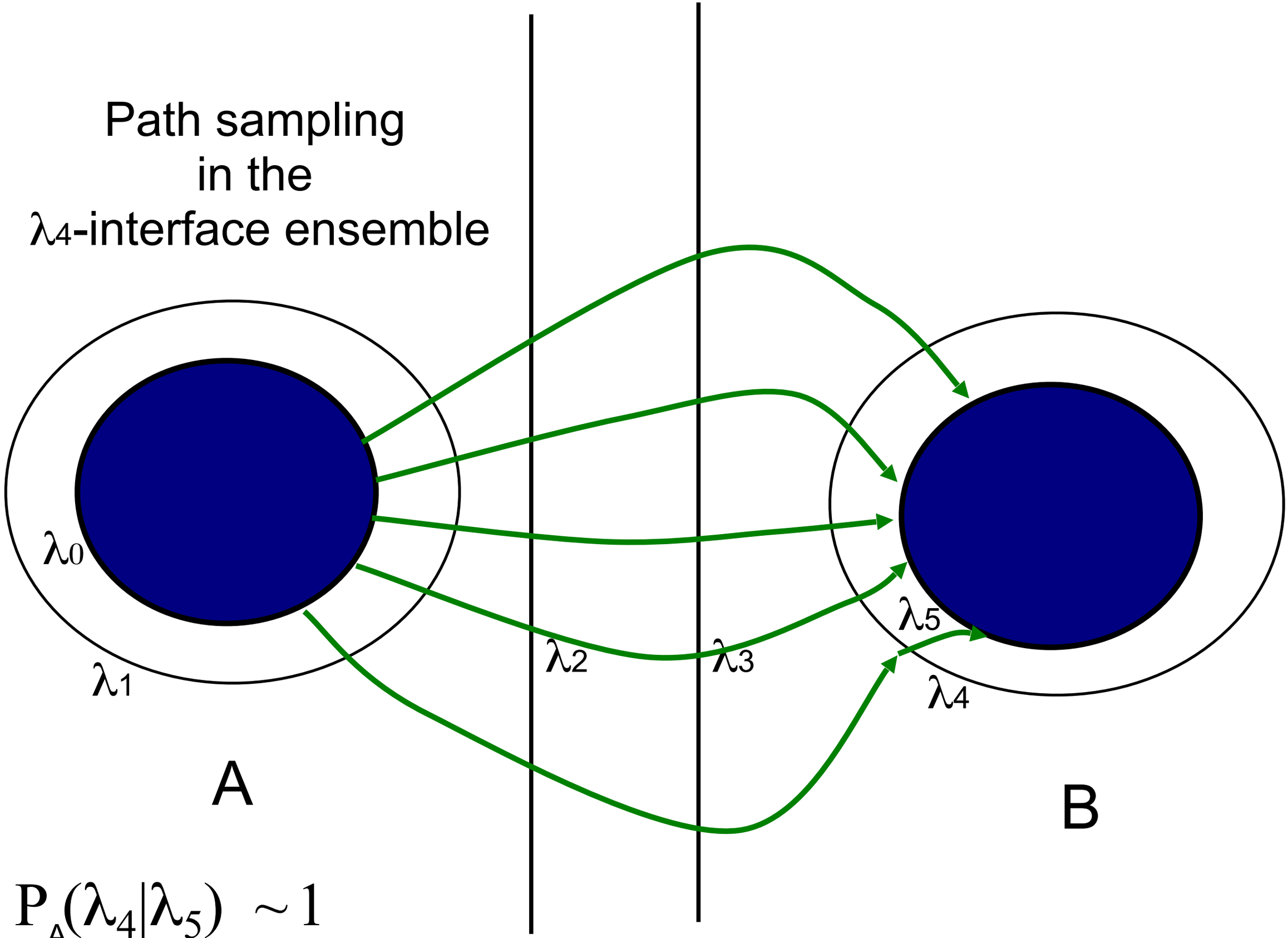
Path sampling
in the
λ_4-interface ensemble



Path sampling
in the
λ_4-interface ensemble

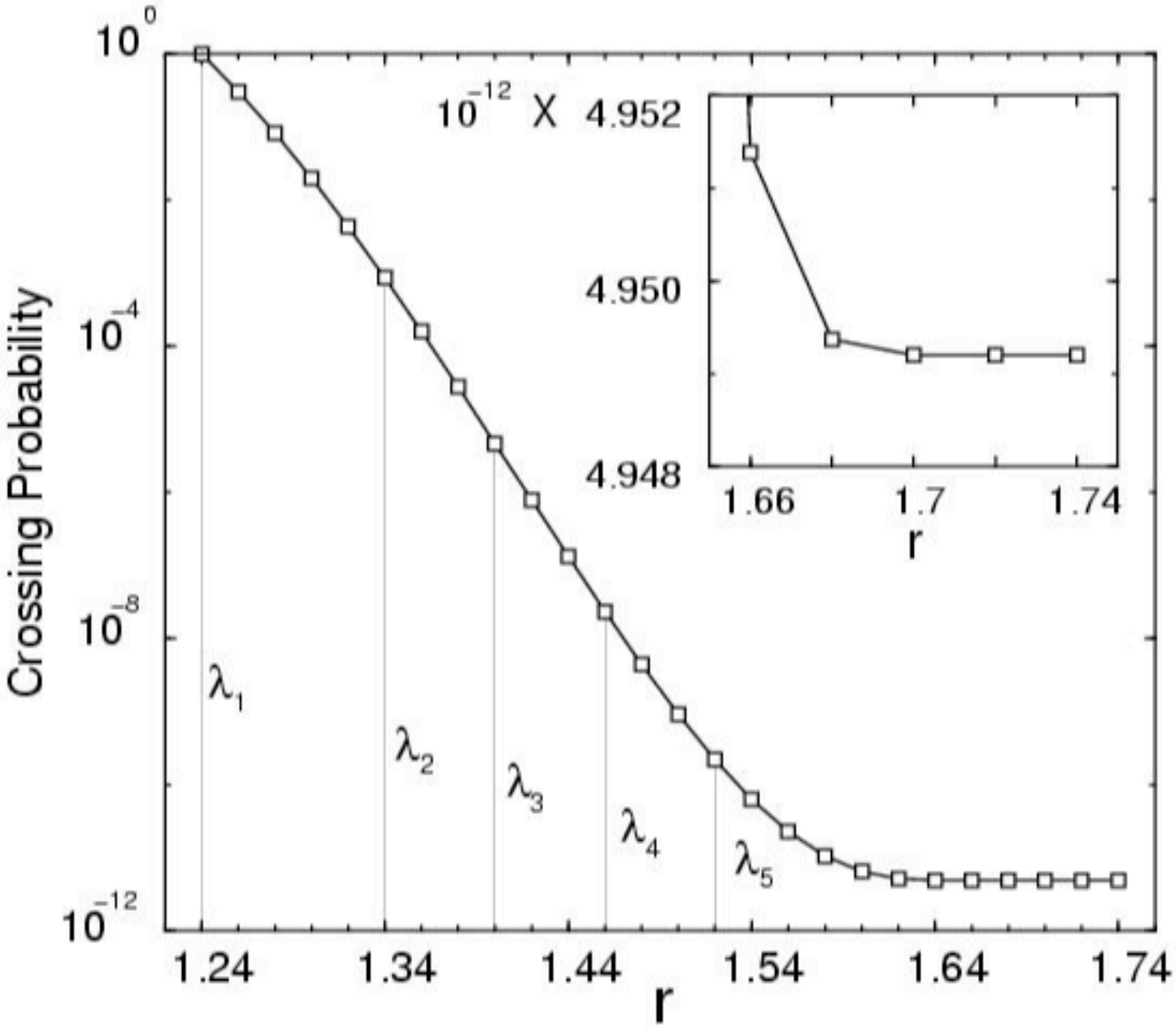


Path sampling in the λ_4-interface ensemble



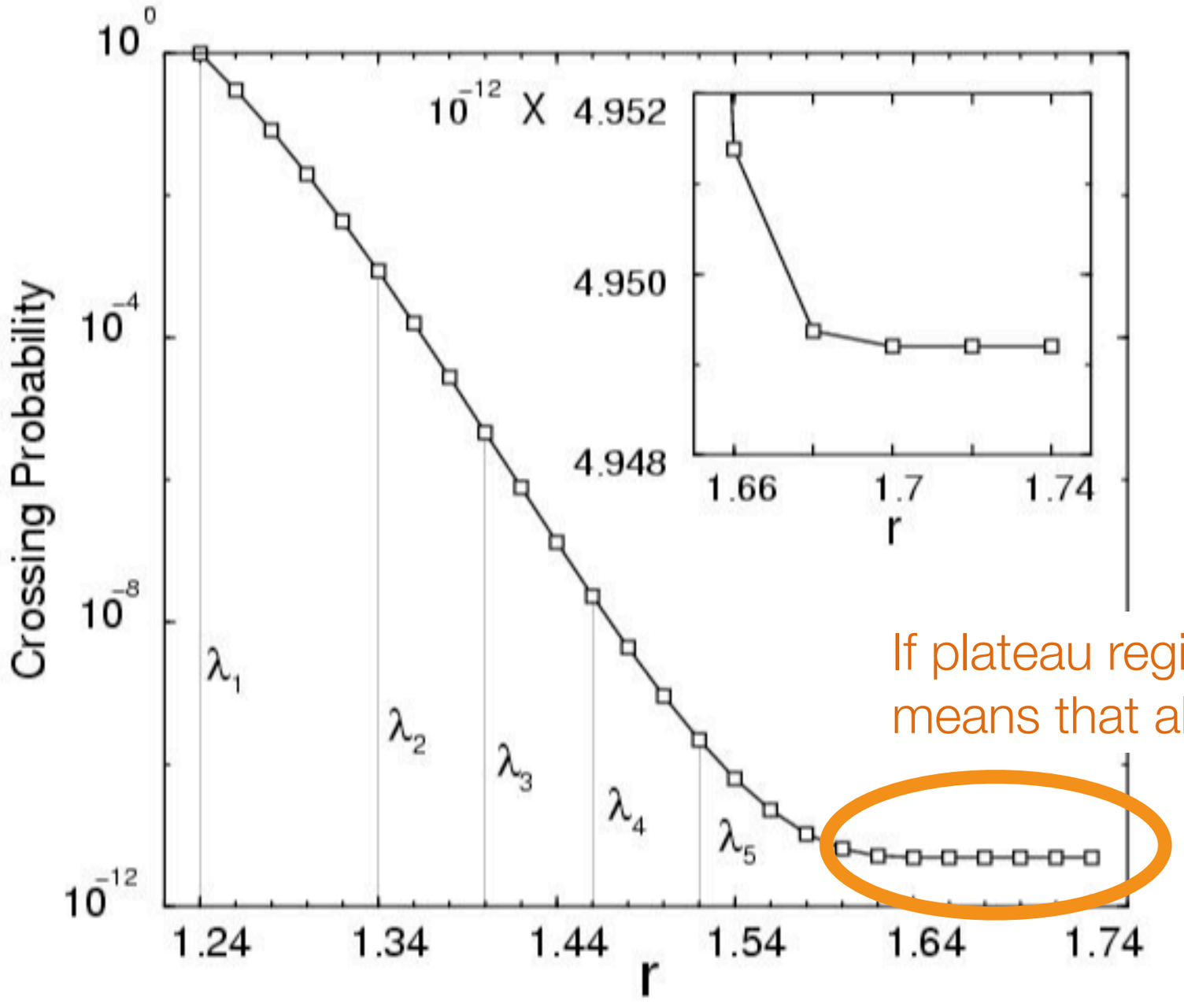
$$P_A(\lambda_4|\lambda_5) \sim 1$$

Construction of the overall crossing probability



…. times the flux through interface 1 yields the rate !

Construction of the overall crossing probability



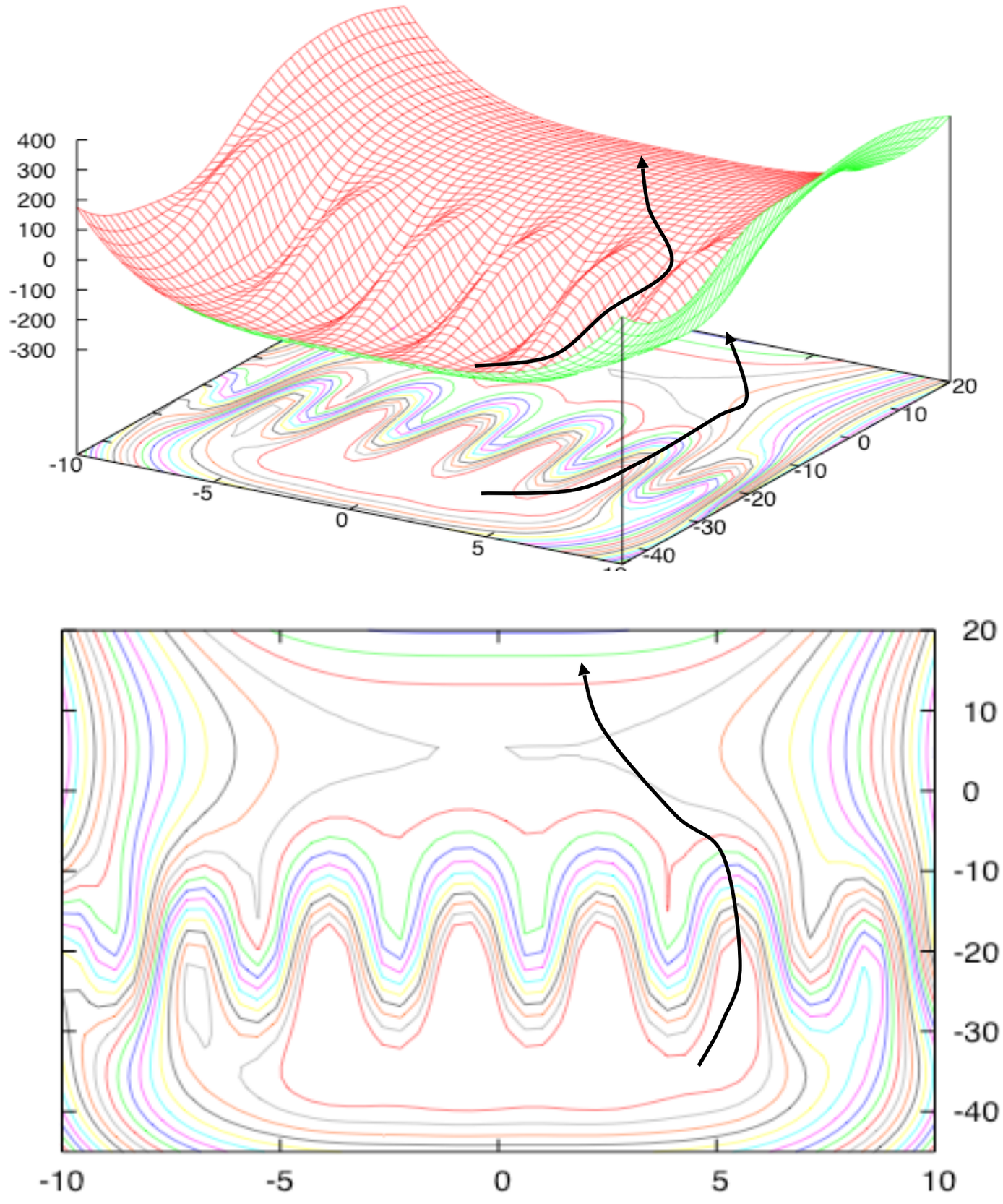
If plateau region is reached it means that all paths reach B

…. times the flux through interface 1 yields the rate !

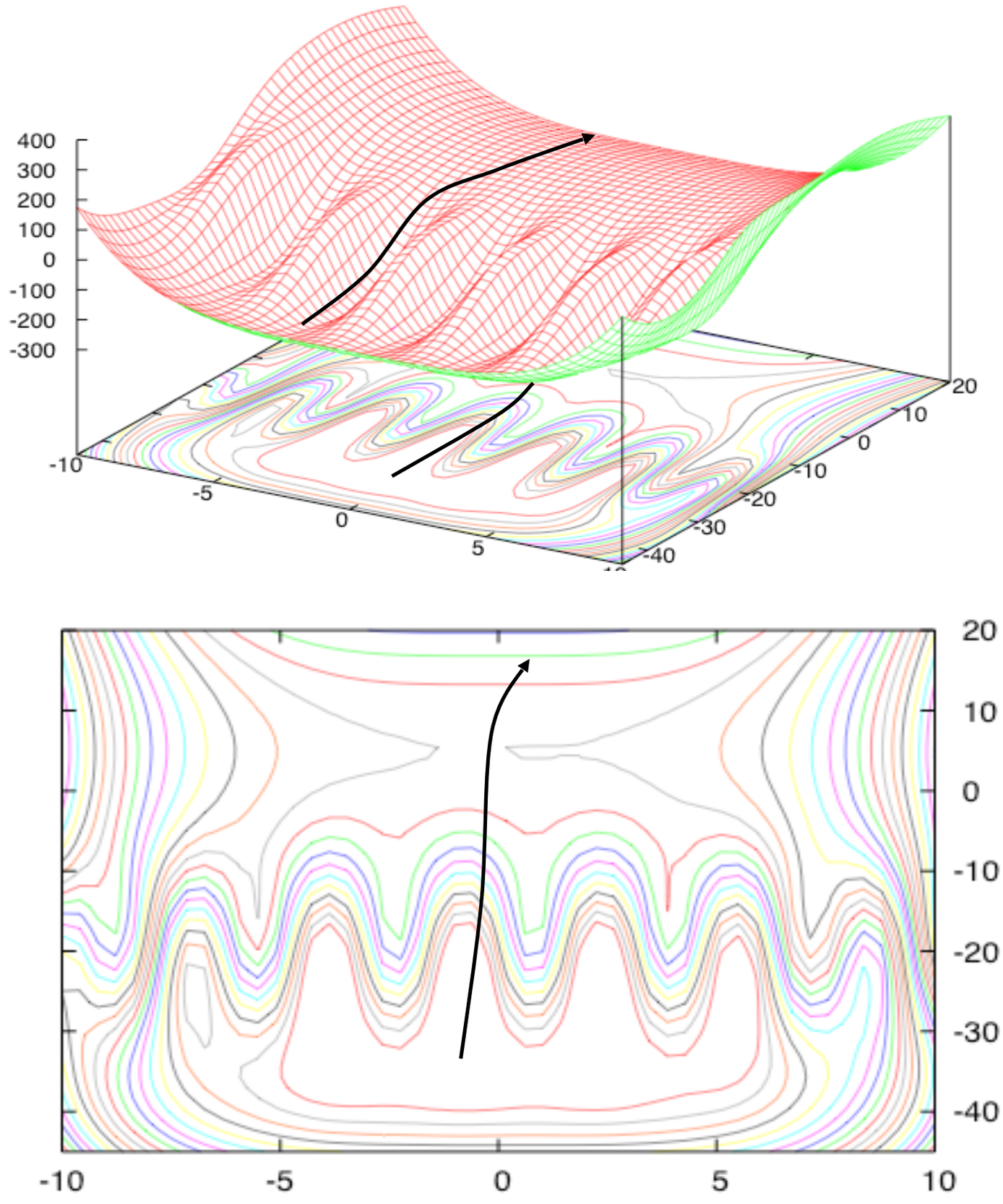
TIS compared to TPS

- TIS uses flexible path lengths.
- Only shooting moves and possibly time-reversal moves. The shifting moves are redundant.
- Faster convergence (no cancellation between positive and negative terms).
- Complete new algorithm based on the interface crossing condition instead US approach applied on the endpoint of the path.
- Always faster than the original TPS rate calculation algorithm.

The problem of multiple reaction channels

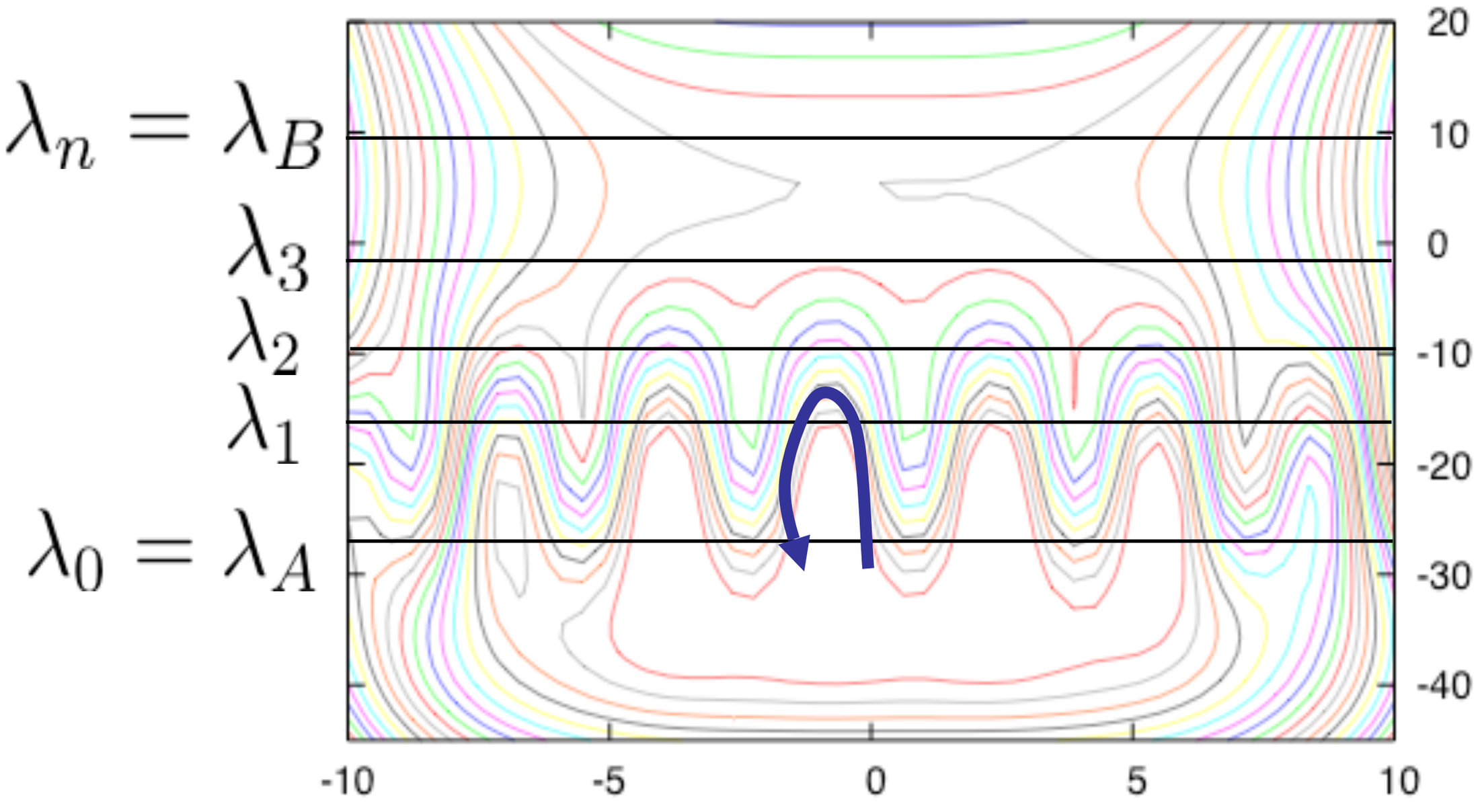


The problem of multiple reaction channels



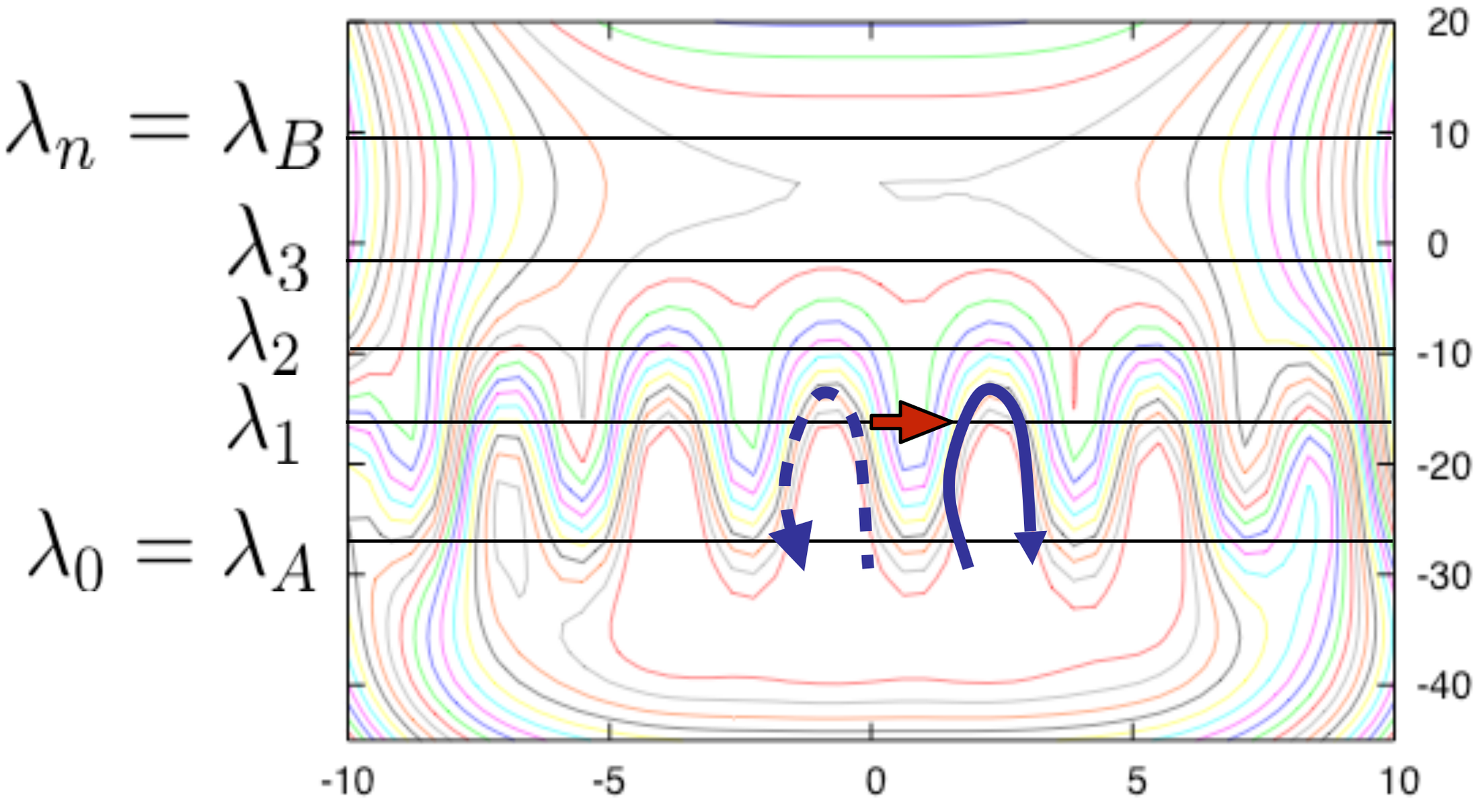
The problem of multiple reaction channels

For example: the [1+] path ensemble gets easily stuck.



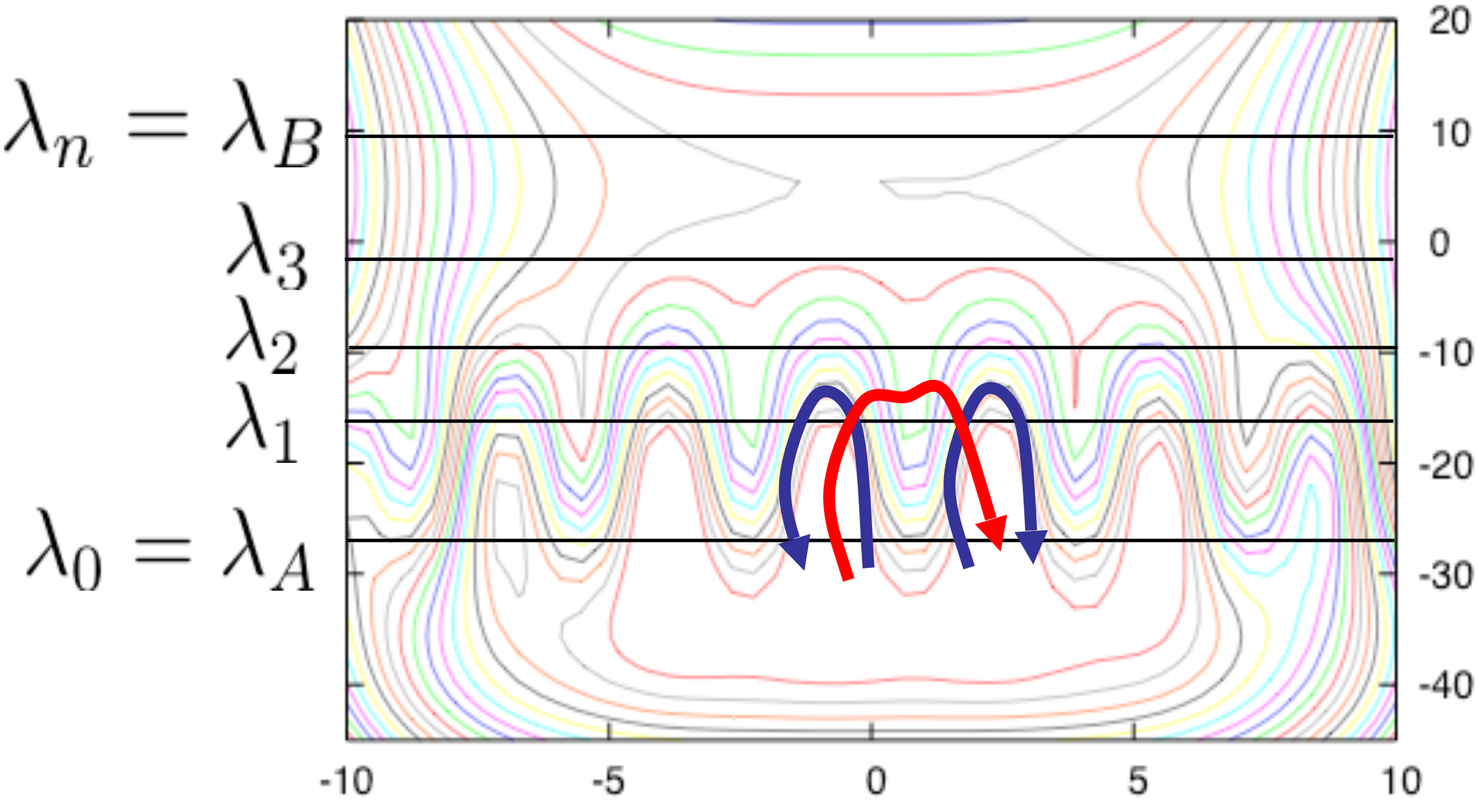
The problem of multiple reaction channels

Even after many MC path moves, I will probably never sample the right path via shooting moves.



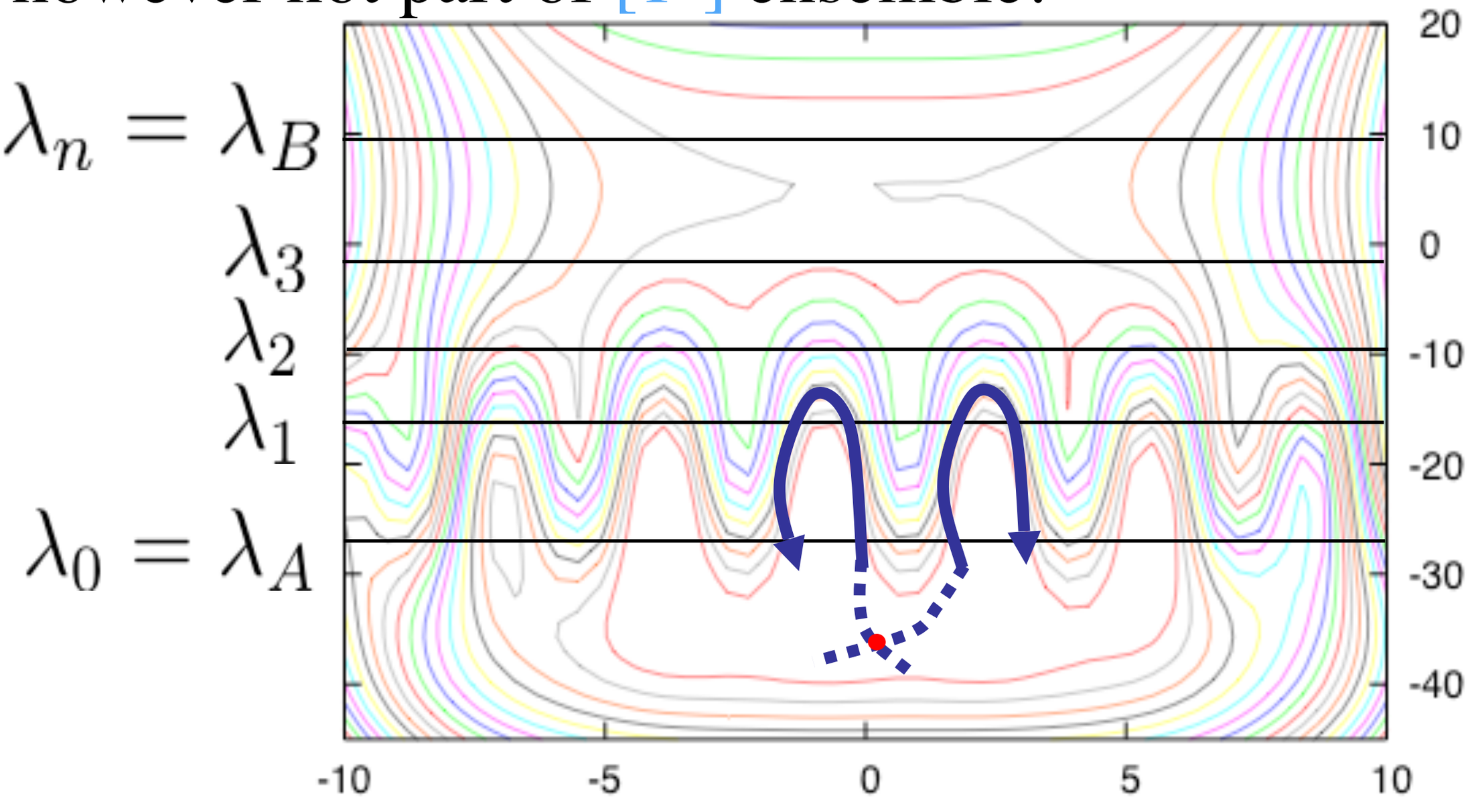
The problem of multiple reaction channels

This would require to generate an intermediate connecting path (red) with very high energy.



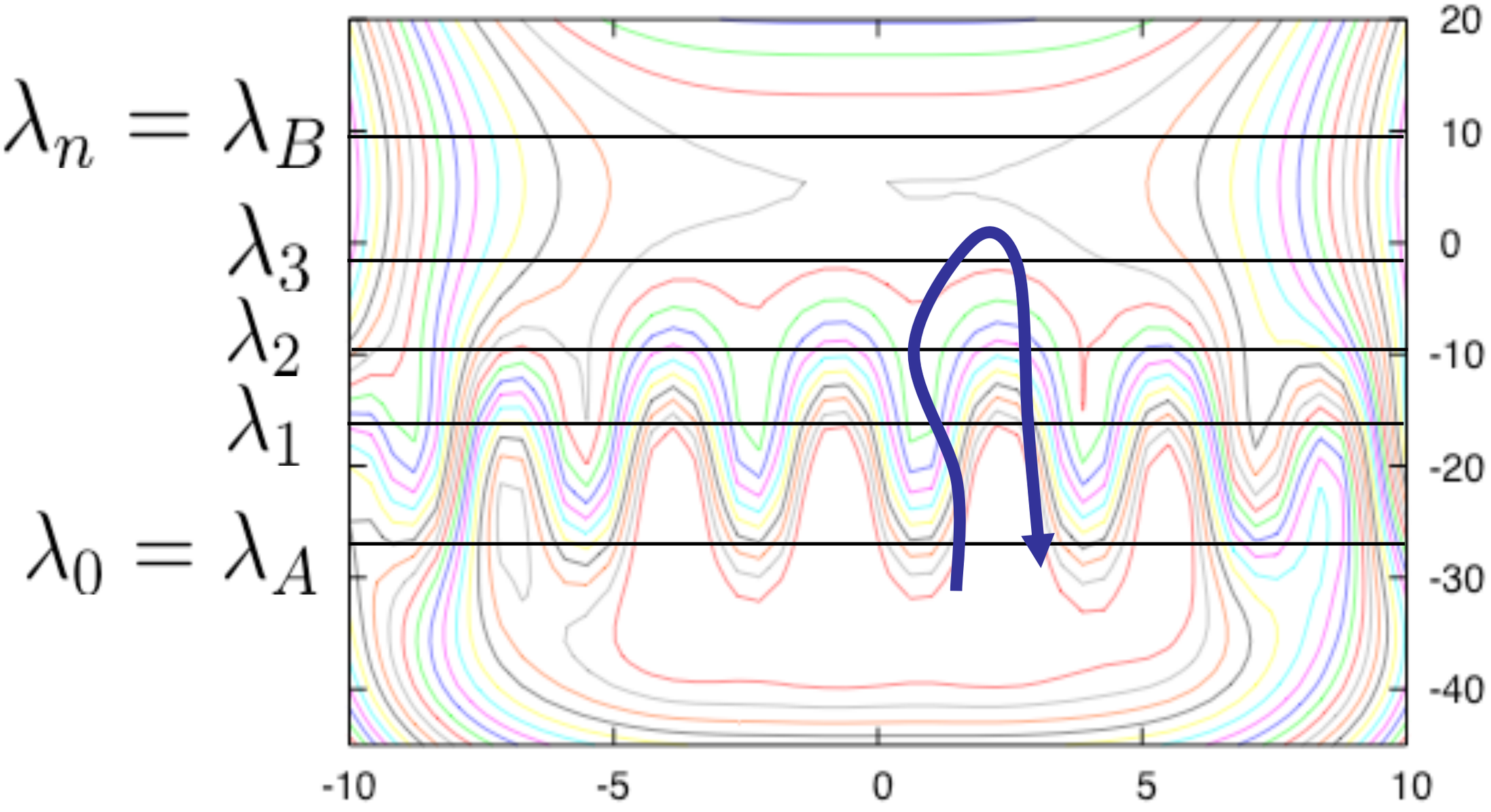
The problem of multiple reaction channels

Alternatively, it would help if we could access the parallel channel via a shooting within state A. This is however not part of [1+] ensemble.



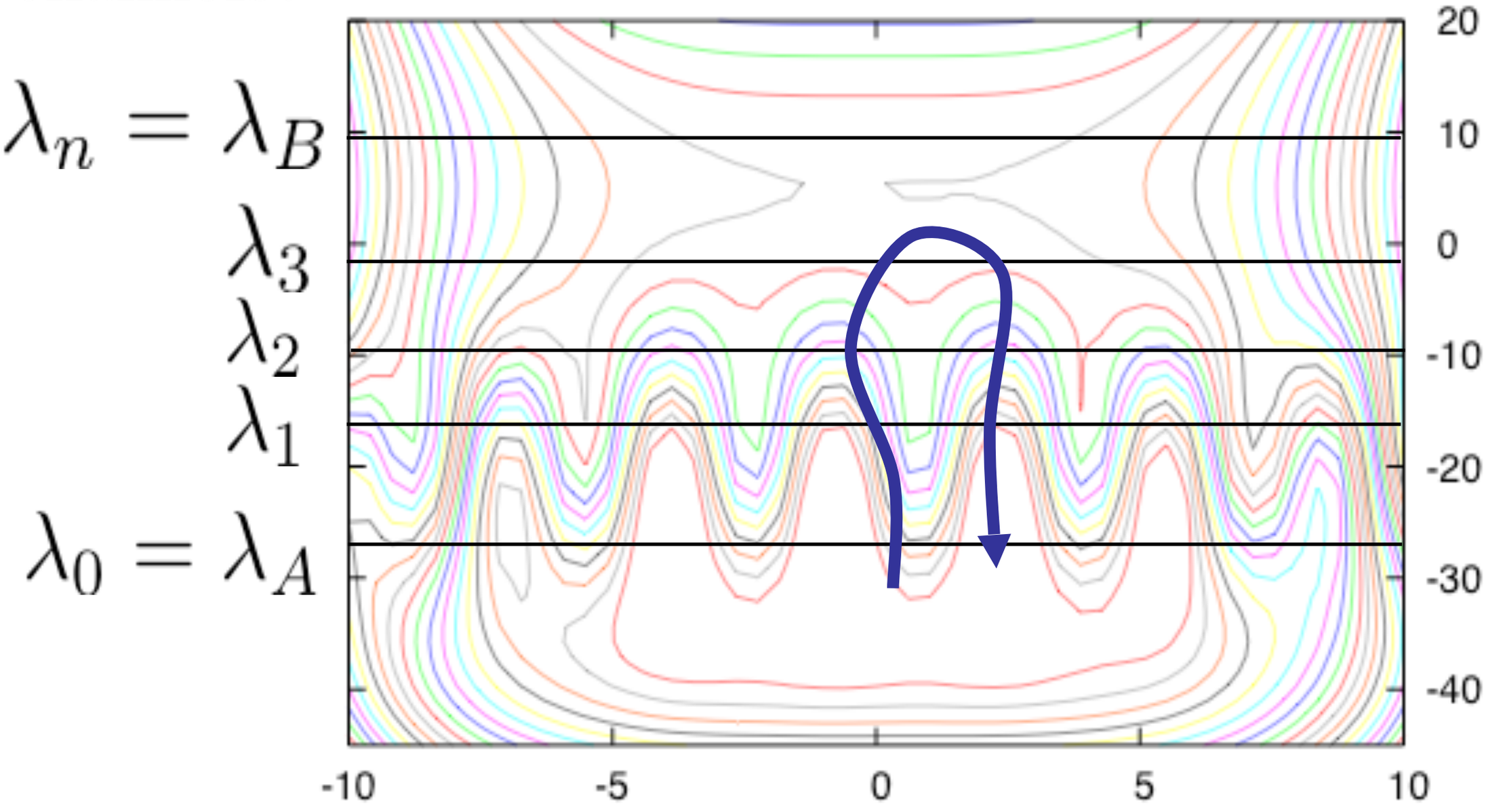
The problem of multiple reaction channels

On the other hand, the paths in the [3+] ensemble are higher in energy and can easily move across different channels!



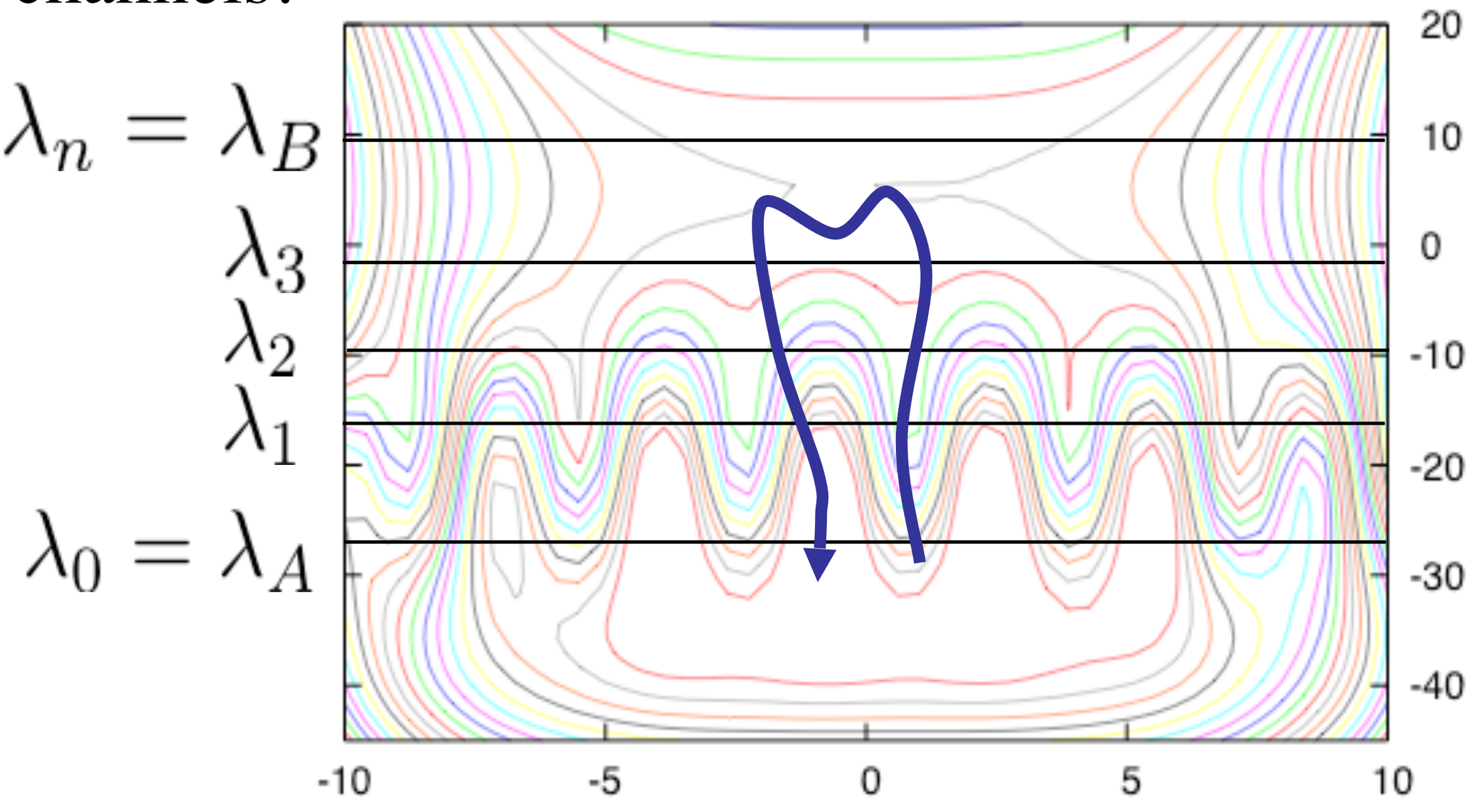
The problem of multiple reaction channels

On the other hand, the paths in the [3+] ensemble are higher in energy and can easily move across different channels!



The problem of multiple reaction channels

On the other hand, the paths in the [3+] ensemble are higher in energy and can easily move across different channels!

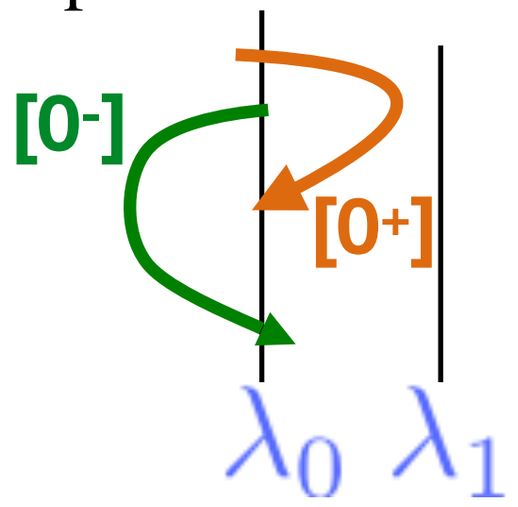


RETIS: a new TIS algorithm based on Path Swapping (Replica Exchange)

- We replace the MD simulation by another "internal" path ensemble [0-].

- Then flux is calculated by:

$$f_A = \left(\Delta t \left\langle L^{[0^-]} \right\rangle + \Delta t \left\langle L^{[0^+]} \right\rangle \right)^{-1}$$

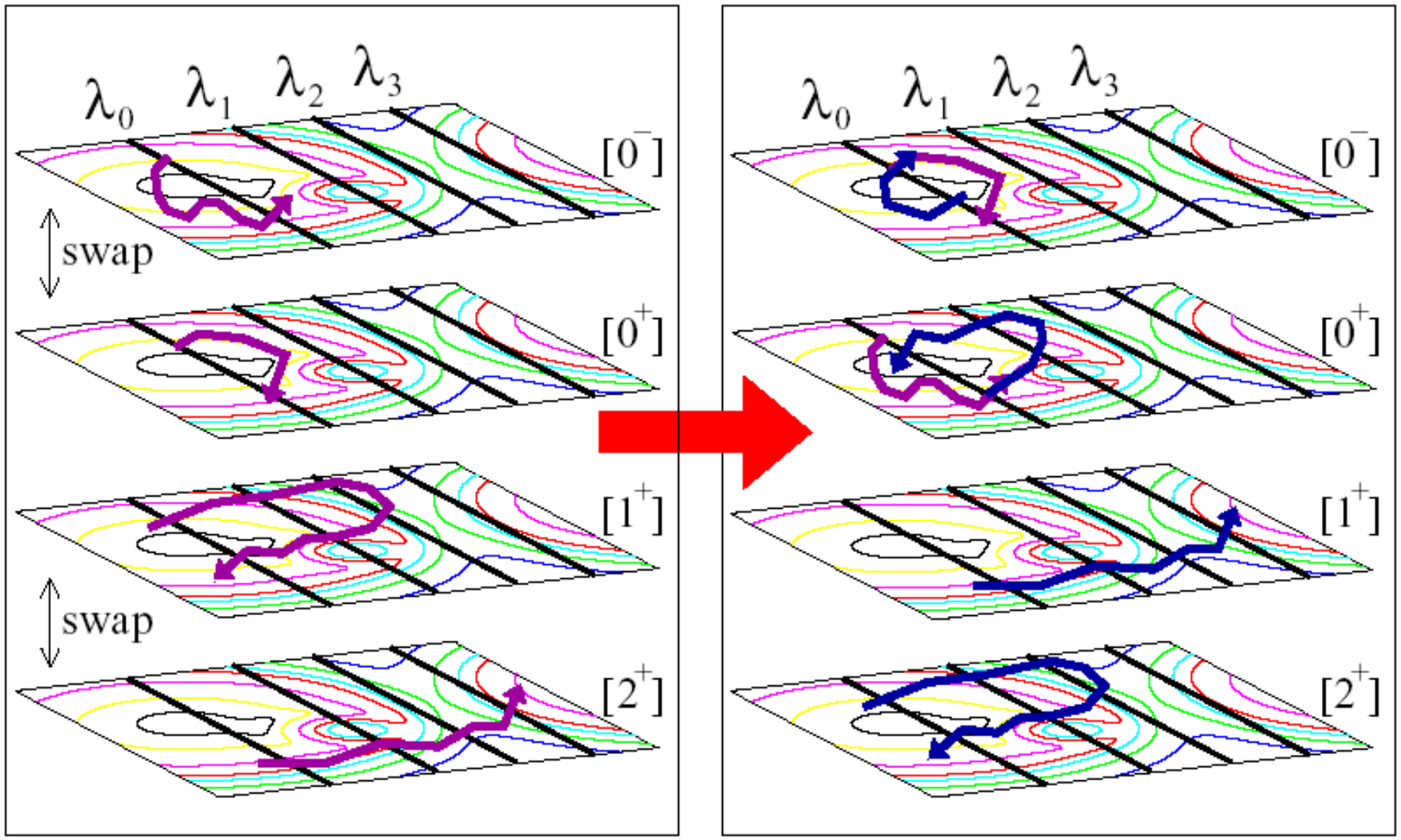


- Path in the [i+] are not stopped when crossing λ_{i+1}, but continued until hitting λ_A (=λ_0) or λ_B (=λ_n).

- Now, we can use techniques known from replica-exchange/parallel-tempering and swap trajectories from one ensemble to the other

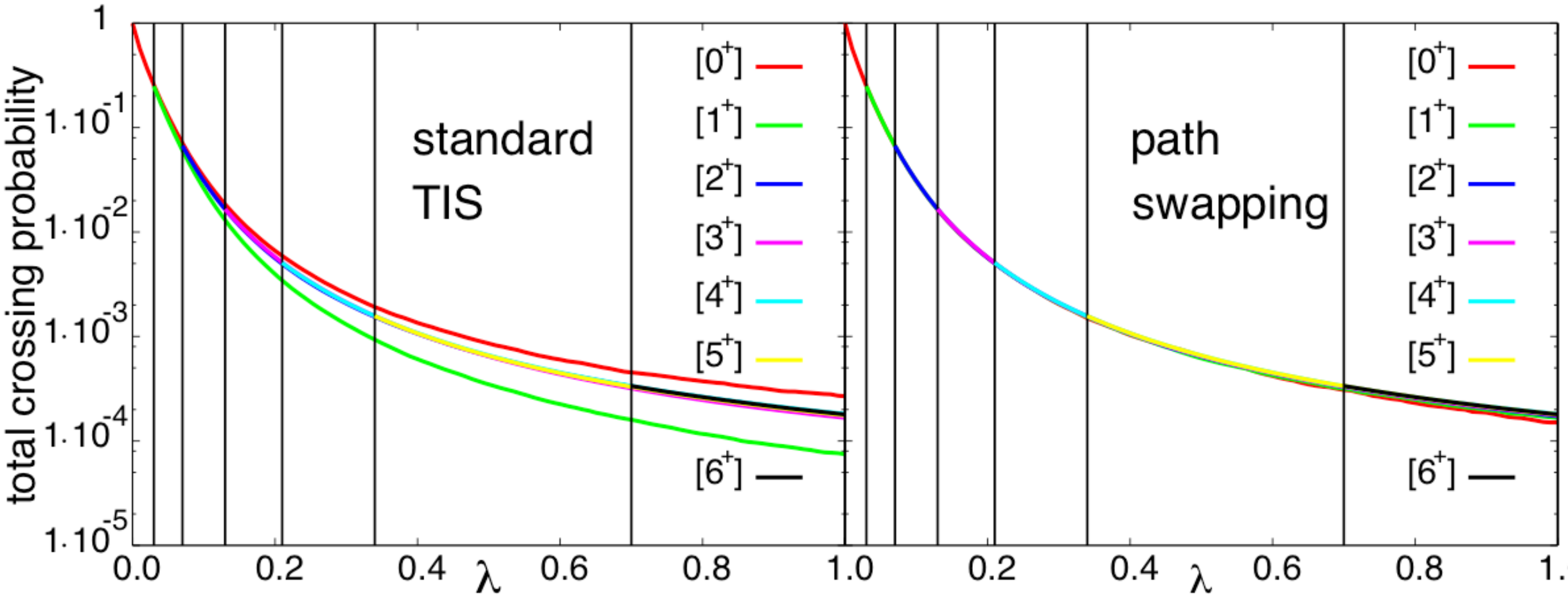
TvE, PRL 98, 268301 (2007)

Parallel Path Swapping



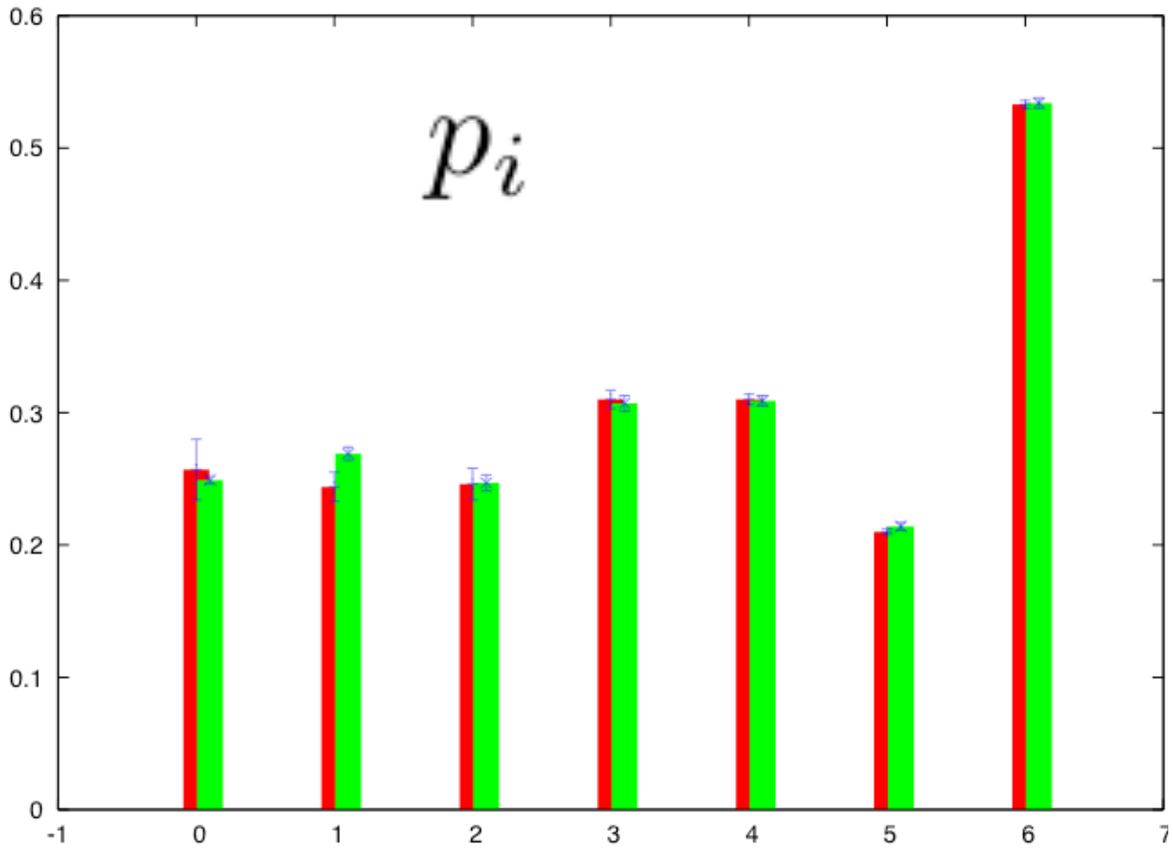
TvE, PRL 98, 268301 (2007)

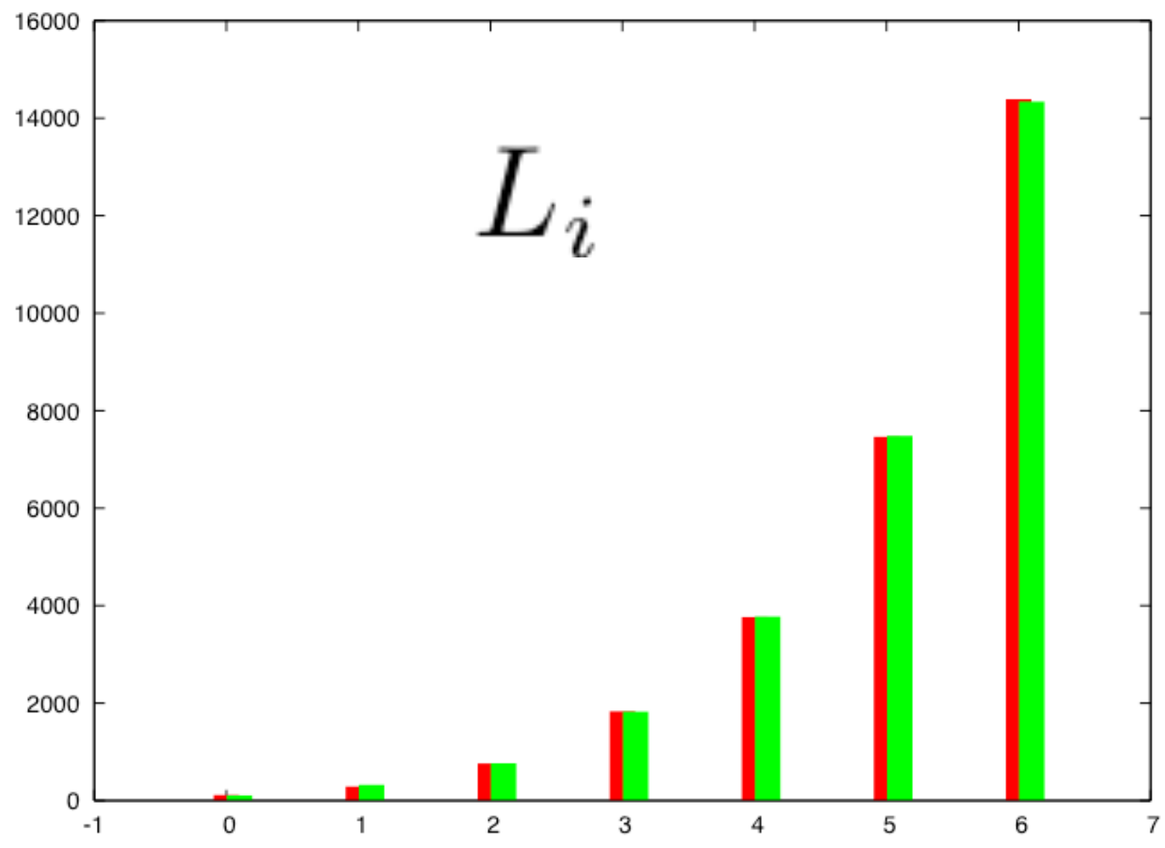
Illustration of improved convergency of the crossing probability by the RETIS algorithm for a model (PBD) of DNA denaturation



TvE, PRL 98, 268301 (2007)

Results for p_i and L_i

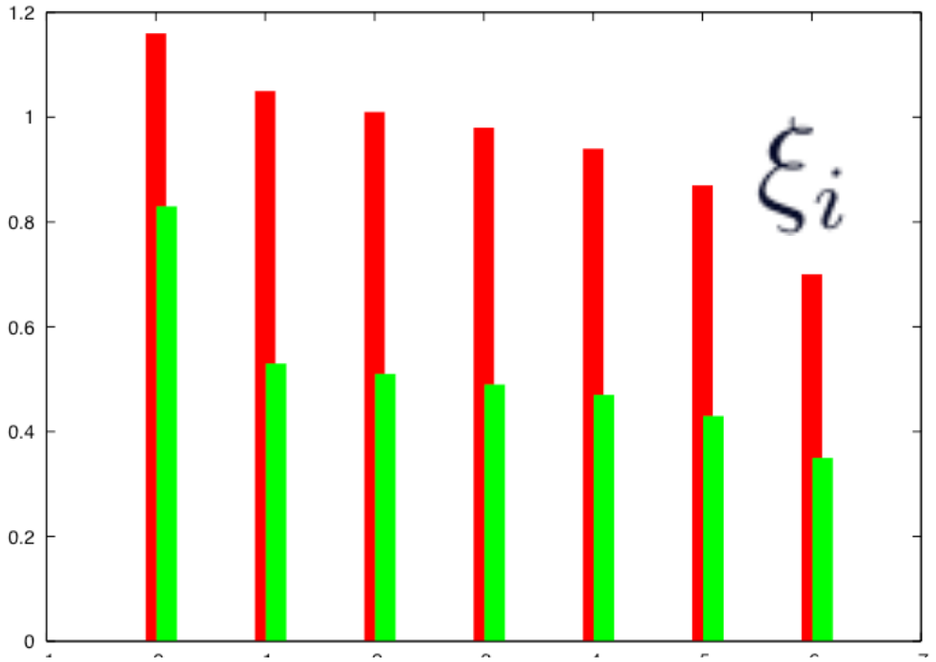


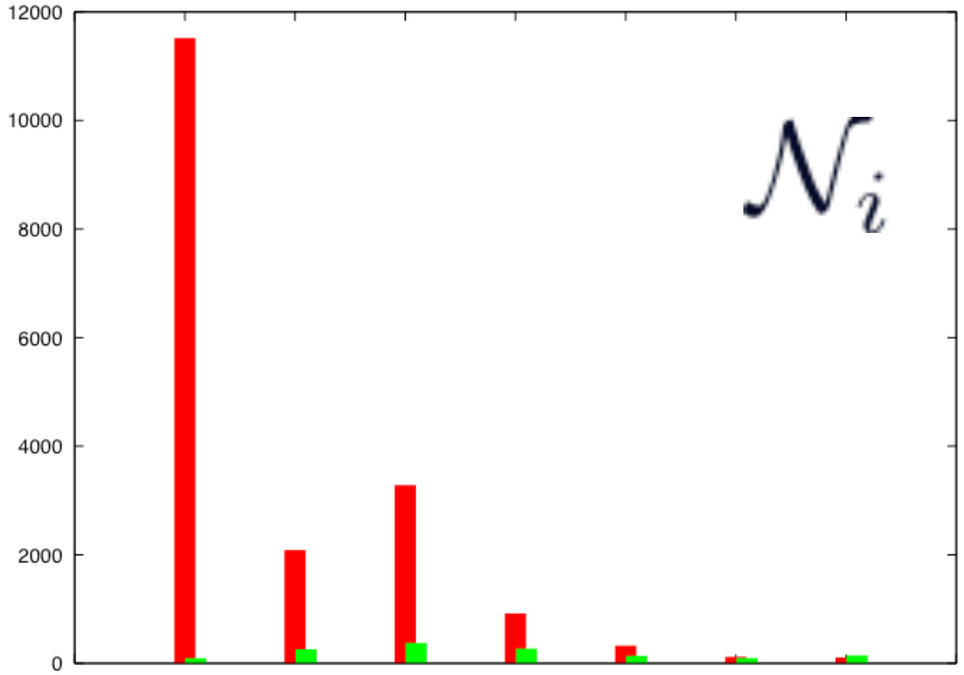


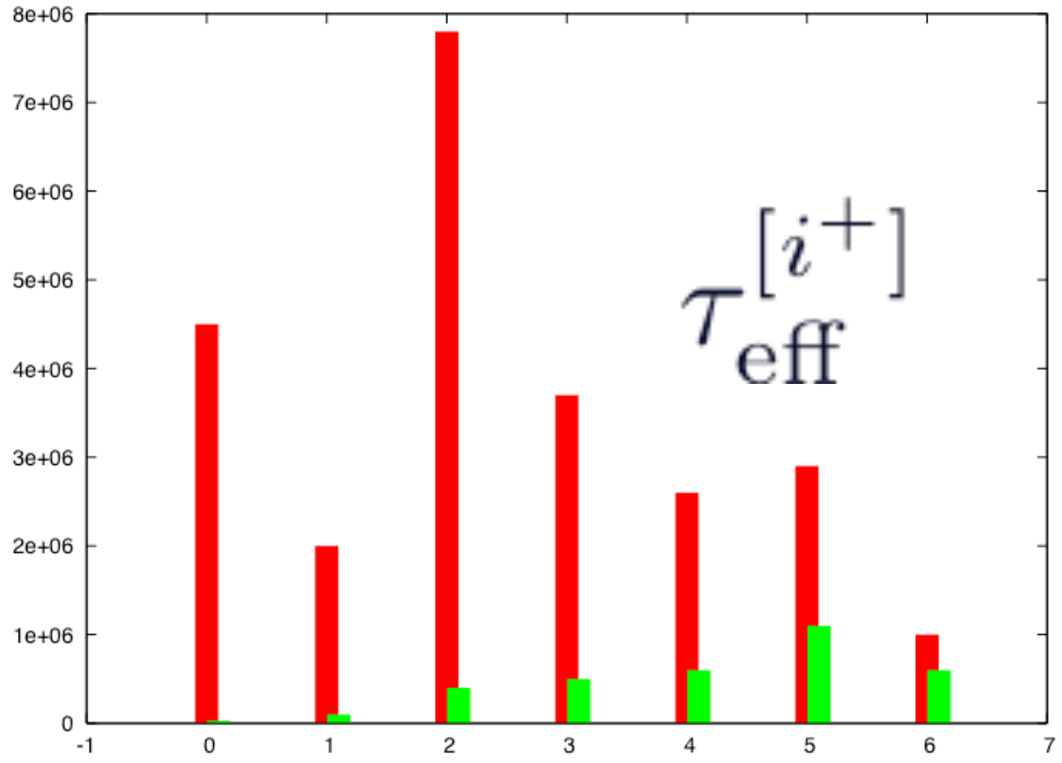
Standard TIS

TIS with swapping

Results for ξ_i, $\mathcal{N}_i$, and $\tau_{\text{eff}}^{[i^+]}$







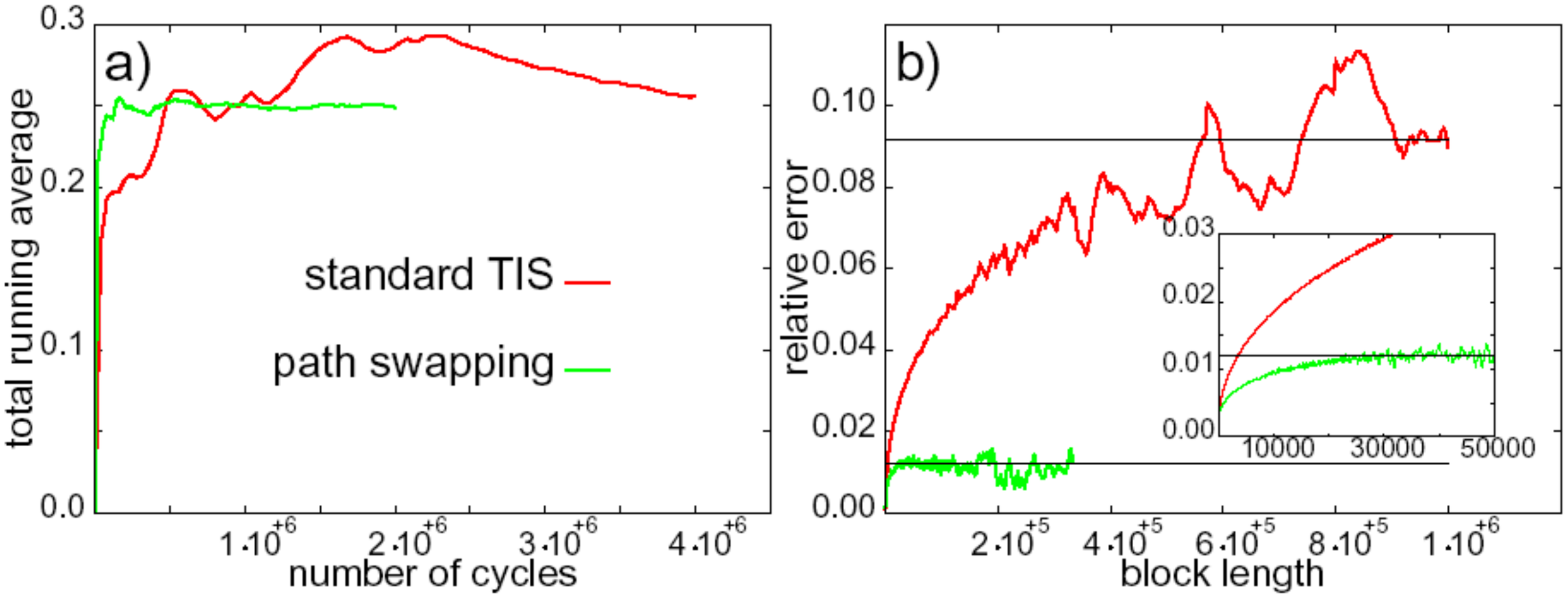
Standard TIS

TIS with swapping

$$\tau_{\text{eff}}^{\text{tot}} = 149.1 \cdot 10^7$$

$$\tau_{\text{eff}}^{\text{tot}} = 6.5 \cdot 10^7$$

Total running average and block averaging results



$[0^+]$ path ensemble results

NTNU
Norwegian University of Science and Technology

Titus S. van Erp

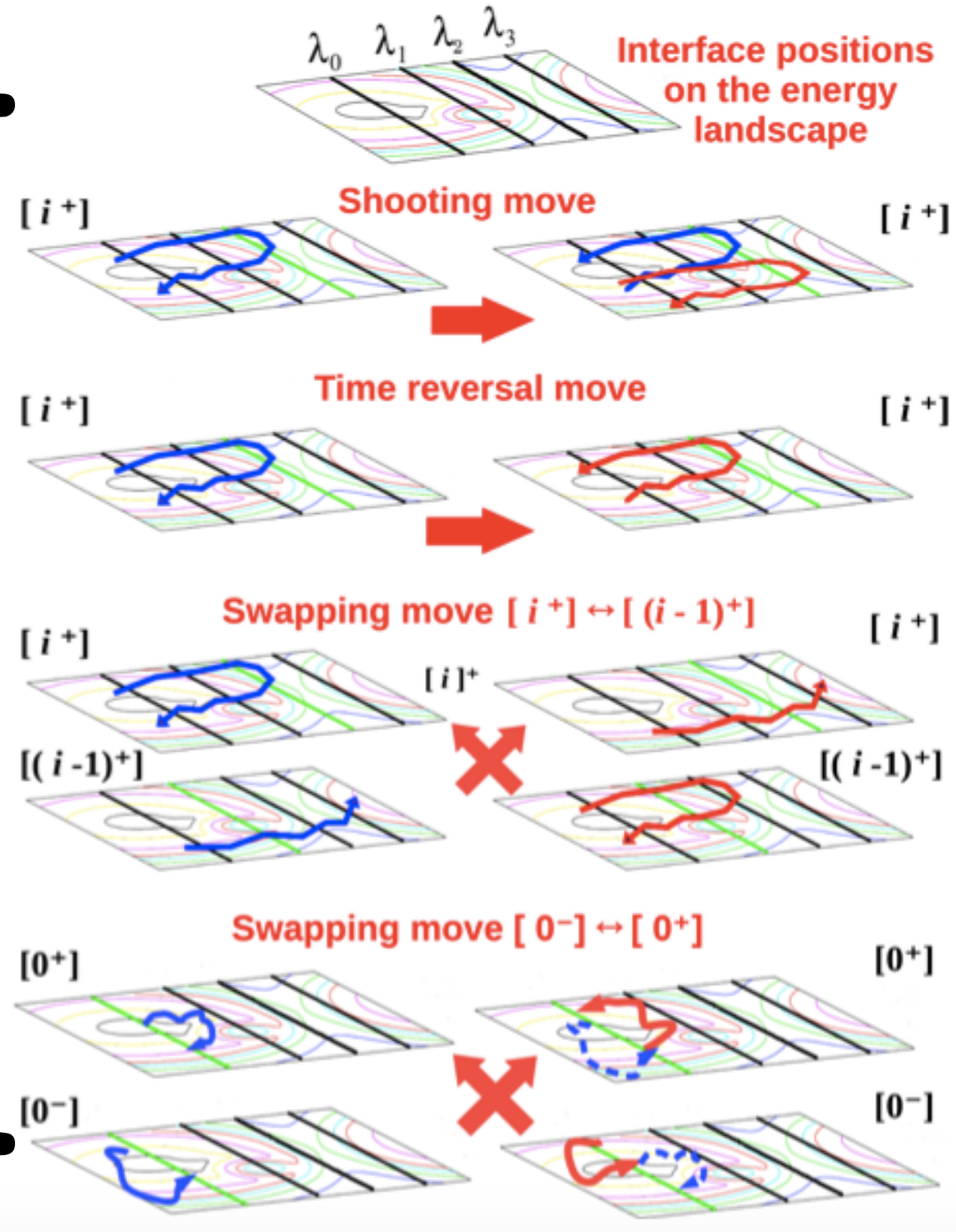
Replica Exchange TIS:

T. S. Van Erp, Phys. Rev. Lett. 2007

Cabriolu, Refsnes, Bolhuis, van Erp
J. Chem. Phys. 2017

MC moves:

- No MD part
- Instead [0$^-$] path ensemble
- Paths are always continued till A or B are reached
- Swap-Moves



Replica Exchange TIS

- Faster decorrelation than TIS and provides cheap (free) path swaps between path ensembles.

- But more difficult to implement than TIS… but now we have finally open software packages that have done it for you! PyRETIS, OPS (open path sampling)

- Unlike TIS, it is not embarrassingly parallel (PyRETIS uses MD engines that can run parallel, path ensembles are updated sequentially).

- Significantly more efficient than TIS especially in complex systems and when multiple reaction tubes/channels exist. It was 20 times faster than TIS for a study on DNA denaturation using the mesoscopic Peyrard-Bishop-Dauxois model.

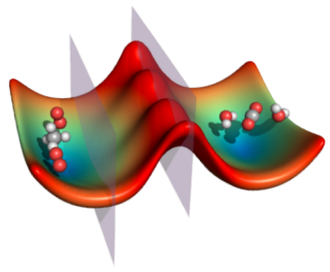
NTNU

Norwegian University of Science and Technology

Titus S. van Erp

PyRETIS 1.0.0 About Installing Getting started Input file Site Page Search

Welcome to PyRETIS!



PyRETIS

— rare events in Python

PyRETIS is a Python library for **rare event molecular simulations** with emphasis on methods based on transition interface sampling and replica exchange transition interface sampling.

PyRETIS: A well-done, medium-sized python library for rare events,
J. Comp. Chem., **38**, 2439-2451, (2017), Anders Lervik, Enrico Riccardi, and Titus S. van Erp

Lorentz center

Center for Scientific Workshops in All Disciplines

Current Workshop | Overview Back | Home | Search | Contact

Transition Path Sampling Simulations via PyRETIS: Theory and Application of Rare Events Methods to Compute Transition and Reaction Rates

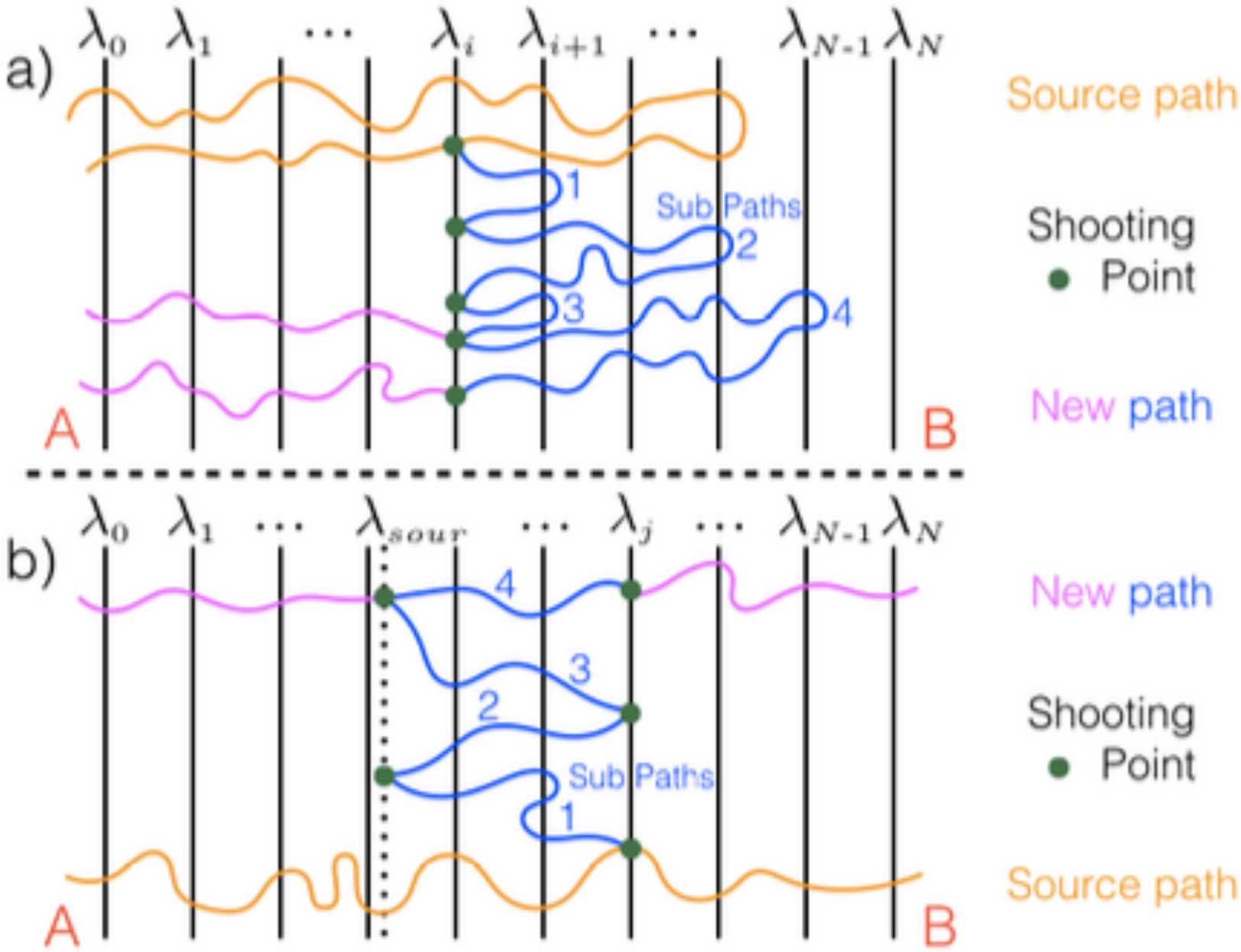
from 11 Mar 2019 through 15 Mar 2019

NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

Can we do better than shooting? Yes we can!

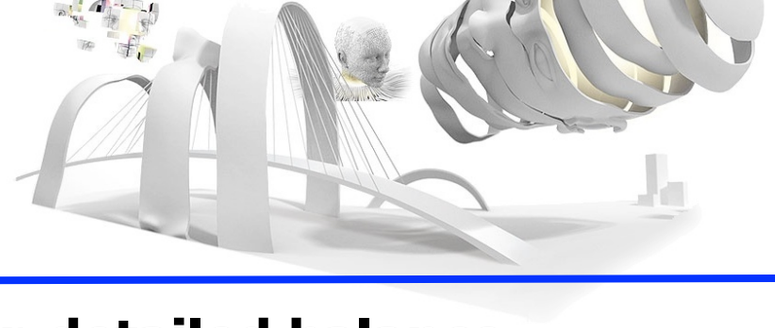
E. Riccardi, O. Dahlen, and T. S. Van Erp, J. Phys. Chem. Lett. 2017



Acceptance rule based on Super-detailed balance

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

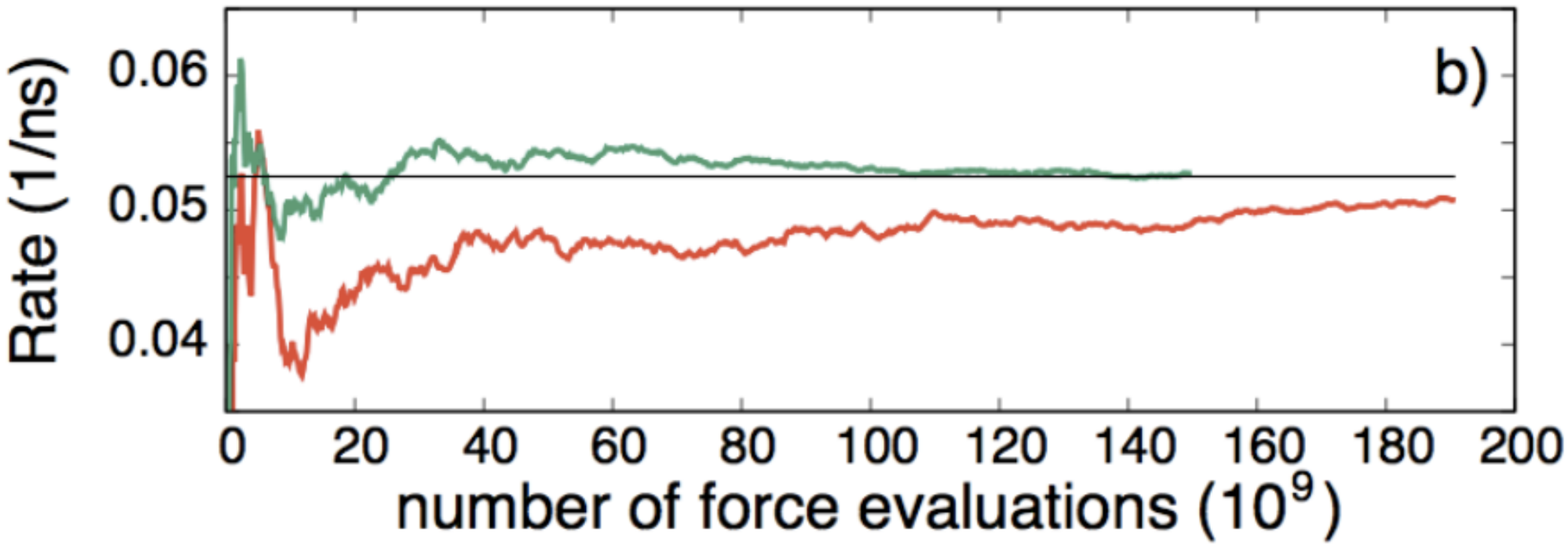
and some other tricks ……(for the details see Riccardi, Dahlen, van Erp, JPCL 2017)



Acceptance rule based on Super-detailed balance

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

and some other tricks ……(for the details see Riccardi, Dahlen, van Erp, JPCL 2017)



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

Stone skipping was found 12 times faster

TIS related methods

- Partial Path TIS
- Milestoning
- Forward Flux Sampling
- TS-PPTIS

Key papers:

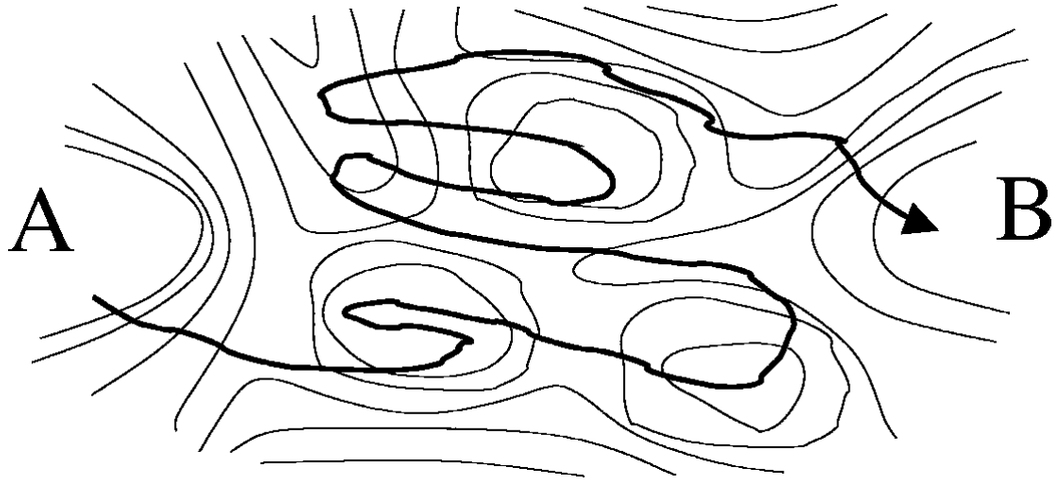
PPTIS: *Rate constants for diffusive processes by partial path sampling,* J. Chem. Phys. 120, 4055 (2004); Moroni, Bolhuis, van Erp, https://doi.org/10.1063/1.1644537

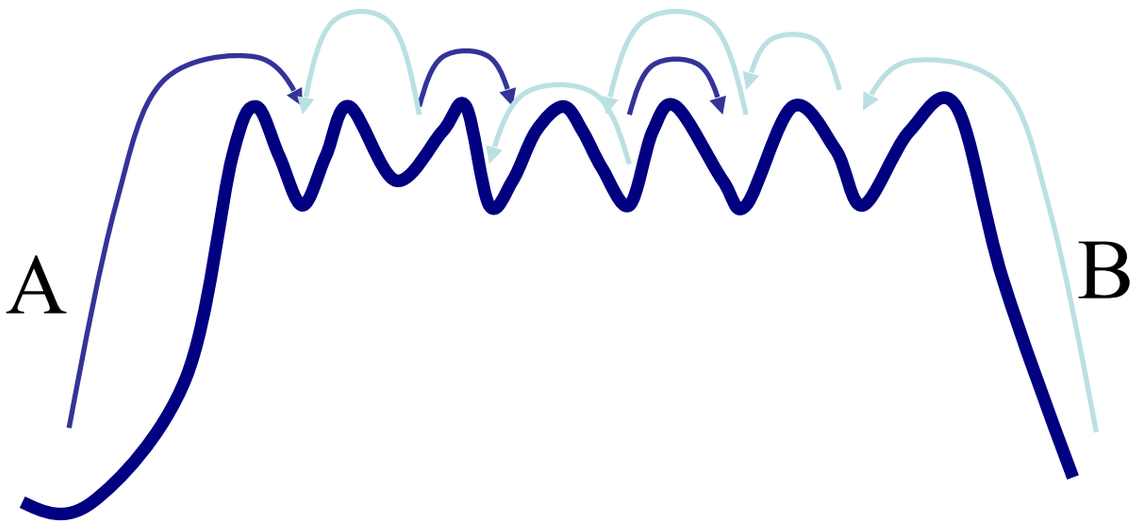
Milestoning: *Computing time scales from reaction coordinates by milestoning* J. Chem. Phys. 120, 10880 (2004) Faradjian, Elber, https://doi.org/10.1063/1.1738640

FFS: *Sampling Rare Switching Events in Biochemical Networks,* Phys. Rev. Lett. 94, 018104 (2005), Allen, Warren, ten Wolde, https://doi.org/10.1103/PhysRevLett.94.018104

TS-PPTIS: *Efficient Numerical Reconstruction of Protein Folding Kinetics with Partial Path Sampling and Pathlike Variables,* Phys. Rev. Lett. 110, 108106 (2013) https://doi.org/10.1103/PhysRevLett.110.108106

Diffusive barriers: Partial Path TIS





TIS and TPS paths can be very long for diffusive barriers

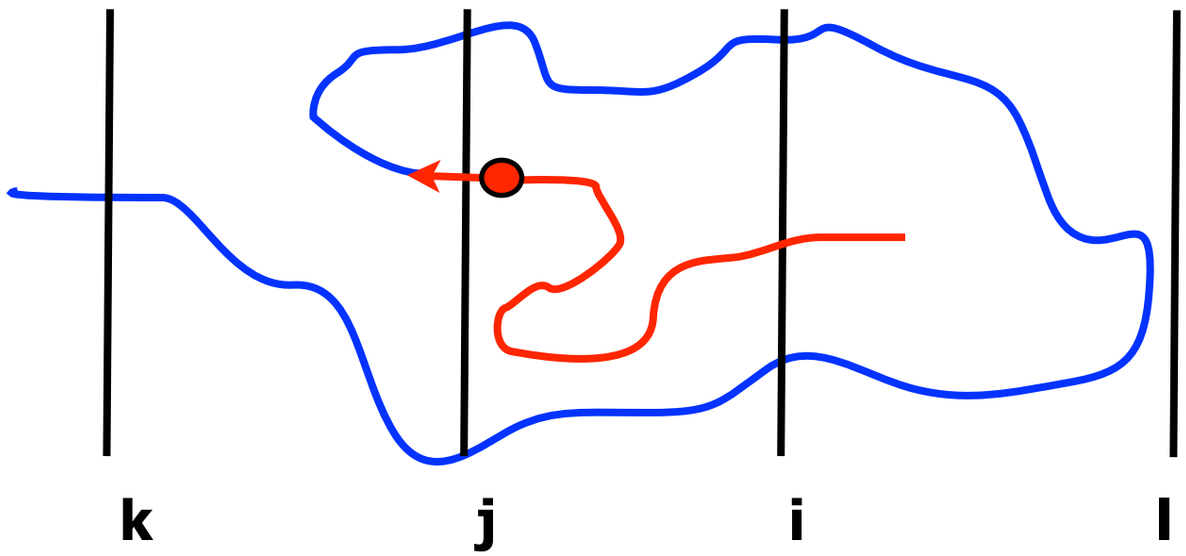
- assume complete memory loss after some time
- if hopping probabilities known overall rate constant k_{AB} can be found
- Hopping probabilities can be calculated by partial path TIS

NTNU

Norwegian University of
Science and Technology

The conditional crossing probability

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



NTNU

Norwegian University of Science and Technology

The conditional crossing probability

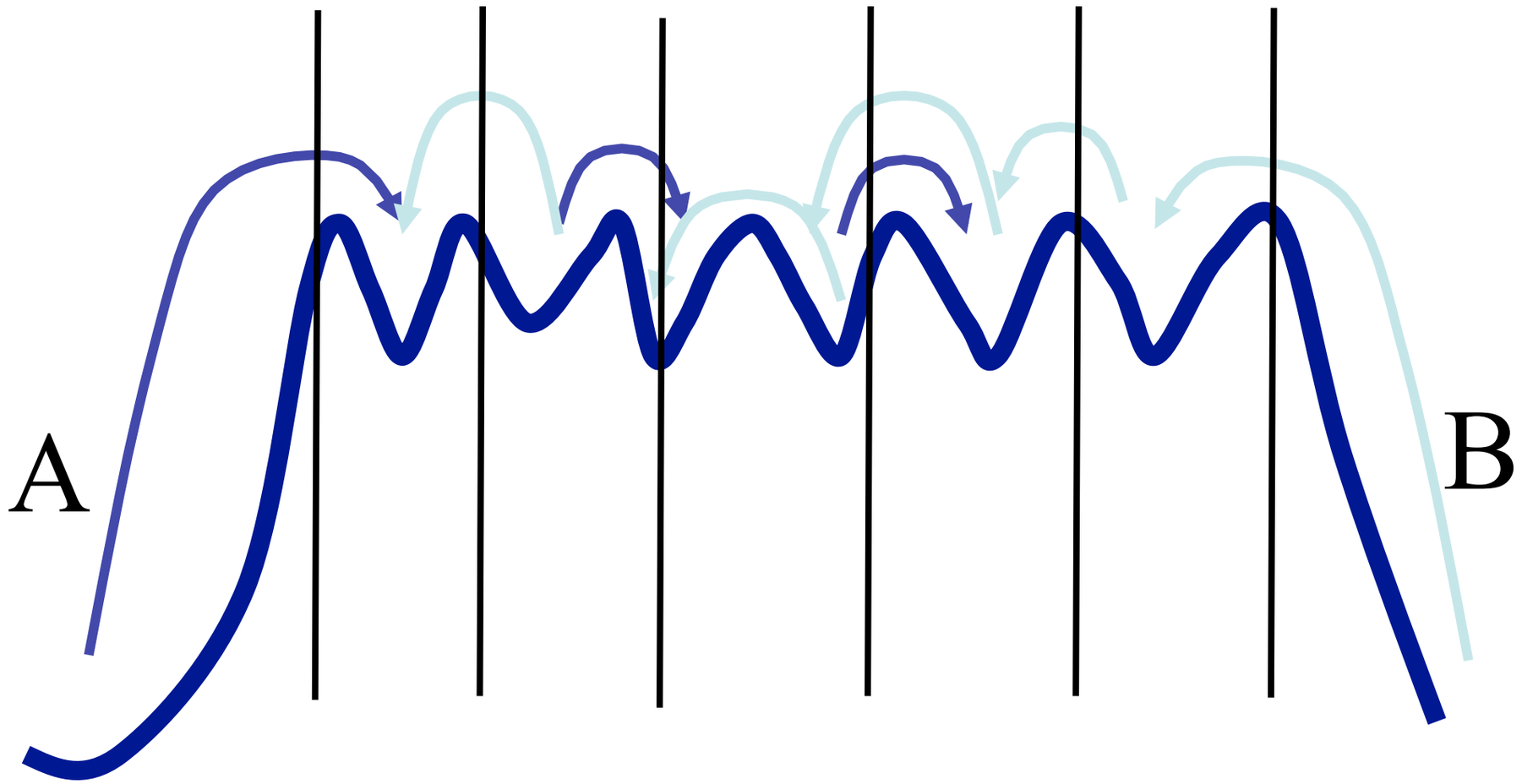
$$P\binom{k}{l}\Big|\binom{j}{i}$$

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

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

Partial Path TIS

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



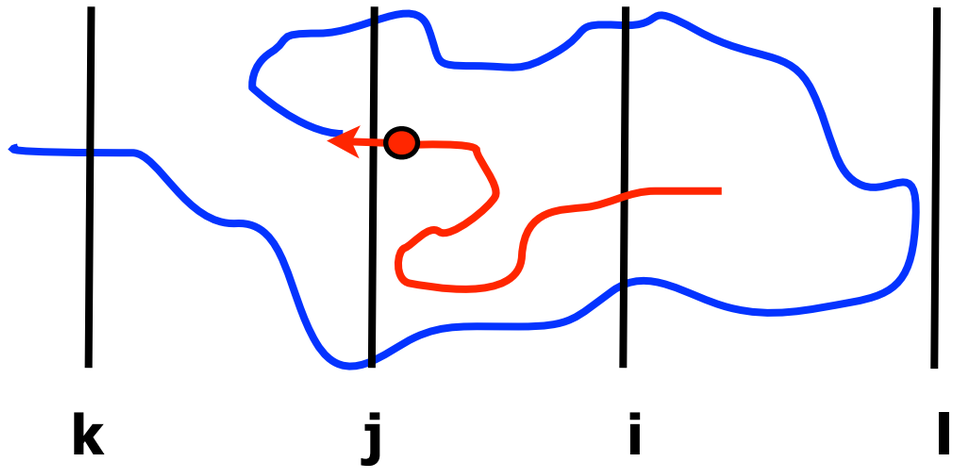
Memory-loss assumption: $P(^{l}_{k}|^{j}_{j\pm m}) \approx P(^{l}_{k}|^{j}_{j\pm 1})$

NTNU

Norwegian University of Science and Technology

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

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

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

which fulfill the following relations:

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

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

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

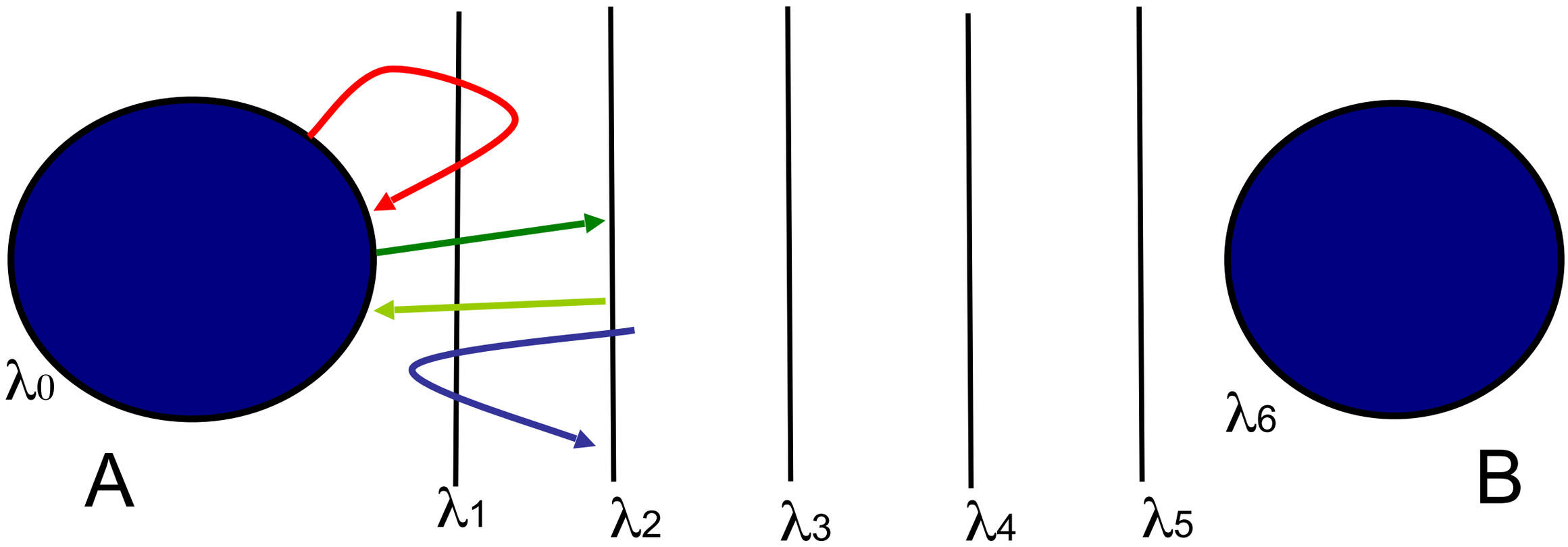
$$P_j^{-} = \frac{p_{j-1}^{\mp} P_{j-1}^{-}}{p_{j-1}^{\pm} + p_{j-1}^{=} P_{j-1}^{-}}.$$

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

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

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

Partial Path sampling



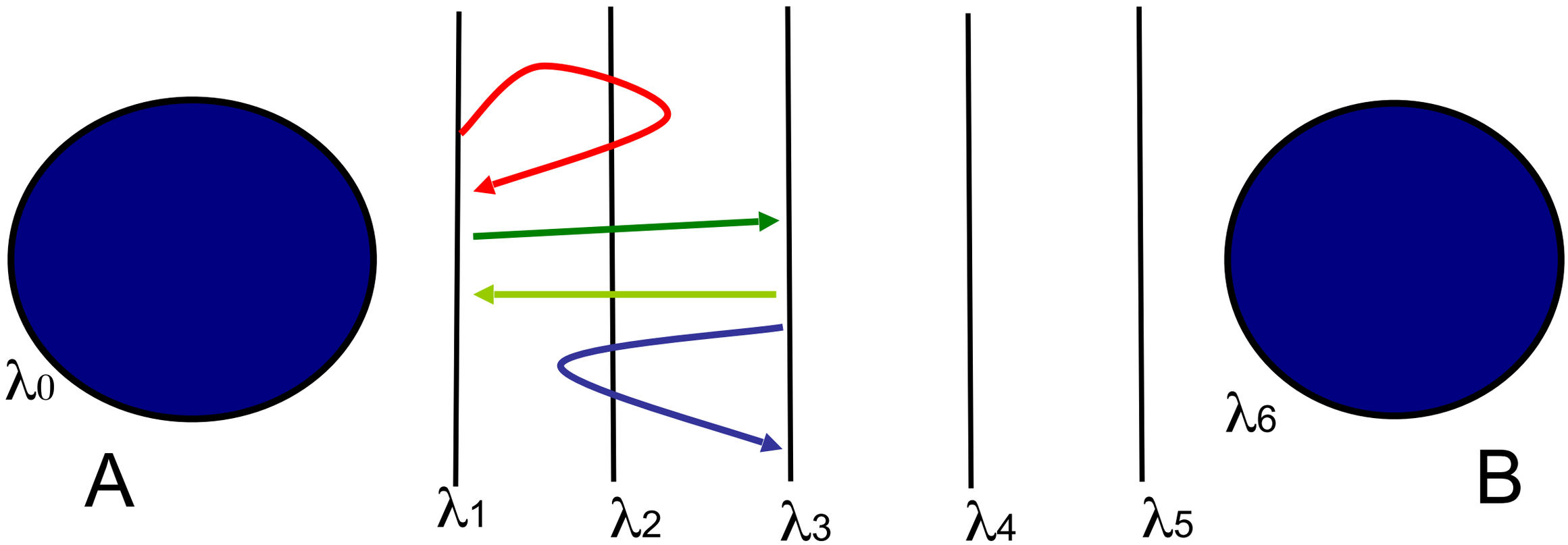
$$p_1^{\pm} = \frac{\#p_{02}}{\#p_{00}+\#p_{02}}$$

$$p_1^{\mp} = \frac{\#p_{20}}{\#p_{22}+\#p_{20}}$$

$$p_1^{=} = 1 - p_1^{\pm}$$

$$p_1^{\ddagger} = 1 - p_1^{\mp}$$

Partial Path sampling



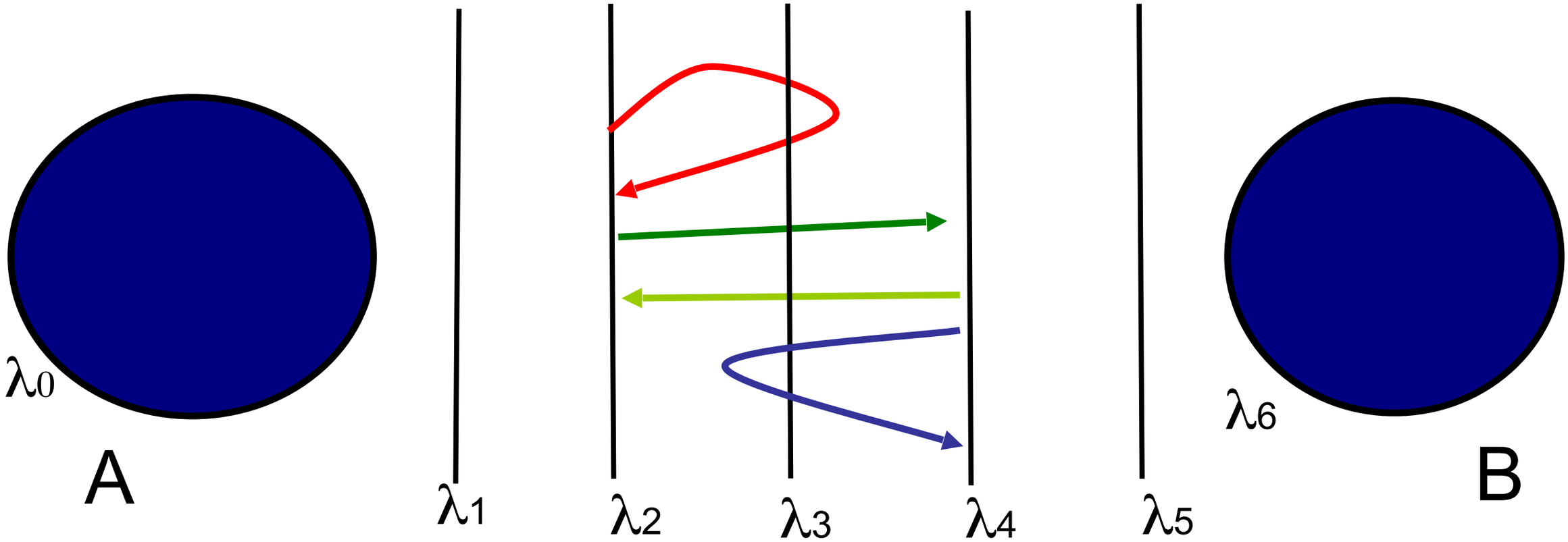
$$p_2^{\pm} = \frac{\#p_{13}}{\#p_{11}+\#p_{13}}$$

$$p_2^{\mp} = \frac{\#p_{31}}{\#p_{33}+\#p_{31}}$$

$$p_2^{=} = 1 - p_2^{\pm}$$

$$p_2^{\ddagger} = 1 - p_2^{\mp}$$

Partial Path sampling



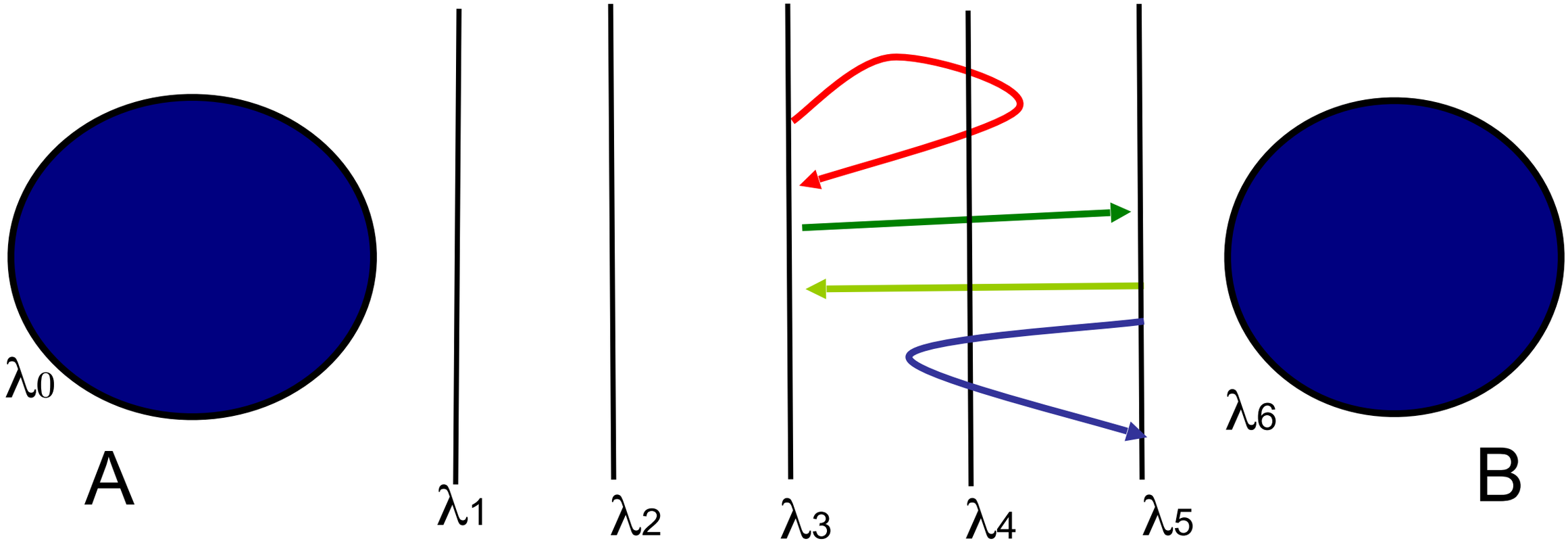
$$p_3^{\pm} = \frac{\#p_{24}}{\#p_{22}+\#p_{24}}$$

$$p_3^{\mp} = \frac{\#p_{42}}{\#p_{44}+\#p_{42}}$$

$$p_3^{=} = 1 - p_3^{\pm}$$

$$p_3^{\ddagger} = 1 - p_3^{\mp}$$

Partial Path sampling



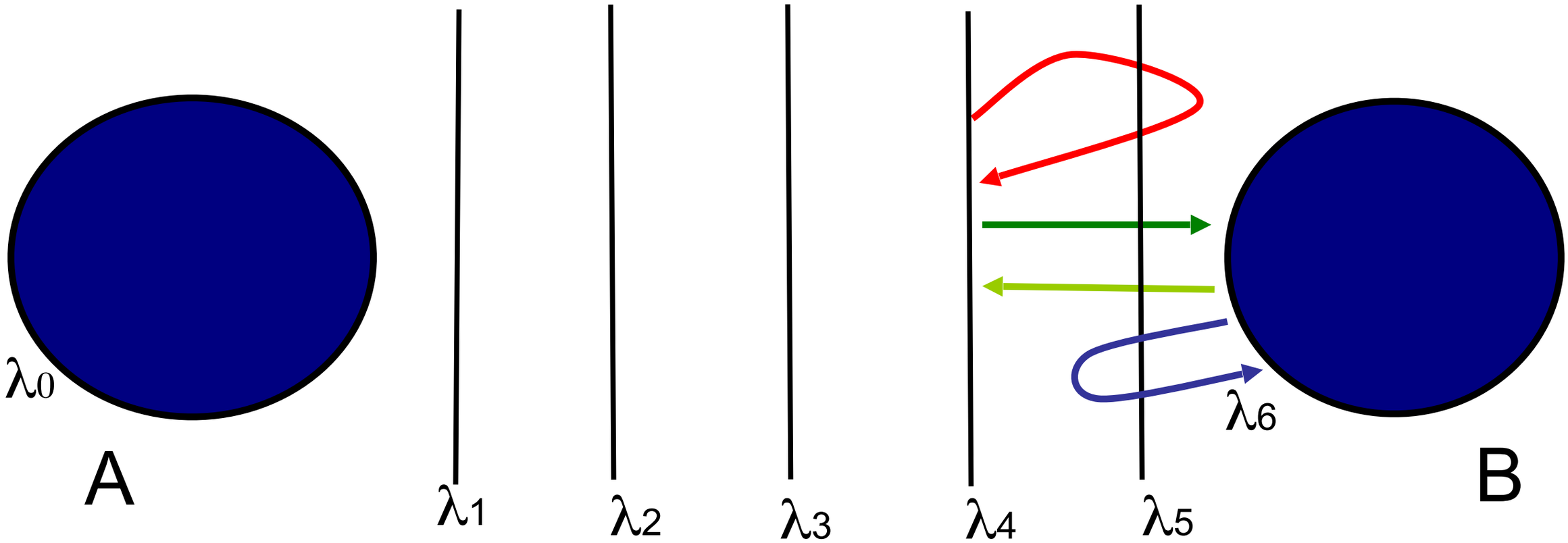
$$p_4^{\pm} = \frac{\#p_{35}}{\#p_{33}+\#p_{35}}$$

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

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

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

Partial Path sampling



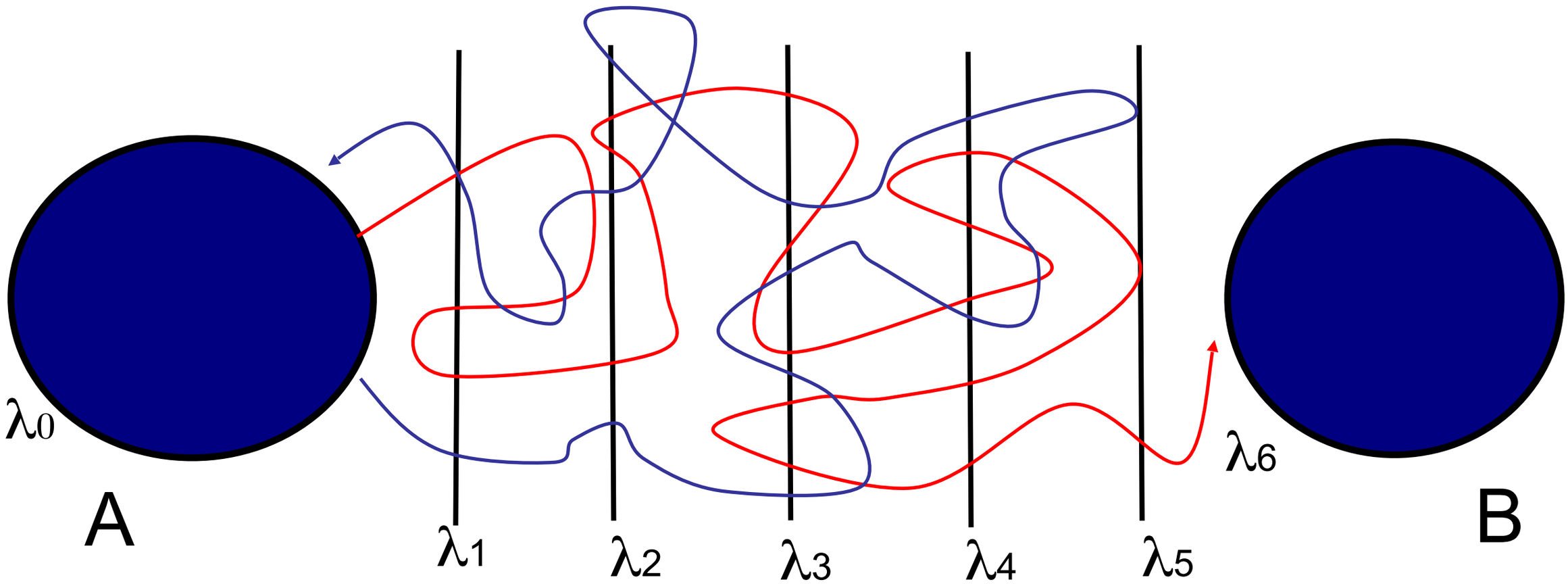
$$p_5^{\pm} = \frac{\#p_{46}}{\#p_{44}+\#p_{46}}$$

$$p_5^{\mp} = \frac{\#p_{64}}{\#p_{66}+\#p_{64}}$$

$$p_5^{=} = 1 - p_5^{\pm}$$

$$p_5^{\ddagger} = 1 - p_5^{\mp}$$

Partial Path sampling

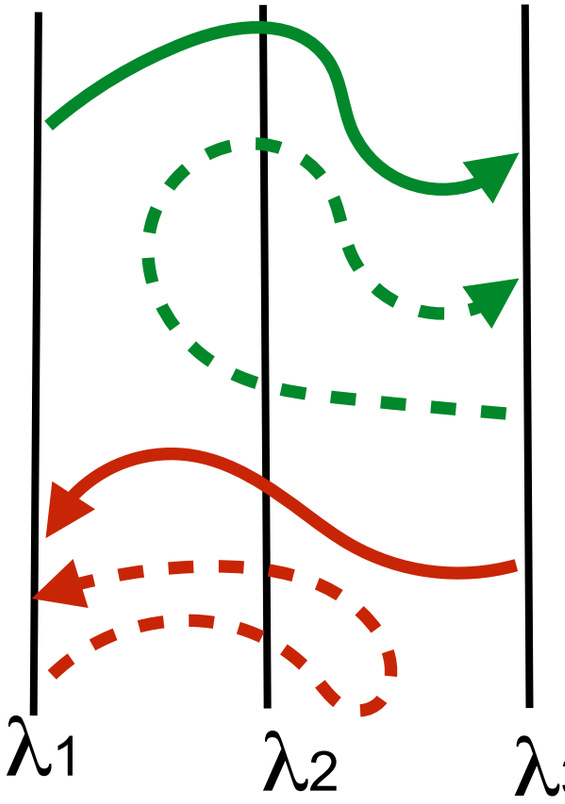


$$P_1^+ = P_1^- = 1,$$

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

While TIS and RETIS are exact, PPTIS gives an approximation to the rate. Pathways are much shorter. PPTIS is particularly good for diffusive barrier crossings for which full paths are very long and the memory loss assumption is a very good approximation.

Differences between PPTIS and Milestoning



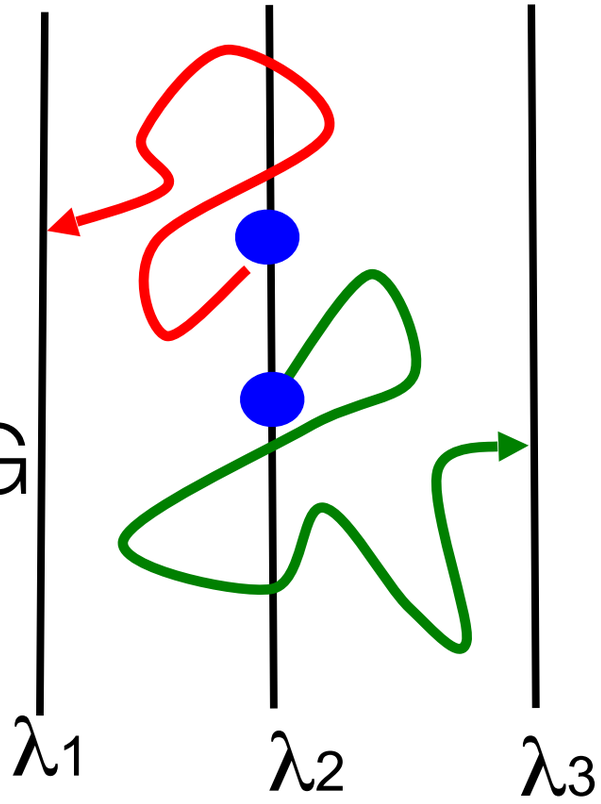
PPTIS

- Trajectories from shooting/time-rev.
- Spatial memory:

$$p_i^{\pm} \neq p_i^{\ddagger} \qquad p_i^{\mp} \neq p_i^{=}$$

Is important if there are barriers orthogonal to the RC, inertia effects.

MILESTONING



- Trajectories released from eq. distribution
- Time-memory

$$p_{1\to 3}(t) = \int_0^t p_{1\to 2}(t')p_2^{+}(t-t')\mathrm{d}t'$$

Important if there is no clear separation of time-scales (i.e. when the rate is ill-defined)

Forward Flux Sampling (FFS):

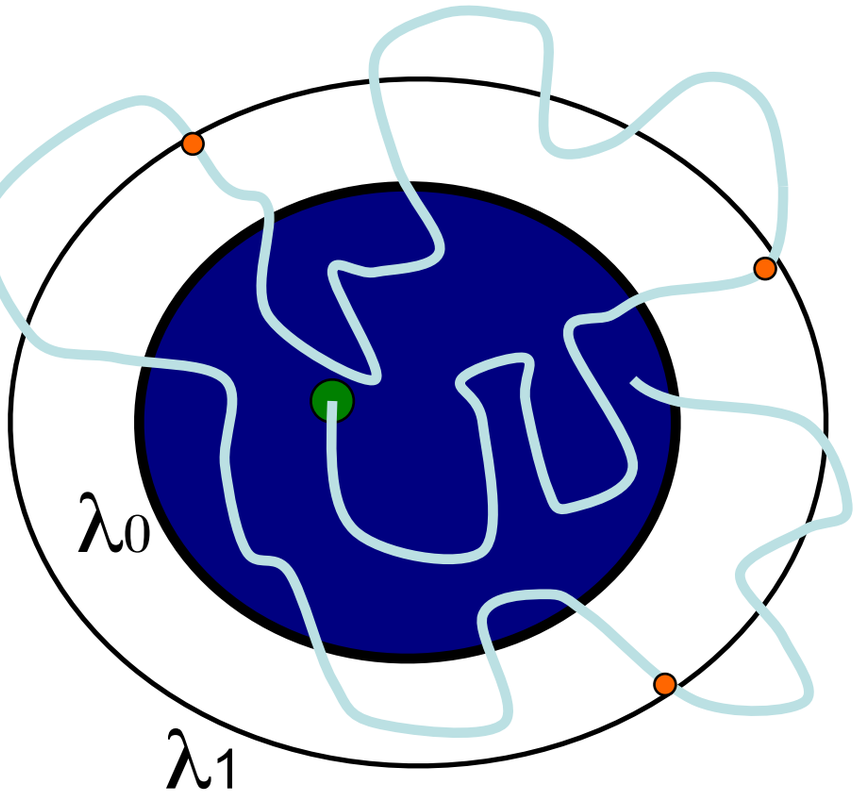
- Based on the same rate equations as TIS:

$$k_{AB} = f_A \prod_{i=0}^{n-1} P_A(\lambda_{i+1} \mid \lambda_i)$$

- But instead of Metropolis-Hastings MC it uses *splitting* to generate paths

FFS

MD run to determine effective escape flux through interface 1

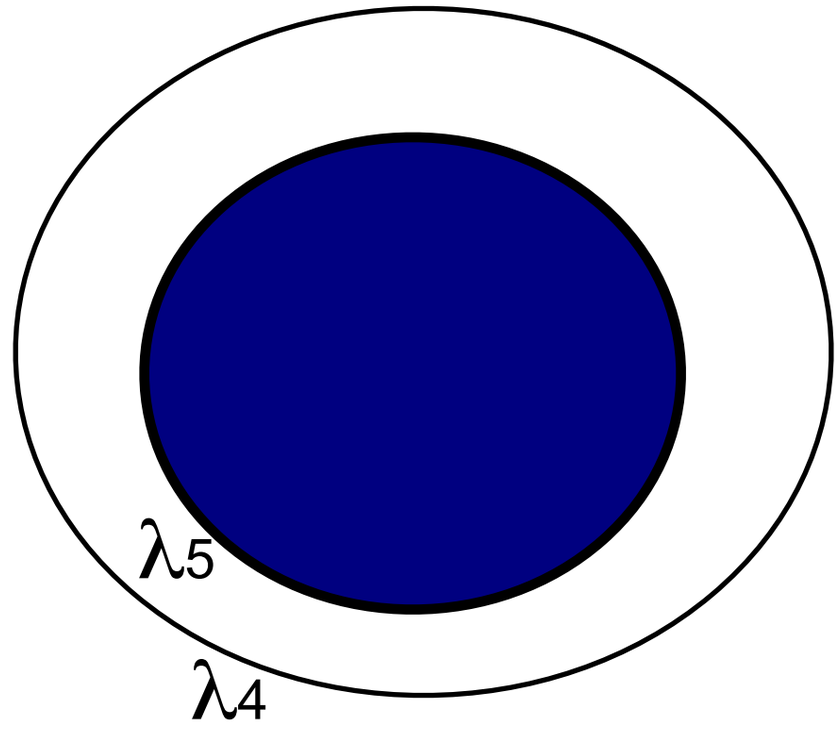


A

$$\Phi = \frac{1}{\Delta T} \frac{N_C}{N_{MD}}$$

λ2

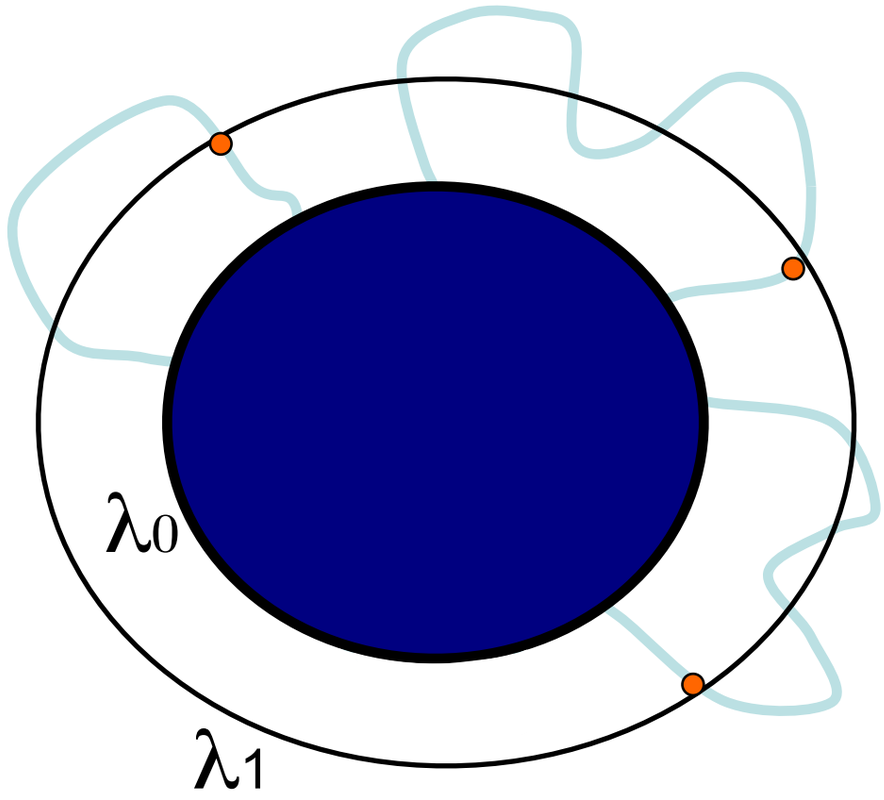
λ3



B

FFS

Store the effective first crossings (full phase points) with λ_1

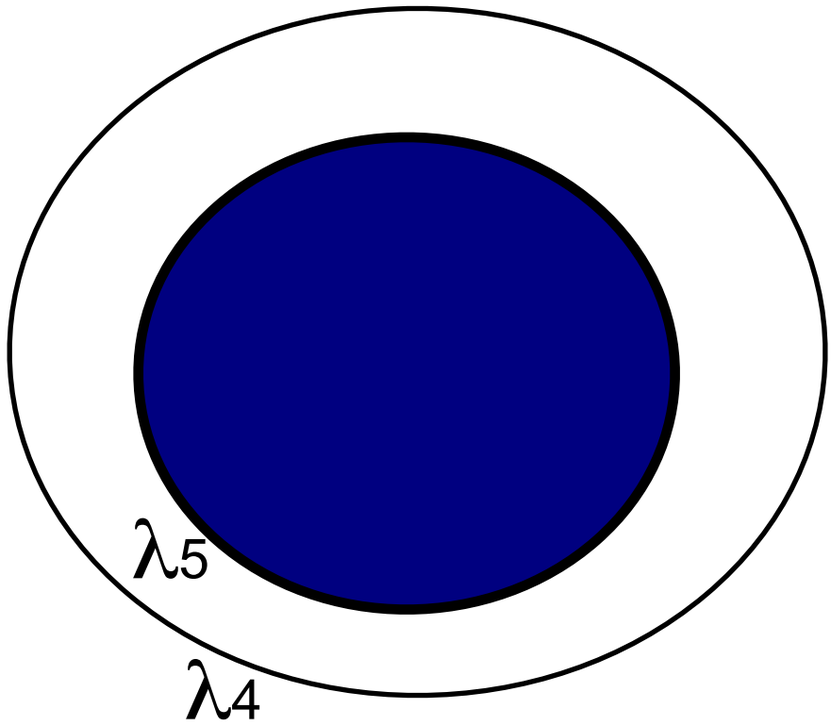


A

Randomly pick a point and release an MD trajectory from it until it hits λ_0 or λ_2

λ2

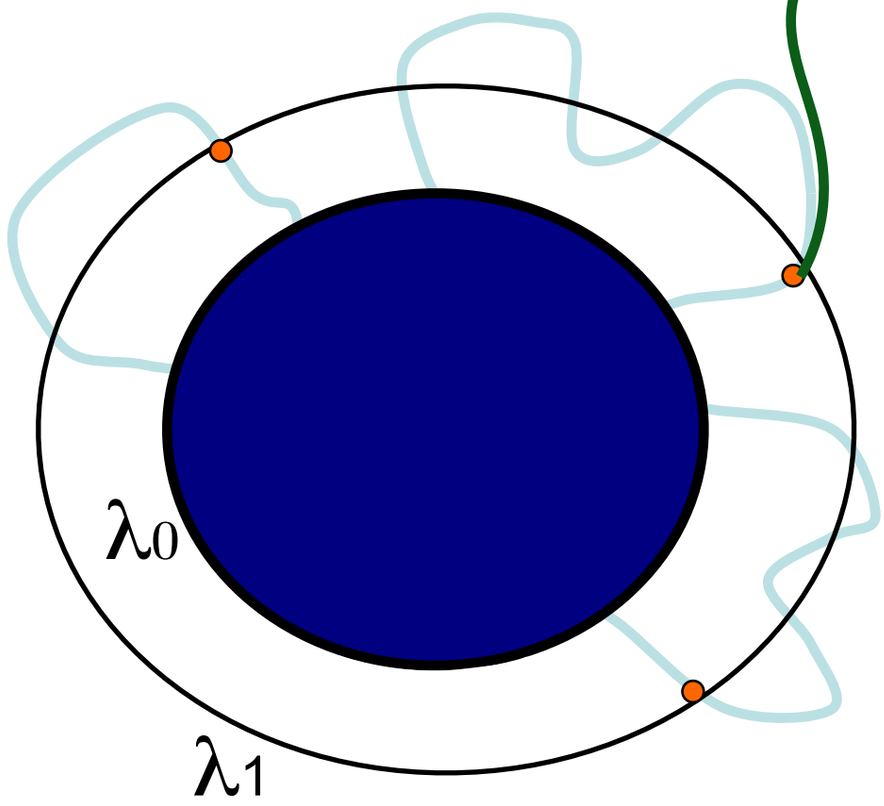
λ3



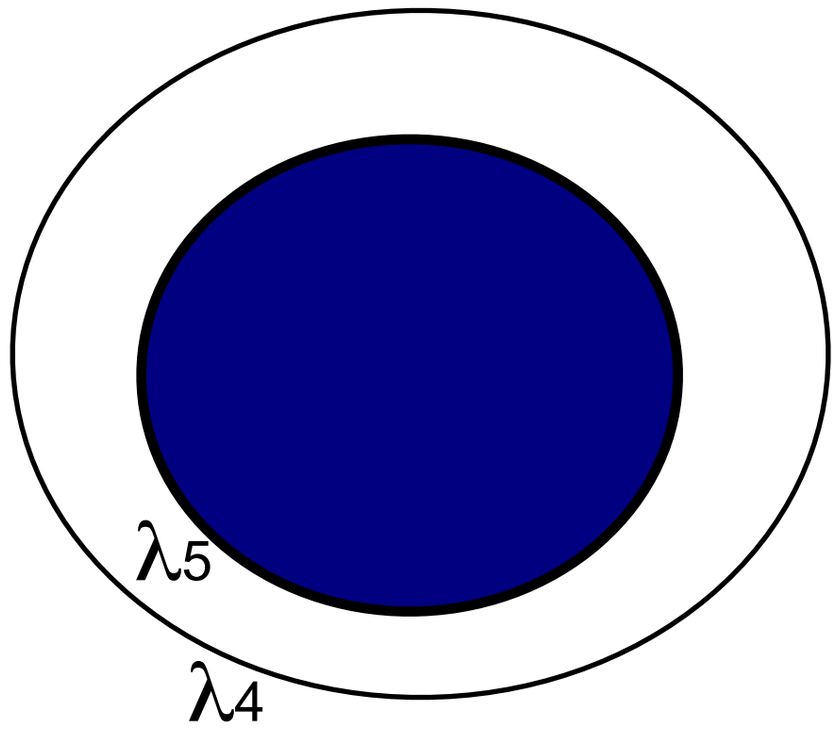
B

FFS

Store the effective first crossings (full phase points) with λ_1



A



B

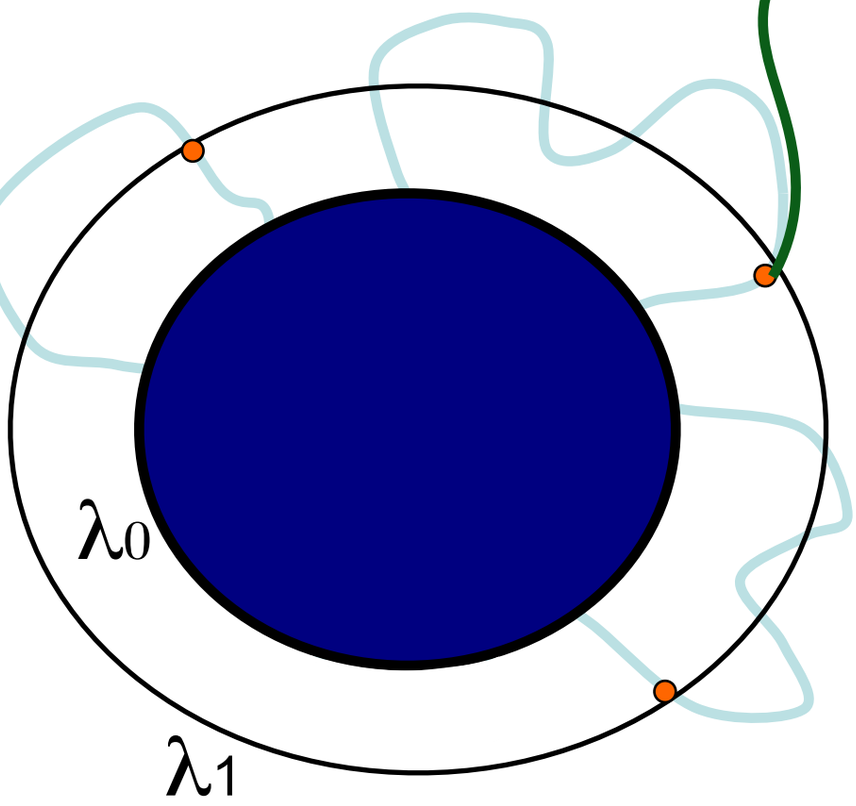
λ2

λ3

Randomly pick a point and release an MD trajectory from it until it hits λ_0 or λ_2

FFS

Store the effective first crossings (full phase points) with λ_1

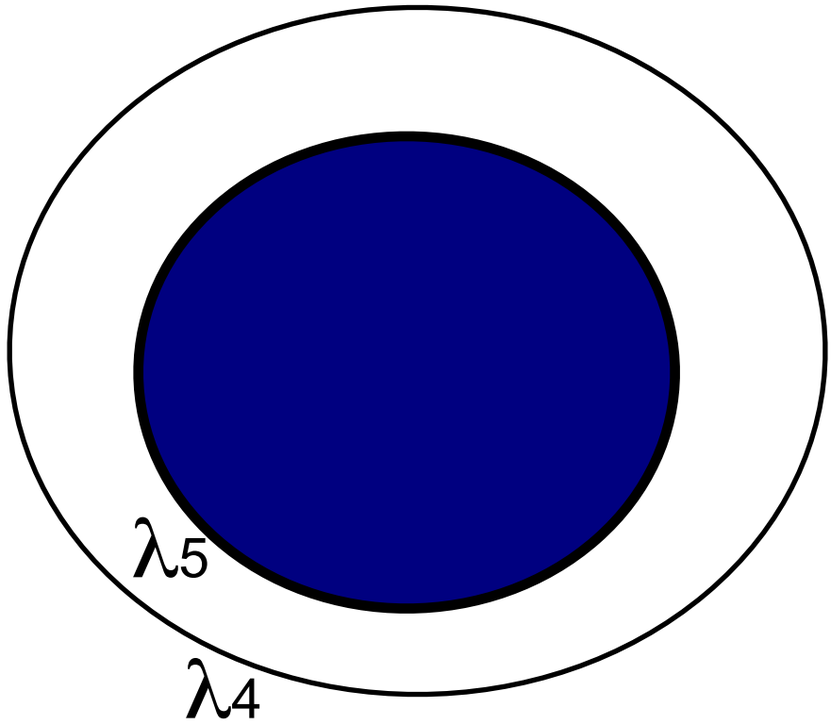


A

Repeat many times

λ2

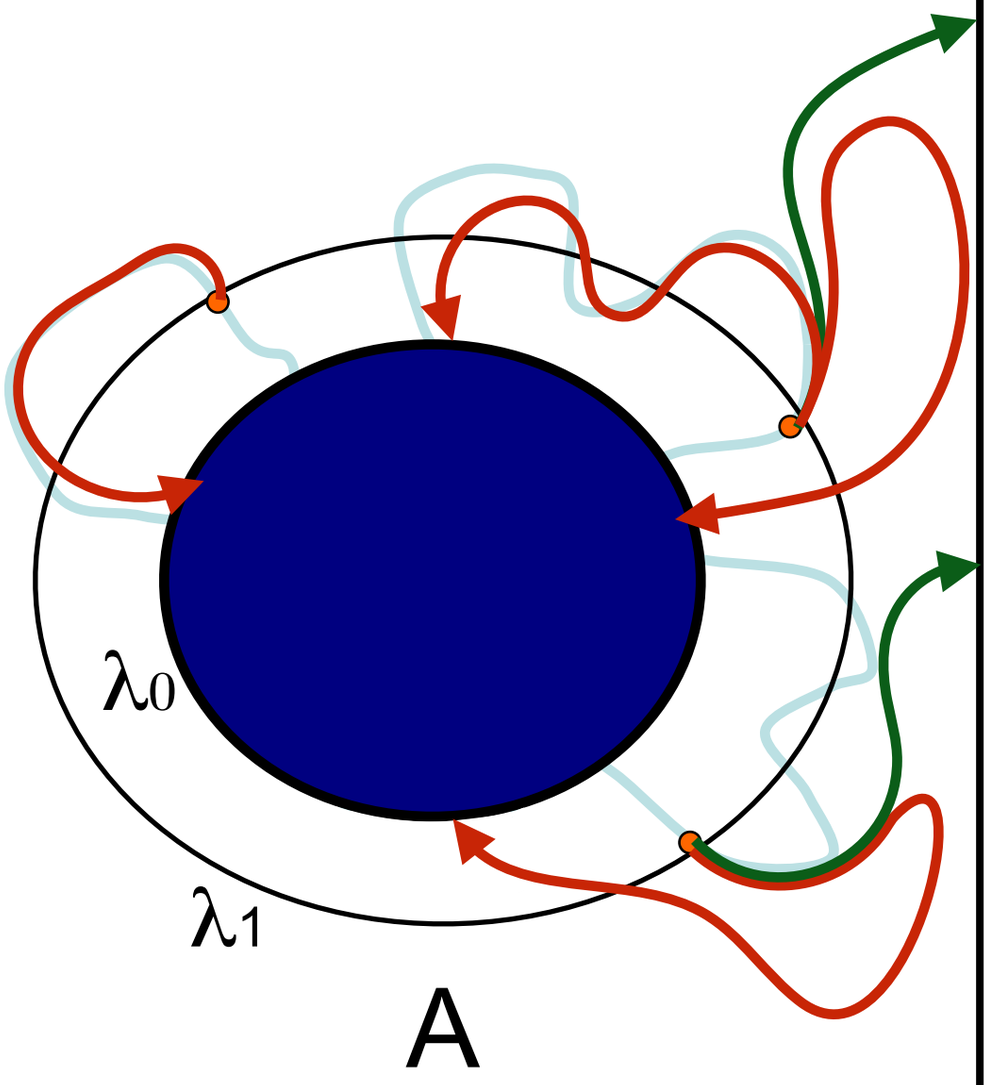
λ3



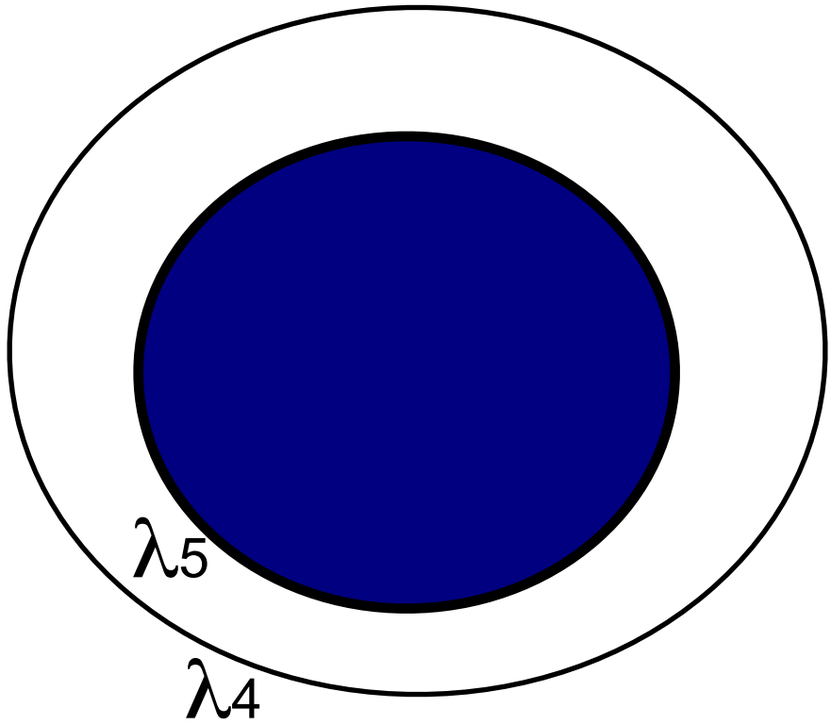
B

FFS

Store the effective first crossings (full phase points) with λ_1



A



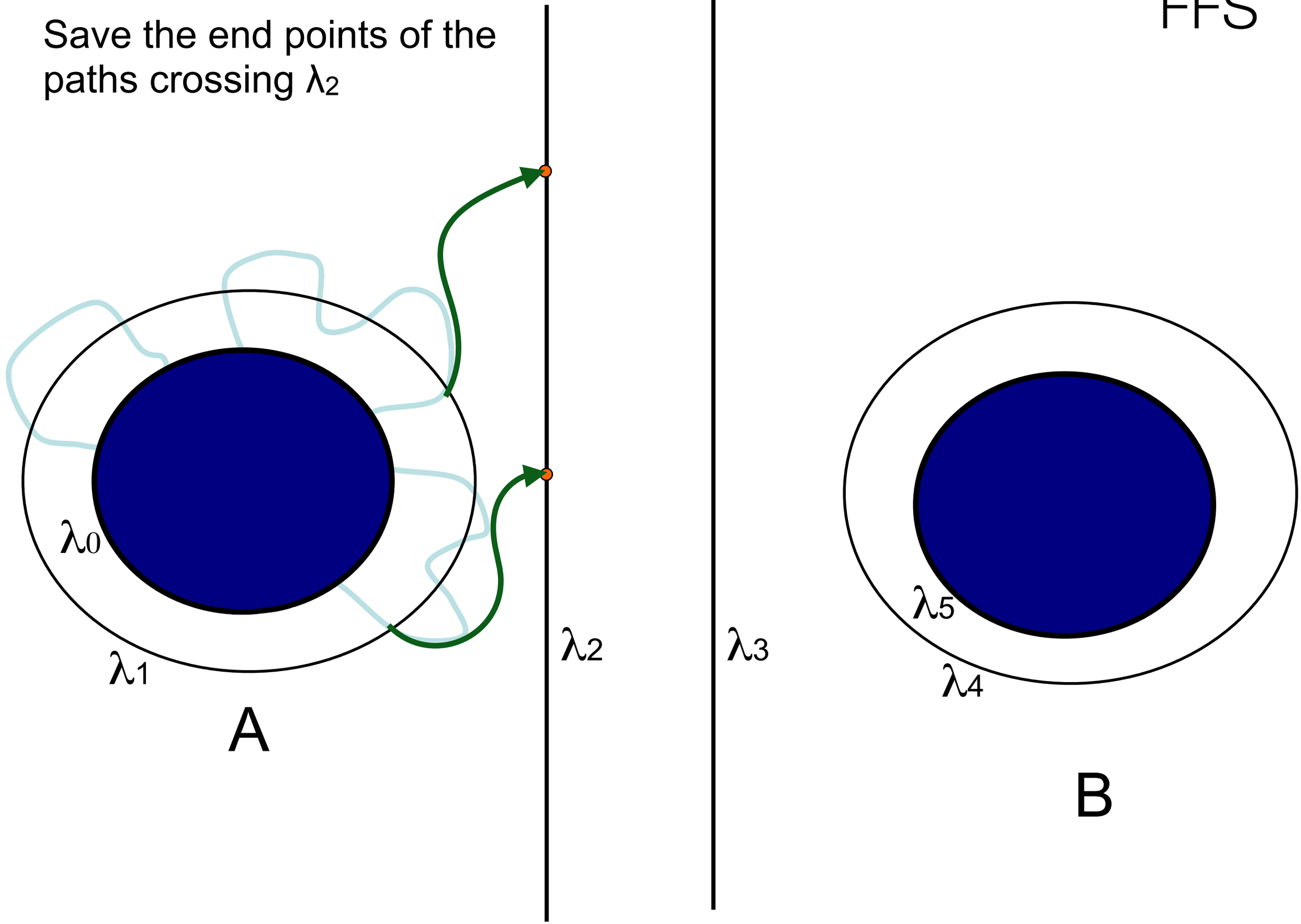
B

Repeat many times

$$P_A(\lambda_2|\lambda_1) = \frac{\#p_{12}}{\#p_{10}+\#p_{12}}$$

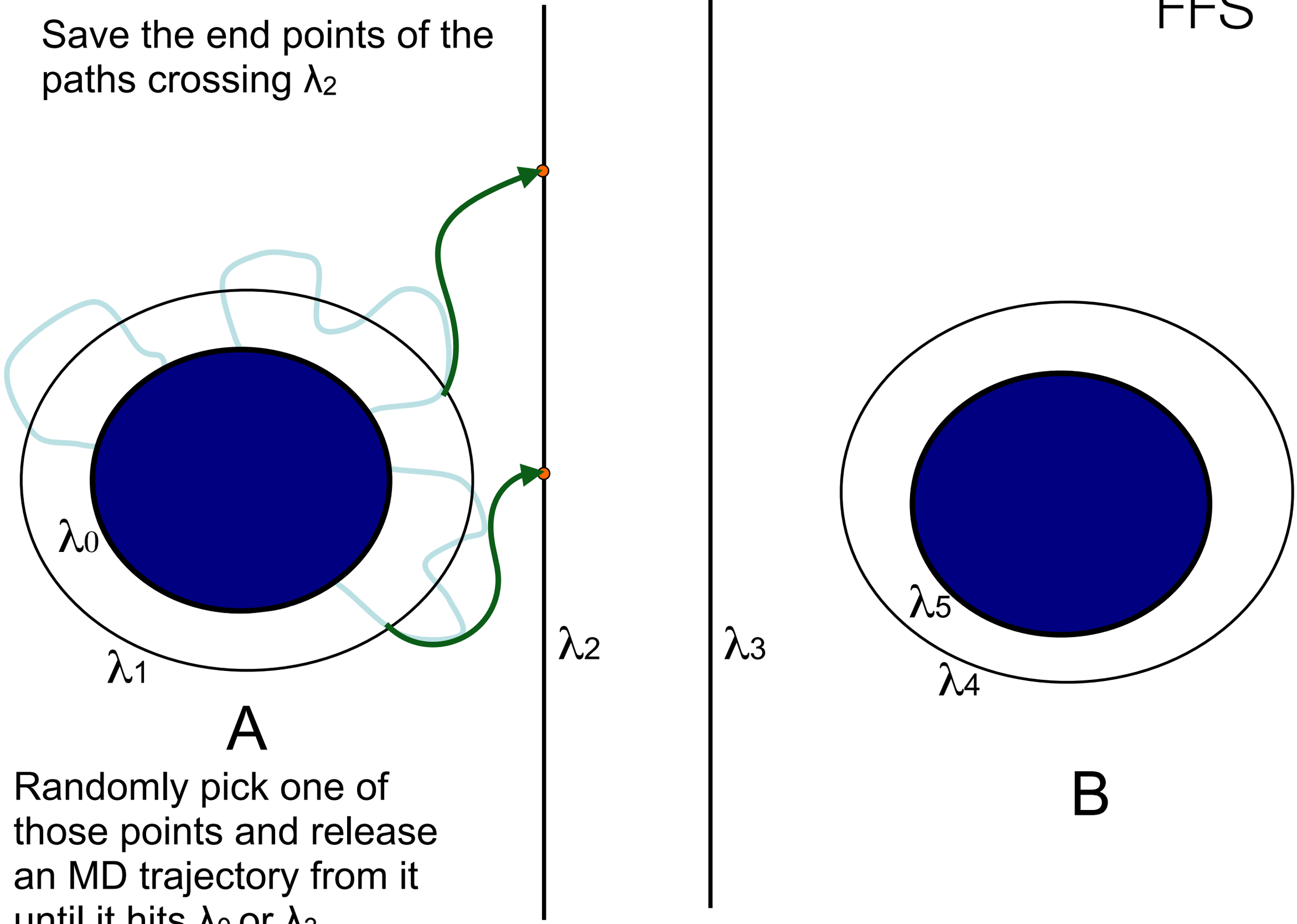
FFS

Save the end points of the paths crossing λ_2



FFS

Save the end points of the paths crossing λ_2



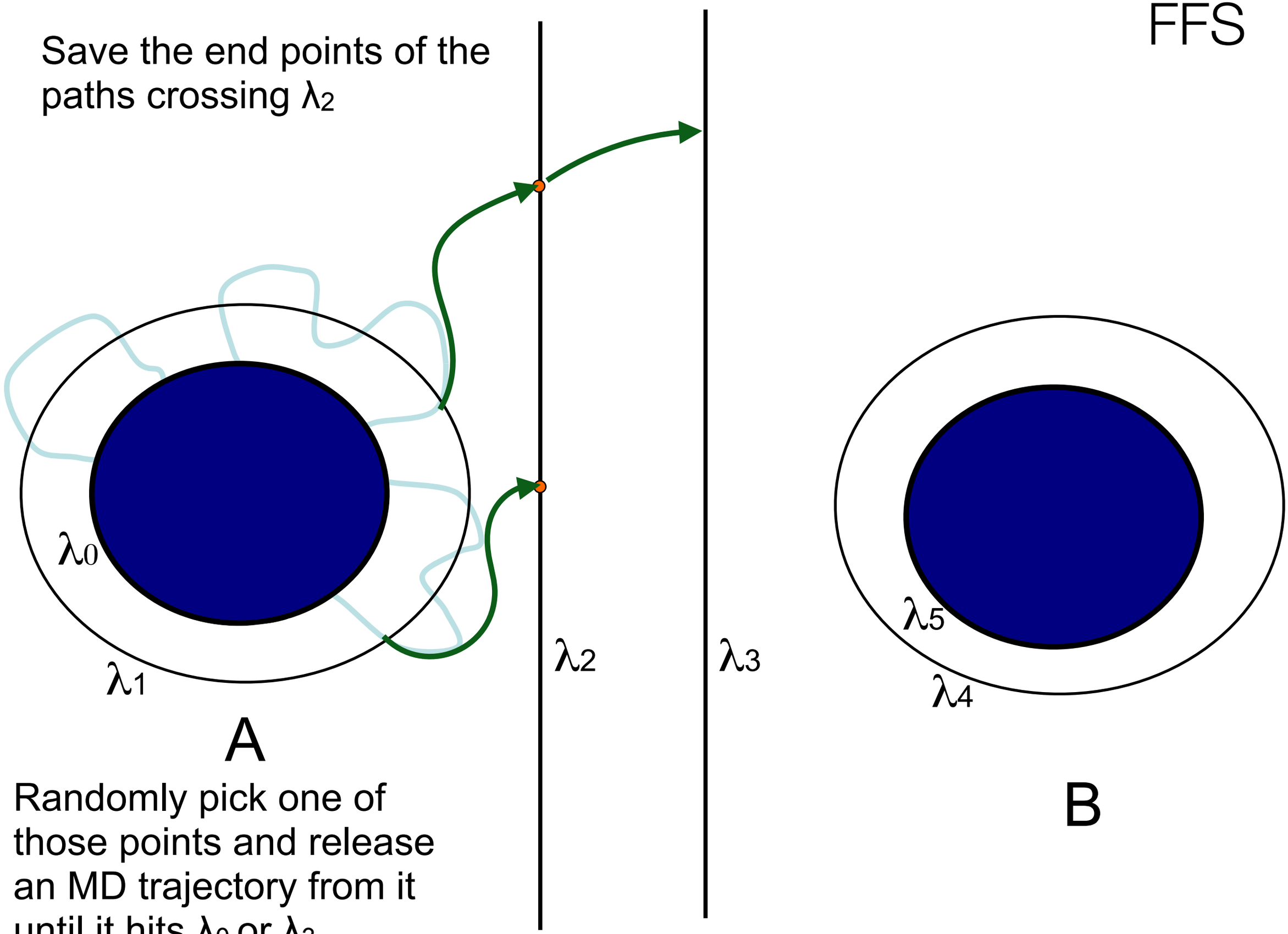
A

Randomly pick one of those points and release an MD trajectory from it until it hits λ_0 or λ_3

B

FFS

Save the end points of the paths crossing λ_2

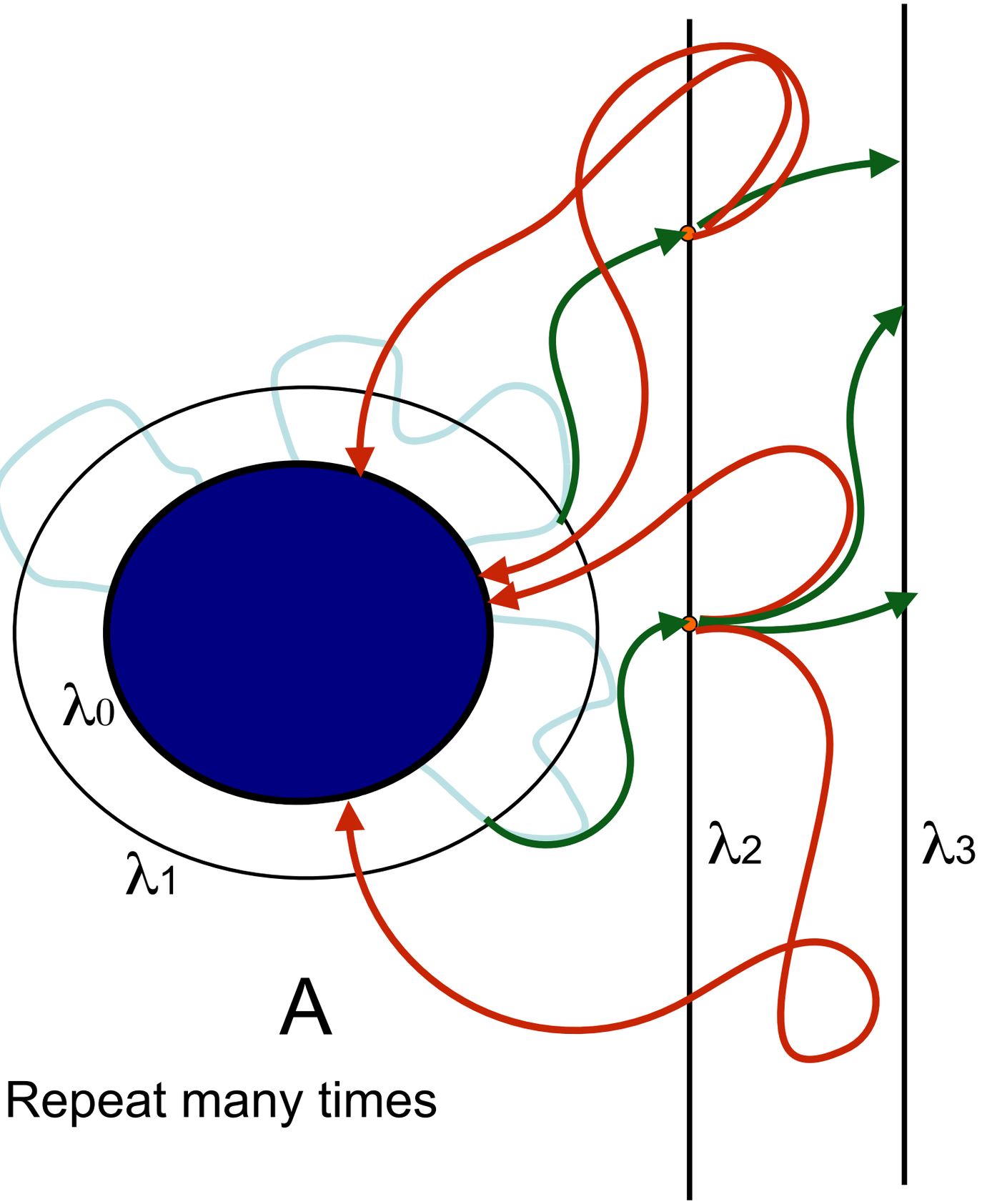


A

Randomly pick one of those points and release an MD trajectory from it until it hits λ_0 or λ_3

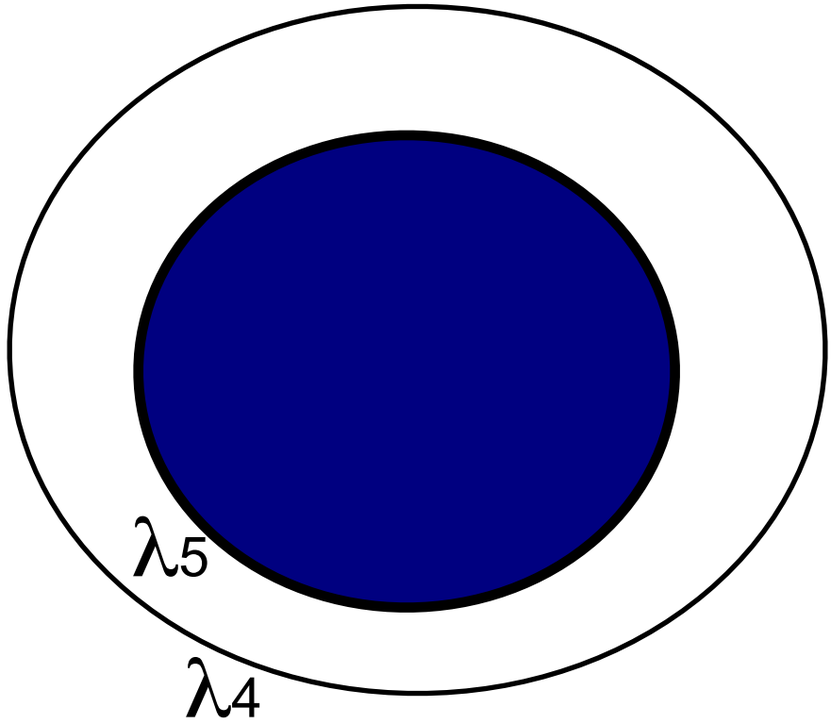
B

FFS



A

Repeat many times

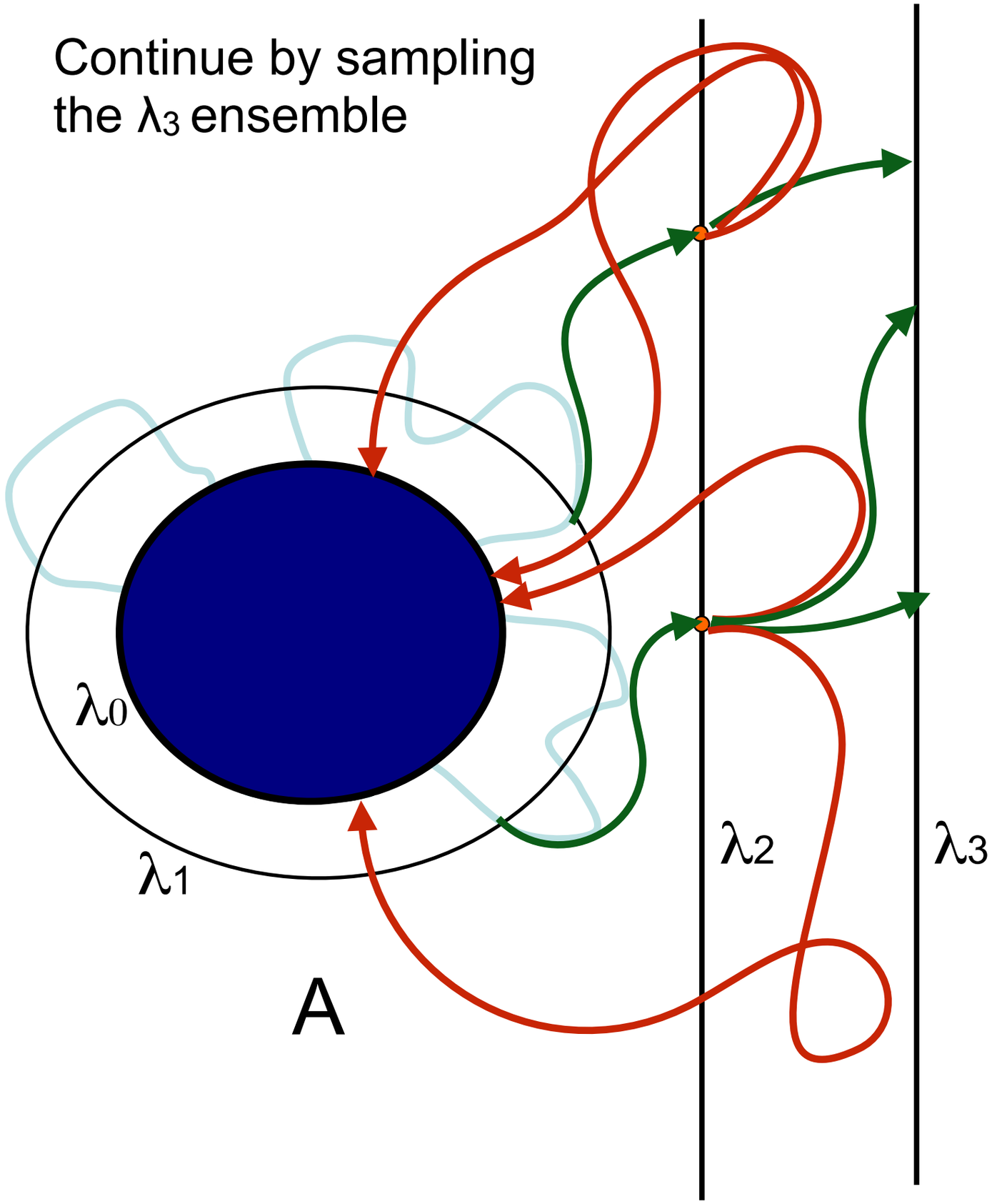


B

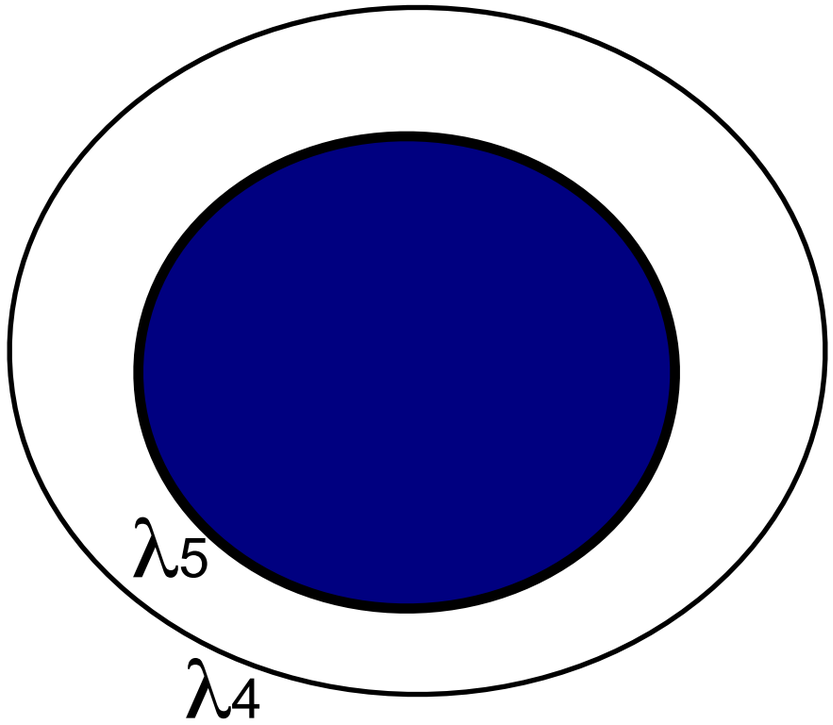
$$P_A(\lambda_3|\lambda_2) = \frac{\#p_{23}}{\#p_{20}+\#p_{23}}$$

FFS

Continue by sampling the λ_3 ensemble



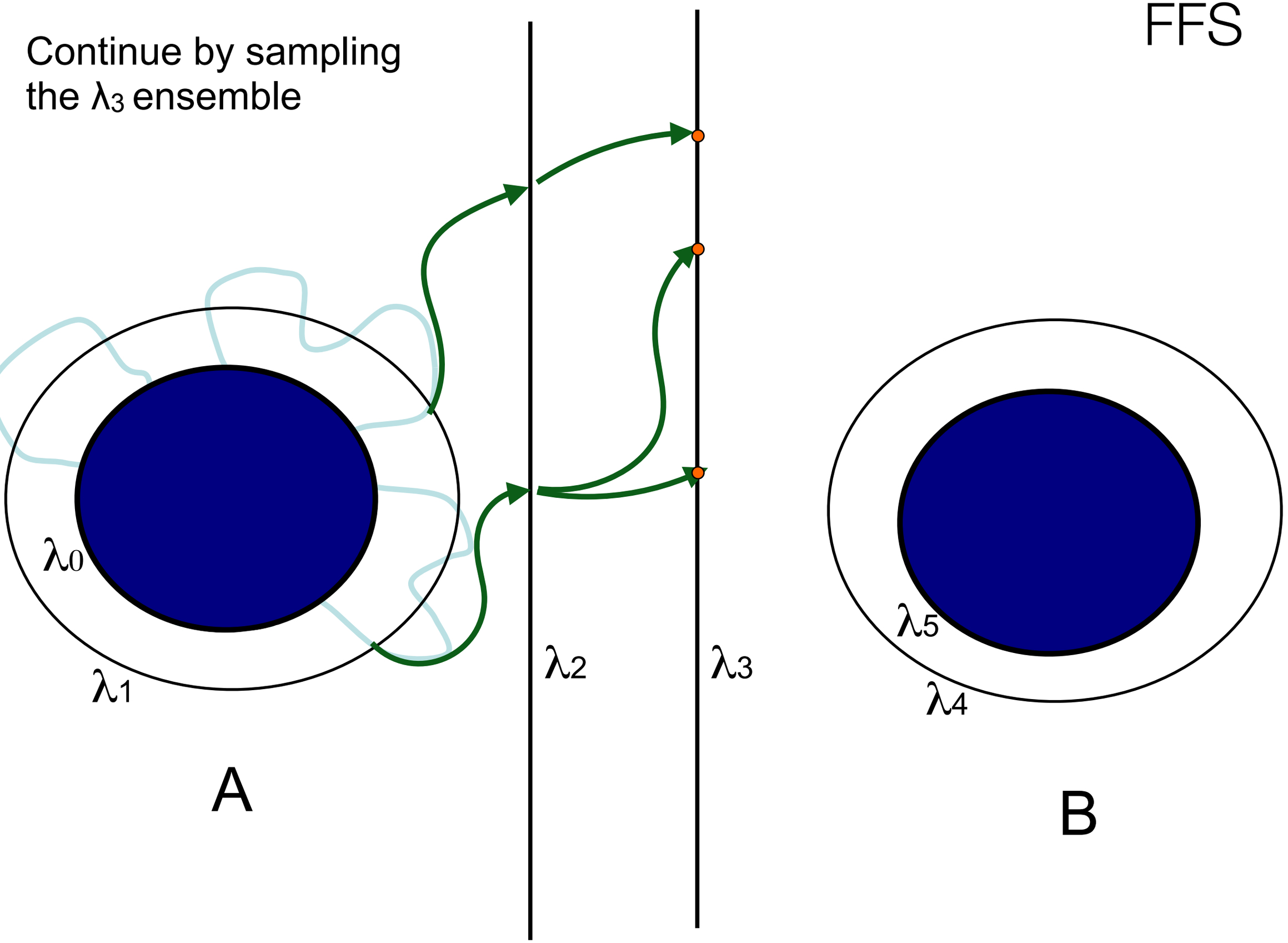
A



B

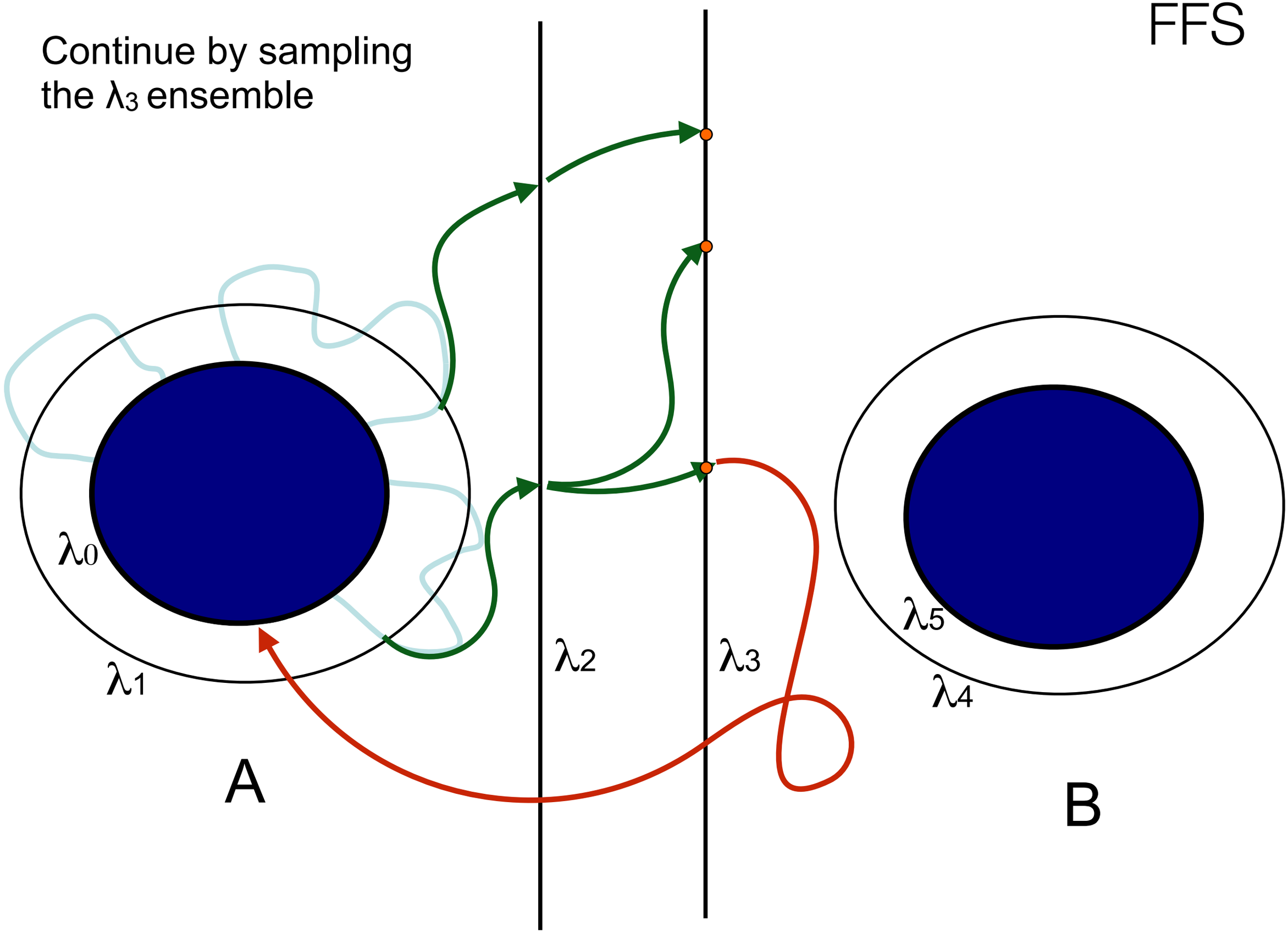
FFS

Continue by sampling the λ_3 ensemble



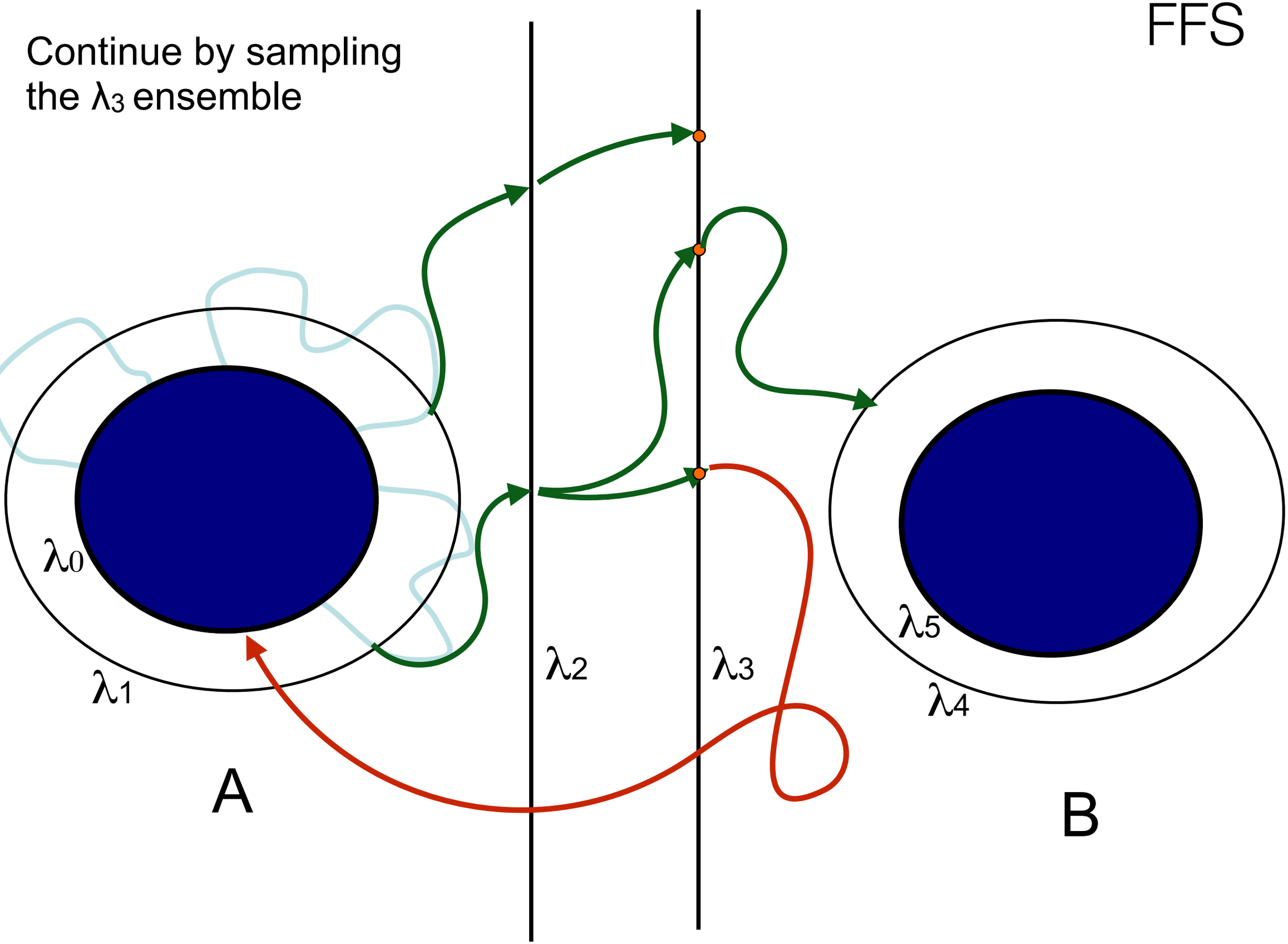
FFS

Continue by sampling the λ_3 ensemble



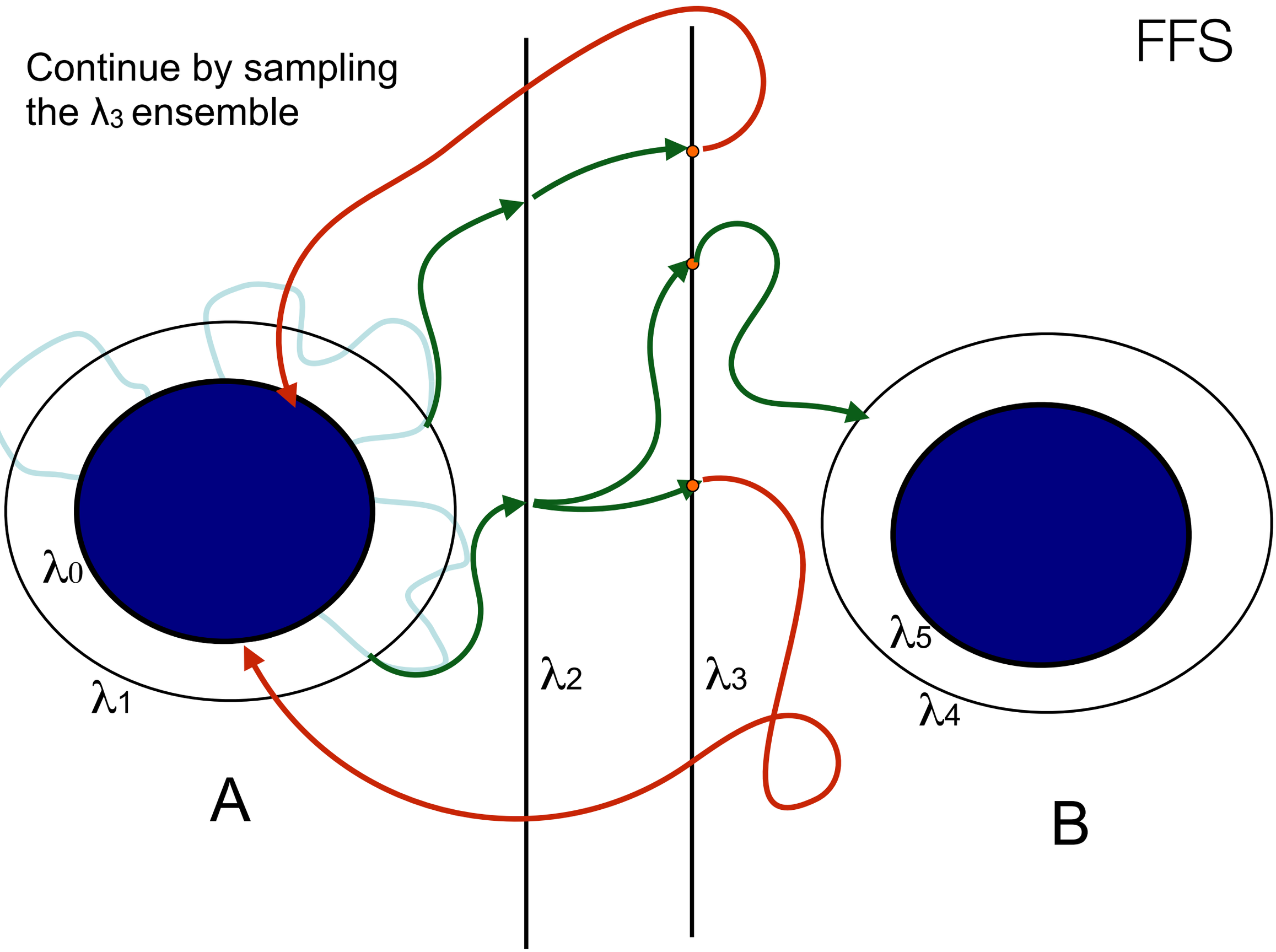
FFS

Continue by sampling the λ_3 ensemble



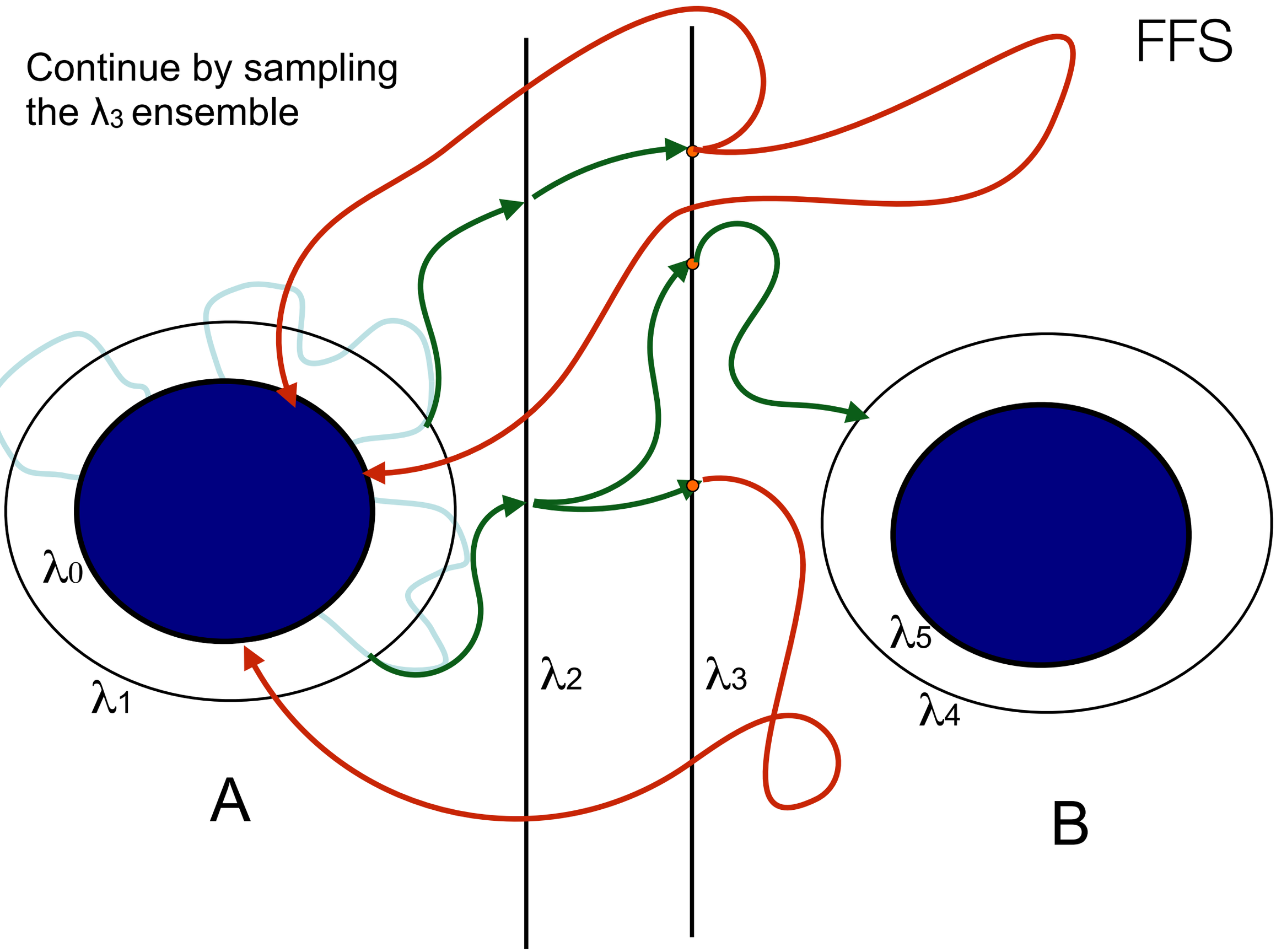
FFS

Continue by sampling the λ_3 ensemble



FFS

Continue by sampling the λ_3 ensemble



Continue by sampling
the λ_3 ensemble

FFS

λ_0

λ_1

λ_2

λ_3

λ_4

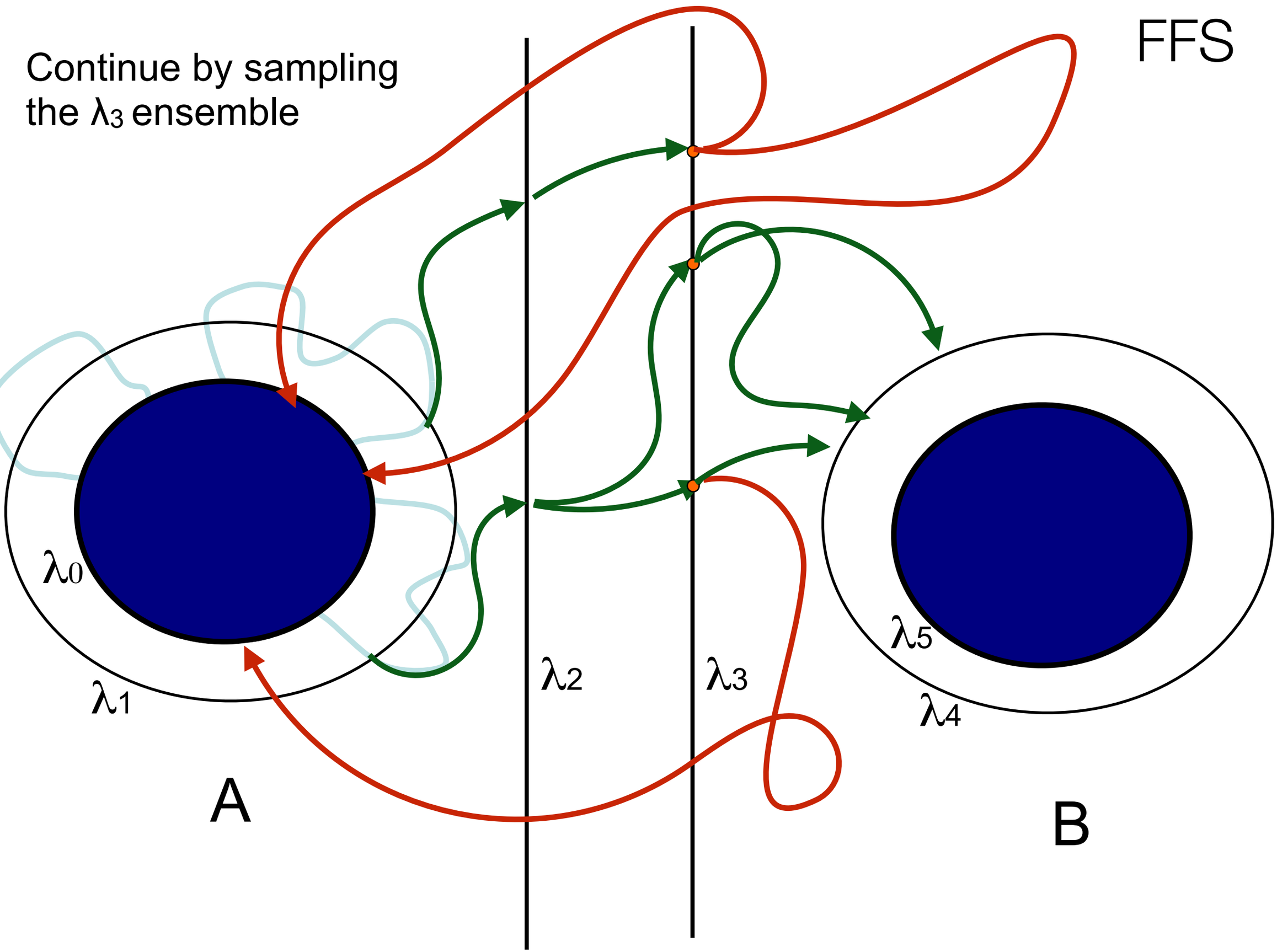
λ_5

A

B

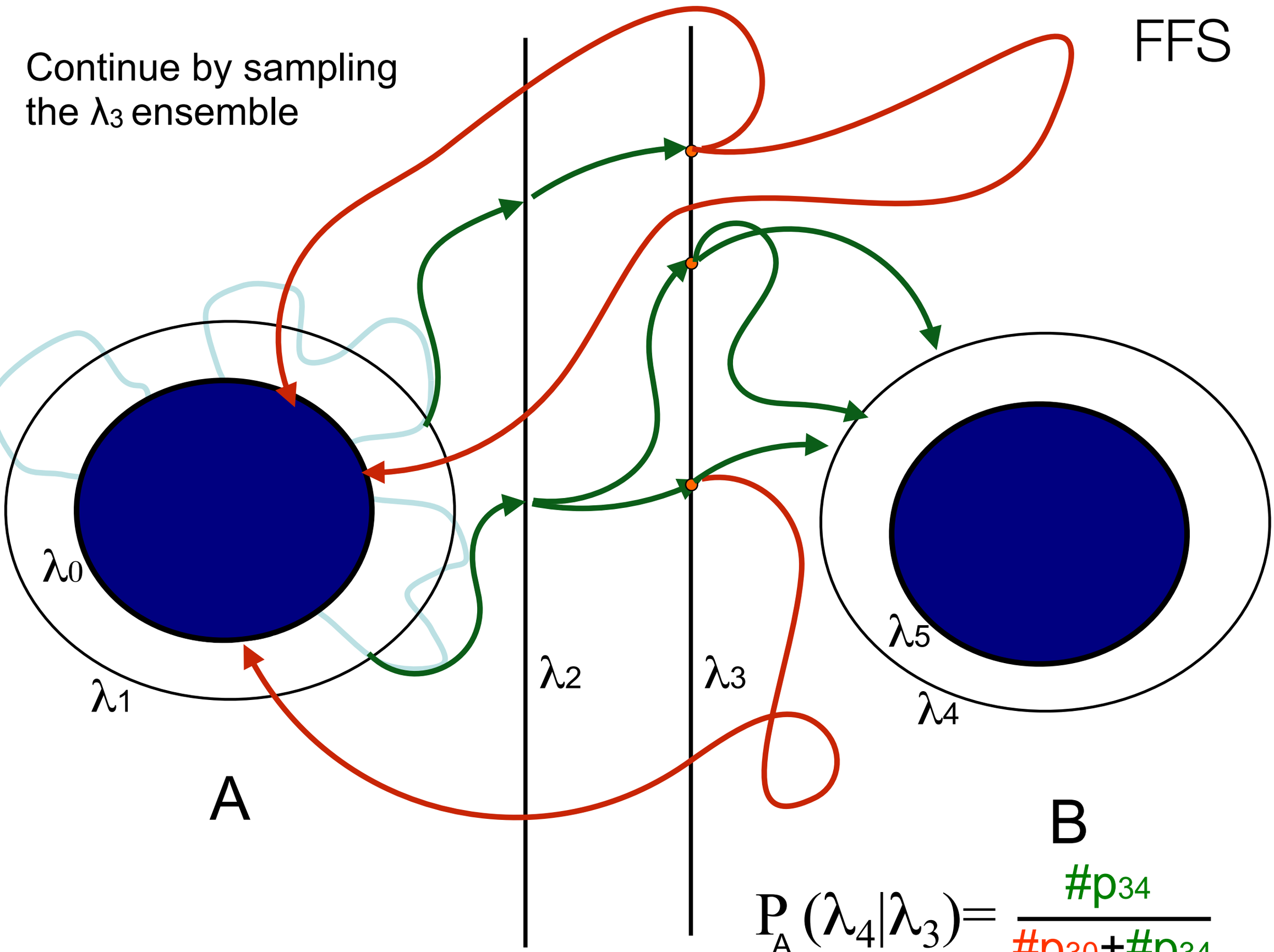
FFS

Continue by sampling the λ_3 ensemble



FFS

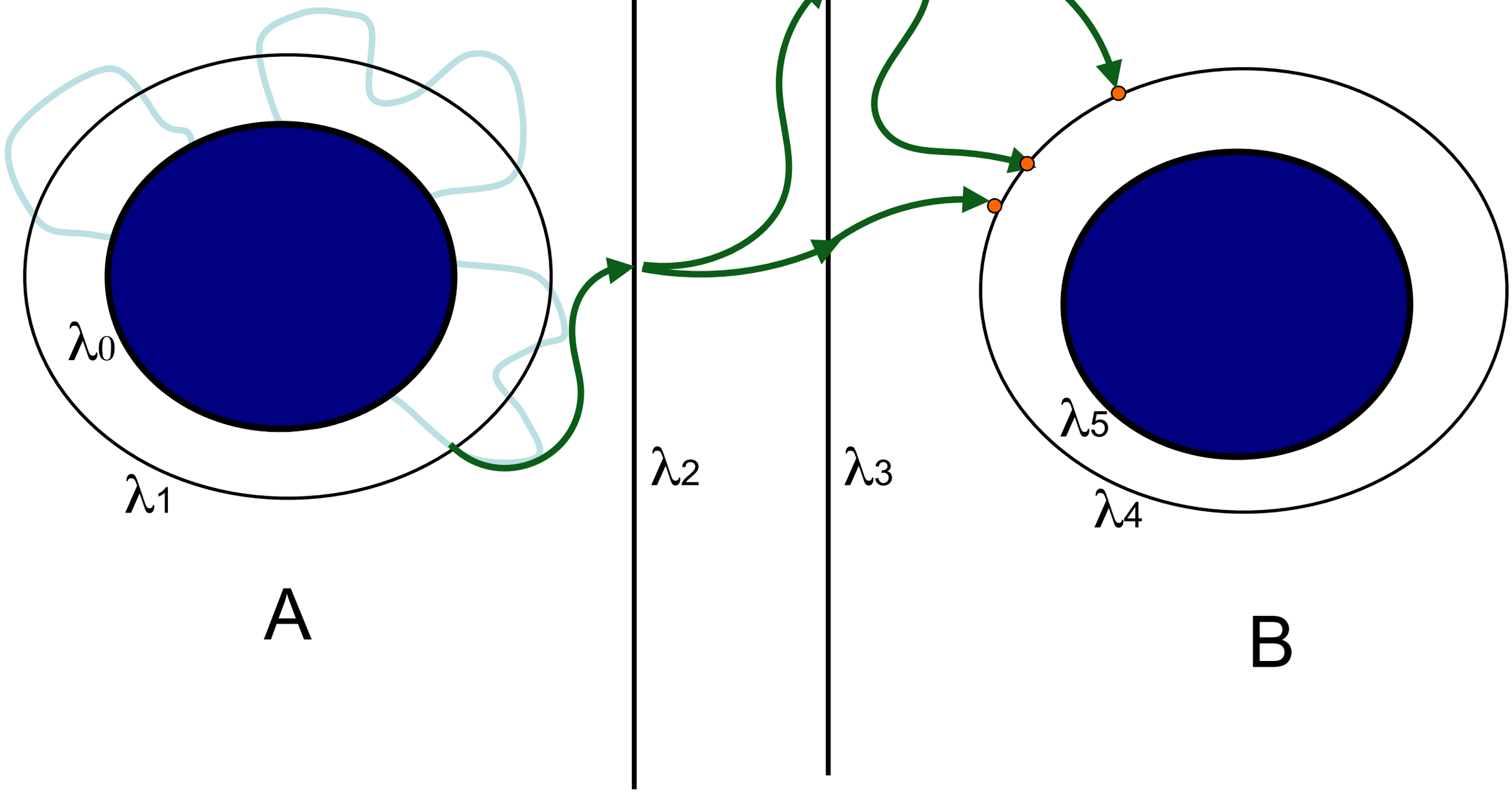
Continue by sampling the λ_3 ensemble



$$P_A(\lambda_4|\lambda_3) = \frac{\#p_{34}}{\#p_{30}+\#p_{34}}$$

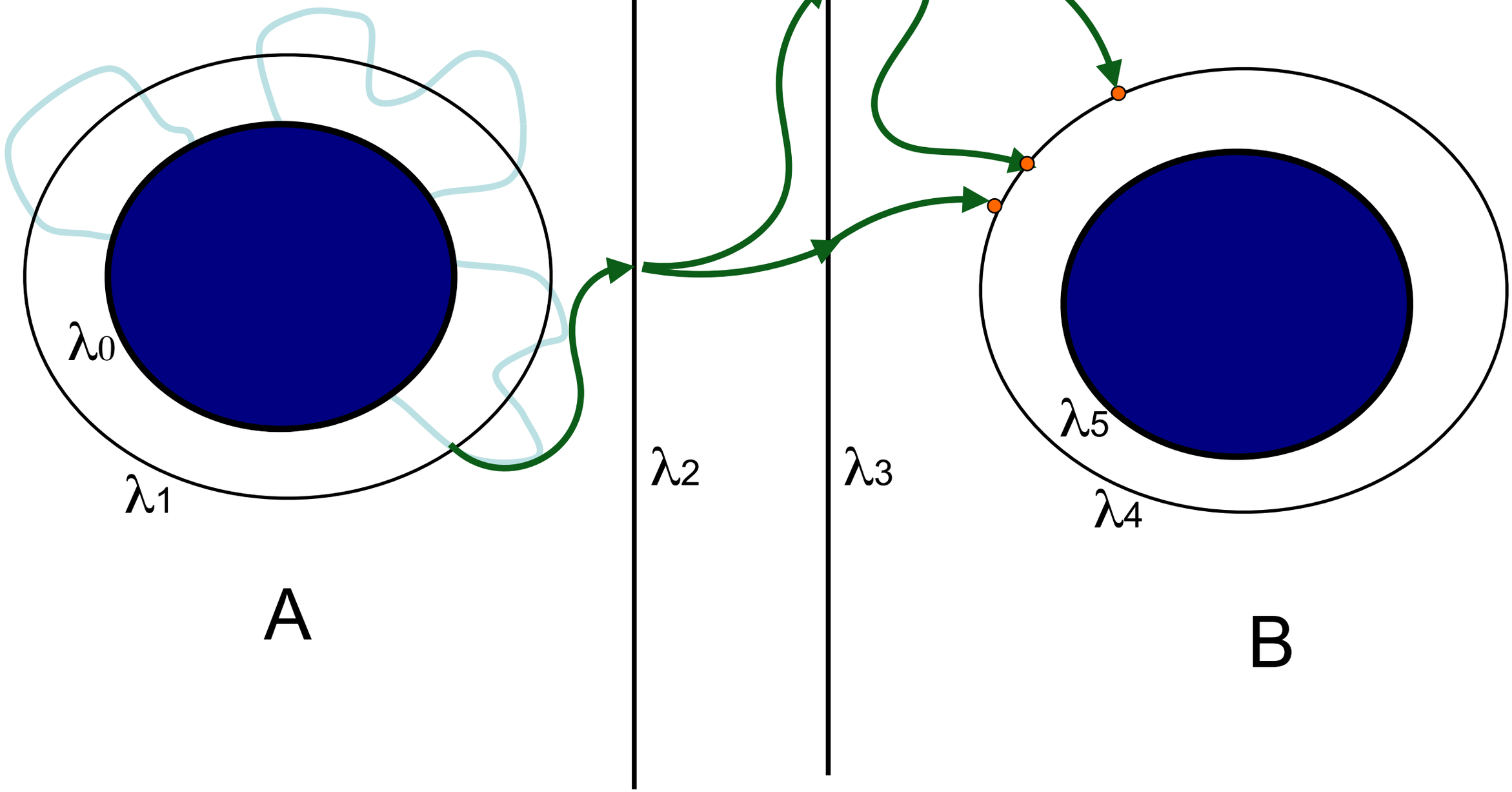
Continue by sampling
the λ4 ensemble

FFS



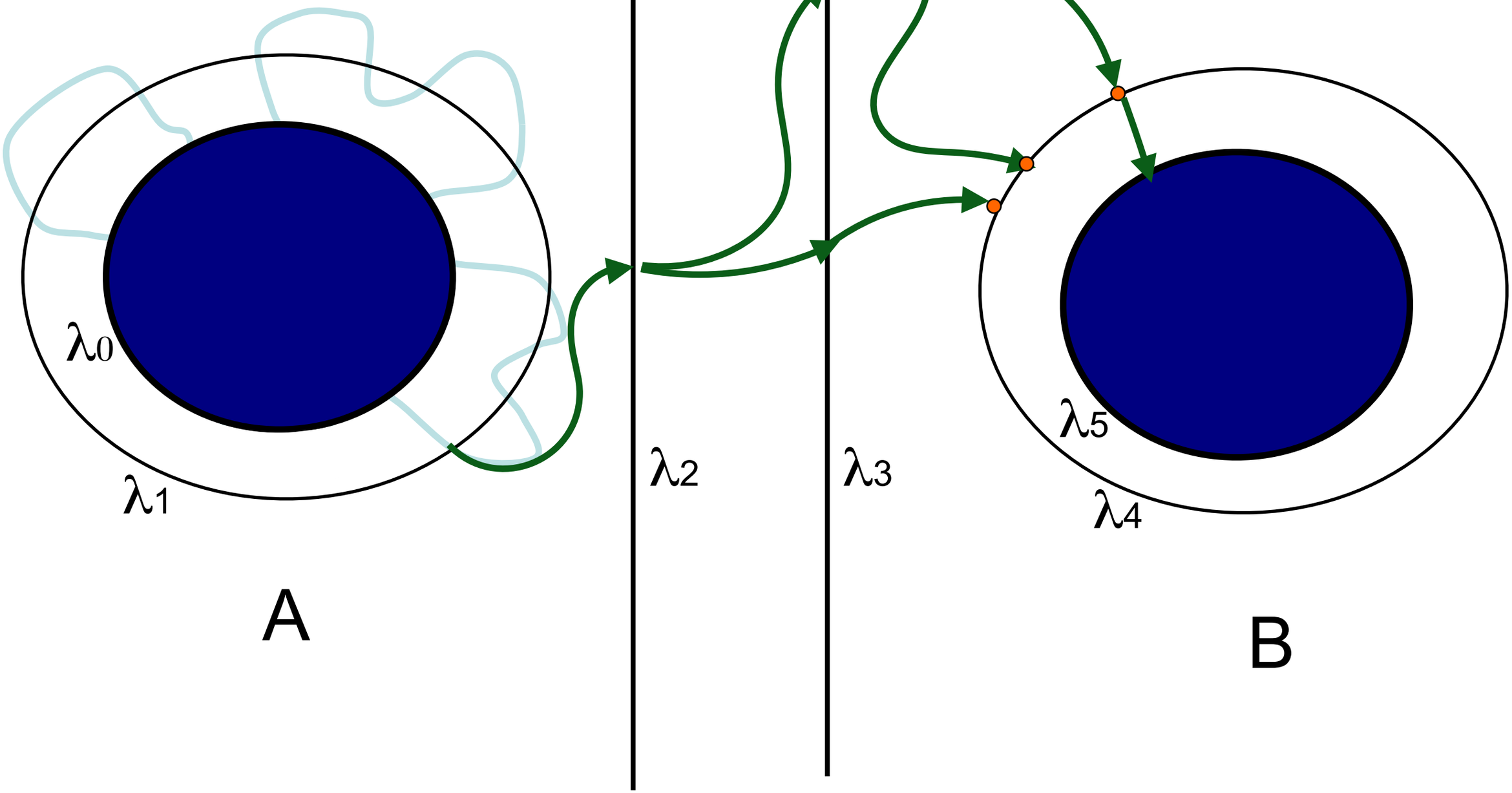
Continue by sampling
the λ4 ensemble

FFS



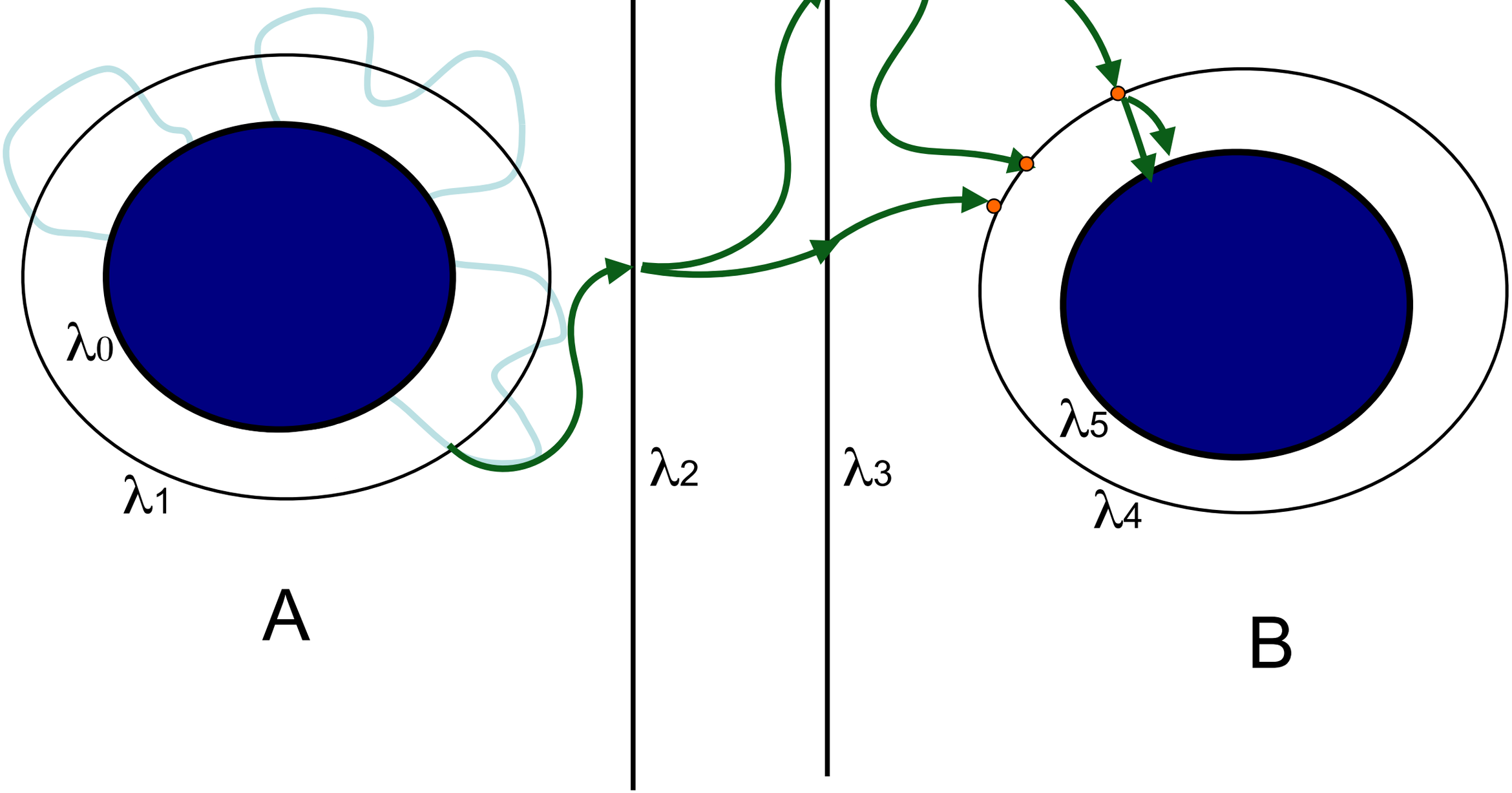
Continue by sampling
the λ4 ensemble

FFS



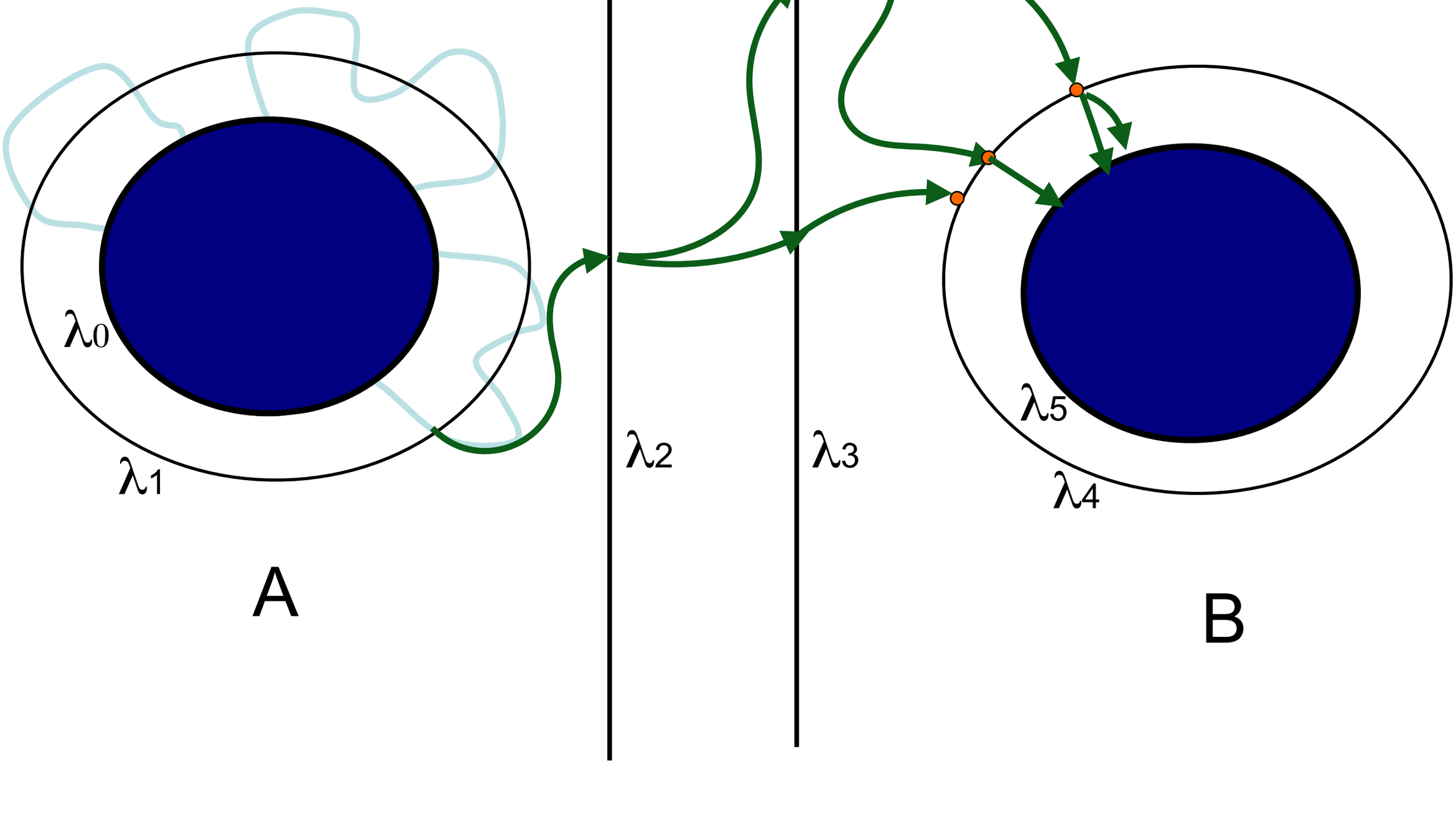
Continue by sampling
the λ4 ensemble

FFS



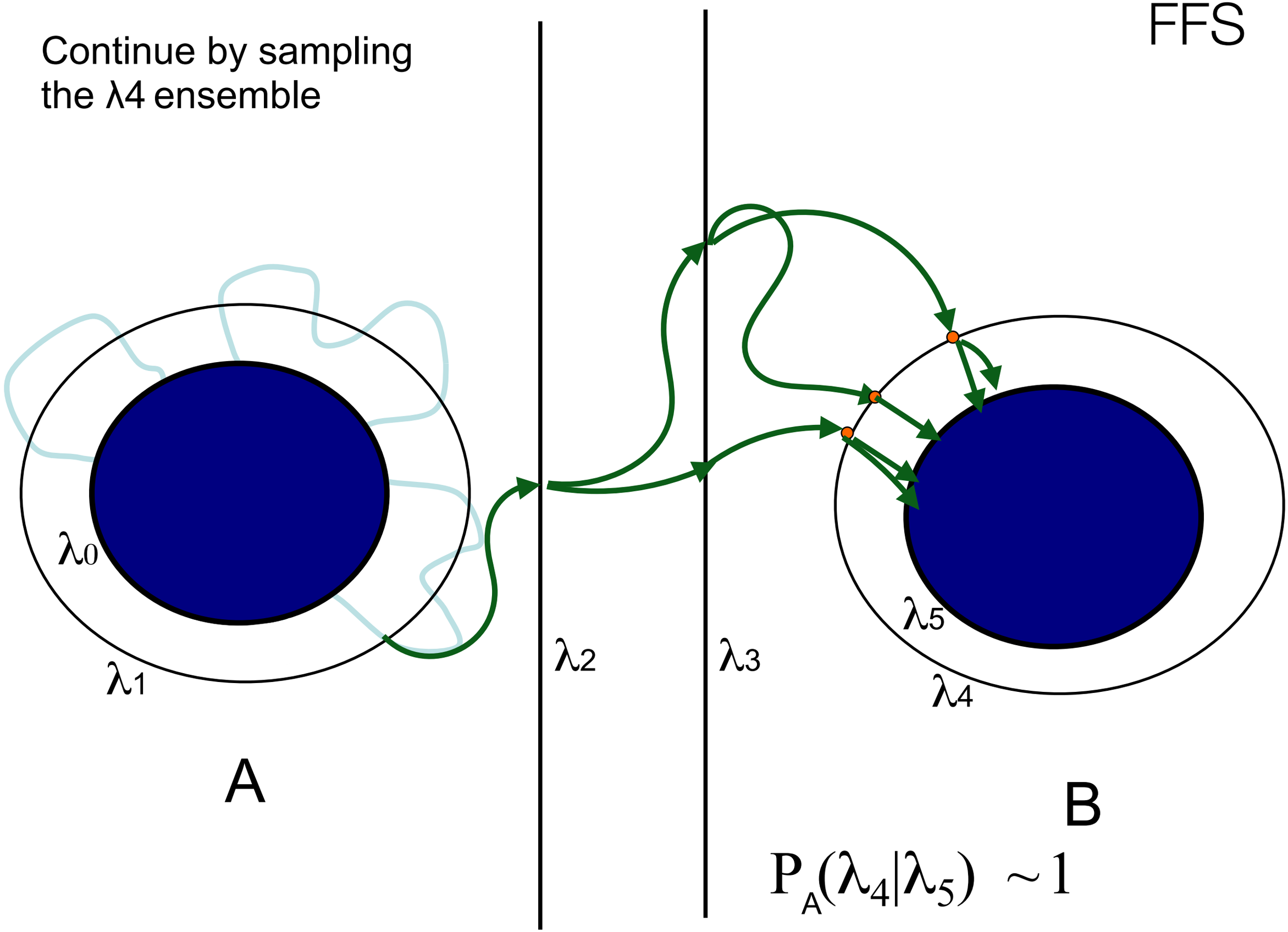
Continue by sampling
the λ4 ensemble

FFS



FFS

Continue by sampling the λ4 ensemble



$$P_A(\lambda_4|\lambda_5) \sim 1$$

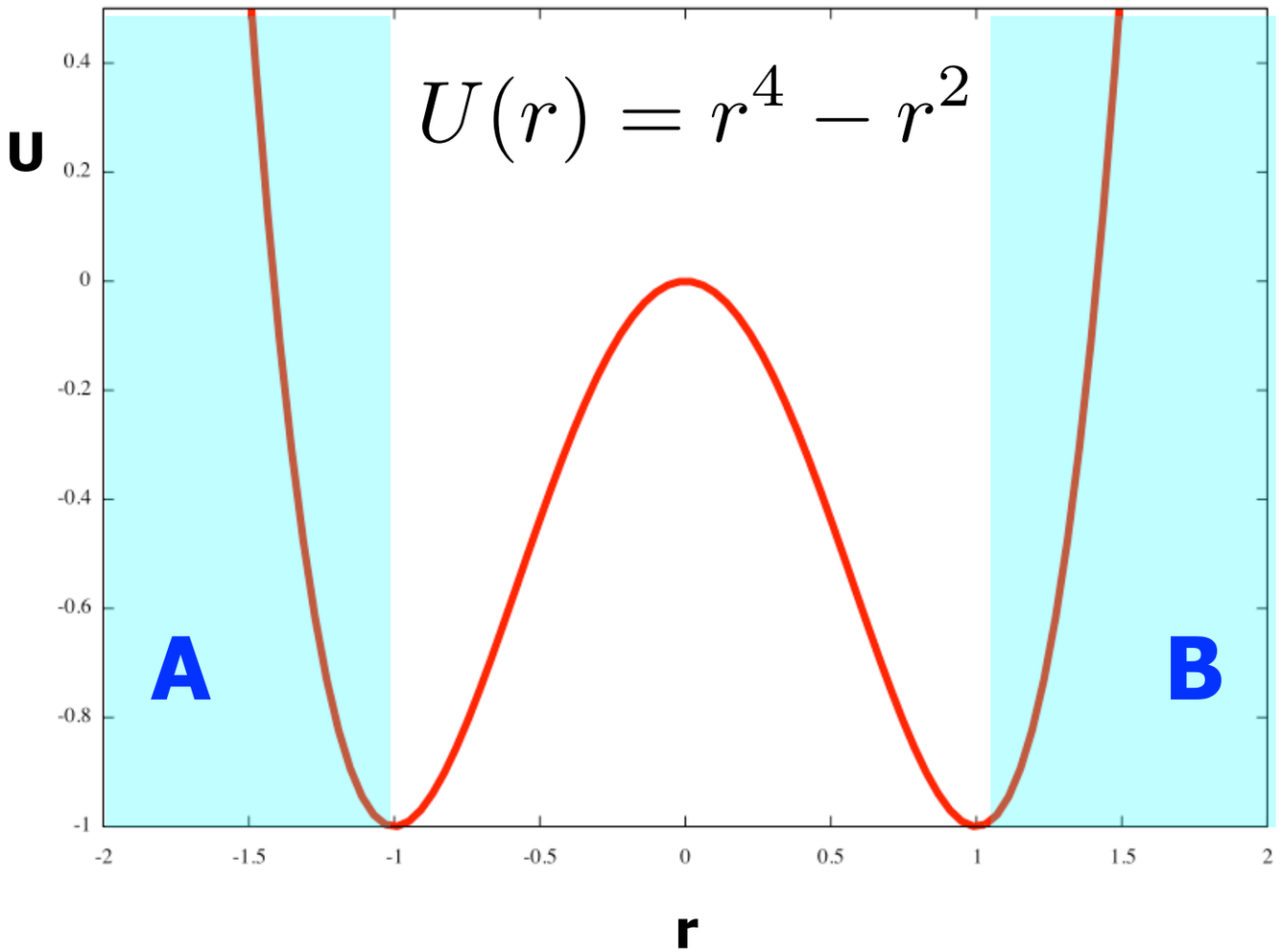
Pros and cons of Forward Flux Sampling:

- Many trajectories relative to the number of MD steps
- But many of these trajectories are correlated
- Can be used for non-equilibrium dynamics for which the phase-space density is unknown

- Can’t be use with deterministic dynamics
- Exact like TIS/RETIS
- Efficiency of FFS is even more than the Reactive Flux method affected by a bad choice of RC (e.g. hysteresis) and configuration space might not be sufficient for defining a functioning RC
- Results might look very well converged even if they are not!

Titus van Erp

Failure of FFS on a 1D example!



Langevin Dynamics

$$k_B = m = 1$$

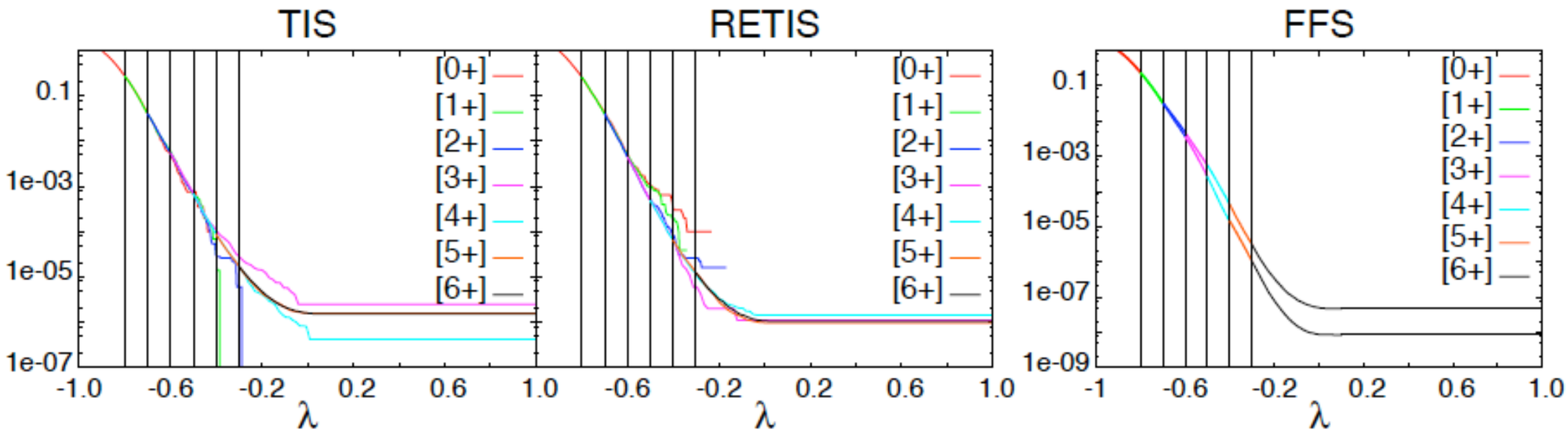
$$\gamma = 0.3,\ T = 0.7$$

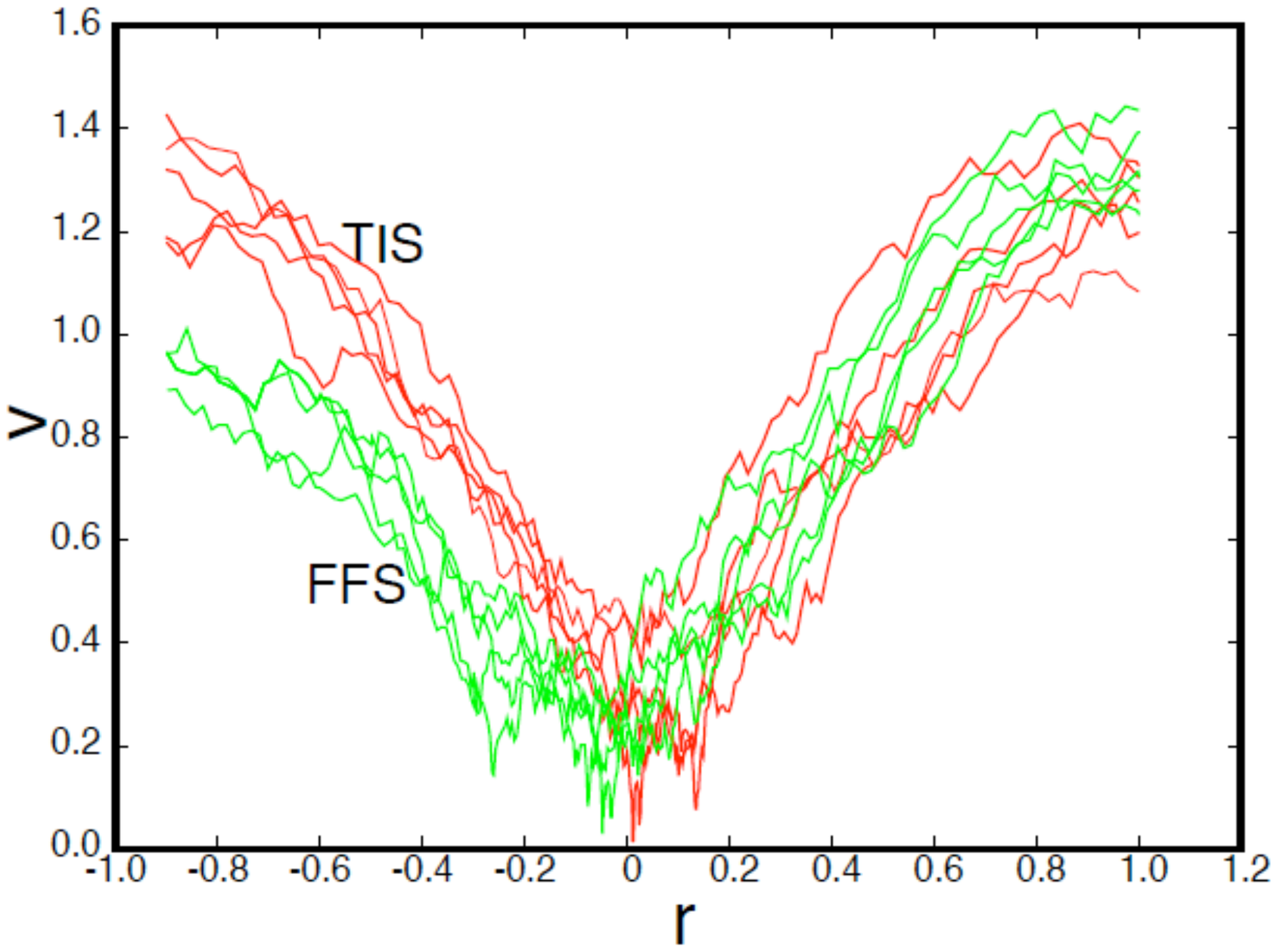
8 interfaces

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

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

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	$0.263 \pm 1\%$	$1.04 \cdot 10^{-6} \pm 19\%$	$2.73 \cdot 10^{-7} \pm 19\%$
RETIS	$0.265 \pm 1\%^*$	$1.05 \cdot 10^{-6} \pm 25\%^*$	$2.79 \cdot 10^{-7} \pm 25\%^*$
FFS (long MD run)	$0.263 \pm 1\%$	$4.69 \cdot 10^{-8} \pm 6\%^*$	$1.23 \cdot 10^{-8} \pm 6\%^*$
FFS (short MD run)	$0.259 \pm 2\%$	$8.45 \cdot 10^{-9} \pm 9\%^*$	$2.18 \cdot 10^{-9} \pm 9\%^*$





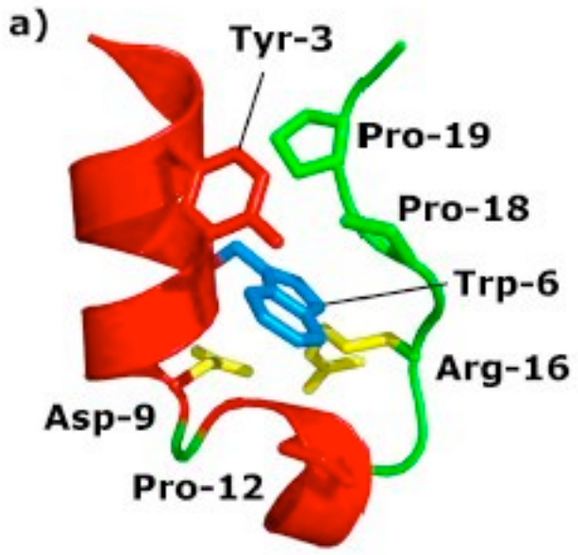
FFS have asymmetric velocity profile on the symmetric barrier!

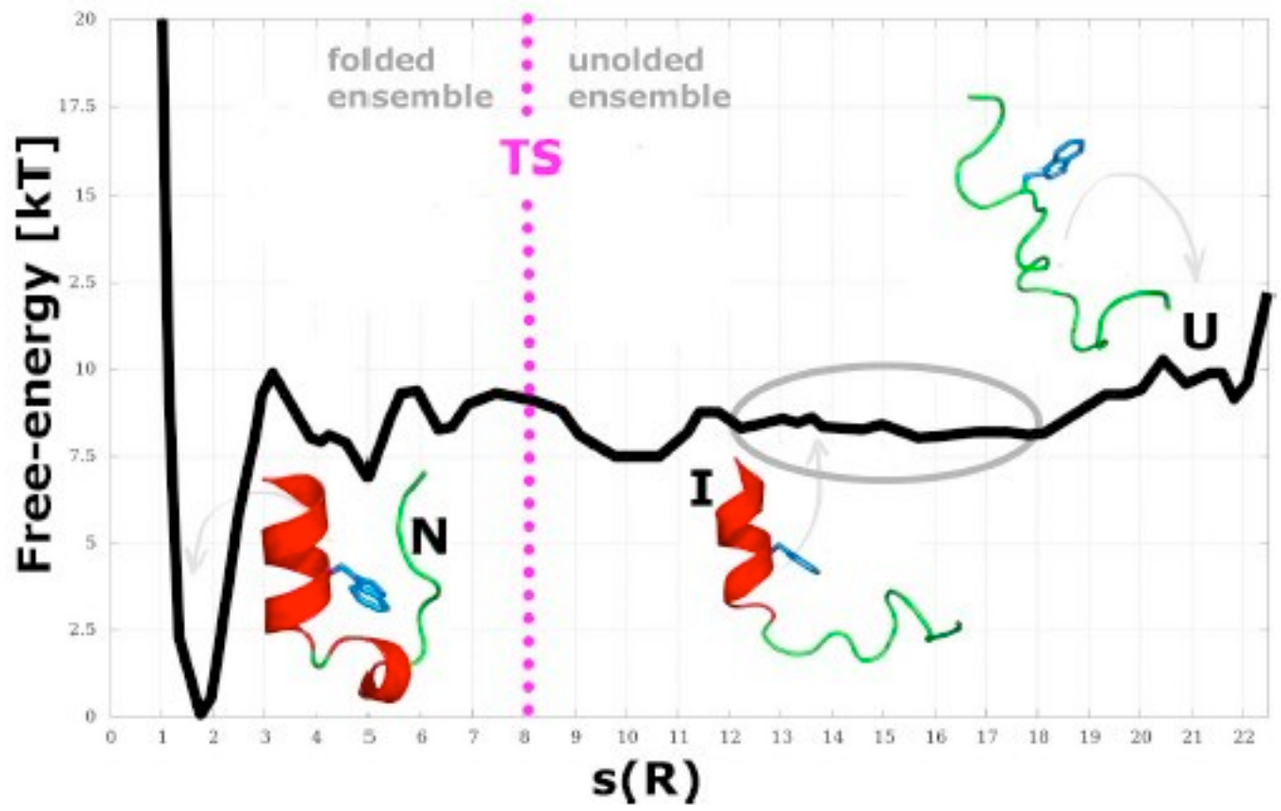
NTNU

Norwegian University of Science and Technology

Transition State PPTIS

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





Applied to the Trp-cage miniprotein folding

PPTIS can be used in combination with free energy calculations in order to calculate small transmission coefficients on a diffusive barrier

NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

Application:
Ab Initio autoionization of water

Analysis:
Method to analyse reaction mechanism:
The predictive Power Method

NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

Application: Ab Initio autoionization of water

REPORTS

Autoionization in Liquid Water

Phillip L. Geissler,[1] Christoph Dellago,[1,2] David Chandler,[1*]
Jürg Hutter,[3†] Michele Parrinello[3]

Science AAAS

2001

Destabilization of ions is due rare electric field fluctuations which arises primarily from long-range electrostatic interactions. **Local** properties such as ion coordination number and the presence of specific hydrogen bonds, **fail** to account for the bond-destabilizing fluctuation in our simulations.

On the recombination of hydronium and hydroxide ions in water

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

PNAS

2011

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

Titus S. van Erp

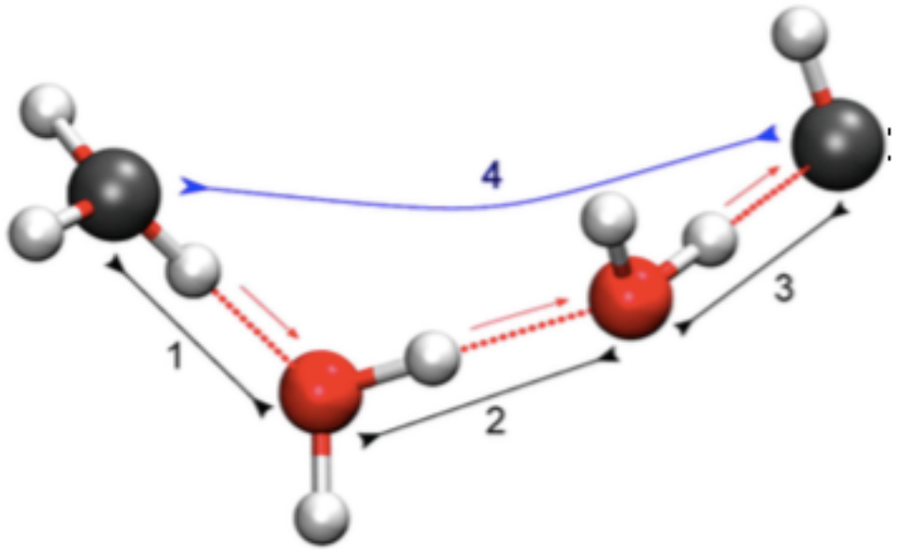
Application: Ab Initio autoionization of water

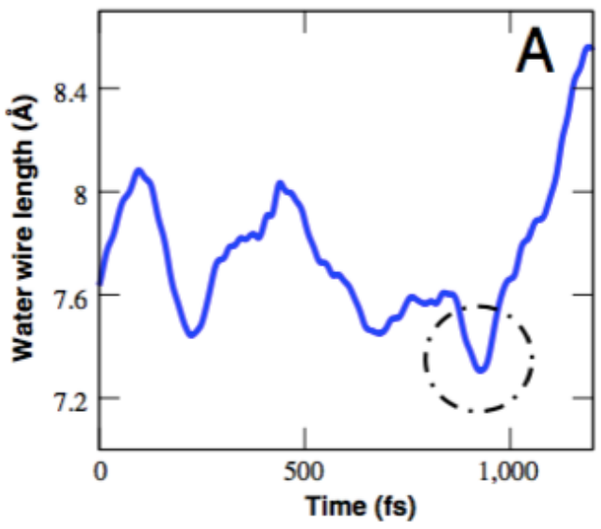
On the recombination of hydronium and hydroxide ions in water

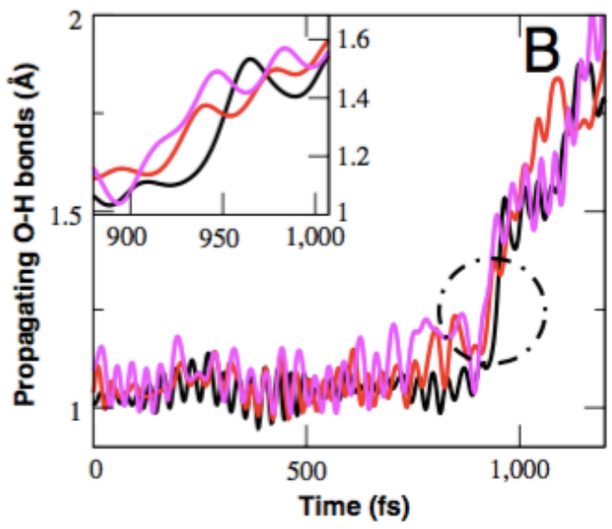
PNAS

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.



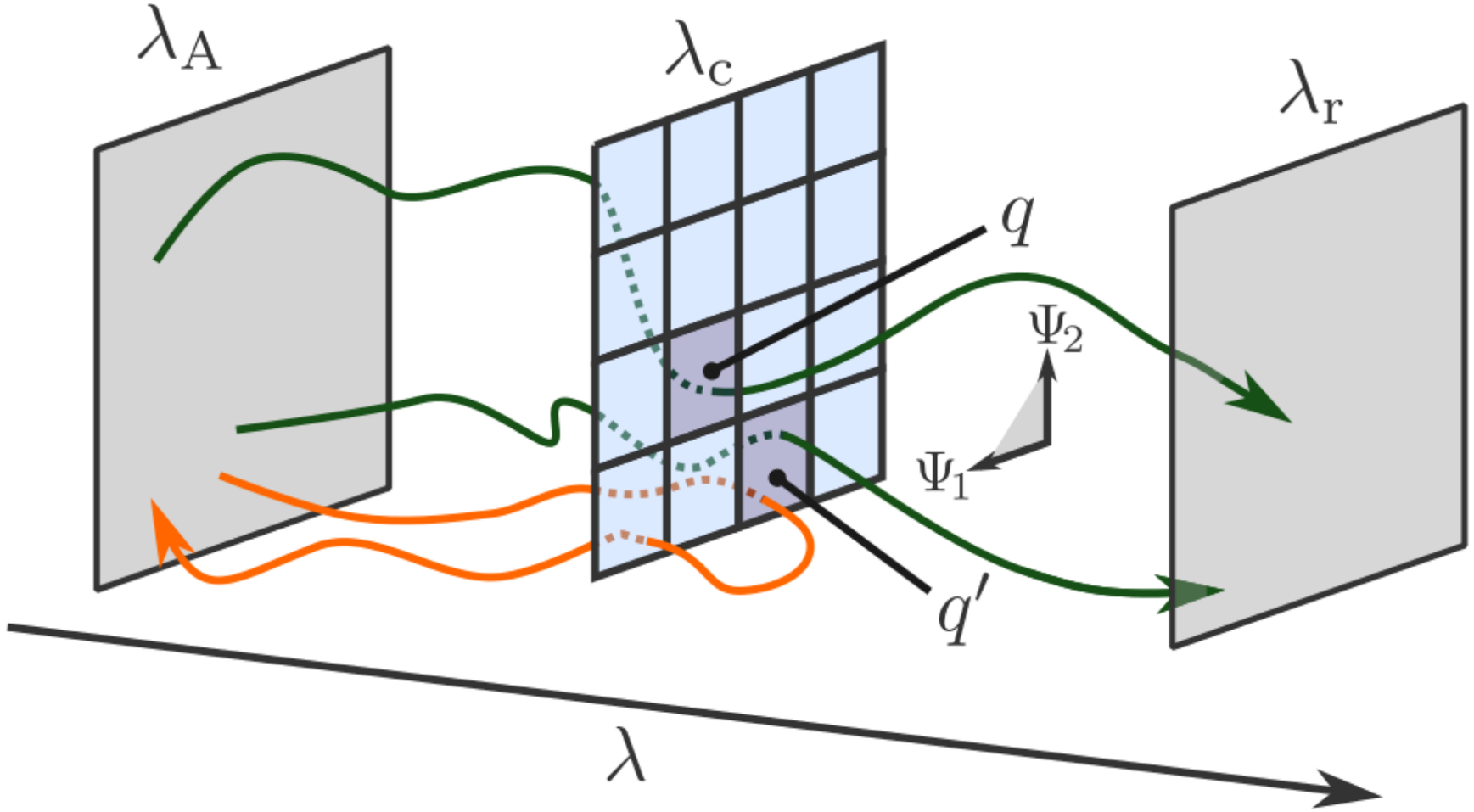




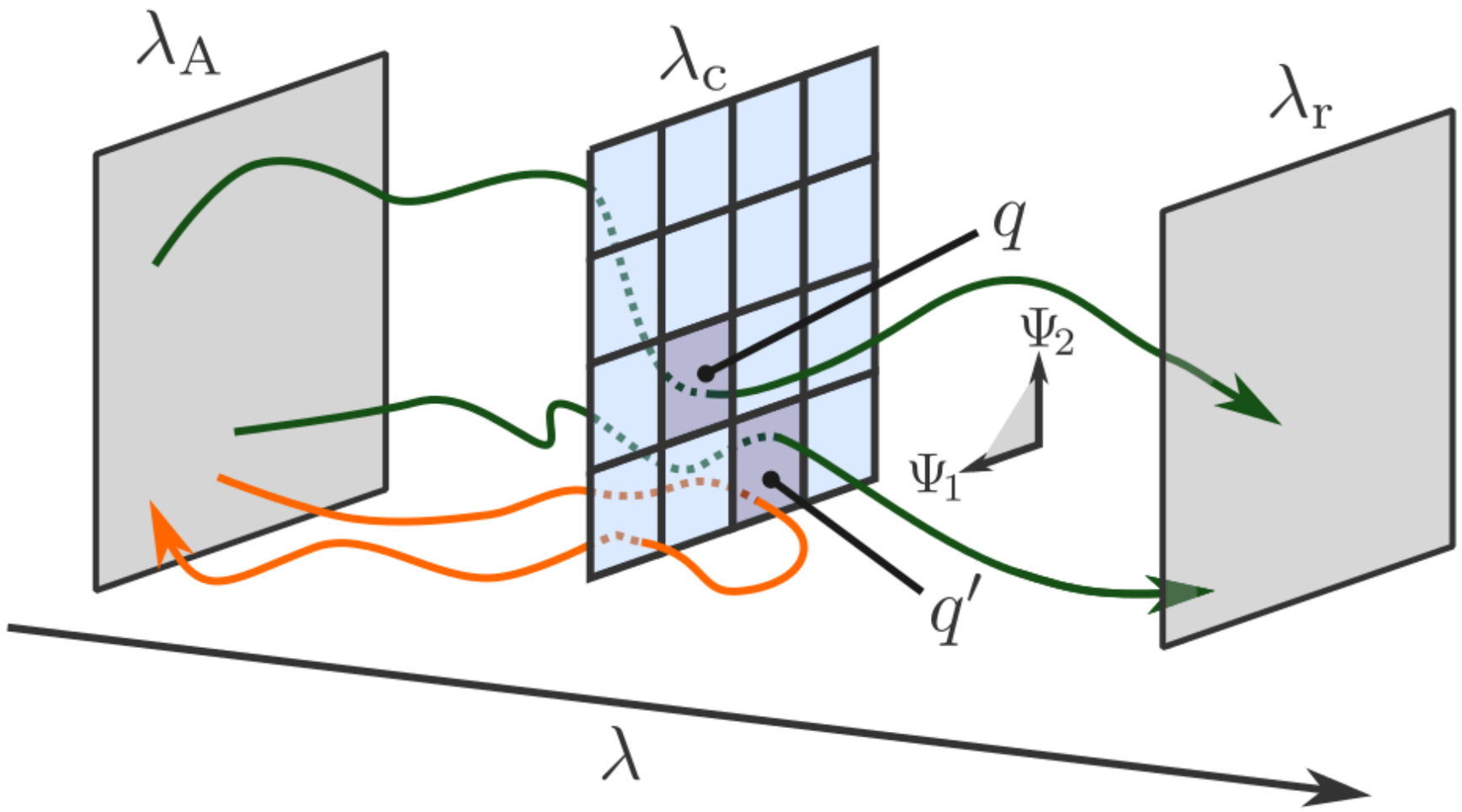
Analyzing complex reaction mechanisms using path sampling

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

J. Chem. Theory Comput. **12**, 5398-5410, (2016)



Analysis of paths: Predictive power method



Consider the set of trajectories starting at λ_A, crossing λ_c at least once, and ending at either λ_A or λ_B.

t_q: the fraction of trajectories passing through bin q in the λ_c surface.

r_q: the fraction of trajectories passing through bin q and cross λ_r.

u_q: the fraction of trajectories passing through bin q but do not reach λ_r.

Consider the set of trajectories starting at λ_A, crossing λ_c at least once, and ending at either λ_A or λ_B.

t_q: the fraction of trajectories passing through bin q in the λ_c surface.

r_q: the fraction of trajectories passing through bin q and cross λ_r.

u_q: the fraction of trajectories passing through bin q but do not reach λ_r.

Relations:

$$t_q = u_q + r_q \qquad \sum_q t_q = 1$$

$$\sum_q r_q = \mathcal{P}_A(\lambda_r|\lambda_c) \qquad \sum_q u_q = 1 - \mathcal{P}_A(\lambda_r|\lambda_c)$$

If $r_q/t_q = 1$ or $r_q = 0$ for each bin $q \Rightarrow$ **optimal predictive ability**

$$t_q = u_q + r_q \qquad \sum_q t_q = 1$$

$$\sum_q r_q = \mathcal{P}_A(\lambda_r|\lambda_c) \qquad \sum_q u_q = 1 - \mathcal{P}_A(\lambda_r|\lambda_c)$$

If $r_q/t_q = 1$ or $r_q = 0$ for each bin $q \Rightarrow$ **optimal predictive ability**

We define now an overall measure $\mathcal{T}$ of predictive power as a weighted average of r_q/t_q over q where each bin is weighted with the fraction of reactive trajectories passing through q.

$$\begin{aligned}
\mathcal{T} &\equiv \sum_q \left(\frac{r_q}{\sum_v r_v} \right) \frac{r_q}{t_q} = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q^2}{t_q} = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q(t_q - u_q)}{t_q} \\
&= \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q r_q - \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q u_q}{t_q} = 1 - \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q u_q}{t_q} = 1 - \mathcal{S}
\end{aligned}$$

$$\tau \equiv \sum_q \left(\frac{r_q}{\sum_v r_v} \right) \frac{r_q}{t_q} = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q^2}{t_q} = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q(t_q - u_q)}{t_q}$$

$$= \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q r_q - \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q u_q}{t_q} = 1 - \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \sum_q \frac{r_q u_q}{t_q} = 1 - \mathcal{S}$$

In continuous space, S is the overlap integral of the reactive and unreactive distributions.

$$\mathcal{S}_A^{\lambda_c,\lambda_r}[\Psi^N] = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \int \left(\frac{r^{\lambda_c,\lambda_r}(\Psi^N) u^{\lambda_c,\lambda_r}(\Psi^N)}{t^{\lambda_c}(\Psi^N)} \right) \mathrm{d}\Psi^N$$

The highest possible predictive ability is obtained by the collective variables that minimize the overlap

$$\mathcal{S}_{A,0}^{\lambda_c,\lambda_r} = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \min_{\Psi^N} \left[\int \left(\frac{r^{\lambda_c,\lambda_r}(\Psi^N) u^{\lambda_c,\lambda_r}(\Psi^N)}{t^{\lambda_c}(\Psi^N)} \right) \mathrm{d}\Psi^N \right]$$

$$\mathcal{S}_A^{\lambda_c,\lambda_r}[\Psi^N] = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \int \left(\frac{r^{\lambda_c,\lambda_r}(\Psi^N) u^{\lambda_c,\lambda_r}(\Psi^N)}{t^{\lambda_c}(\Psi^N)} \right) \mathrm{d}\Psi^N$$

The highest possible predictive ability is obtained by the collective variables that minimize the overlap

$$\mathcal{S}_{A,0}^{\lambda_c,\lambda_r} = \frac{1}{\mathcal{P}_A(\lambda_r|\lambda_c)} \min_{\Psi^N} \left[\int \left(\frac{r^{\lambda_c,\lambda_r}(\Psi^N) u^{\lambda_c,\lambda_r}(\Psi^N)}{t^{\lambda_c}(\Psi^N)} \right) \mathrm{d}\Psi^N \right]$$

If the orthogonal collective variables have **no correlation** with reactivity:

$$\begin{aligned} r^{\lambda_c,\lambda_r}(\Psi^N) &= \mathcal{P}_A(\lambda_r|\lambda_c)\, t^{\lambda_c}(\Psi^N), \qquad u^{\lambda_c,\lambda_r}(\Psi^N) = [1 - \mathcal{P}_A(\lambda_r|\lambda_c)]\; t^{\lambda_c}(\Psi^N). \\ \Rightarrow \mathcal{S}_A^{\lambda_c,\lambda_r}[\Psi^N] &= 1 - \mathcal{P}_A(\lambda_r|\lambda_c), \quad \Rightarrow \mathcal{T}_A^{\lambda_c,\lambda_r} = \mathcal{P}_A(\lambda_r|\lambda_c). \end{aligned}$$

If the orthogonal CVs have **maximum deterministic correlation**:

$$\begin{aligned} & r^{\lambda_c,\lambda_r}(\Psi^N) = t^{\lambda_c}(\Psi^N) \quad \text{or} \quad r^{\lambda_c,\lambda_r}(\Psi^N) = 0 \quad \text{for all values of} \quad \Psi^N \\ & \Rightarrow \mathcal{S}_A^{\lambda_c,\lambda_r}[\Psi^N] = 0 \quad \text{and} \quad \mathcal{T}_A^{\lambda_c,\lambda_r}[\Psi^N] = 1 \end{aligned}$$

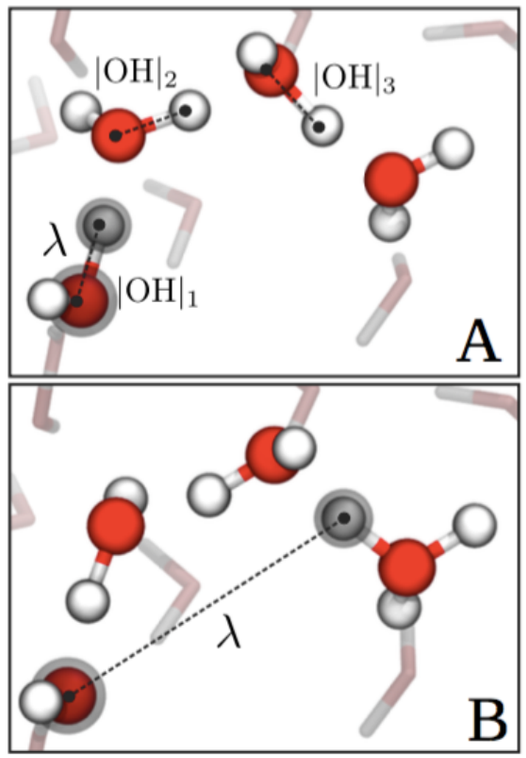
Auto-ionization of water: a RETIS/CP2K study

(PNAS, **115**, E4569, (2018), Moqadam, Lervik, Riccardi, Venkatraman, Bjørn K. Alsberg, and van Erp)

RETIS combined with DFT-based dynamics (CP2K, BLYP-functional)

32 water molecules

20 interfaces, last one unreachable: all trajectories start and end with pure water



+- 20,000 trajectories in each path ensemble of which:
10,000 Replica Exchange, 5000 time-reversal, 5000 shootings
Each ensemble had at least 1000 accepted shooting moves

Reaction coordinate λ

if only H2O present: λ=MAX{|O-H|}
(the largest OH-bond in the system)

if H3O+, OH- present: λ=MIN{$|O_{OH^-}H_{H3O^+}|$}
(smallest distance between hydroxide-oxygen and one of the three hydronium-hydrogens)

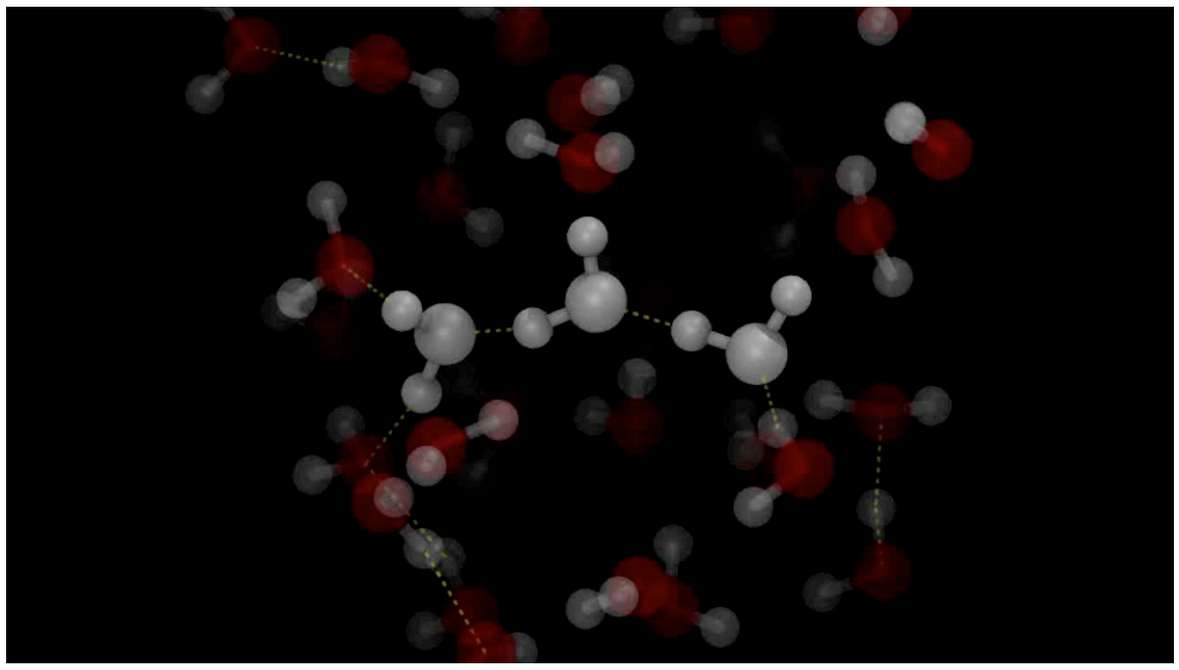
Titus S. van Erp

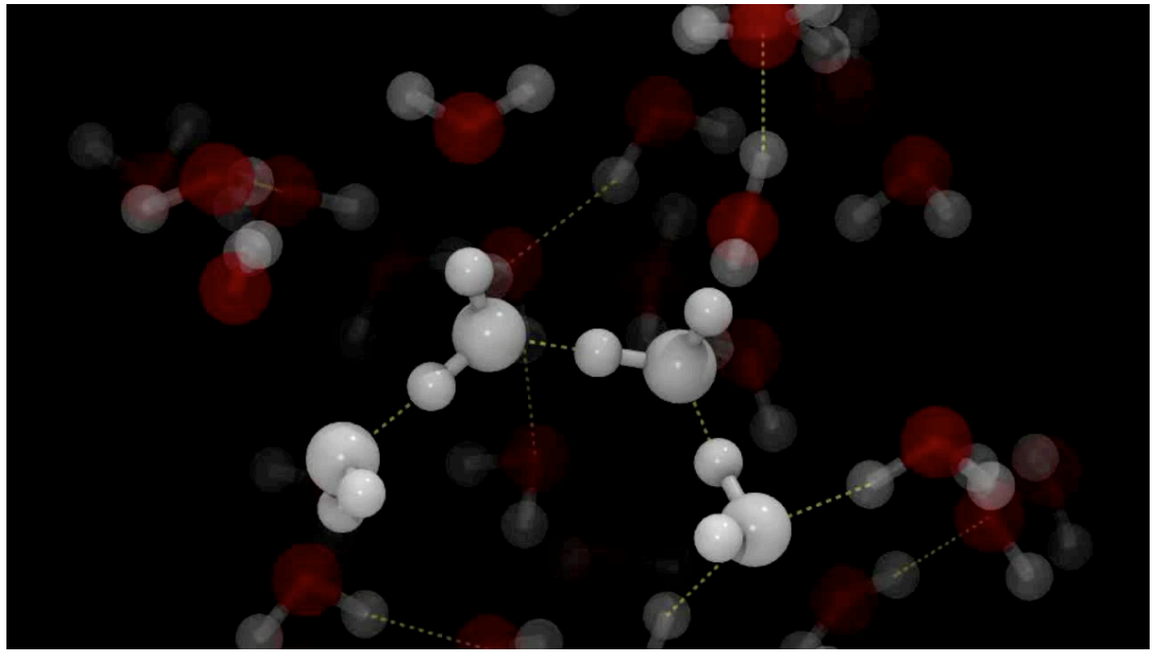
Application: Ab Initio autoionization of water

Auto-ionization of water: a RETIS/CP2K study

(PNAS, **115**, E4569, (2018), Moqadam, Lervik, Riccardi, Venkatraman, Bjørn K. Alsberg, and van Erp)

Short trajectories with hydrogen wire of 3 and 4 waters:

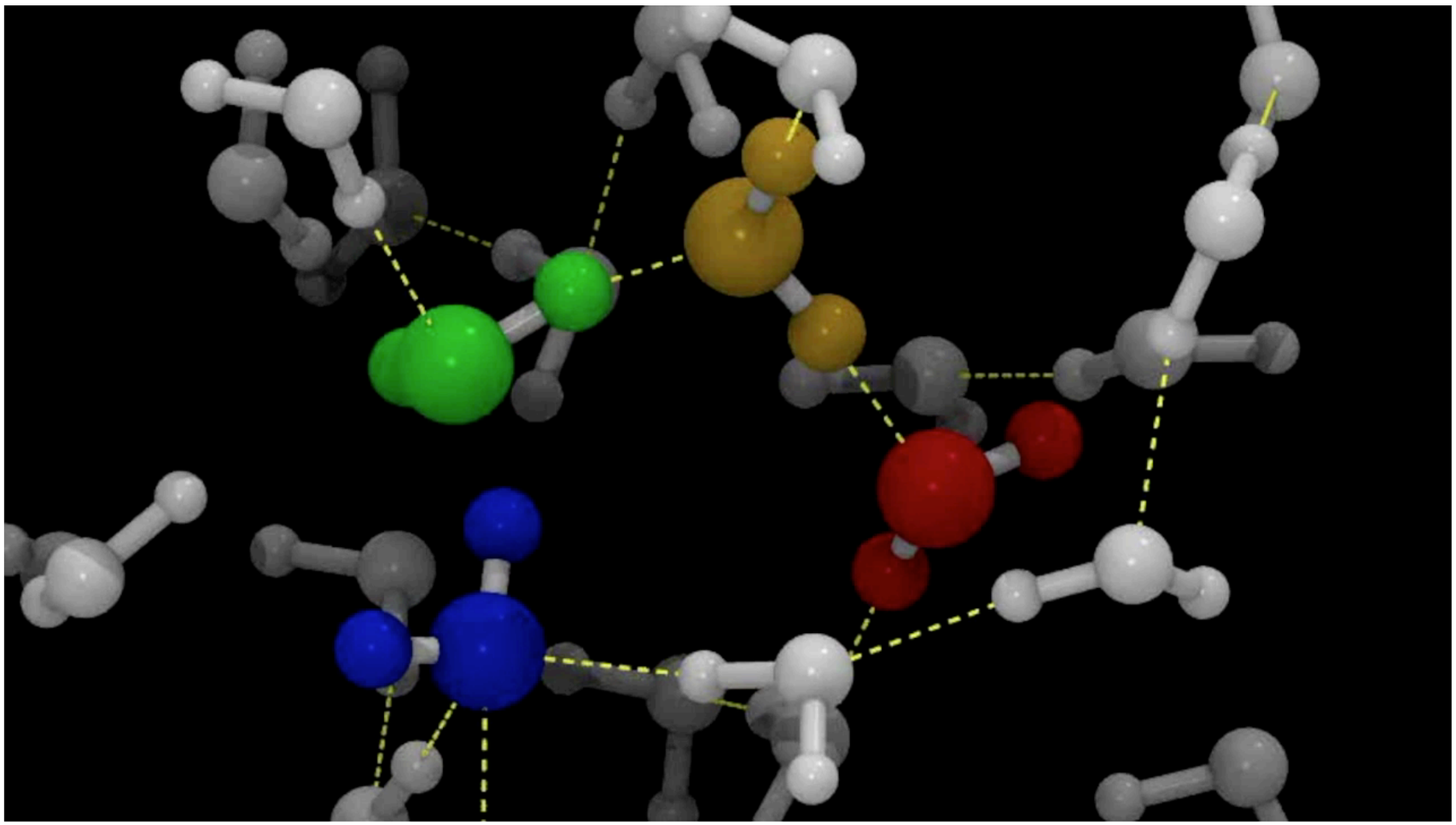




Auto-ionization of water: a RETIS/CP2K study

(PNAS, **115**, E4569, (2018), Moqadam, Lervik, Riccardi, Venkatraman, Bjørn K. Alsberg, and van Erp)

long trajectory with hydrogen swap

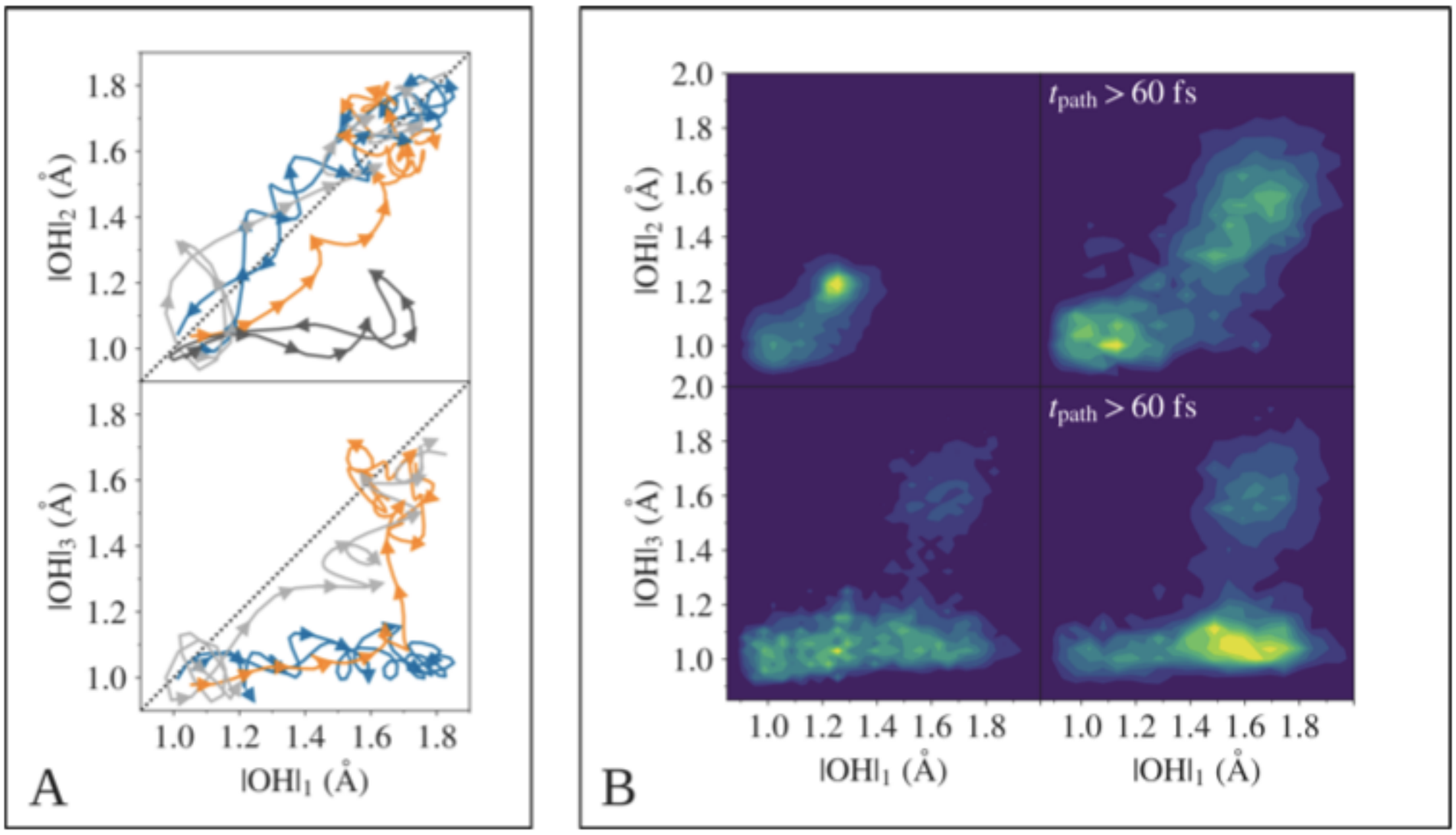


NTNU

Norwegian University of Science and Technology

Titus S. van Erp

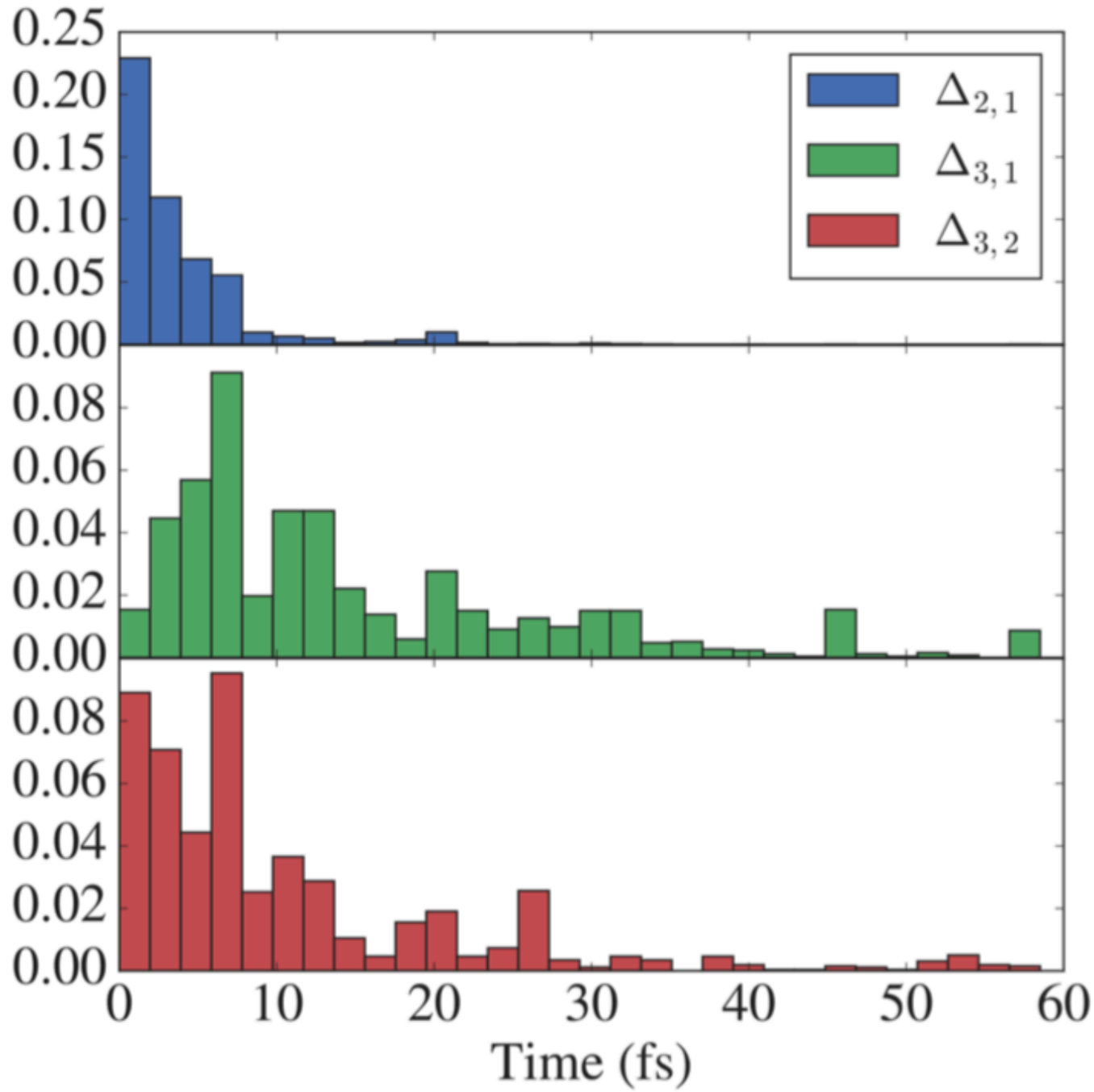
Application: Ab Initio autoionization of water



1st and 2nd proton jump almost always go concerted. The 3rd proton jump mainly follows up in a step-wise fashion

Titus S. van Erp

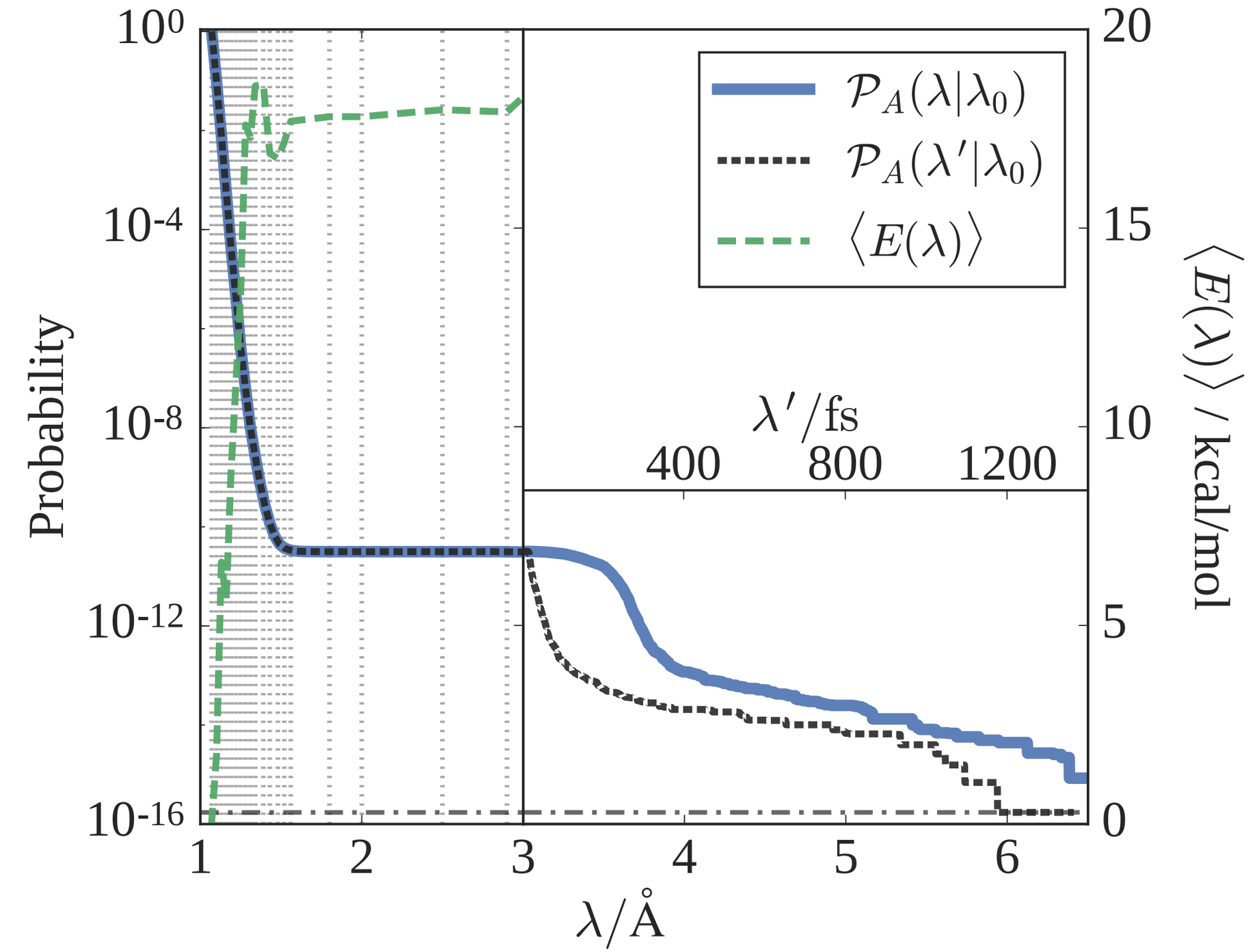
Application: Ab Initio autoionization of water



NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

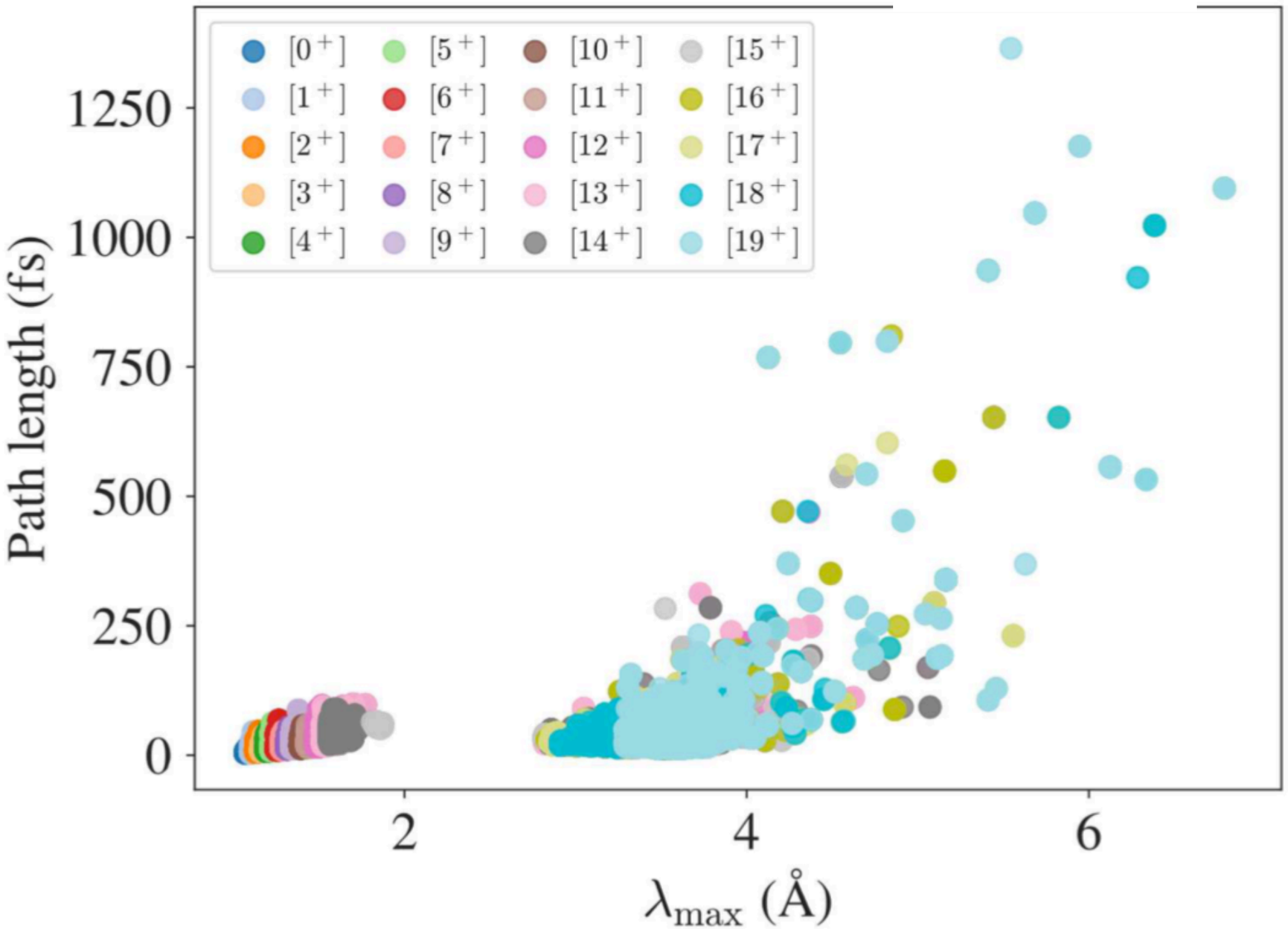
Application: Ab Initio autoionization of water

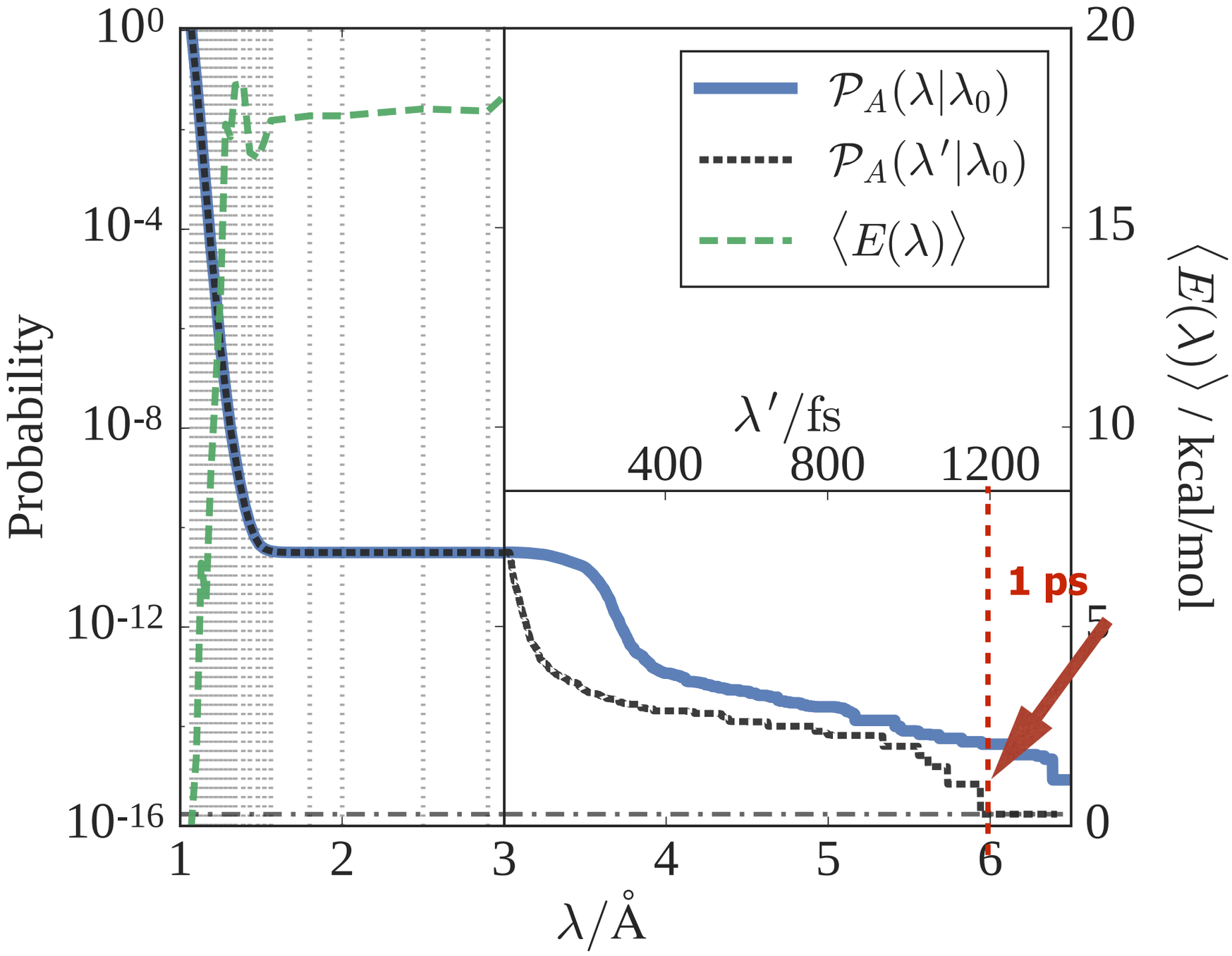


NTNU
Norwegian University of
Science and Technology

Titus S. van Erp

Application: Ab Initio autoionization of water




10^0
10^{-4}
10^{-8}
10^{-12}
10^{-16}
Probability
20
15
10
5
0
$\langle E(\lambda) \rangle$ / kcal/mol
$\mathcal{P}_A(\lambda|\lambda_0)$
$\mathcal{P}_A(\lambda'|\lambda_0)$
$\langle E(\lambda) \rangle$
λ'/fs
400
800
1200
1 ps
1
2
3
4
5
6
λ/Å

Comparison with experimental data?

Untersuchungen über die Kinetik der Neutralisation. I

Von *M. EIGEN* und *L. DE MAEYER*

Aus dem Max-Planck-Institut für Physikalische Chemie, Göttingen

(Vorgetragen von M. Eigen anläßlich der 54. Hauptversammlung der Deutschen Bunsen-Gesellschaft für physikalische Chemie e.V. in Goslar am 21. Mai 1955)

Mit Hilfe rechteckiger Hochspannungsimpulse werden Messungen der Zeitabhängigkeit des Dissoziationsfeldeffekts in sehr reinem Wasser (der spezifischen Leitfähigkeit $5{,}7 \cdot 10^{-8}\ \Omega^{-1}\ \mathrm{cm}^{-1}$ bei 25° C) durchgeführt. Aus der gefundenen Relaxation lassen sich die Geschwindigkeitskonstanten für die Gleichgewichtseinstellung:

$$H^+ + OH^- \underset{k_D}{\overset{k_A}{\rightleftharpoons}} H_2O$$

bestimmen.
Man erhält für die Geschwindigkeitskonstante der Neutralisationsreaktion:

$$k_A = 1{,}3\ (\pm\ 0{,}2) \cdot 10^{11}\ \text{liter/mol sec bei } 25°\ \text{C}.$$

Für die Konstante der Dissoziationsreaktion ergibt sich der Wert:

$$k_D = 2{,}6 \cdot 10^{-5}\ \text{sec}^{-1}.$$

Der Mechanismus der Neutralisationsreaktion, die nach den vorliegenden Ergebnissen eine der schnellsten Lösungsreaktionen ist, wird eingehend diskutiert.
Die beschriebene Meßanordnung ist allgemein für die Untersuchung zeitabhängiger Feldeffekte in flüssigen und festen Systemen (Halbleiter) geeignet.

Alternative threshold-free definition
Proton swap condition

$$k = 0.16\ \mathrm{s}^{-1}$$

Eigen & Maeyer:

$k=2.6\text{x}10^{-5}\ \text{s}^{-1}$:

each water dissociates
once per 10 h
E_{act}=15.5-16.5 kcal/mol

Our our simulations:

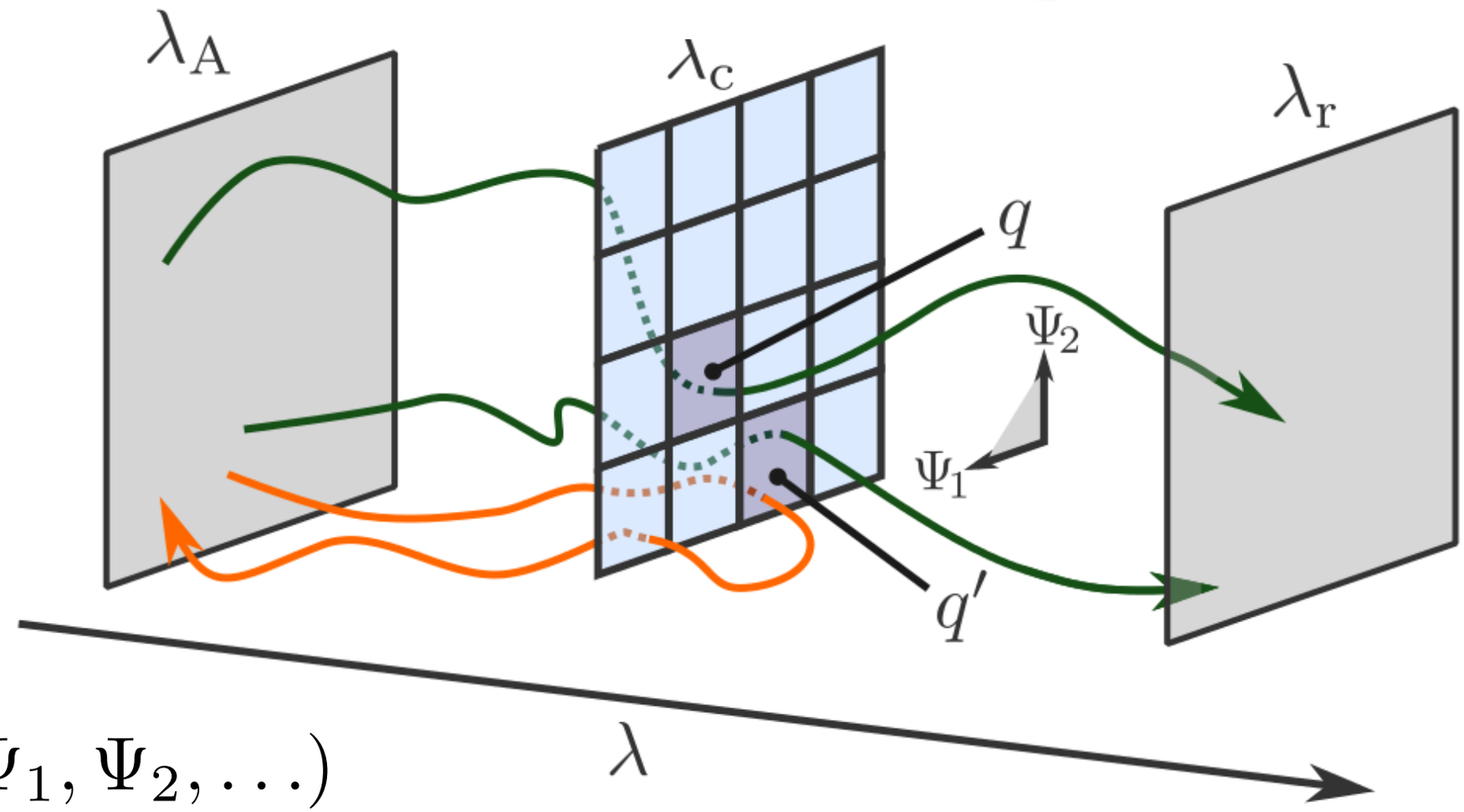
$$\text{flux} = 2.9 \cdot 10^{12}\ \mathrm{s}^{-1}$$

$$\mathcal{P}_A = 4.0 \cdot 10^{-15}$$

$$k = \text{flux} \cdot \mathcal{P}_A = 1.1 \cdot 10^{-2}\ \mathrm{s}^{-1}$$

each water dissociates
once per 1.5 minutes
(1 ps lifetime threshold)
E_{act}=17.8 kcal/mol

Application: Ab Initio autoionization of water



$$\xi = (\Psi_1, \Psi_2, \ldots)$$

$$\mathcal{T}_A^{\lambda^c,\lambda^r} = 1 - \frac{1}{\mathcal{P}_A(\lambda^r|\lambda^c)} \int \frac{r^{\lambda^c,\lambda^r}(\xi) u^{\lambda^c,\lambda^r}(\xi)}{r^{\lambda^c,\lambda^r}(\xi) + u^{\lambda^c,\lambda^r}(\xi)} \mathrm{d}\xi$$

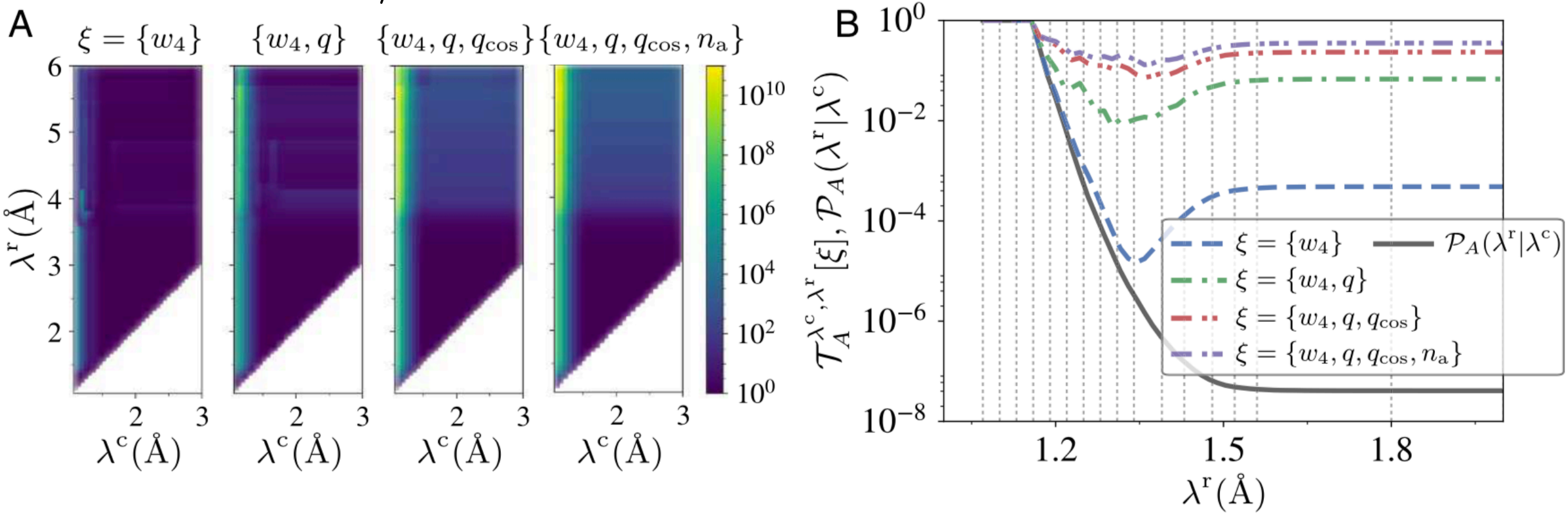
"predictive power" J. Chem. Theory Comput. **12**, 5398-5410, (2016)

If $\lambda^c << \lambda^r$, 'reactivity' is a rare event and r(ξ) is small -> **Path reweighting** (Rogal J, Lechner W, Juraszek J, Ensing B, Bolhuis PG (2010). J Chem Phys 133:174109.)

NTNU
Norwegian University of Science and Technology
Titus S. van Erp

Application: Ab Initio autoionization of water

$\mathcal{T}/\mathcal{P}$



w_4: H-bond wire(*)
n_a: number of accepted H-bonds
q: measure of tetrahedral ordering around O^{λ}, the oxygen with largest OH bond(**)
q_{cos}: alignment between 2 nearest H-bonds

(*) the hydrogen bond wire is the shortest wire containing the species O^{λ} and i − 1 other water species at the first point in time when λ is greater than a given threshold value, λ_t=1.15 Å

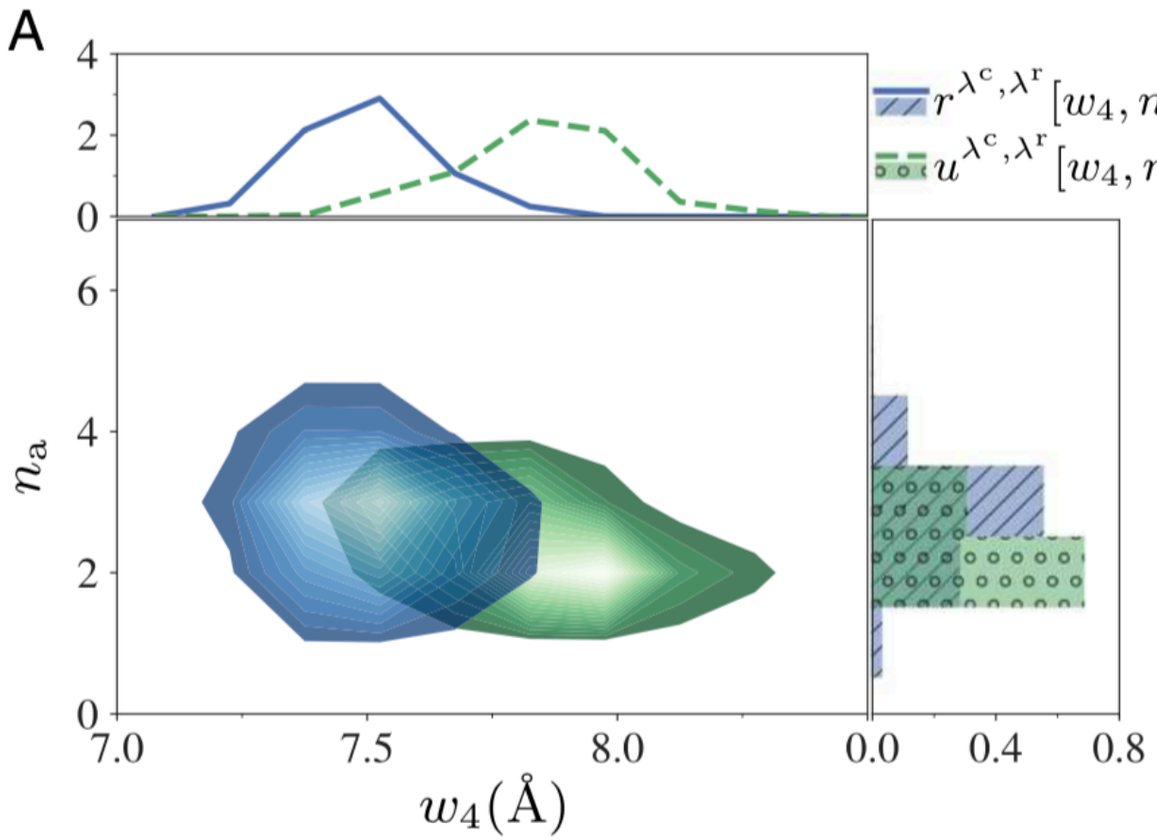
(**) $$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos\psi_{jk} + \frac{1}{3}\right)^2$$

NTNU

Norwegian University of Science and Technology

Titus S. van Erp

Application: Ab Initio autoionization of water



r and **u** are normalized in this figure! (Otherwise **r** looks like a zero flat-line)

Consider two cases

i) If 7.15<w4 <7.6 and at the same time na =3, the probability for a reactive event is $3.6 \cdot 10^{-6}$

This is 58 larger than from a random point at λ^c .

ii) If (ii) w4 < 7.3 and simultaneously na = 4, the chance increases to 0.15.

But 45% of reactive trajectories cross λ^c through region i) and only 0.6% through region ii)

The predictive power **T** weights the contributions of i) and ii) accordingly

⇒ i) has 75 higher weight than ii)

Can we use Machine Learning to provide best candidates for CV to be tested in the predictive power method?

- Every odd path ensemble used as calibration set.
- Every even path ensemble were used for the test set.
- As heavily skewed distributions are difficult to treat with ML, we further omitted the reweighting of the datasets with the statistical weights of the corresponding path ensembles.
- Instead, we applied the ML techniques as a qualitative approach to find new parameters that could be tested quantitatively within the predictive power method.
- To avoid over-interpretation we restricted the complexity of the ML decision process and imposed a maximum of four order parameters in the predictive power method.

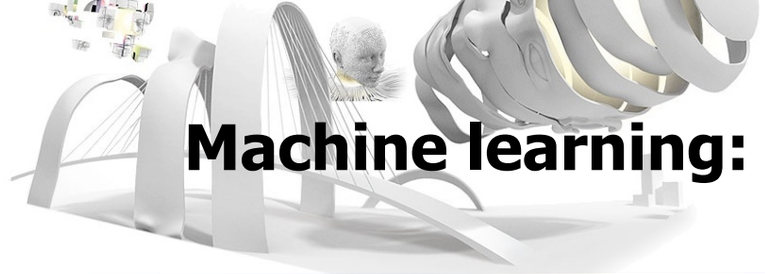
Can we use Machine Learning to provide best candidates for CV to be tested in the predictive power method?

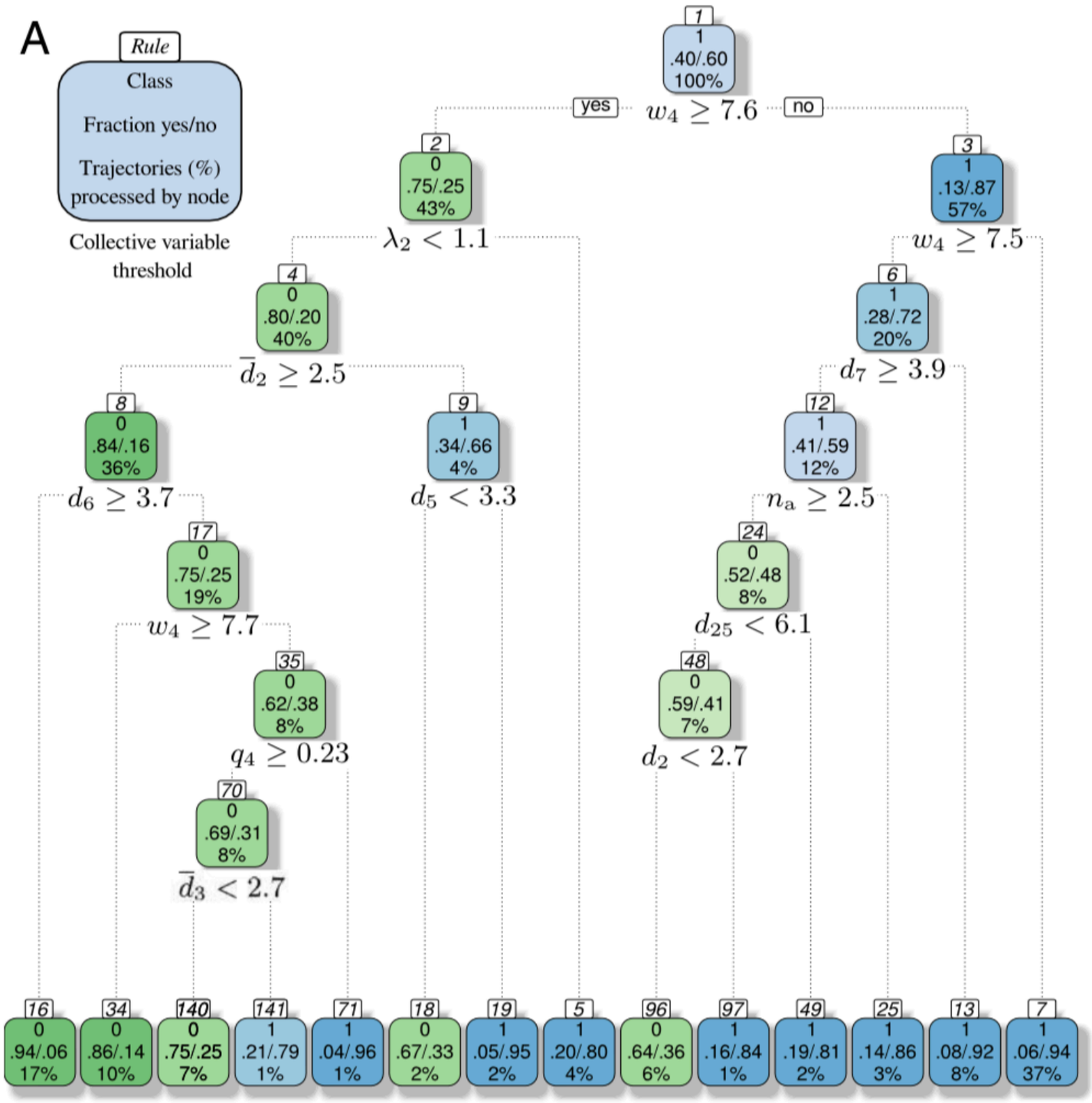
- Our ML model: single-tree–based decision models based on classification and regression decision trees (CART)
- We considered 138 collective variables consisting of oxygen– oxygen distances; oxygen–hydrogen distances for initially bound water molecules; all angles formed by O^{λ} and its four closest oxygen neighbors; and the Steinhardt order parameters of orders 3, 4, and 6. In addition, the order parameters already considered were added.

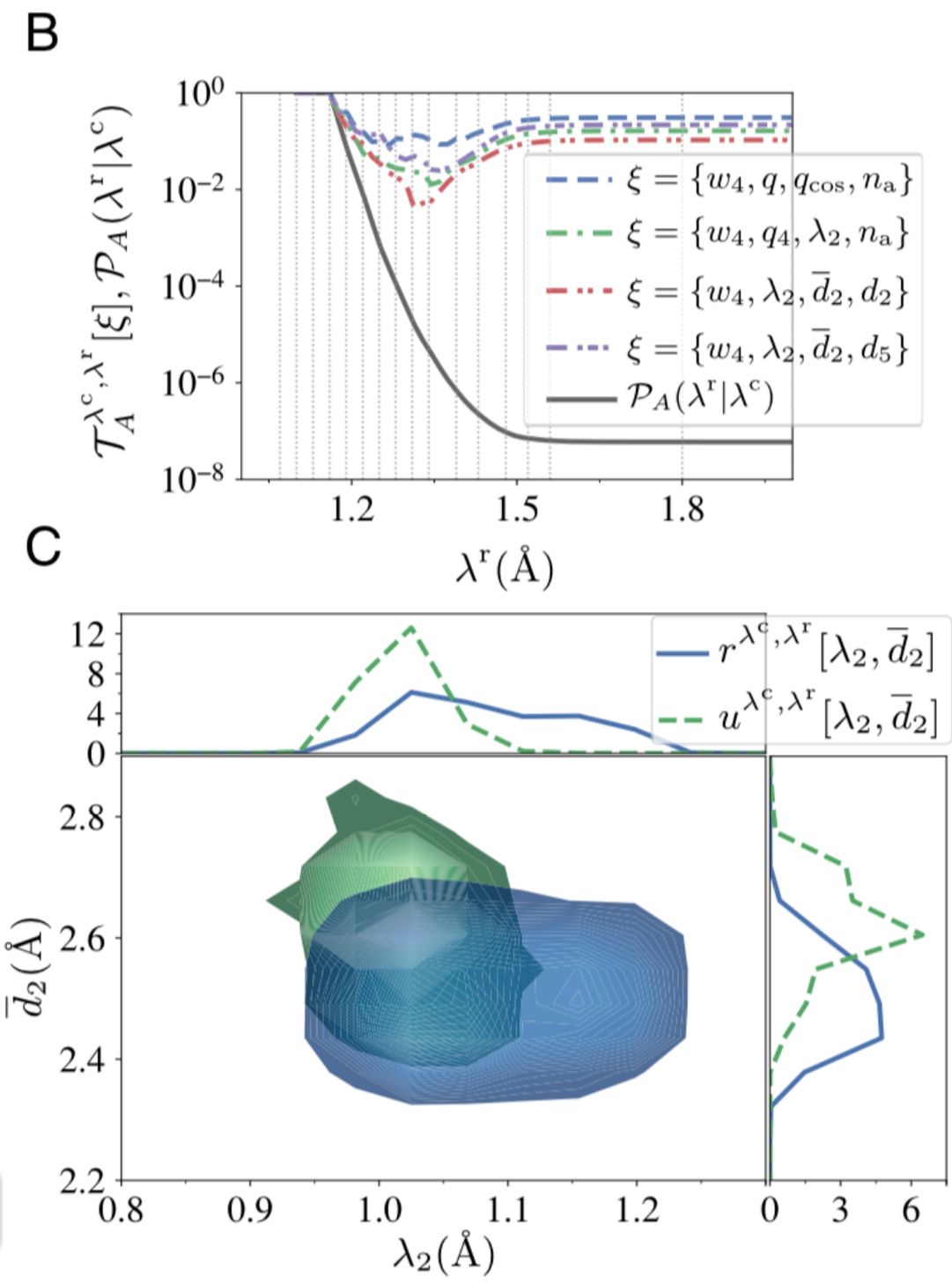
Titus S. van Erp

Machine learning: Identify Reaction Triggers

NTNU
Norwegian University of Science and Technology







Conclusions

- Using RETIS we can reach the minute/hour timescale using ab-initio MD in the case of water dissociation

- We designed a new approach to analyze the data of path sampling simulations 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** but **not a sufficient** condition for the initiation of water splitting.

- Other initiation triggers are the number of accepted hydrogen bonds, local distortion from tetrahedral order, and possibly other **local or non-local** properties.

- Machine Learning techniques can identify initiation triggers that are not easily found by intuition or looking at movies. Though in this case it did not beat human intuition/effort.

- Yet, the Machine Learning approach did find all important parameters very fast. Possible improvements are now being studied (can we do without intuition/hyper-parameters?)

- Can we use the knowledge of what triggers the reaction for designing catalytic strategies?