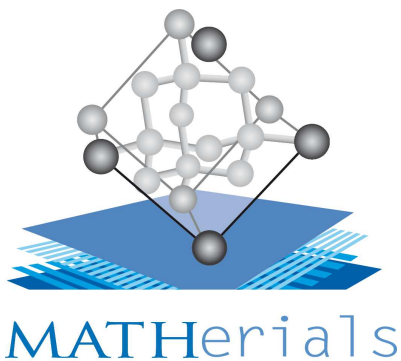
Algorithms for computational statistical physics
Part 2: Sampling metastable dynamics

T. Lelièvre

CERMICS - Ecole des Ponts ParisTech & Matherials project-team - INRIA



ICTS, August 2017

Introduction

Remember the dynamics:

- *Langevin* dynamics:

$$\begin{cases} d\boldsymbol{X}_t = M^{-1}\boldsymbol{P}_t\,dt \\ d\boldsymbol{P}_t = -\nabla V(\boldsymbol{X}_t)\,dt - \gamma M^{-1}\boldsymbol{P}_t\,dt + \sqrt{2\gamma\beta^{-1}}d\boldsymbol{W}_t \end{cases}$$

  where $\gamma > 0$ and $\beta = (k_B T)^{-1}$.

- *overdamped Langevin* (or gradient) dynamics:

$$d\boldsymbol{X}_t = -\nabla V(\boldsymbol{X}_t)\,dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}_t.$$

Introduction

These dynamics are used to compute macroscopic quantities:

(i) Thermodynamic quantities (averages wrt μ of some observables): stress, heat capacity, free energy,...

$$\mathbb{E}_\mu(\varphi(\boldsymbol{X})) = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \simeq \frac{1}{T} \int_0^T \varphi(\boldsymbol{X}_t)\, dt.$$

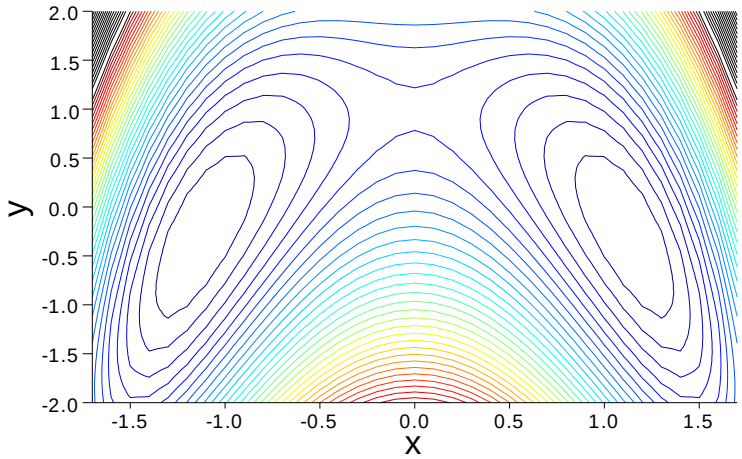
(ii) Dynamical quantities (averages over trajectories): diffusion coefficients, viscosity, transition rates,...

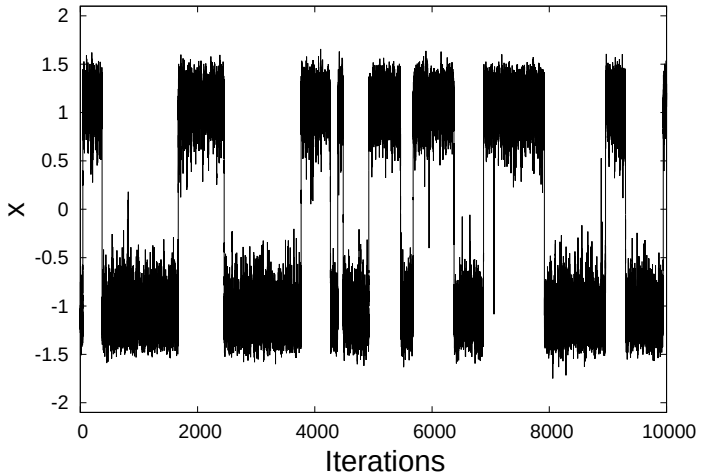
$$\mathbb{E}(\mathcal{F}((\boldsymbol{X}_t)_{t\geq 0})) \simeq \frac{1}{M} \sum_{m=1}^{M} \mathcal{F}((\boldsymbol{X}_t^m)_{t\geq 0}).$$

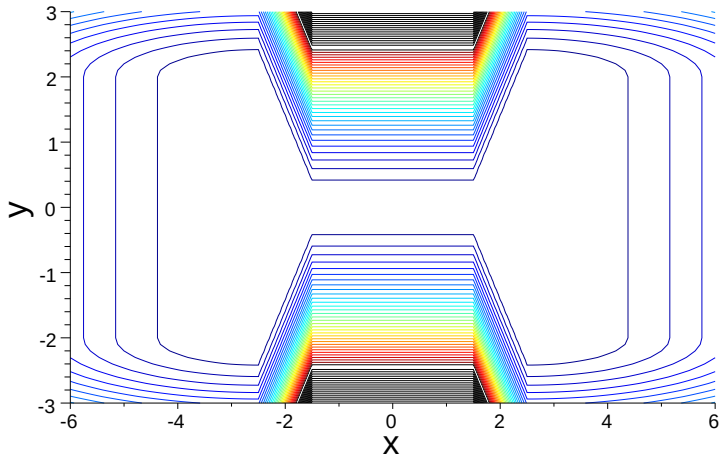
Difficulties: (i) high-dimensional problem ($N \gg 1$); (ii) $\boldsymbol{X}_t$ is a metastable process and μ is a multimodal measure.

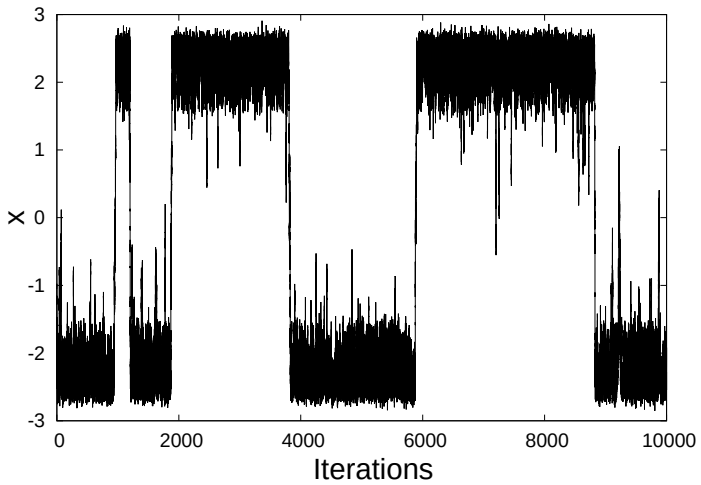
Metastability: energetic and entropic barriers

A two-dimensional schematic picture







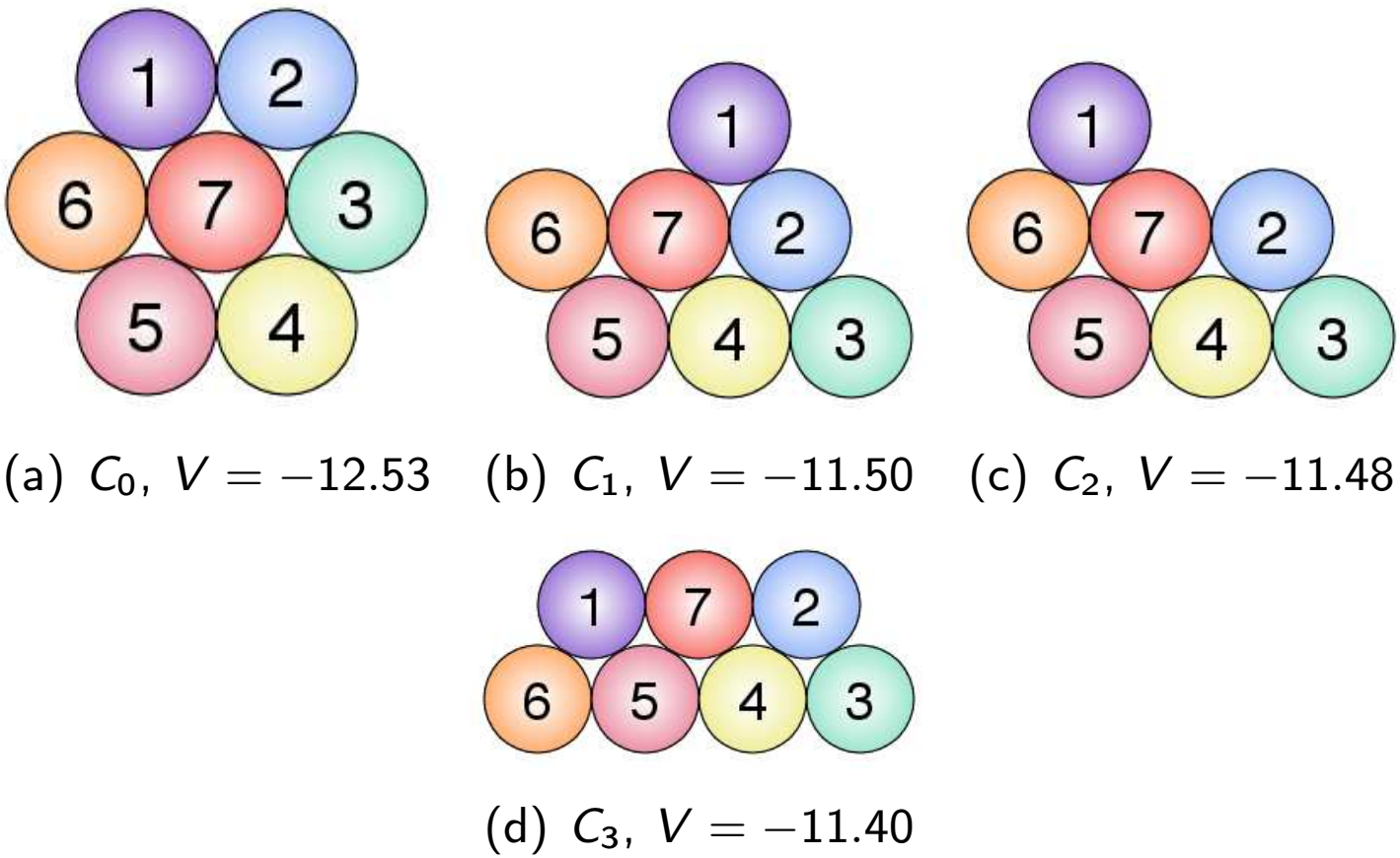


$\longrightarrow$ • Slow convergence of trajectorial averages

• Transitions between metastable states are rare events

A toy example in material sciences

The 7 atoms Lennard Jones cluster in 2D.



(a) C_0, $V = -12.53$ (b) C_1, $V = -11.50$ (c) C_2, $V = -11.48$

(d) C_3, $V = -11.40$

Figure: Low energy conformations of the Lennard-Jones cluster.

$\longrightarrow$ simulation

Introduction

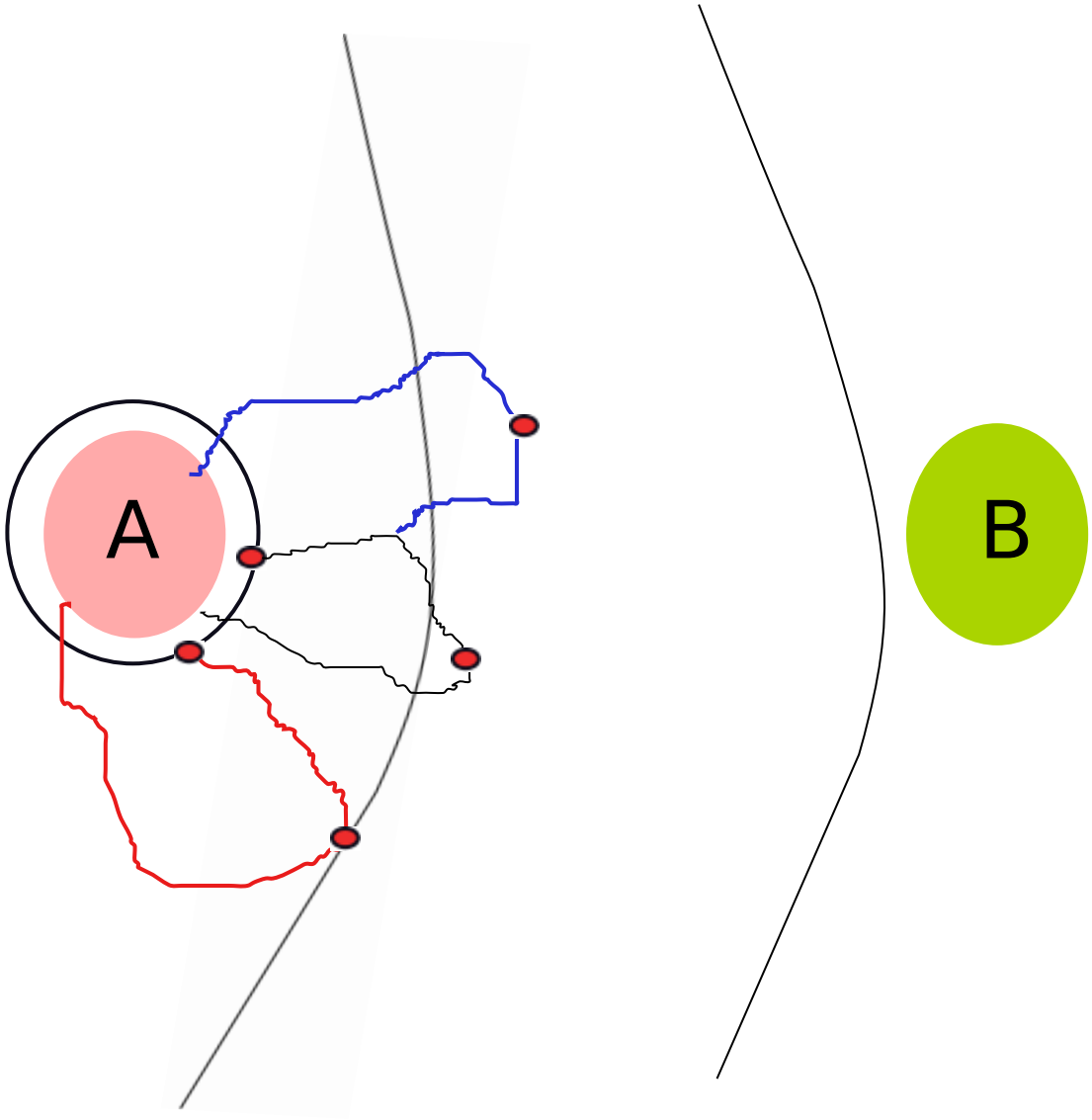
For computing thermodynamics quantities, there is a clear classification of available methods, and the difficulties are now well understood (in particular for free energy computations, see for example [TL, Rousset, Stoltz, 2010]). On the opposite, computing efficiently dynamical quantities remains a challenge.

Outline of the talk:

1. Adaptive Multilevel Splitting methods: Towards efficient sampling of reactive paths. *Rare event simulation.*
2. Accelerated dynamics: These methods have been proposed by A.F. Voter to generate efficiently metastable dynamics. *Mathematical tool: Quasi Stationary Distributions.*

Underlying question: how to properly define and quantify metastability ? Various answers: (i) rate of convergence to equilibrium; (ii) exit time from metastable states; (iii) decorrelation time; (iv) asymptotic variance of estimators.

Splitting strategies



Multilevel splitting

We would like to sample trajectories between two given metastable states A and B. The main assumption is that we are given a smooth one dimensional function $\xi : \mathbb{R}^d \to \mathbb{R}$ (s.t. $|\nabla \xi| \neq 0$) which "indexes" the transition from A to B in the following sense:

$$A \subset \{x \in \mathbb{R}^d, \xi(x) < z_{\min}\} \text{ and } B \subset \{x \in \mathbb{R}^d, \xi(x) > z_{\max}\},$$

where $z_{\min} < z_{\max}$, and $\Sigma_{z_{\min}}$ (resp. $\Sigma_{z_{\max}}$) is "close" to ∂A (resp. ∂B).

Example: $\xi(x) = \|x - x_A\|$ where $x_A \in A$ is a reference configuration in A. We are interested in the event $\{\tau_A < \tau_B\}$, starting from an initial condition on $\Sigma_{z_{\min}}$, where

$$\tau_A = \inf\{t > 0, \boldsymbol{X}_t \in A\}, \quad \tau_B = \inf\{t > 0, \boldsymbol{X}_t \in B\}$$

and

$$\tau_z = \inf\{t > 0, \xi(\boldsymbol{X}_t) > z\}.$$

Multilevel splitting

Question: How to compute dynamical quantities using ξ ? More precisely, we consider: (a) Reactive trajectories and (b) Transition times between the two metastable states A and B.

We propose a multilevel splitting approach [Kahn, Harris, 1951] [Rosenbluth, 1955] to discard failed trajectories and branch trajectories approaching the rare set. We focus on an adaptive variant [Cérou, Guyader, 2007] [Cérou, Guyader, TL, Pommier, 2011]: the Adaptive Multilevel Splitting (AMS) algorithm.

Remark: The algorithm can be seen as a kind of adaptive Forward Flux Sampling [Allen, Valeriani, Ten Wolde, 2009]. It is also related to the Interface Sampling Method [Bolhuis, van Erp, Moroni 2003] and the Milestoning method [Elber, Faradjian 2004]. See the review paper [Bolhuis, Dellago, 2009]. Another splitting technique in MD: weighted ensemble method [Zuckerman, 2010].

Reactive trajectory

A reactive trajectory between two metastable sets A and B is a piece of equilibrium trajectory that leaves A and goes to B without going back to A in the meantime [Hummer,2004] [Metzner, Schütte, Vanden-Eijnden, 2006].

Difficulty: A trajectory leaving A is more likely to go back to A than to reach B.

Splitting algorithm: basic idea

The idea of splitting algorithms (FFS, TIS, RESTART, ...) is to write the rare event

$$\{\tau_B < \tau_A\}$$

as a sequence of nested events: for $z_{\min} < z_1 < \ldots < z_{\max}$,

$$\{\tau_{z_1} < \tau_A\} \supset \{\tau_{z_2} < \tau_A\} \supset \ldots \supset \{\tau_{z_{\max}} < \tau_A\} \supset \{\tau_B < \tau_A\}$$

and to simulate the successive *conditional events*: for $k = 1, 2, \ldots,$

$$\{\tau_{z_q} < \tau_A\} \text{ knowing that } \{\tau_{z_{q-1}} < \tau_A\}.$$

It is then easy to build an unbiased estimator of

$$\mathbb{P}(\tau_B < \tau_A) = \mathbb{P}(\tau_{z_1} < \tau_A)\mathbb{P}(\tau_{z_2} < \tau_A | \tau_{z_1} < \tau_A) \ldots \mathbb{P}(\tau_B < \tau_A | \tau_{z_{\max}} < \tau_A)$$

Splitting algorithm: adaptive level computation

Problem: How to choose the intermediate levels $(z_q)_{q\geq 1}$?

It is easy to check, for a given number of intermediate levels, the optimum in terms of variance is attained if

$$\mathbb{P}(\tau_{z_q} < \tau_A | \tau_{z_{q-1}} < \tau_A) \text{ is constant .}$$

This naturally leads to adaptive versions (AMS, nested sampling) where the levels are determined by using *empirical quantiles*: choose $k < n$, and given n trajectories $(\boldsymbol{X}^m_{t\wedge\tau_A})_{t>0,m=1,\ldots,n}$ in the event $\{\tau_{z_{q-1}} < \tau_A\}$, choose z_q so that

$$\mathbb{P}(\tau_{z_q} < \tau_A | \tau_{z_{q-1}} < \tau_A) \simeq \left(1 - \frac{k}{n}\right).$$

The level z_q is the k-th order statistics of $\sup_{t\geq 0} \xi(\boldsymbol{X}^m_{t\wedge\tau_A})$:

$$\sup_{t\geq 0} \xi(\boldsymbol{X}^{(1)}_{t\wedge\tau_A}) < \ldots < \sup_{t\geq 0} \xi(\boldsymbol{X}^{(k)}_{t\wedge\tau_A}) =: z_q < \ldots < \sup_{t\geq 0} \xi(\boldsymbol{X}^{(n)}_{t\wedge\tau_A}).$$

AMS: estimator of the rare event probability (1/2)

Let Q_{iter} be the number of iterations to reach the level $z_{\max}$:

$$Q_{\mathrm{iter}} = \min\{q \geq 0, z_q > z_{\max}\}$$

(where z_0 is the k-th order statistics of the n initial trajectories). Then, one obtains the estimator:

$$\left(1 - \frac{k}{n}\right)^{Q_{\mathrm{iter}}} \simeq \mathbb{P}(\tau_{z_{\max}} < \tau_A).$$

AMS: estimator of the rare event probability (2/2)

At iteration Q_{iter}, one has an ensemble of n trajectories starting from $\Sigma_{z_{\min}}$ and such that $\tau_{z_{\max}} < \tau_A$. Thus

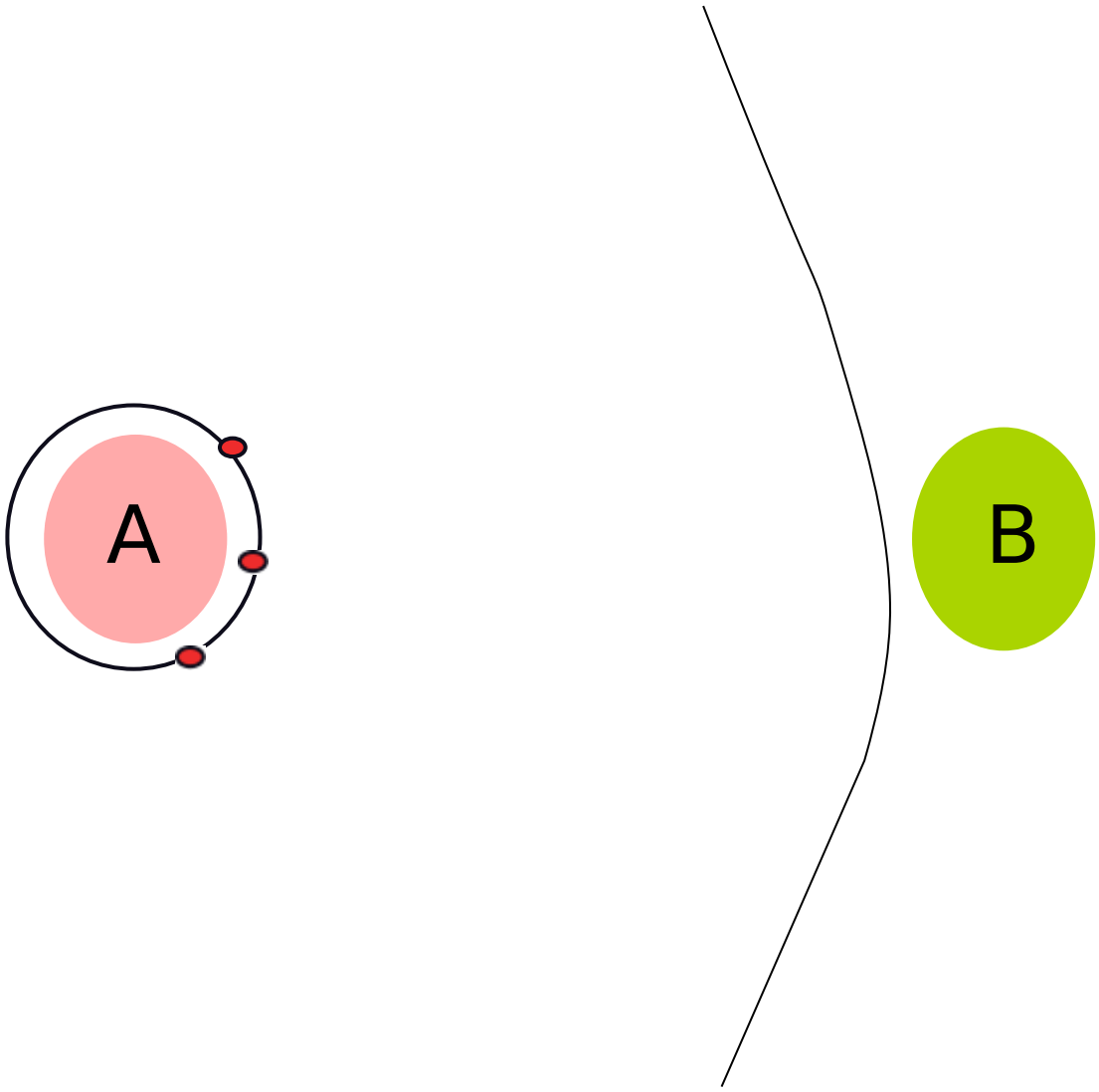
$$\hat{p}_{\text{corr}} := \frac{1}{n} \sum_{\ell=1}^{n} 1_{\{T_B(\boldsymbol{X}^{\ell,Q_{\text{iter}}}) < T_A(\boldsymbol{X}^{\ell,Q_{\text{iter}}})\}} \simeq \mathbb{P}(\tau_B < \tau_A | \tau_{z_{\max}} < \tau_A).$$

$\hat{p}_{\text{corr}}$ is the proportion of trajectories reaching B before A at the last iteration Q_{iter}.

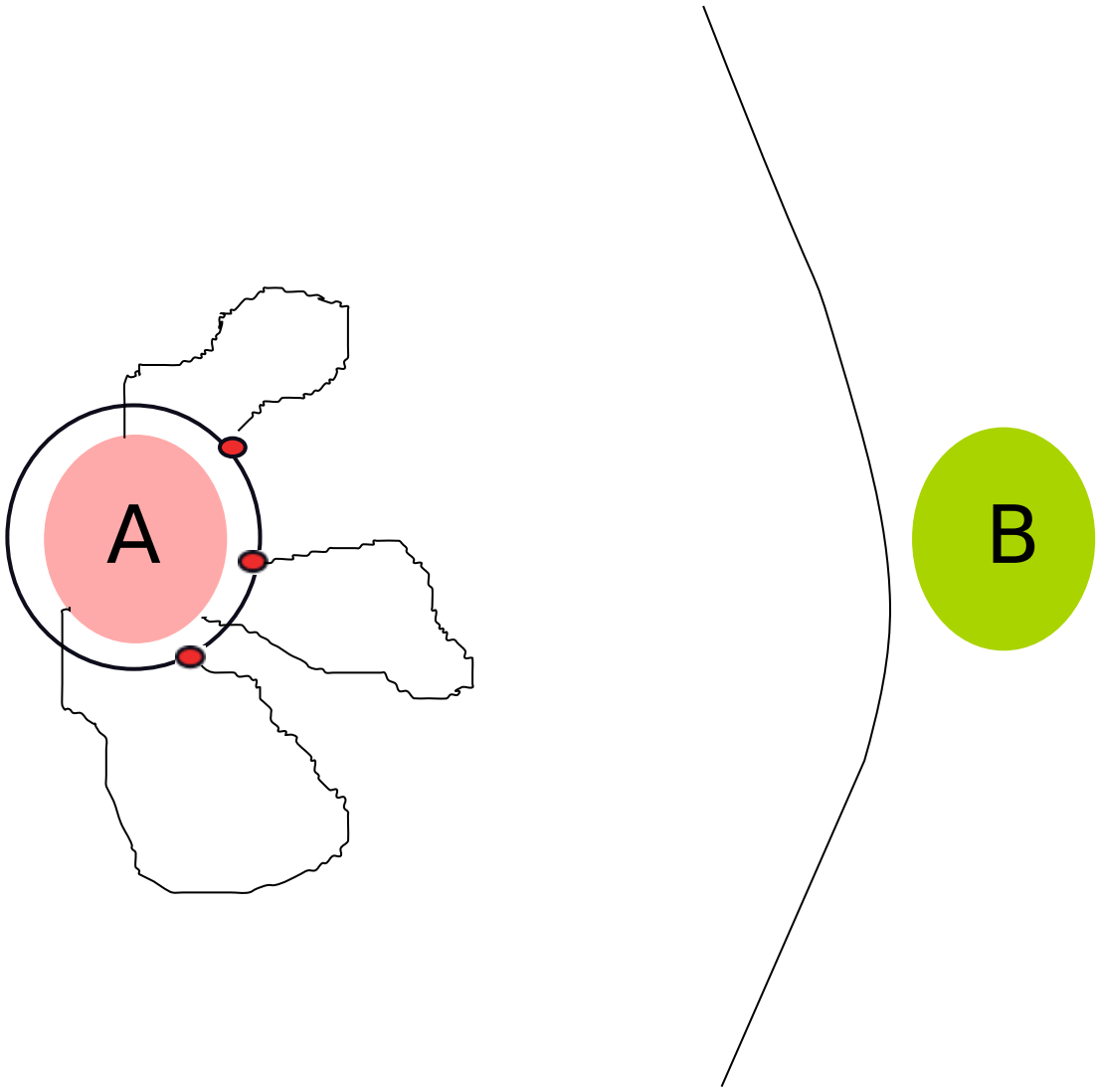
Therefore, an estimator of $\mathbb{P}(\tau_B < \tau_A)$ is

$$\left(1 - \frac{k}{n}\right)^{Q_{\text{iter}}} \hat{p}_{\text{corr}}.$$

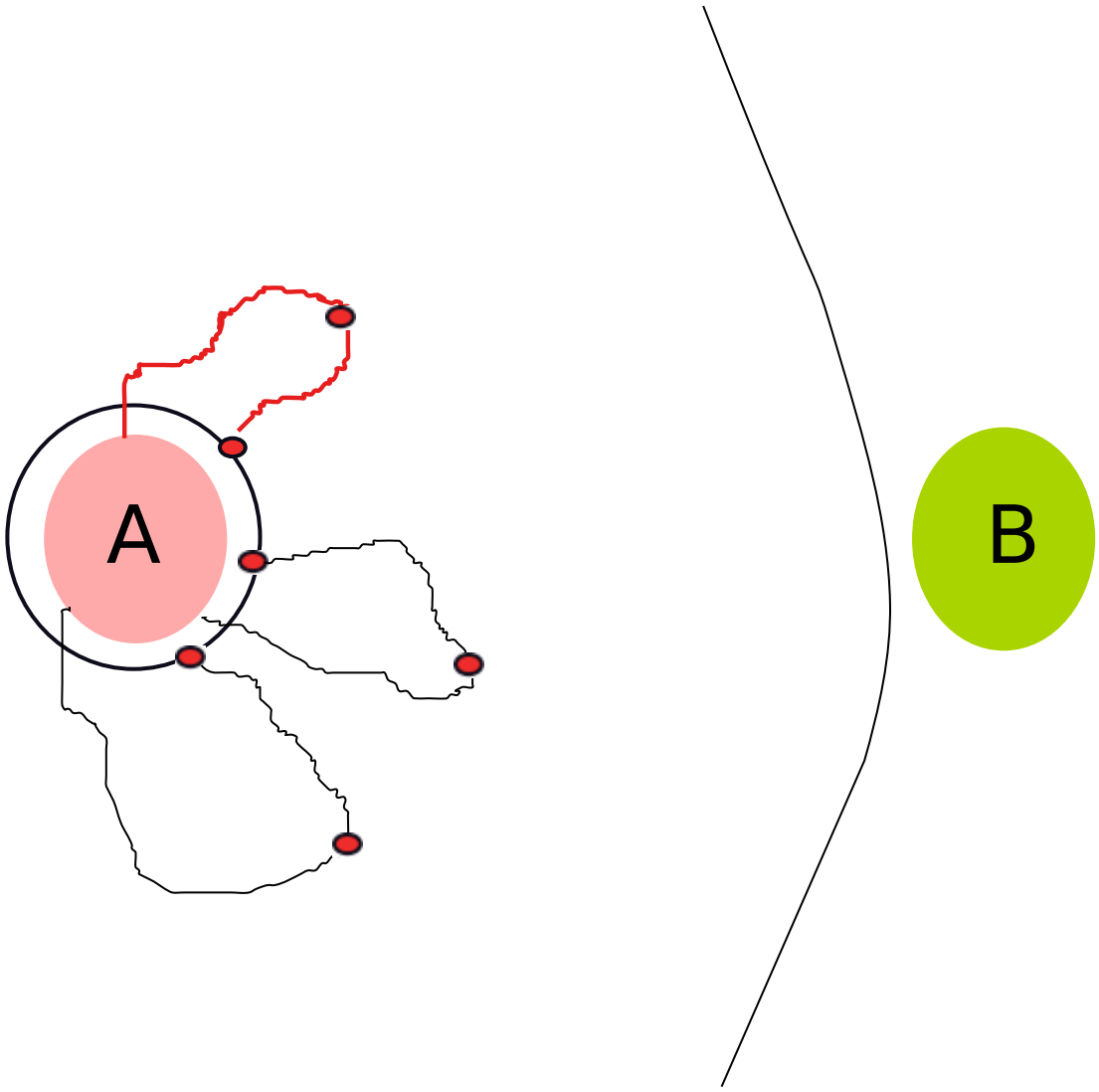
AMS Algorithm



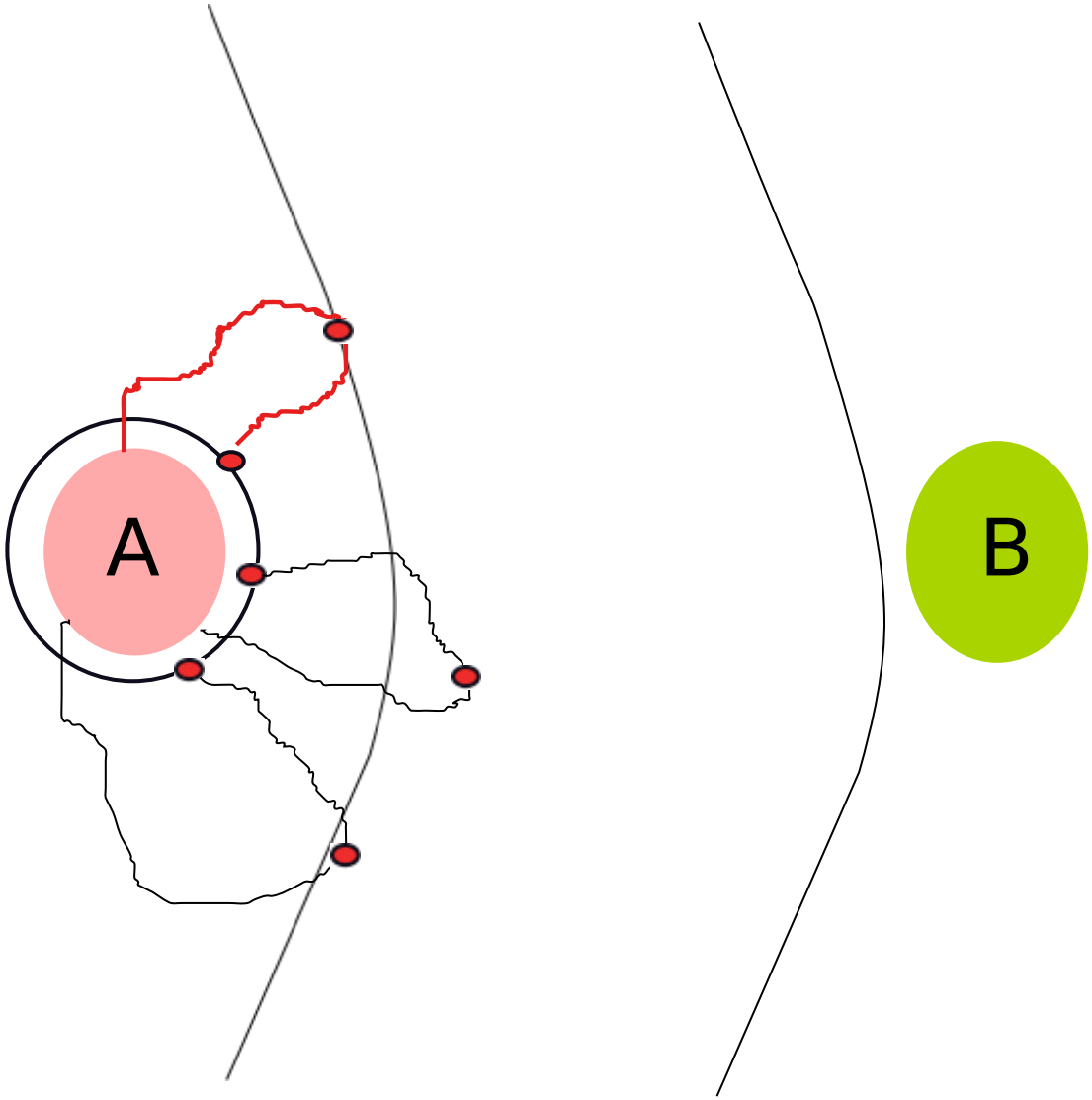
AMS Algorithm



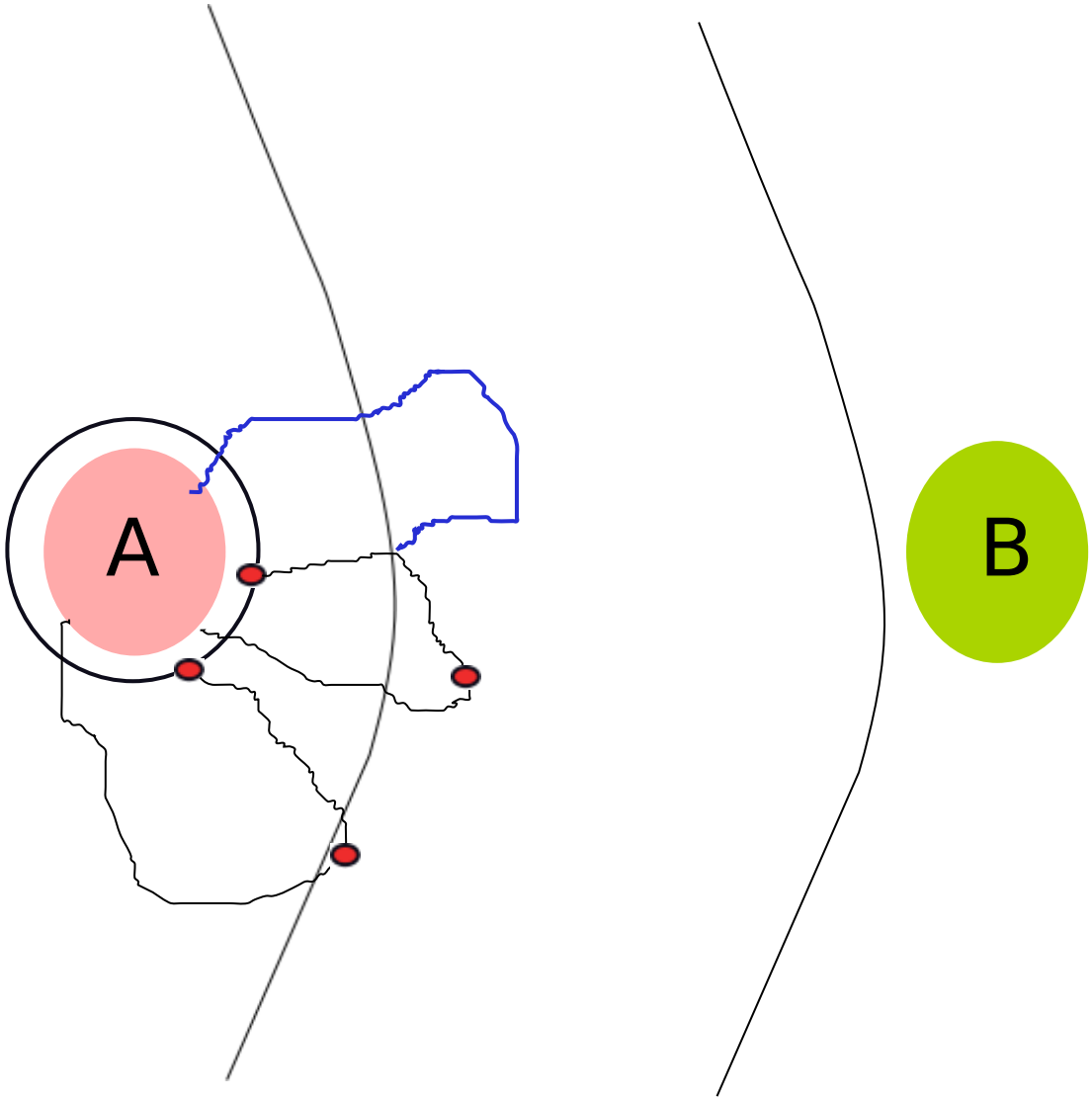
AMS Algorithm



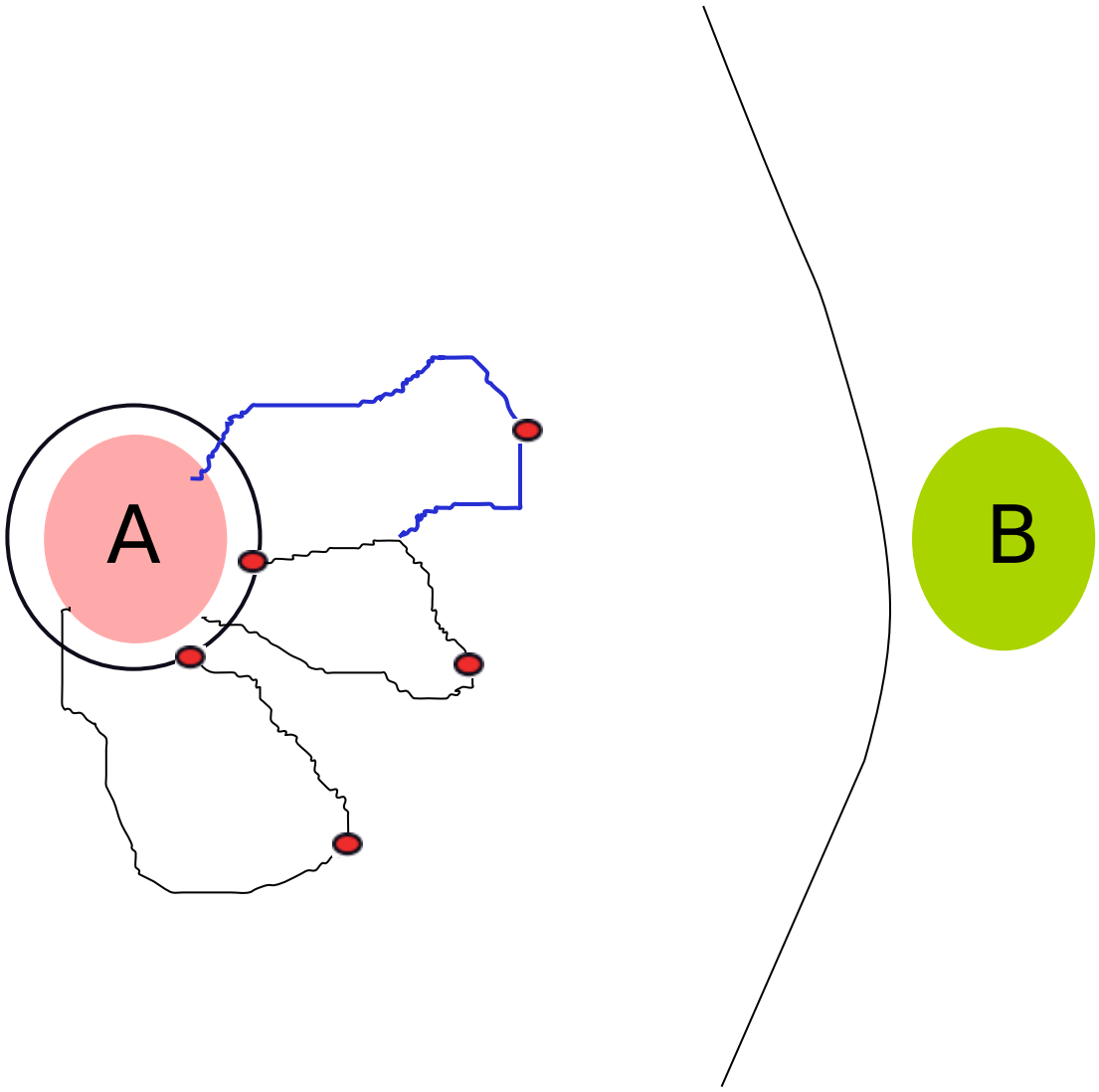
AMS Algorithm



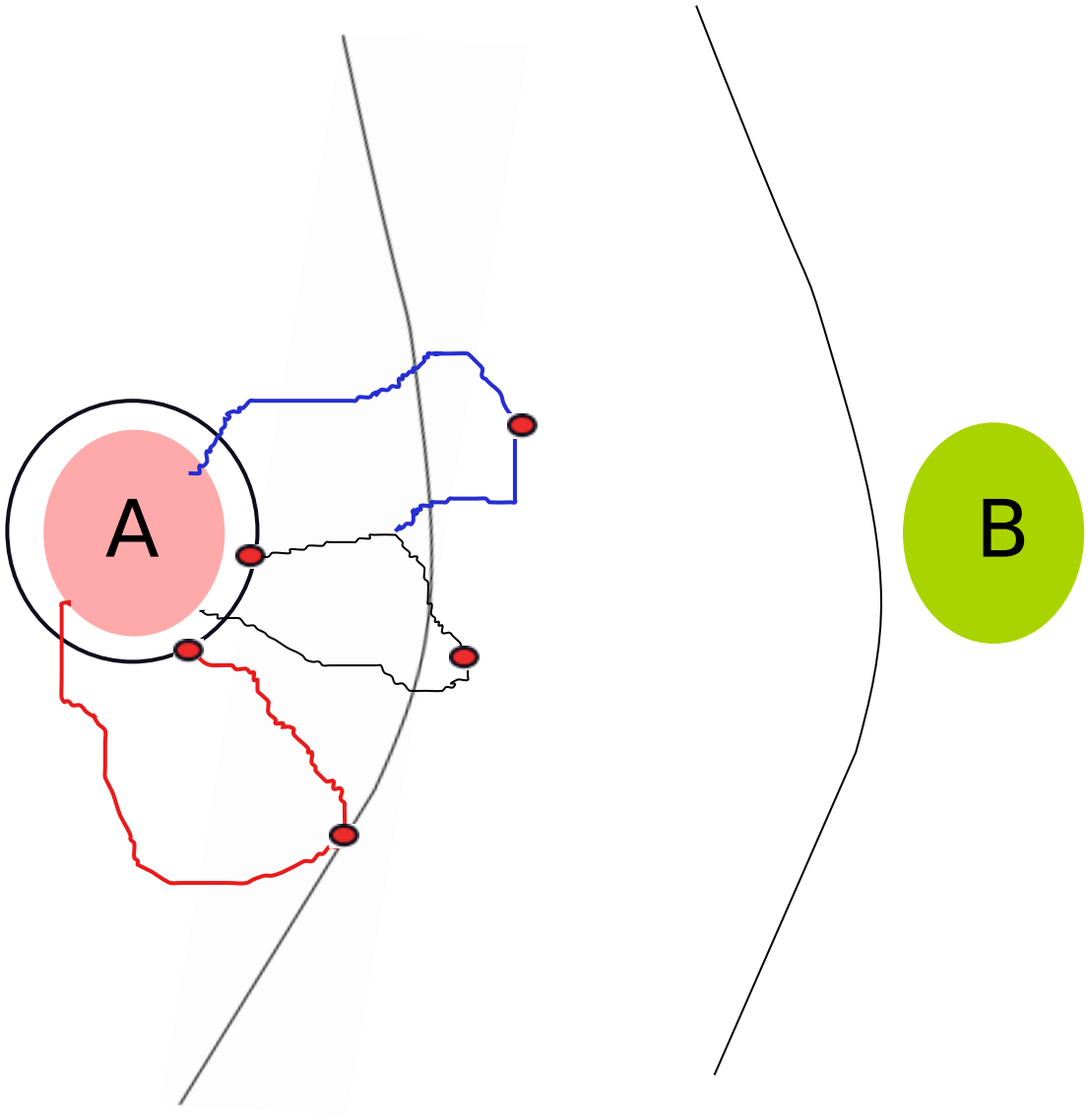
AMS Algorithm



AMS Algorithm



AMS Algorithm



AMS Algorithm: the case of Markov chains

In practice, the dynamics are *discrete in time* and thus, it may happen that more than k trajectories are such that

$$\sup_{t\geq 0} \xi(\boldsymbol{X}^{\ell}_{t\wedge\tau_A}) \leq \sup_{t\geq 0} \xi(\boldsymbol{X}^{(k)}_{t\wedge\tau_A}) =: z_q$$

In this case, all the trajectories with maximum level smaller or equal than z_q should be discarded.

The actual estimator of $\mathbb{P}(\tau_B < \tau_A)$ thus reads:

$$\hat{p} = \left(1 - \frac{K_1}{n}\right) \dots \left(1 - \frac{K_{Q_{\text{iter}}}}{n}\right) \hat{p}_{\text{corr}}$$

instead of $\left(1 - \frac{k}{n}\right)^{Q_{\text{iter}}} \hat{p}_{\text{corr}}$, where $K_q \geq k$ is the effective number of discarded trajectories at iteration q.

AMS Algorithm: unbiasedness

Theorem [C.-E. Bréhier, M. Gazeau, L. Goudenège, TL, M. Rousset, 2016]: For any choice of ξ, n and k,

$$\mathbb{E}(\hat{p}) = \mathbb{P}(\tau_B < \tau_A).$$

The proof is based on Doob's stopping theorem on a martingale built using filtrations indexed by the level sets of ξ. Actually, this result is proved for general path observables and in a much more general setting.

Practical counterparts:

- The algorithm is easy to parallelize.
- One can compare the results obtained with different reaction coordinates ξ to gain confidence in the results.

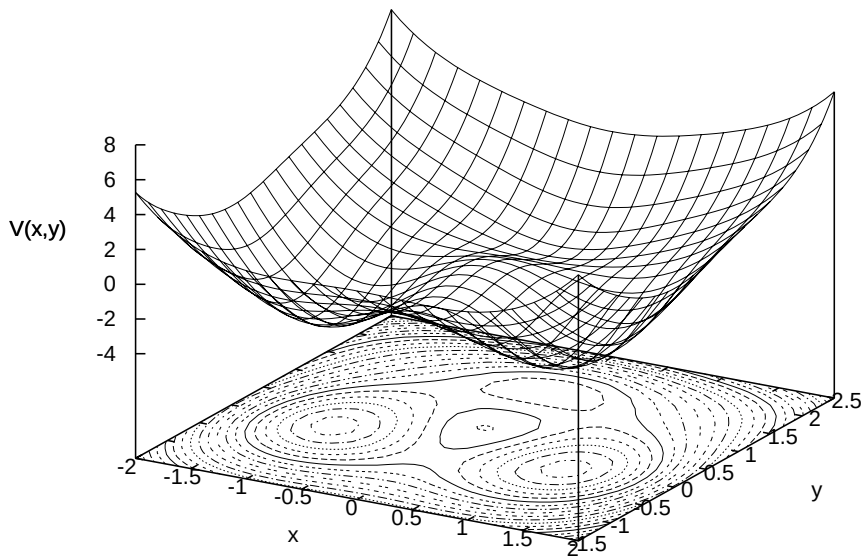
Numerical results: a 2D example

Time-discretization of the overdamped Langevin dynamics:

$$d\boldsymbol{X}_t = -\nabla V(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t$$

with a deterministic initial condition $\boldsymbol{X}_0 = \boldsymbol{x}_0$ and the 2D potential

[Park, Sener, Lu, Schulten, 2003] [Metzner, Schütte and Vanden-Eijnden, 2006]



$$V(x,y) = 3\mathrm{e}^{-x^2-(y-\frac{1}{3})^2} - 3\mathrm{e}^{-x^2-(y-\frac{5}{3})^2} - 5\mathrm{e}^{-(x-1)^2-y^2}$$
$$- 5\mathrm{e}^{-(x+1)^2-y^2} + 0.2x^4 + 0.2\left(y-\frac{1}{3}\right)^4.$$

A 2D example

The interest of this "bi-channel" potential is that, depending on the temperature, one or the other channel is prefered to go from A (around $H_- = (-1, 0)$) to B (around $H_+ = (1, 0)$).

Three reaction coordinates: $\xi^1(x, y) = \|(x, y) - H_-\|$, $\xi^2(x, y) = C - \|(x, y) - H_+\|$ or $\xi^3(x, y) = x$.

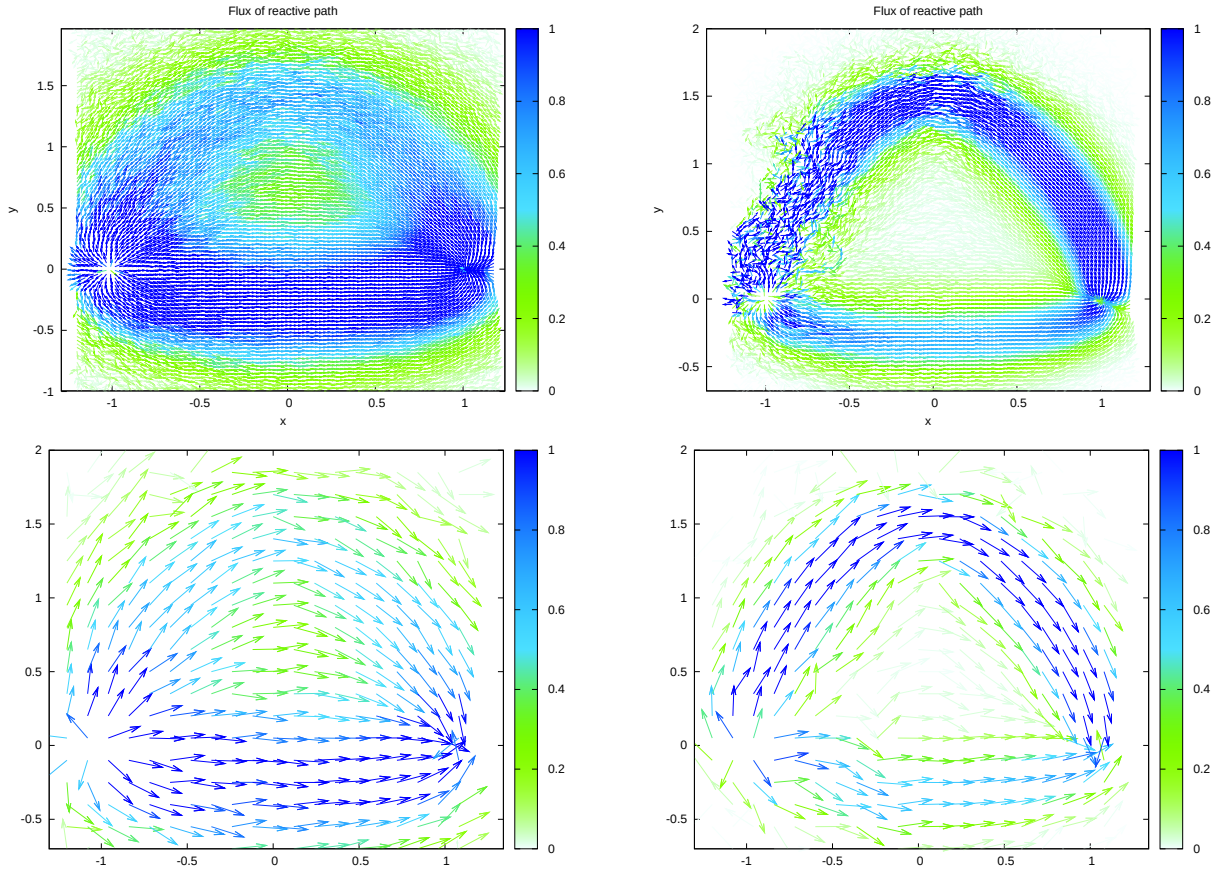
We plot as a function of the number N of independent realizations of AMS, the empirical average

$$\overline{p}_N = \frac{1}{N} \sum_{m=1}^{N} \hat{p}_m$$

together with the associated empirical confidence interval: $[\overline{p}_N - \delta_N/2, \overline{p}_N + \delta_N/2]$ where

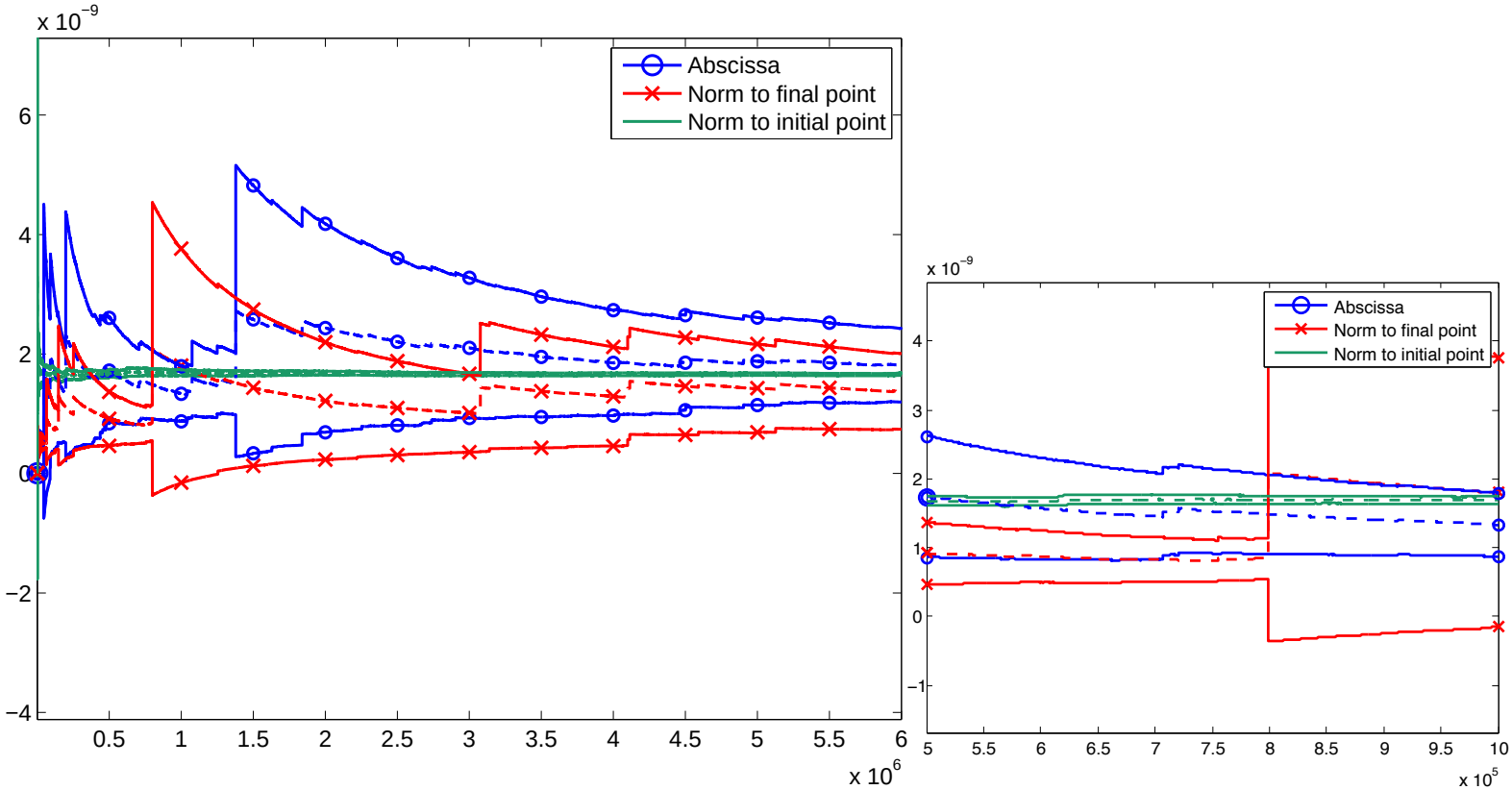
$$\delta_N = 2 \frac{1.96}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{m=1}^{N} (\hat{p}_m)^2 - (\overline{p}_N)^2}$$

A 2D example: flux of reactive trajectories

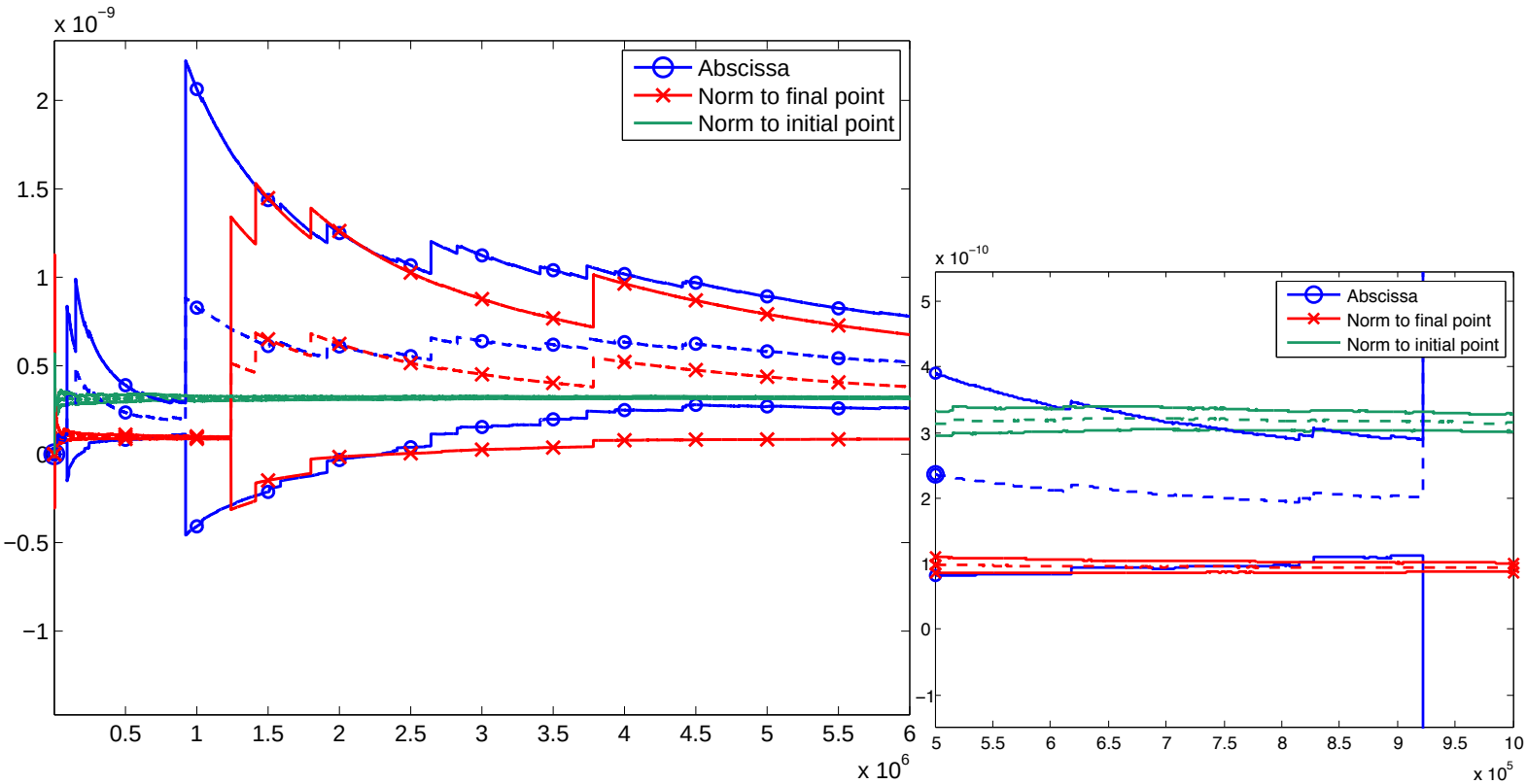


Flux of reactive trajectories, at $\beta = 1.67$ on the left, and $\beta = 6.67$ on the right.

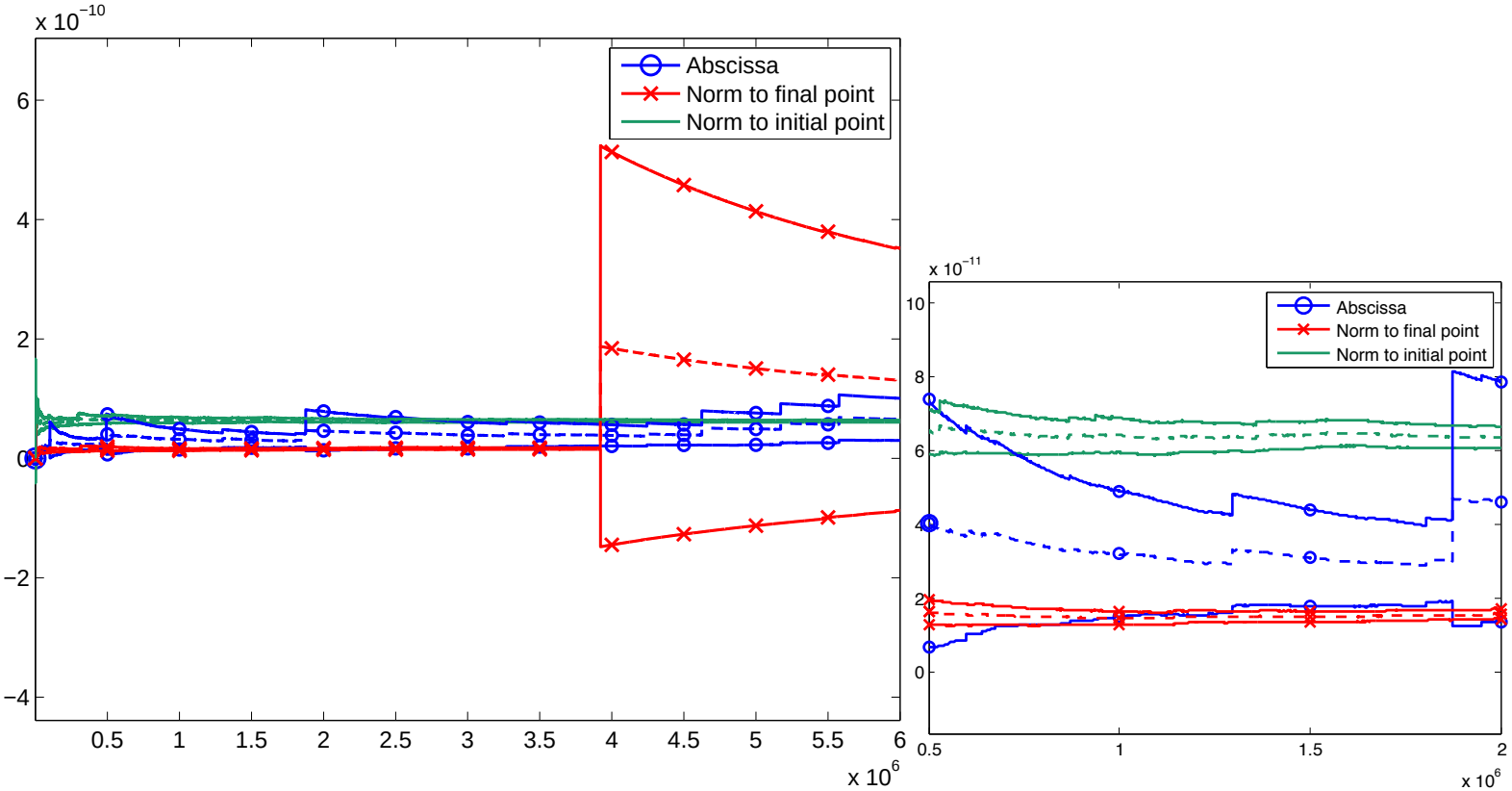
A 2D example: $k = 1$, $n = 100$, $\beta = 8.67$



A 2D example: $k = 1$, $n = 100$, $\beta = 9.33$



A 2D example: $k = 1$, $n = 100$, $\beta = 10$



A 2D example

Observations:

- When N is sufficiently large, confidence intervals overlap.
- For too small values of N, "apparent bias" is observed [Glasserman, Heidelberger, Shahabuddin, Zajic, 1998].
- Fluctuations depend a lot on ξ.

$\longrightarrow$ To gain confidence in the results, check that the estimated quantity is approximately the same for different ξ's.

"Apparent bias" phenomenon

The apparent bias is due to the fact that [Glasserman, Heidelberger, Shahabuddin, Zajic, 1998]:

- Multiple pathways exist to go from A to B.
- Conditionally to reach Σ_z before A, the relative likelihood of each of these pathways depends a lot on z.

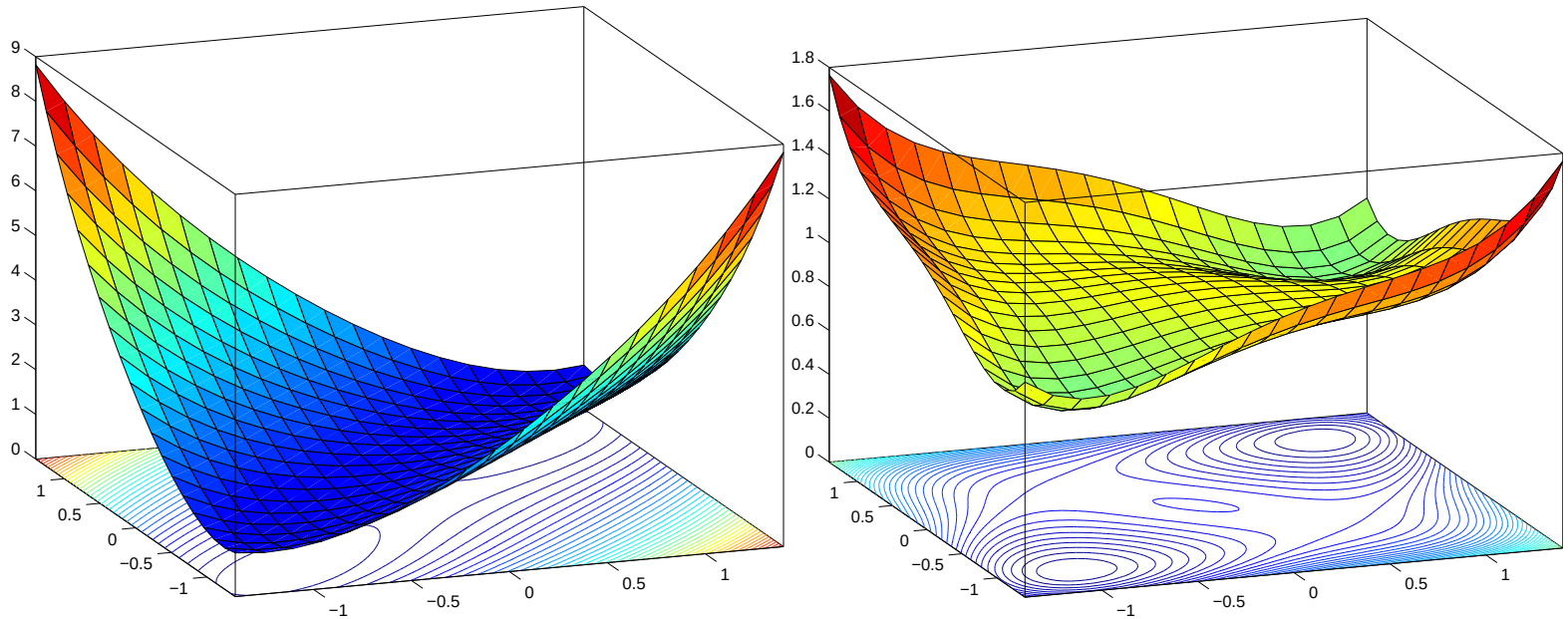
On our example, for small n, we indeed observe that (for ξ^3):

- Most of the time, all replicas at the end go through only one of the two channels (two possible scenarios).
- One of this scenario is rare.
- The values of $\hat{p}$ associated to each of these two scenarios are very different.

This explains the large fluctuations.

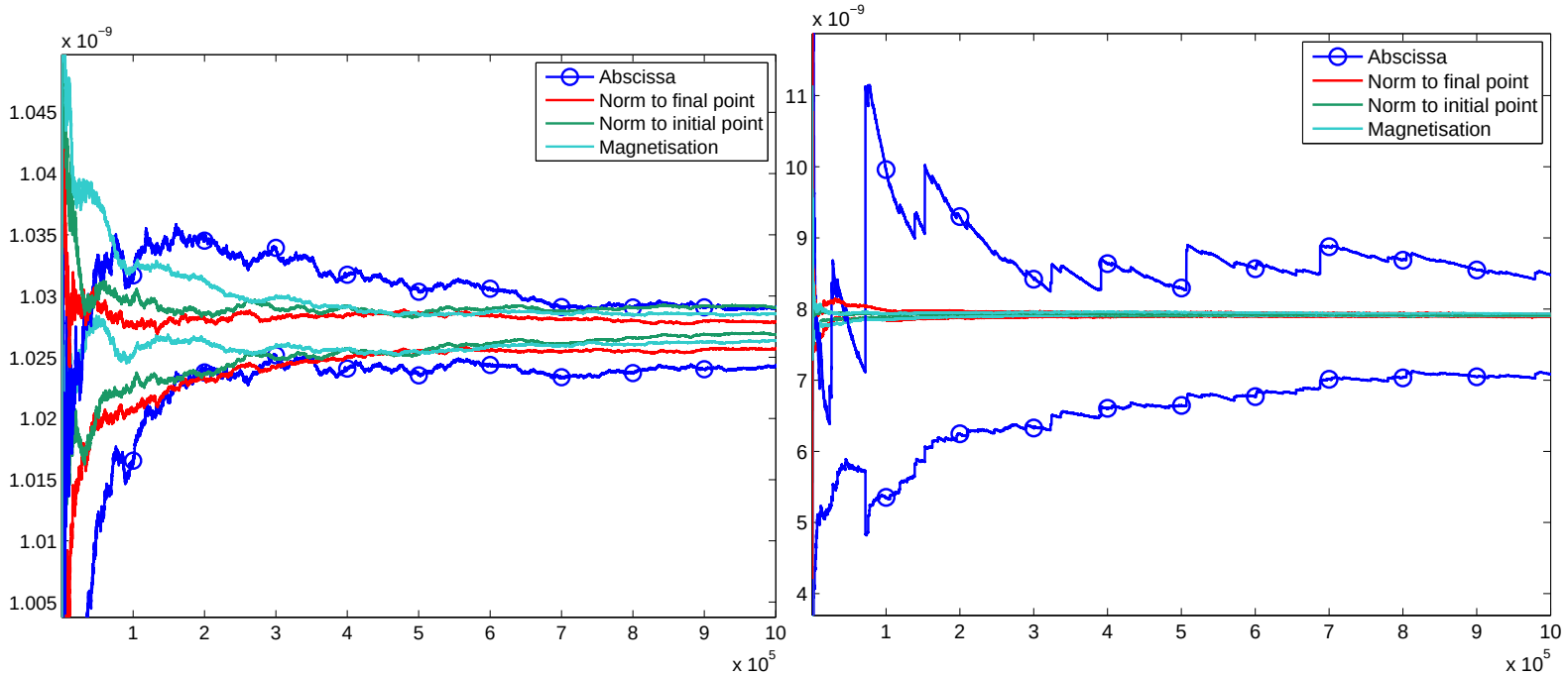
"Apparent bias" phenomenon

Another 2D test case:



Potential $V_\gamma(x, y)$.
Left: $\gamma = 1$ (one channel); right: $\gamma = 0.1$ (two channels).

"Apparent bias" phenomenon



Parameters: $k = 1$, $n = 100$ and $\beta = 80$.

Left: $\gamma = 1$ (one channel). Right: $\gamma = 0.1$ (two channels).

Computing transition times

To use the algorithm to compute transition times, we split a transition path from A to B into: excursions from ∂A to $\Sigma_{z_{\min}}$ and then back to ∂A, and finally an excursion from ∂A to $\Sigma_{z_{\min}}$ and then to B. Assuming that A is metastable ($p \ll 1$), one obtains that the mean transition time is:

$$\mathbb{E}(T) = \left(\frac{1}{p} - 1\right) \mathbb{E}(T_1 + T_2) + \mathbb{E}(T_1 + T_3)$$

where:

- p is the probability, once $\Sigma_{z_{\min}}$ has been reached, to go to B rather than A (approximated by $\hat{p}_N$) ;
- $\mathbb{E}(T_1 + T_2)$ is the mean time for an excursion from ∂A to $\Sigma_{z_{\min}}$ and then back to ∂A (approximated by DNS) ;
- $\mathbb{E}(T_1 + T_3)$ is the mean time for an excursion from ∂A to $\Sigma_{z_{\min}}$ and then to B (approximated by the AMS algorithm).

A 1D example

We consider the double-well potential:

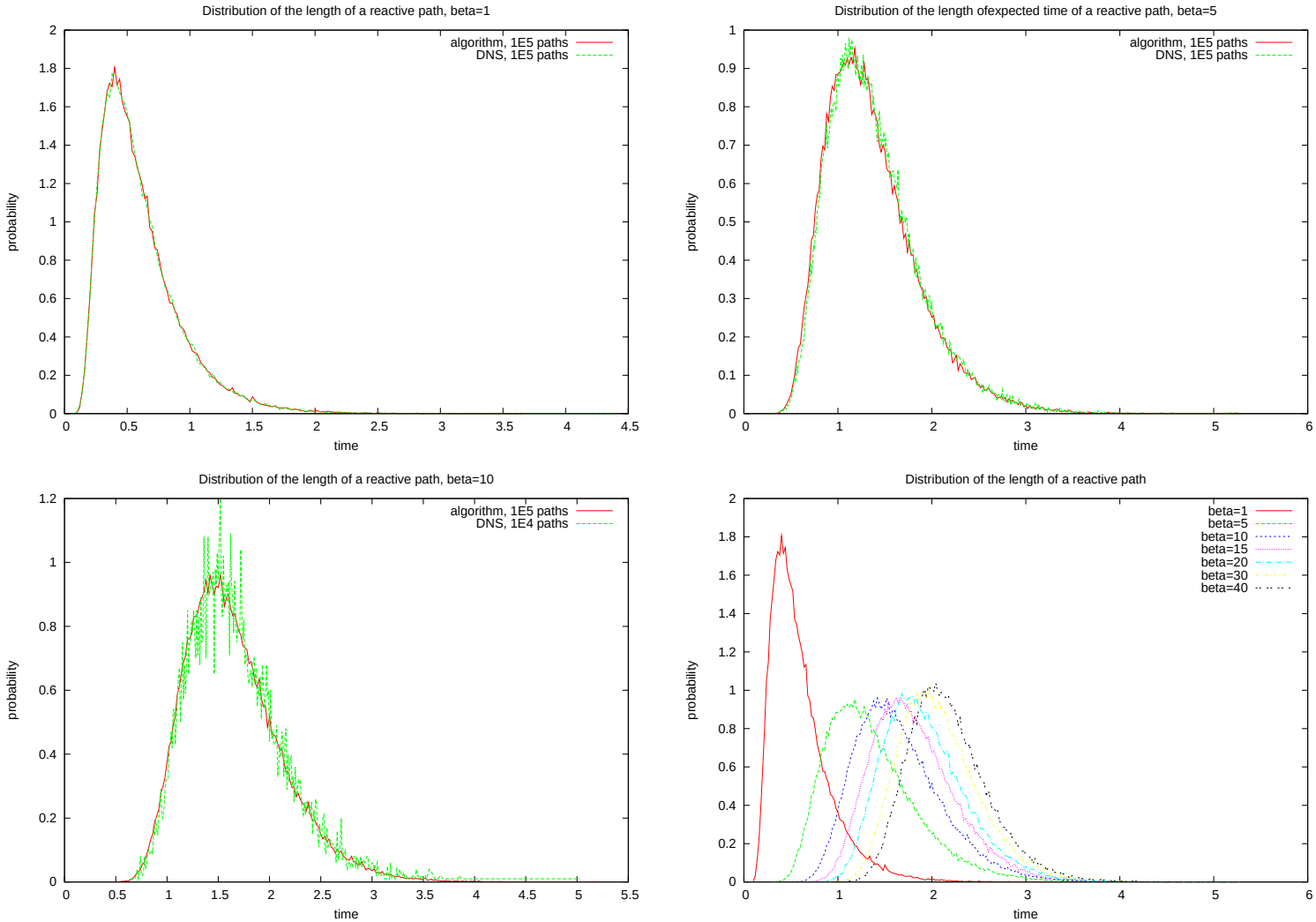
$$V(x) = x^4 - 2x^2,$$

which has two minima at ± 1 and one saddle point at 0.
In this simple one dimensional setting, we set as metastable states $A = \{-1\}$ and $B = \{+1\}$, and the reaction coordinate is taken to be simply

$$\xi(x) = x.$$

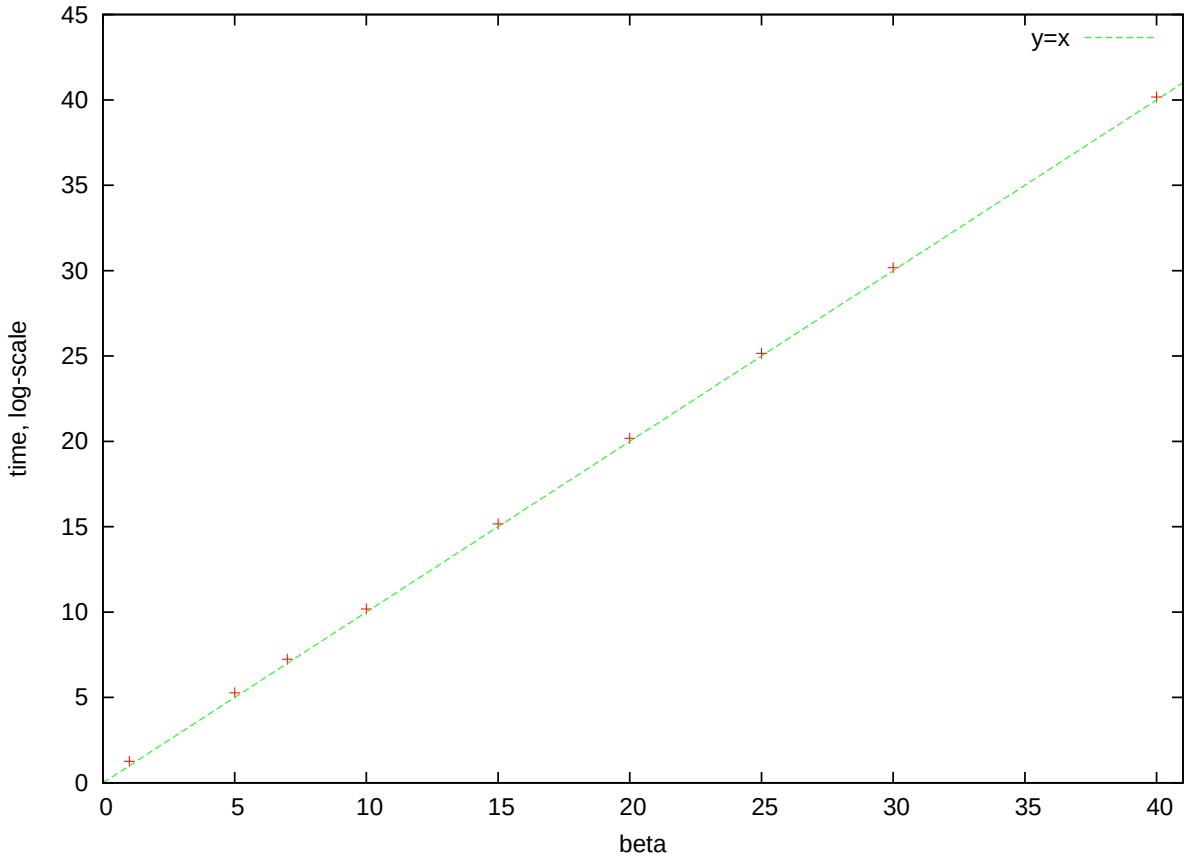
To test the consistency of the algorithm, we compute the distribution of the time-lengths of the reactive paths and compare to DNS (when possible).

A 1D example



Distributions of time-lengths of reactive paths. Comparison with DNS for $\beta = 1, 5, 10$ and distributions for $\beta = 1, 5, 10, 15, 20, 30, 40$.

A 1D example



Comparison of the estimated mean transition times as a function of β with the asymptotic law from large deviation theory:

$$\mathbb{E}(T) \propto \exp(\beta \Delta V)$$

where $\Delta V = 1$ is the height of the energy barrier.

The 2D case

N $\times 10^3$	β	k_{AB} (AMS)	C.I. on k_{AB} (AMS)
2	1.67	$2.03\ 10^{-2}$	$[1.83; 2.22]\ 10^{-2}$
10	1.67	$1.84\ 10^{-2}$	$[1.82; 1.86]\ 10^{-2}$
50	1.67	$1.88\ 10^{-2}$	$[1.87; 1.88]\ 10^{-2}$
100	1.67	$1.89\ 10^{-2}$	$[1.89; 1.90]\ 10^{-2}$
2	6.67	$9.97\ 10^{-8}$	$[7.74; 12.2]\ 10^{-8}$
10	6.67	$9.20\ 10^{-8}$	$[7.71; 10.7]\ 10^{-8}$
50	6.67	$8.88\ 10^{-8}$	$[8.42; 9.34]\ 10^{-8}$
100	6.67	$9.32\ 10^{-8}$	$[9.08; 9.57]\ 10^{-8}$

Estimates of the reaction rate $k_{AB} = 2/\mathbb{E}(T)$, with $\xi = \xi_2$. Values from [Metzner, Schütte, Vanden-Eijnden, 2006] are $k_{AB} = 1.912\ 10^{-2}$ for $\beta = 1.67$ and $k_{AB} = 9.47\ 10^{-8}$ for $\beta = 6.67$.

Recent results using NAMD

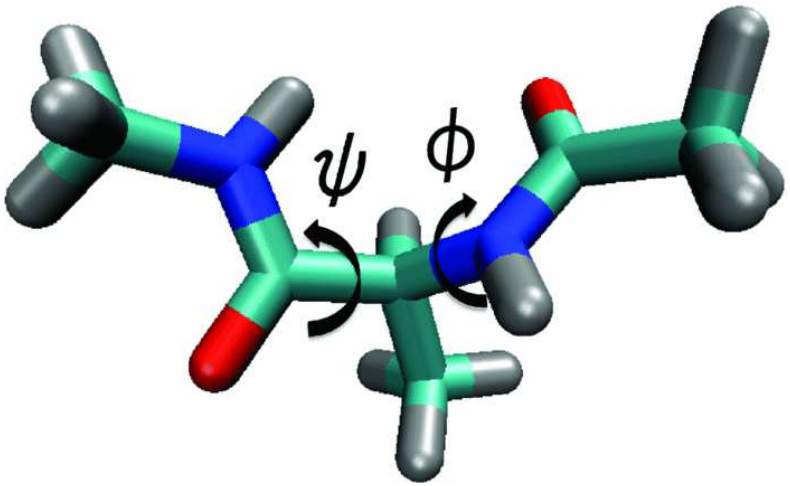
We are currently implementing AMS in the NAMD software

(collaboration with SANOFI, C. Mayne and I. Teo, PhD of L. Silva Lopes).

Three test cases:

- Alanine di-peptide (test case)
- β-cyclodextrin (in progress)
- benzamidine-trypsin dissociation rate

Alanine di-peptide (1/5)

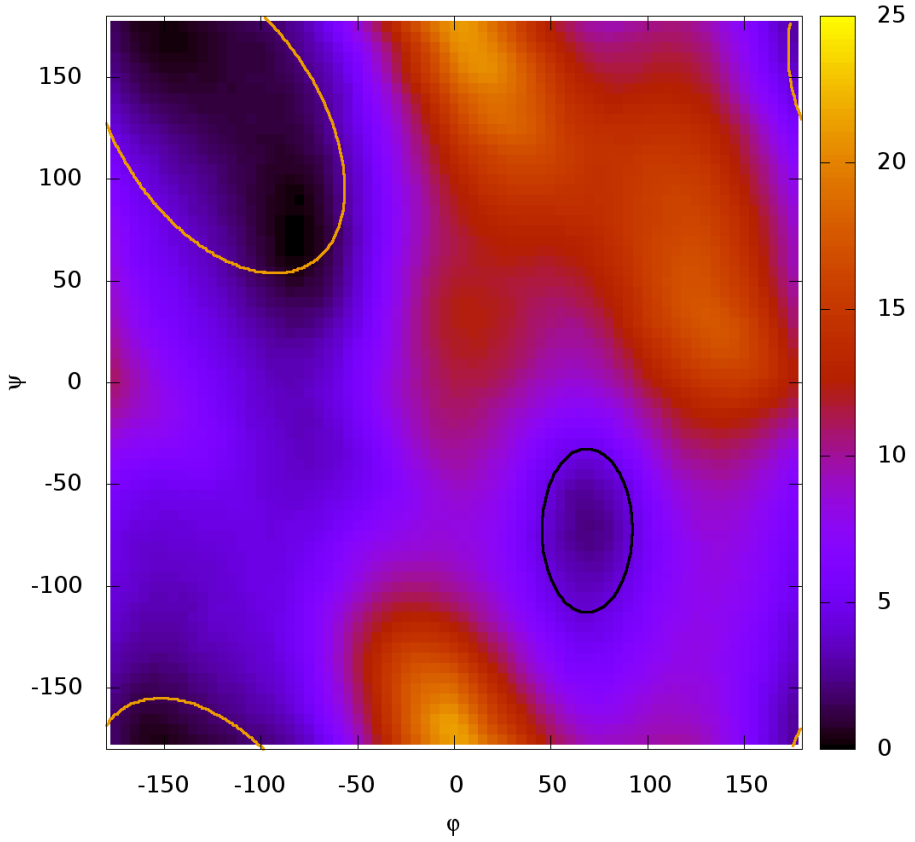


Two reaction coordinates:

- ξ_1 is a continuous piecewise affine function of φ
- $\xi_2(\varphi, \psi) = \min(d_A(\varphi, \psi), 6.4) - \min(d_B(\varphi, \psi), 3.8)$

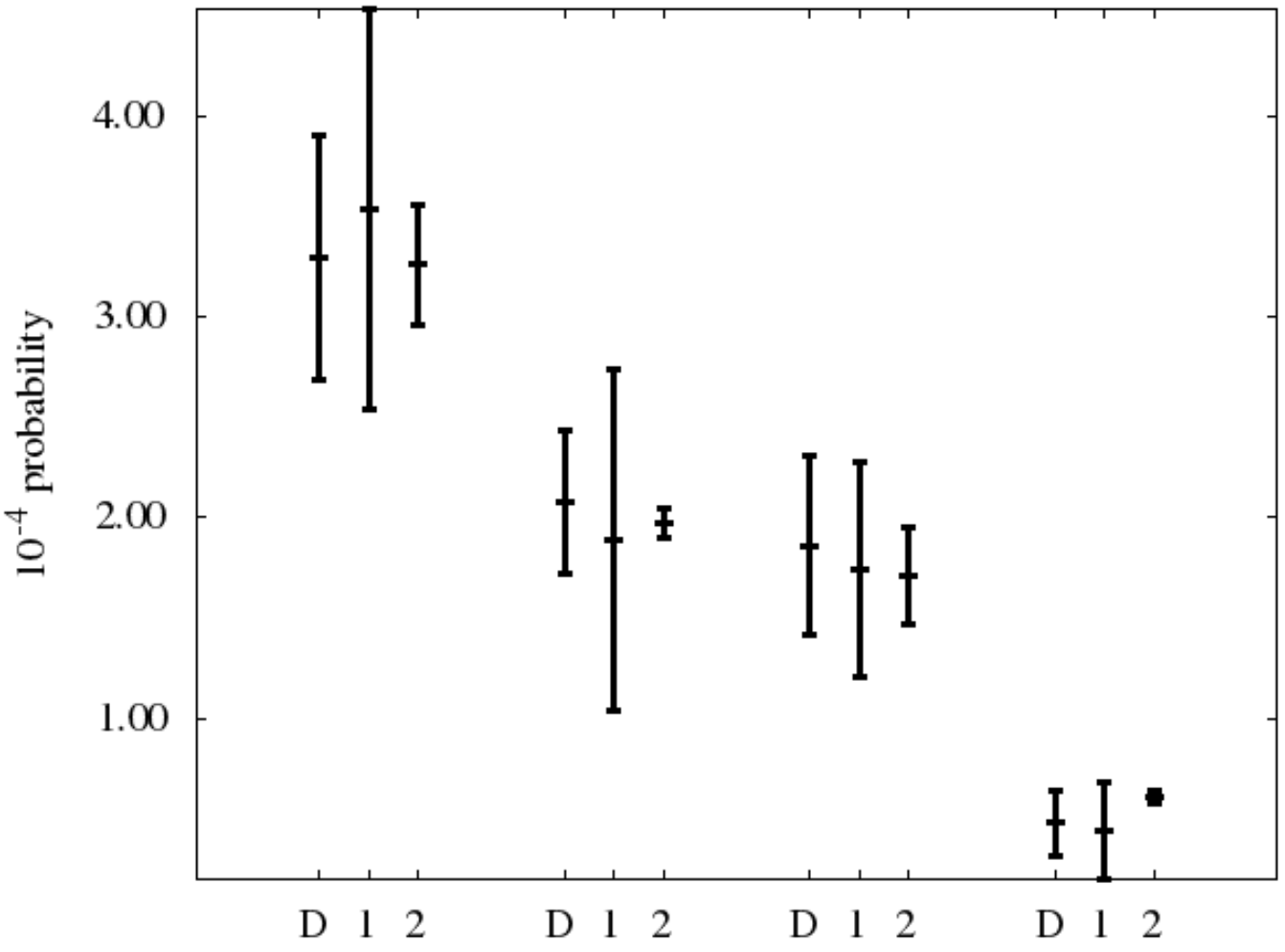
Computational setting: no solvent, force field: CHARMM27. AMS with $n = 500$ to 1000 replicas and $k = 1$.

Alanine di-peptide (2/5)



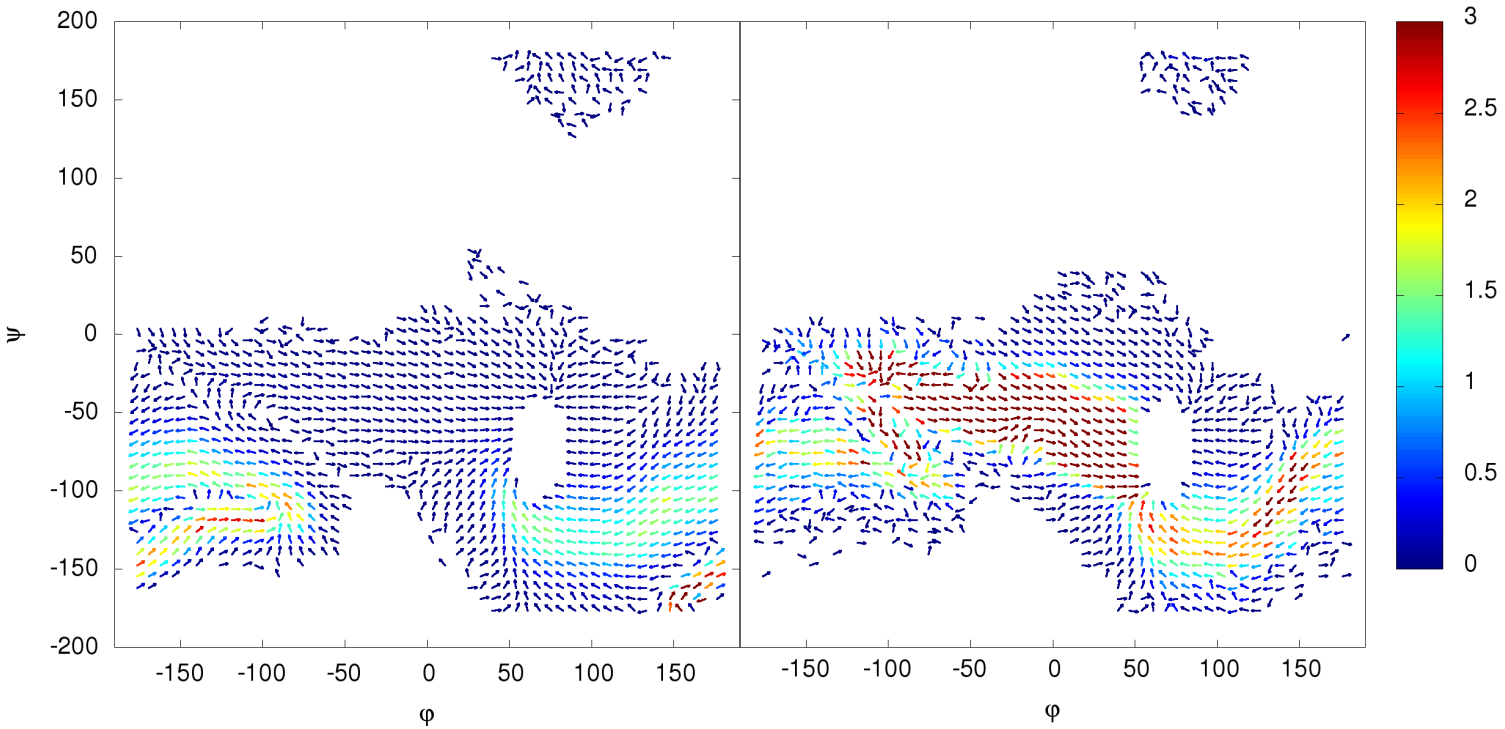
Free energy landscape and zones A (yellow) and B (black).

Alanine di-peptide (3/5)



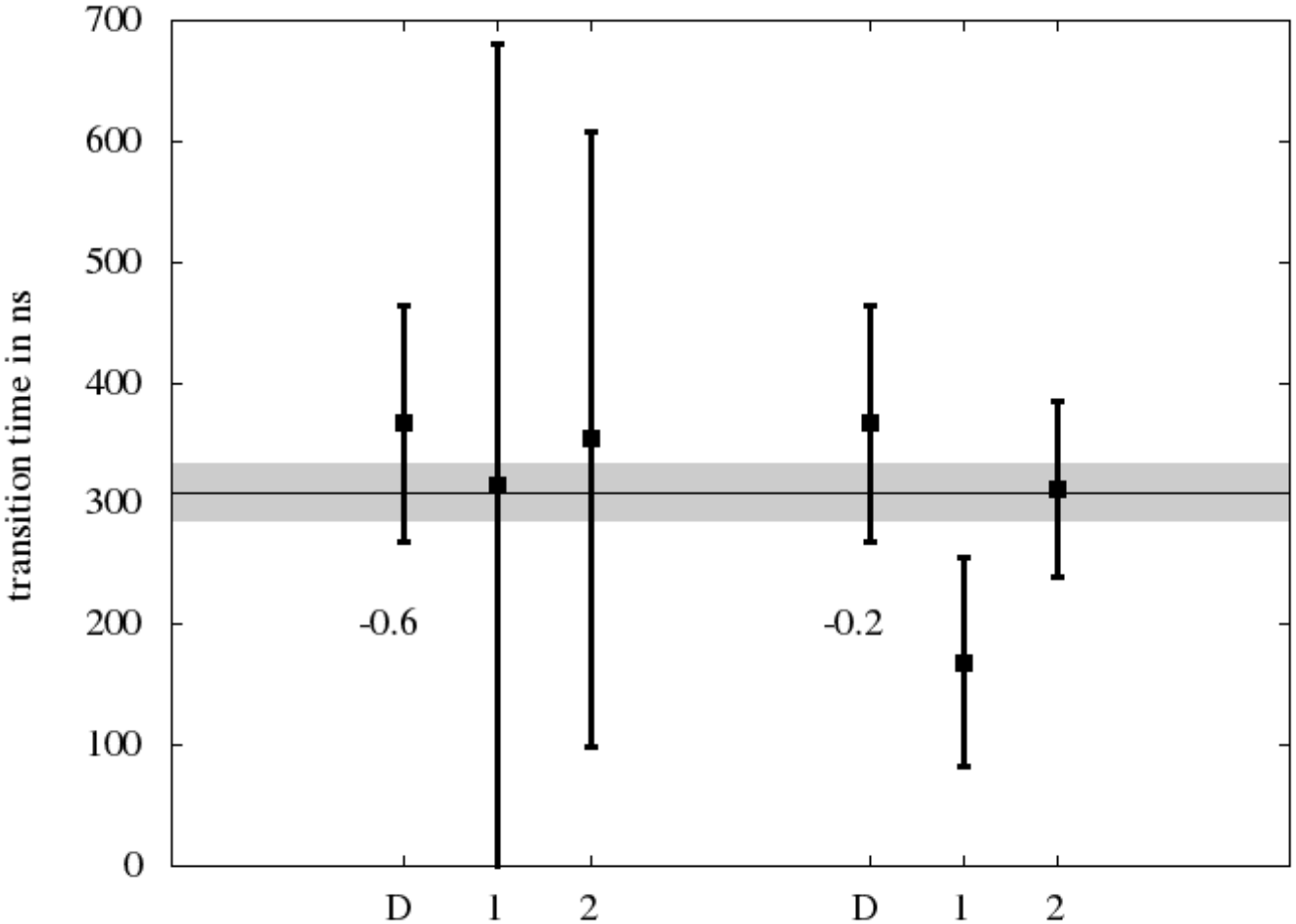
Probability estimations using different initial conditions: D=DNS, 1=ξ_1, 2=ξ_2.

Alanine di-peptide (4/5)



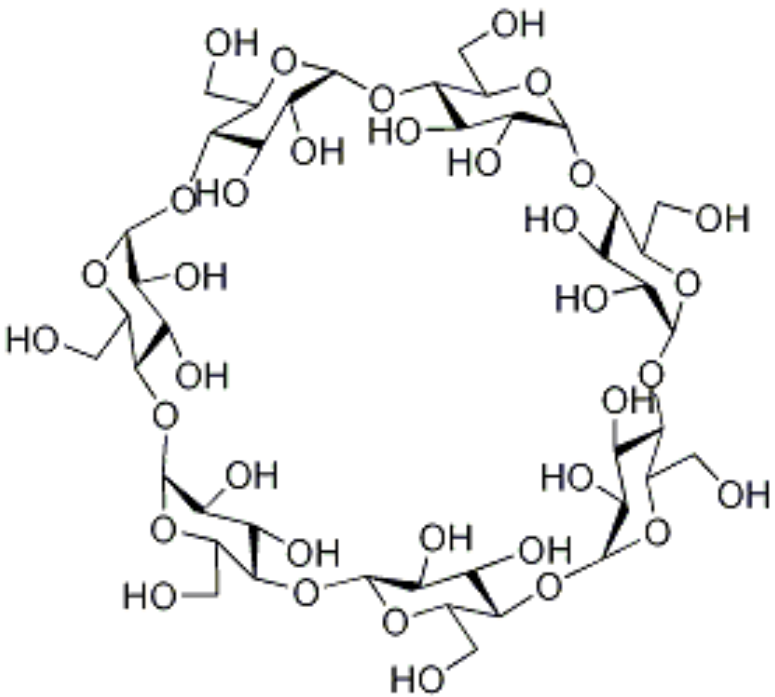
Flux of reactive trajectories, starting from two different initial conditions.

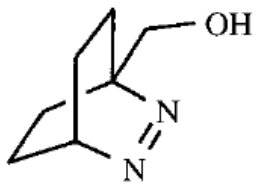
Alanine di-peptide (5/5)



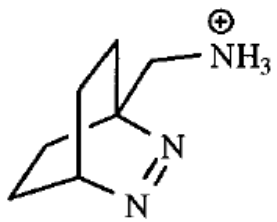
Transition time obtained for two values of z_{min}: D=DNS, 1=ξ_1, 2=ξ_2. Reference value obtained by DNS over a 97 DNS simulations of $2\mu s$.

β-cyclodextrin (1/3)





1d



1f

The β-cyclodextrin with ligands d and f.

β-cyclodextrin (2/3)

$\longrightarrow$ One reactive path obtained using AMS of the ligand 1f leaving the cyclodextrin.

Computational setting: explicit solvent, about 6000 atoms, 2fs time step, AMBER force field, parametrization with GAFF. AMS with $n = 50$ replicas and $k = 1$.

β-cyclodextrin (3/3)

Estimate of the dissocation rate with AMS:

- ligand d: 1.36e+08 ($\pm$ 1.e+08) fs (30 runs of AMS)
- ligand f: 2.65e+06 ($\pm$ 6.23+05) fs (57 runs of AMS)

Experimental values (from [X. Zhang, G. Gramlich, X. Wang and W. M. Nau, J. Am. Chem. Soc., 124 (2) 2002]):

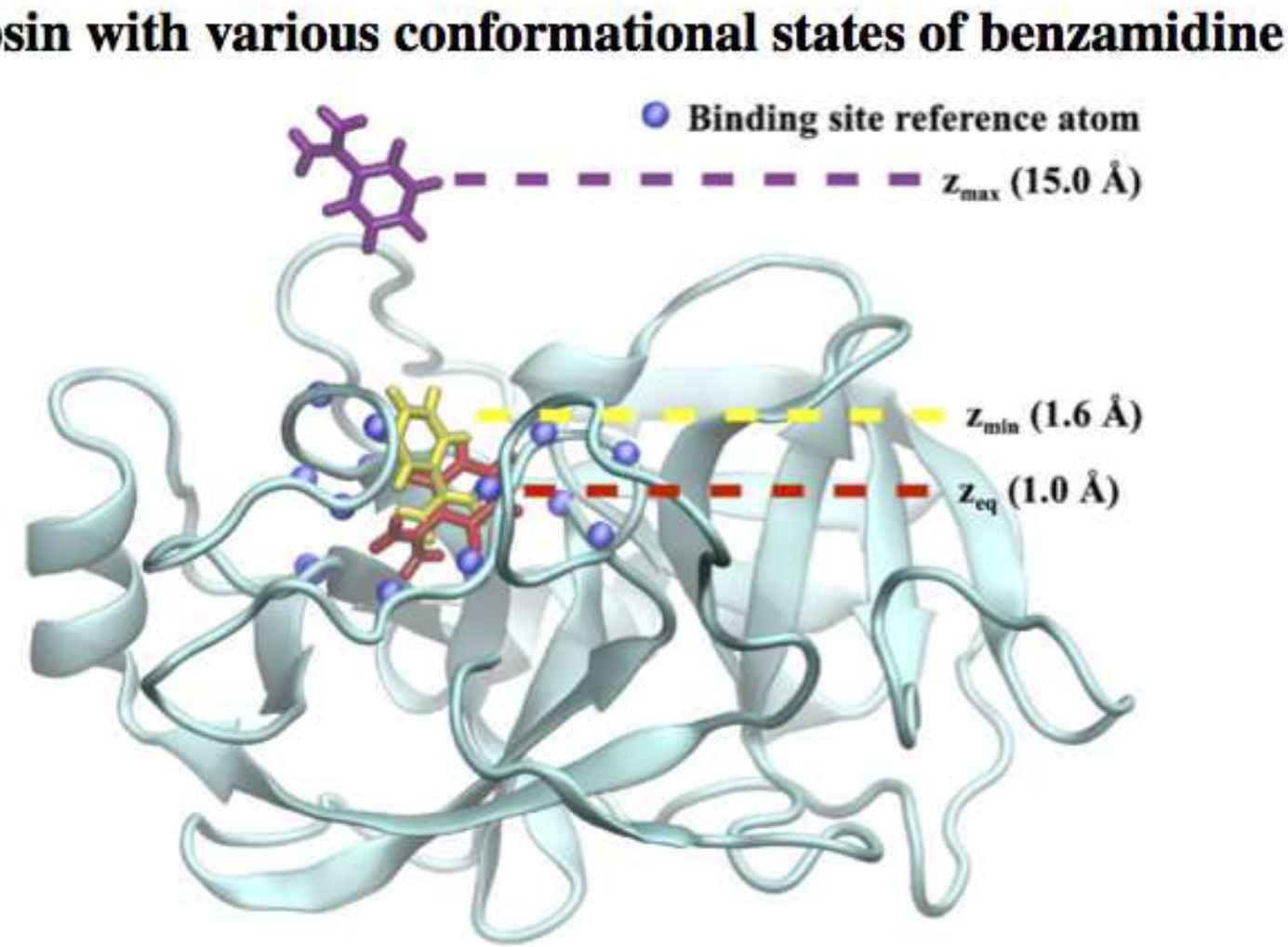
- ligand d: $5.55e+08$ fs
- ligand f: $8.33e+07$ fs

Computational times: AMS gives 50 trajectories, with a total simulation time = 4.4e+07 fs (d) ; 3.5e+07 fs (f).

... to be continued

Benzamidine-trypsin (1/2)

We recently used AMS to estimate the off rate of benzamidine from trypsin [I. Teo, C. Mayne, K. Schulten and TL, 2016].



Benzamidine-trypsin (2/2)

We obtain a dissociation rate $k_{\text{off}} = (260 \pm 240) s^{-1}$ within the same order of magnitude as the experimentally measured rate $(600 \pm 300) s^{-1}$.

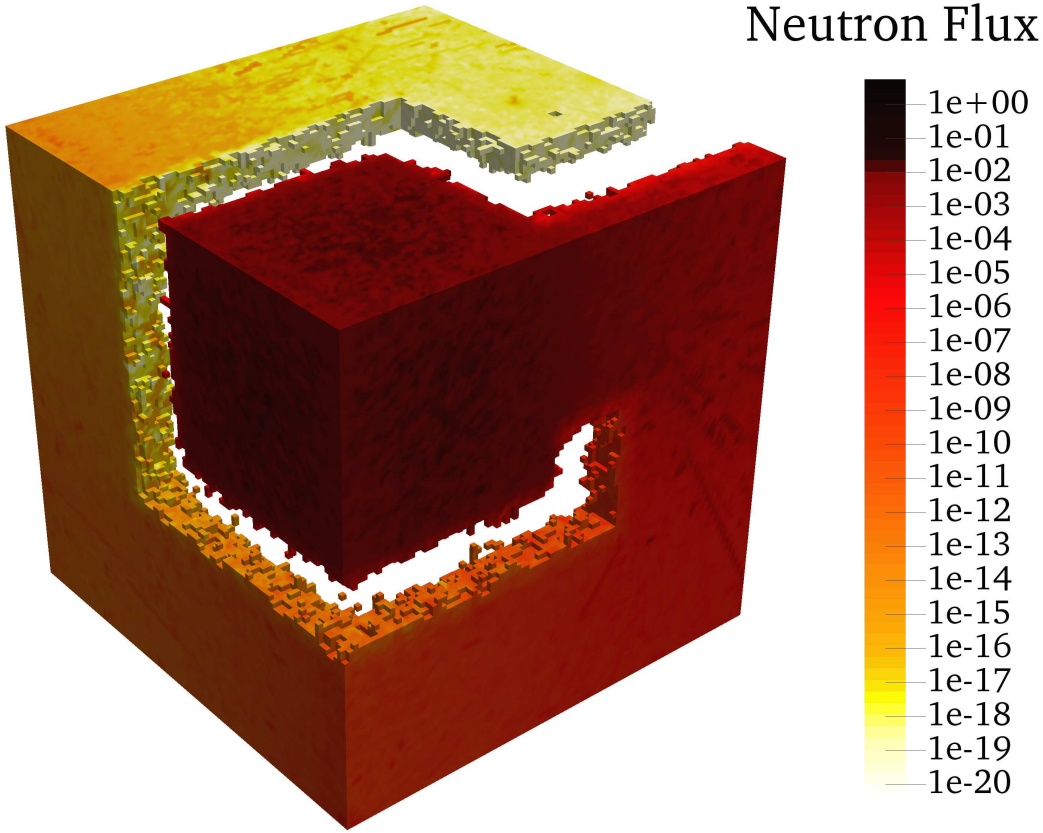
The overall simulation time taken, summed over all 1000 replicas, was $2.1 \mu s$ ($2.3 \mu s$ after including direct MD and steered MD simulations), which is four orders of magnitude shorter than the estimated dissociation time of one event.

The main practical difficulty seems to be the determination of a 'good' domain A.

Computational setting: 68 789 atoms, with 21 800 water molecules, 62 sodium ions, and 68 chloride ions. Water: TIP3P model. CHARMM36 force field, with parameters for benzamidine obtained from the CGenFF force field. NPT conditions, at 298 K and 1 atm Langevin thermostat and barostat settings, using 2 fs time steps. AMS with $n = 1000$ replicas and $k = 1$.

Another application

The AMS algorithm in other context: in collaboration with CEA, AMS is currently implemented in the Tripoli code for nuclear safety application.



Conclusions (1/2)

Practical recommendations:

- A careful implementation of the splitting step leads to unbiased estimators for non-normalized quantities.
- Perform many independent realizations of AMS.
- Use ξ as a numerical parameter.

The algorithm is very versatile:

- Non-intrusivity: the MD integrator is a black box.
- Can be applied both to entropic and energetic barriers.
- Can be adapted to generate trajectories of any stopped process.
- Other models: non-homogeneous Markov process, random fields, ...
- Algorithmic variants: other resampling procedure, additional selection, ...

Conclusions (2/2)

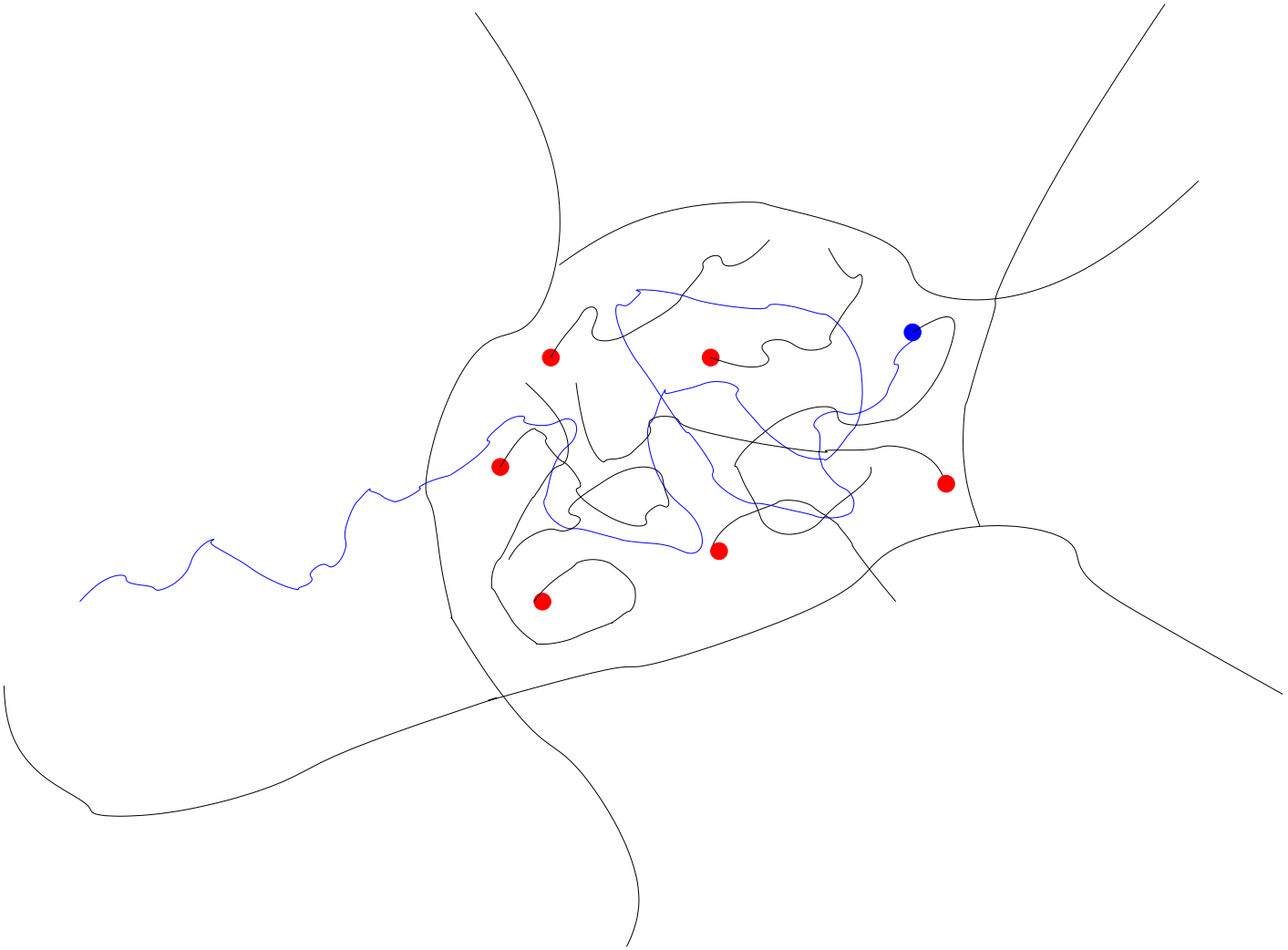
Works in progress:

- Implementation in the NAMD software (collaboration with SANOFI, C. Mayne and I. Teo), and in TRIPOLI (collaboration with CEA).
- Adaptive computation of better and better ξ.
- Analysis of the efficiency as a function of ξ. For optimal choice of ξ, the cost of AMS is (for n large)

$$\left((\log p)^2 - \log p\right)$$

much better than the cost of naive Monte Carlo: $\frac{1-p}{p}$. How does this degrade when ξ departs from the optimal case ?

Accelerated dynamics



Accelerated dynamics

The aim of this part is twofold:

- First, discuss a numerical method to efficiently sample metastable dynamics: the parallel replica method proposed by A. Voter.
- Second, justify rigorously kinetic Monte Carlo models which are used to simulate metastable dynamics over long times using a jump process between metastable states.

Both analysis will be based on the notion of quasi-stationary distribution.

Outline of this part:

- Introduction of the quasi-stationary distribution
- The Parallel Replica algorithm
- Precise asymptotic results of the first exit point density: justifying kinetic Monte Carlo models, the Harmonic Transition State Theory and the Eyring-Kramers formulas.

Accelerated dynamics

The bottom line of the accelerated dynamics proposed by A. Voter in the late 90's is to get efficiently the state-to-state dynamics. Three algorithms: Parallel replica, Hyperdynamics, Temperature Accelerated Dynamics.

Let us consider the overdamped Langevin dynamics:

$$d\boldsymbol{X}_t = -\nabla V(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t$$

and let assume that we are given a mapping

$$\mathcal{S} : \mathbb{R}^d \to \mathbb{N}$$

which to a configuration in $\mathbb{R}^d$ associates a state number. Think of a numbering of the wells of the potential V.

Objective: generate very efficiently a trajectory $(S_t)_{t\geq 0}$ which has (almost) the same law as $(\mathcal{S}(\boldsymbol{X}_t))_{t\geq 0}$.

The Quasi-Stationary Distribution

How to take advantage of metastability to build efficient sampling techniques ?

Let us consider a metastable state W, and

$$T_W = \inf\{t \geq 0, \boldsymbol{X}_t \notin W\}.$$

Lemma: Let $\boldsymbol{X}_t$ start in the well W. Then there exists a probability distribution ν with support W such that

$$\lim_{t\to\infty} \mathcal{L}(\boldsymbol{X}_t | T_W > t) = \nu.$$

Remark: Quantitative definition of a metastable state:

exit time $\gg$ local equilibration time

The Quasi-Stationary Distribution

Property 1: $\forall t > 0$, $\forall A \subset W$,

$$\nu(A) = \frac{\displaystyle\int_W \mathbb{P}(\boldsymbol{X}_t^{\boldsymbol{x}} \in A,\, t < T_W^{\boldsymbol{x}})\, \nu(d\boldsymbol{x})}{\displaystyle\int_W \mathbb{P}(t < T_W^{\boldsymbol{x}})\, \nu(d\boldsymbol{x})}.$$

If $\boldsymbol{X}_0 \sim \nu$ and if $(\boldsymbol{X}_s)_{0 \leq s \leq t}$ has not left the well, then $\boldsymbol{X}_t \sim \nu$.

Property 2: Let $L = -\nabla V \cdot \nabla + \beta^{-1}\Delta$ be the infinitesimal generator of $(\boldsymbol{X}_t)$. Then the density u_1 of ν ($d\nu = u_1(\boldsymbol{x})d\boldsymbol{x}$) is the first eigenfunction of $L^* = \operatorname{div}(\nabla V + \beta^{-1}\nabla)$ with absorbing boundary conditions:

$$\begin{cases} L^* u_1 = -\lambda_1 u_1 \text{ on } W, \\ \quad u_1 = 0 \text{ on } \partial W. \end{cases}$$

The Quasi-Stationary Distribution

Property 3: If $\boldsymbol{X}_0 \sim \nu$ then,

- the first exit time T_W from W is exponentially distributed with parameter λ_1 ;
- T_W is independent of the first hitting point $\boldsymbol{X}_{T_W}$ on ∂W ;
- the exit point distribution is proportional to $-\partial_n u_1$: for all smooth test functions $\varphi : \partial W \to \mathbb{R}$,

$$\mathbb{E}^{\nu}(\varphi(\boldsymbol{X}_{T_W})) = -\frac{\displaystyle\int_{\partial W} \varphi\, \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx}.$$

Link with kinetic Monte Carlo models (1/2)

Starting from the QSD in W, the exit event from W is Markovian: it can be rewritten as one step of a Markov jump process (kinetic Monte Carlo or Markov state model):

Link with kinetic Monte Carlo models (2/2)

Let us introduce $\lambda_1 = 1/\mathbb{E}(T_W)$ and

$$p(i) = \mathbb{P}(\boldsymbol{X}_{T_W} \in \partial W_i) = -\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx}.$$

To each possible exit region ∂W_i is associated a rate $k(i) = \lambda_1 p(i)$.
If $\tau_i \sim \mathcal{E}(k(i))$ are independent, then

- The exit time is $\min(\tau_1, \ldots, \tau_I)$;
- The exit region is $\arg\min(\tau_1, \ldots, \tau_I)$.

Escaping from a metastable state

How to use these properties to design efficient algorithms ?

Assume that the stochastic process remained trapped for a very long time in a metastable state W. How to accelerate the escape event from W, in a statistically consistent way ?

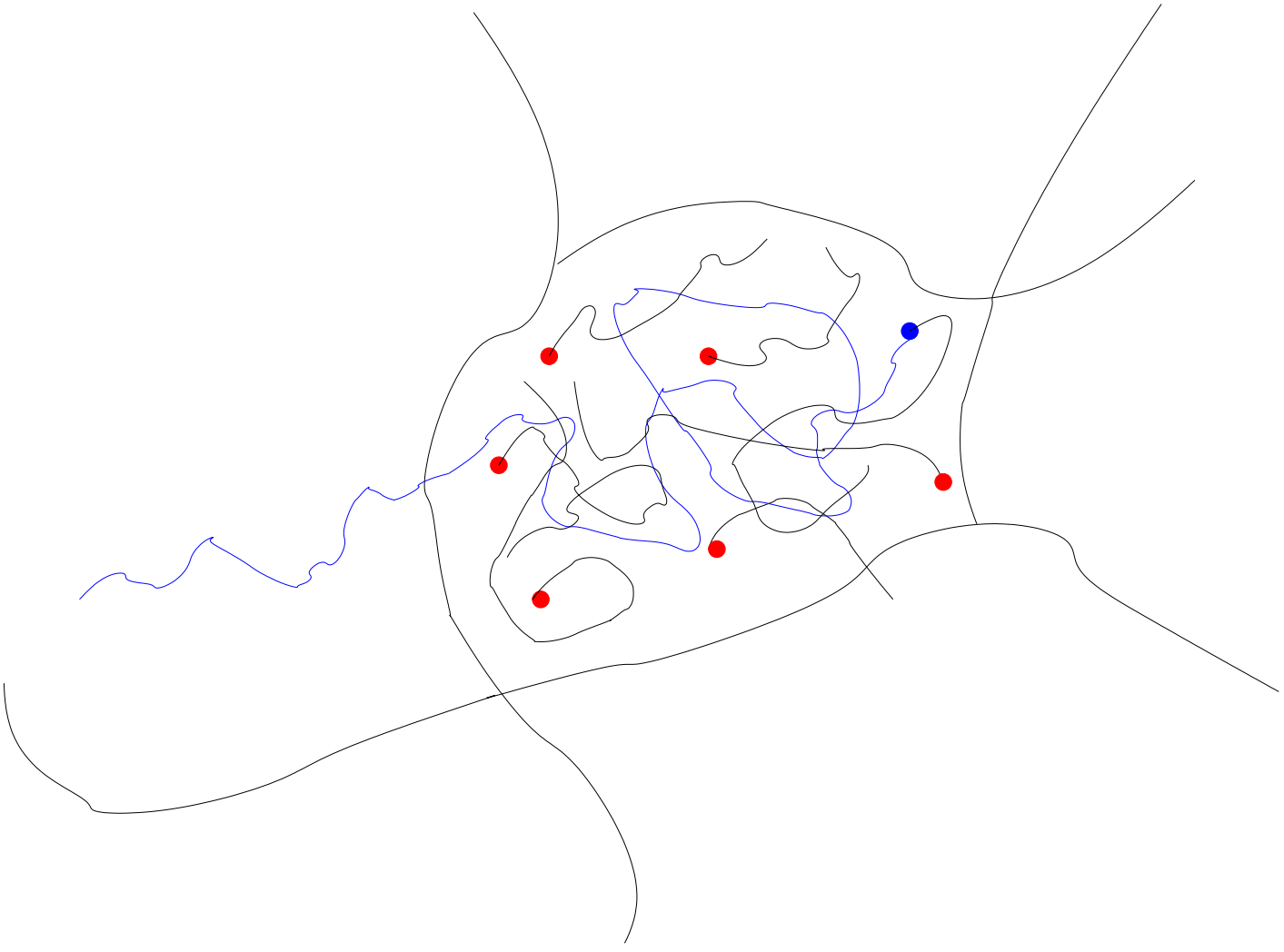
Remark: In practice, one needs to:

- Choose the partition of the domain into (metastable) states;
- Associate to each state an equilibration time (a.k.a. *decorrelation time*).

These are not easy tasks... we will come back to that.

Remark: All the algorithms below equally apply to the Langevin dynamics but the extensions of the mathematical results to the Langevin dynamics are not straightforward...

The Parallel Replica Algorithm

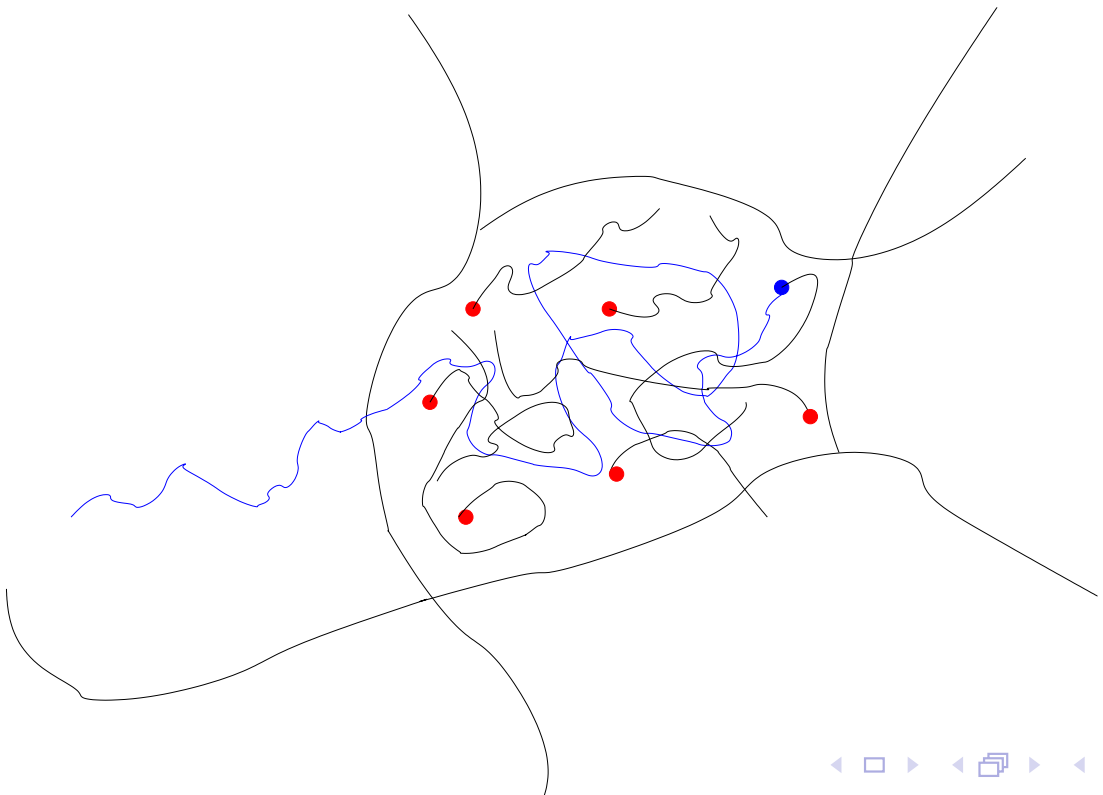


The Parallel Replica Algorithm

Idea: perform many independent exit events in parallel.

Two steps:

- Distribute N independent initial conditions in W according to the QSD ν ;
- Consider the first exit event, and multiply it by the number of replicas.



The Parallel Replica Algorithm

Why is it consistent ?

- Exit time is independent of exit point so that

$$\boldsymbol{X}^{I_0}_{T_W^{I_0}} \overset{\mathcal{L}}{=} \boldsymbol{X}^{1}_{T_W^{1}},$$

  where $I_0 = \arg\min_i(T_W^i)$;

- Exit times are i.i.d. exponentially distributed so that, for all N,

$$N \min(T_W^1, \ldots, T_W^N) \overset{\mathcal{L}}{=} T_W^1.$$

Remark: In practice, discrete time processes are used. Exponential laws become geometric, and one can adapt the algorithm by using the identity [Aristoff, TL, Simpson, 2014]: if τ_i i.i.d. with geometric law,

$$N[\min(\tau_1, \ldots, \tau_N) - 1] + \min[i \in \{1, \ldots, N\}, \tau_i = \min(\tau_1, \ldots, \tau_N)] \overset{\mathcal{L}}{=} \tau_1.$$

The Parallel Replica Algorithm

The full algorithm is in three steps:

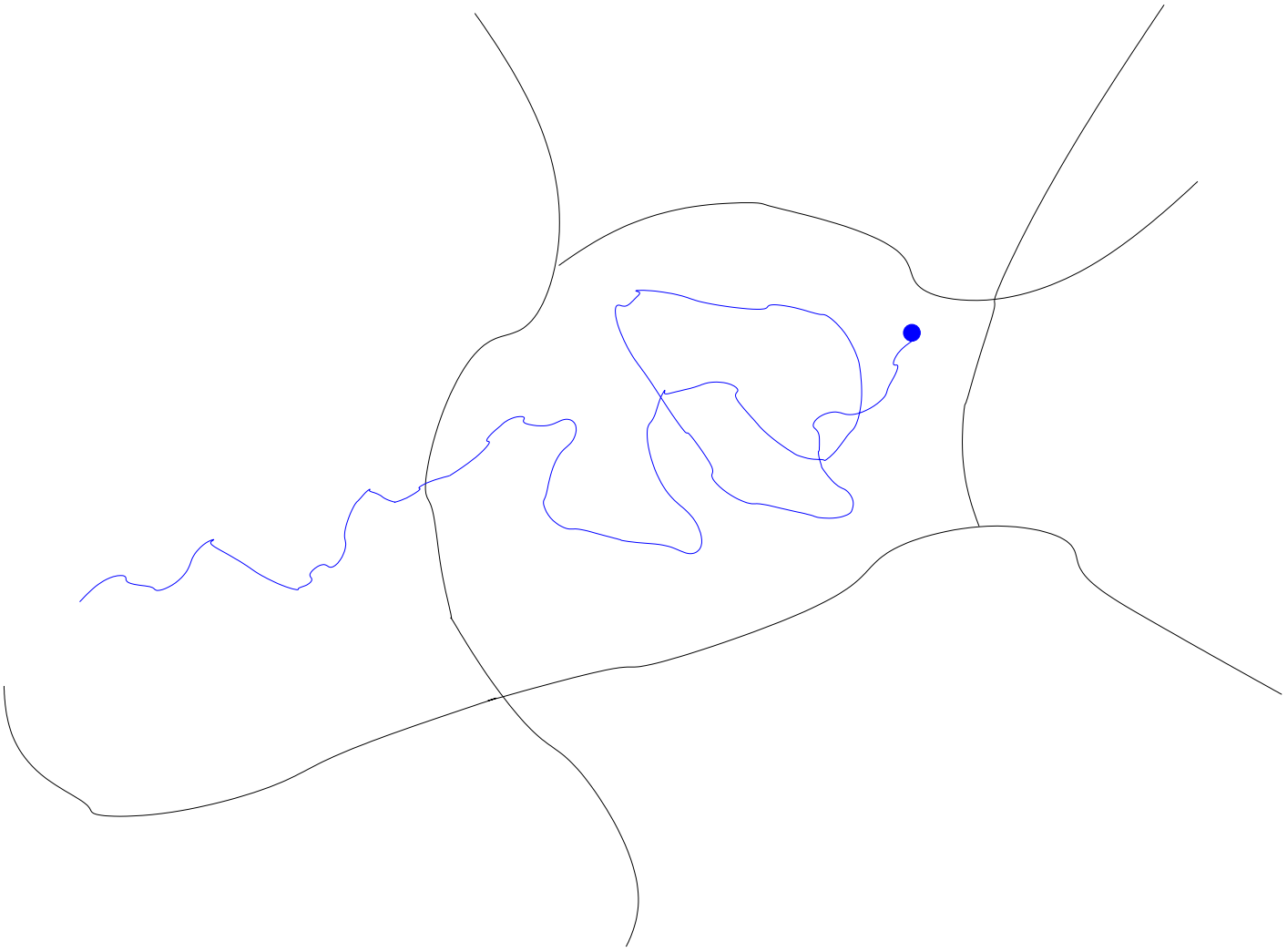
- Decorrelation step
- Dephasing step
- Parallel step

The Parallel Replica Algorithm

Decorrelation step: run the dynamics on a reference walker...

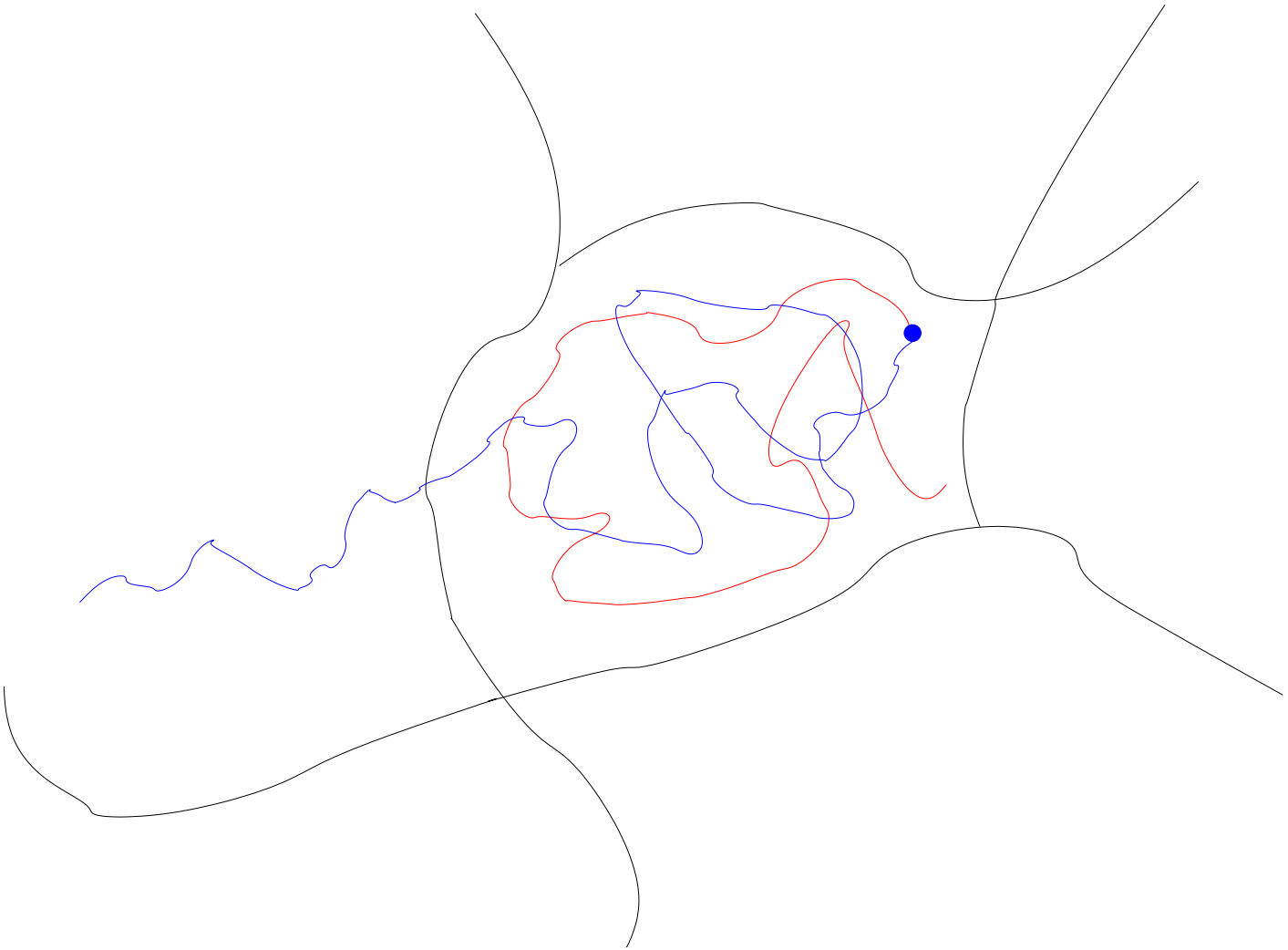
The Parallel Replica Algorithm

Decorrelation step: ... until it remains trapped for a time τ_{corr}.



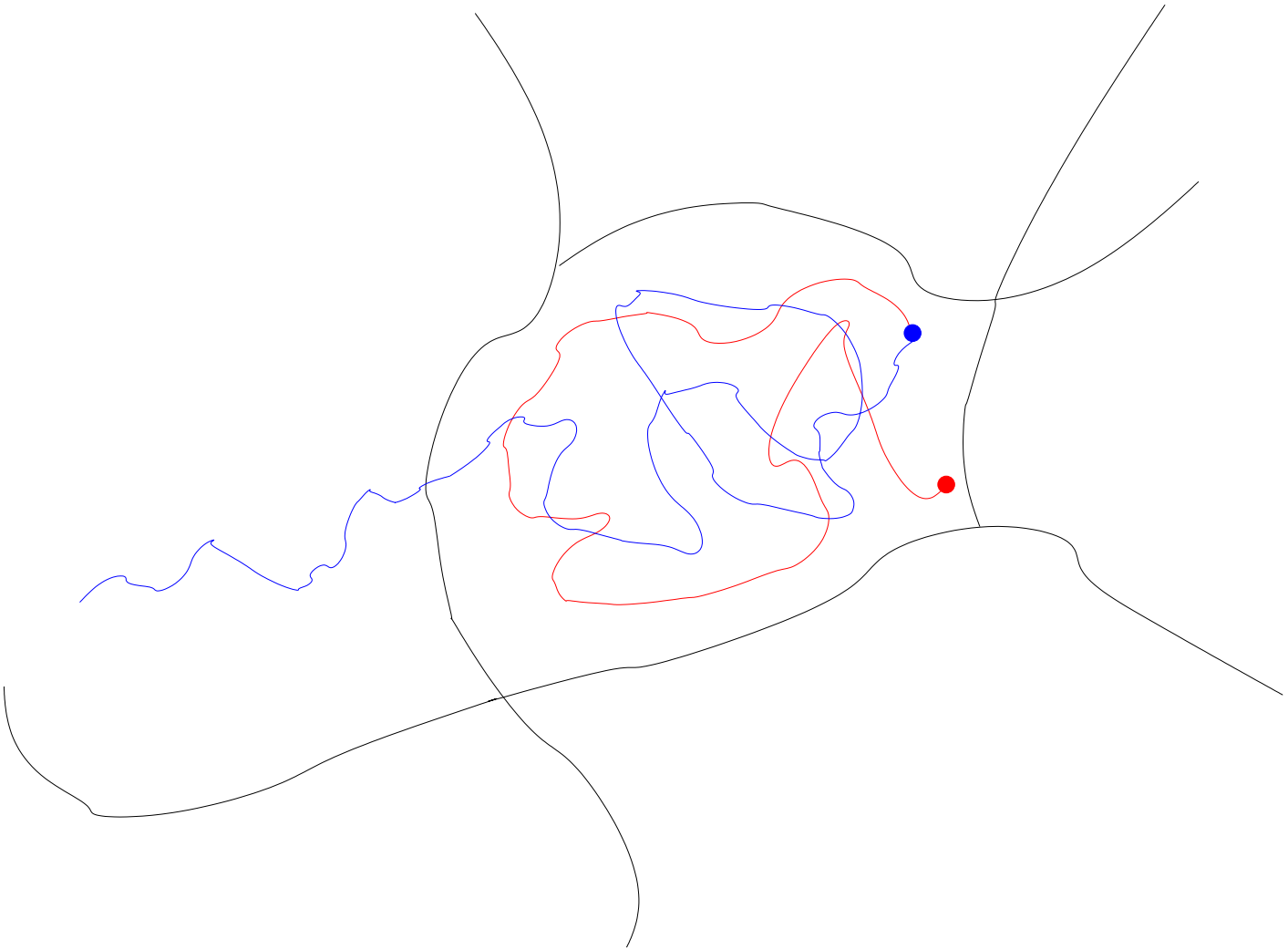
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



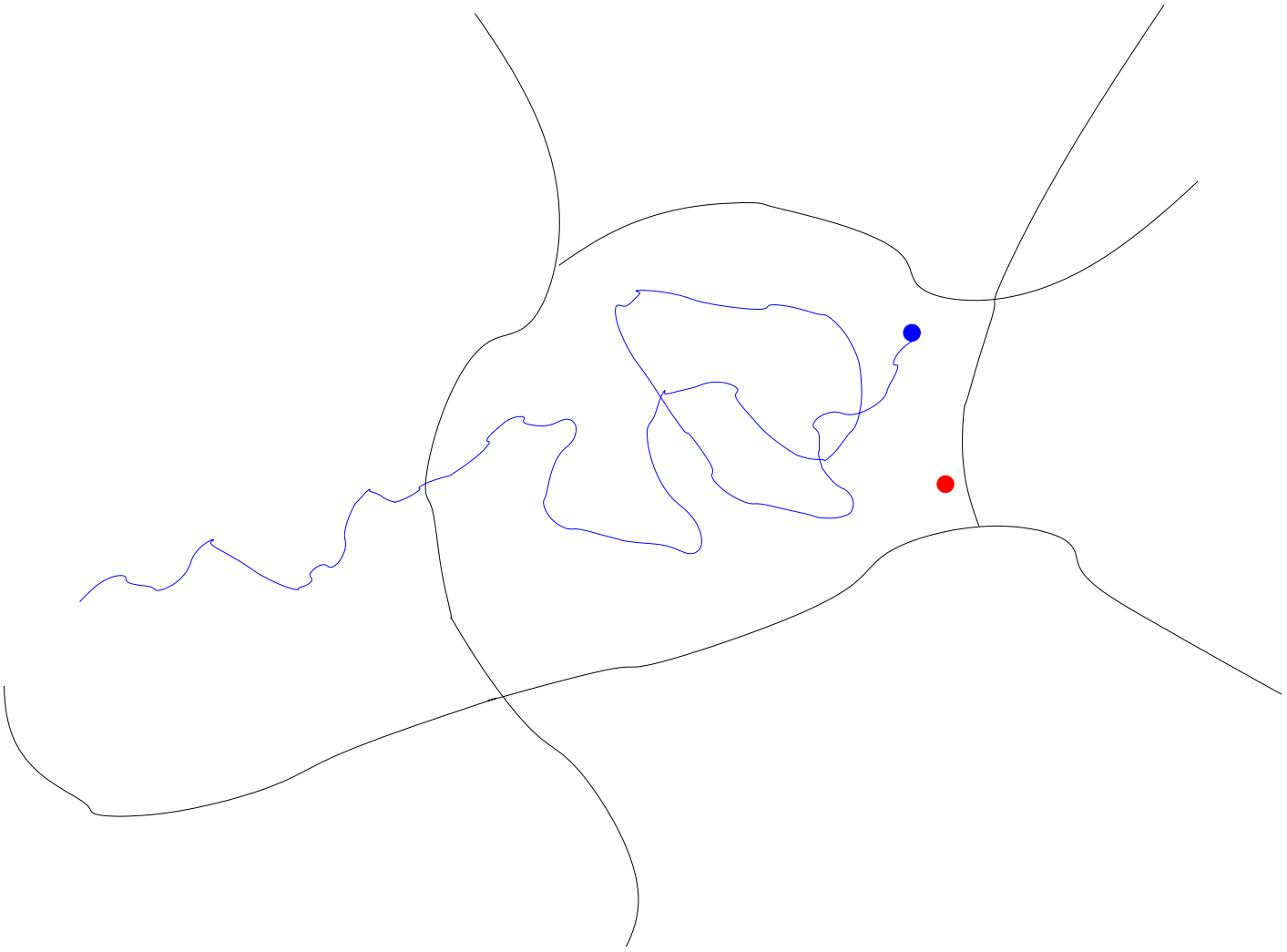
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



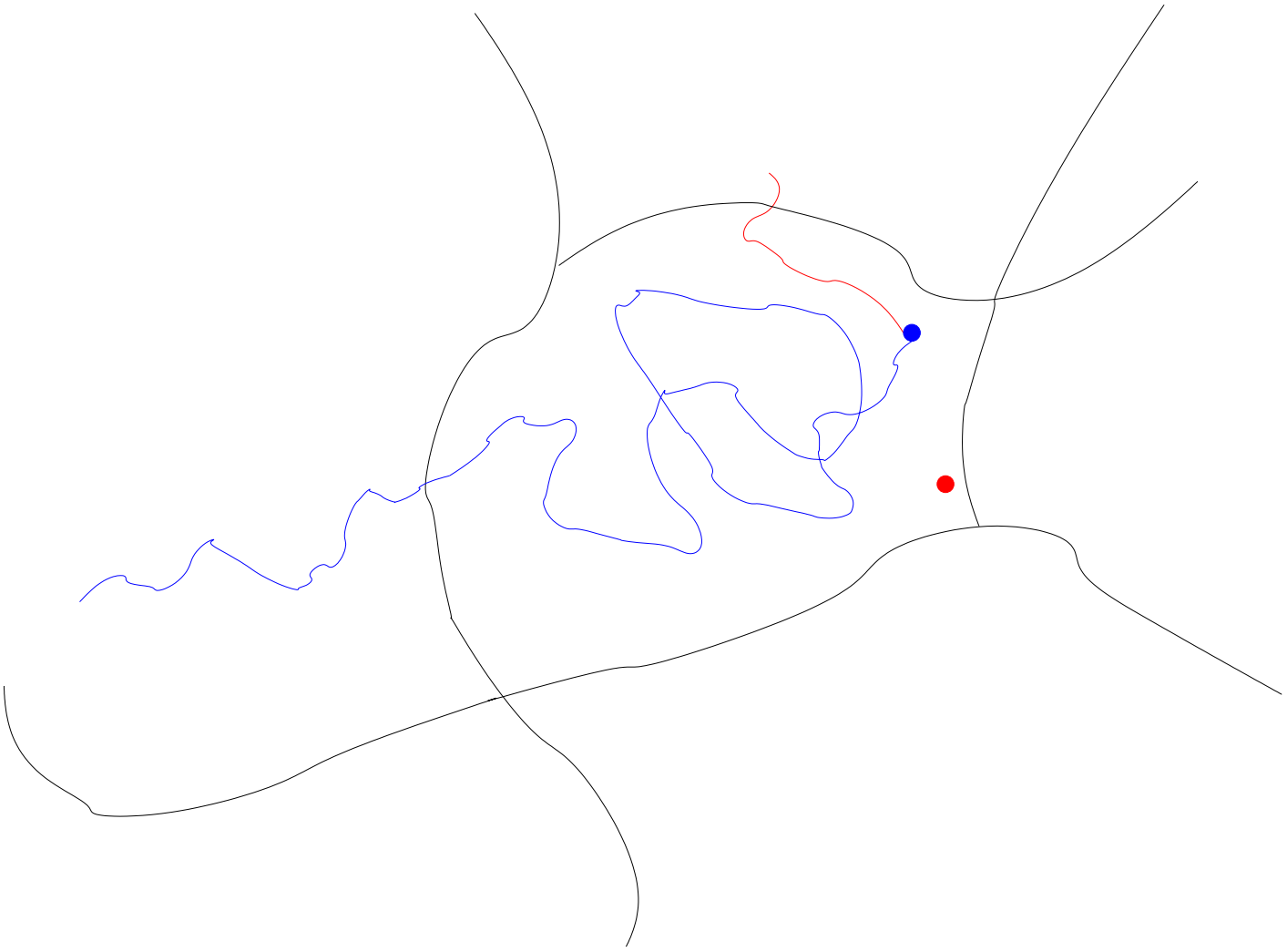
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



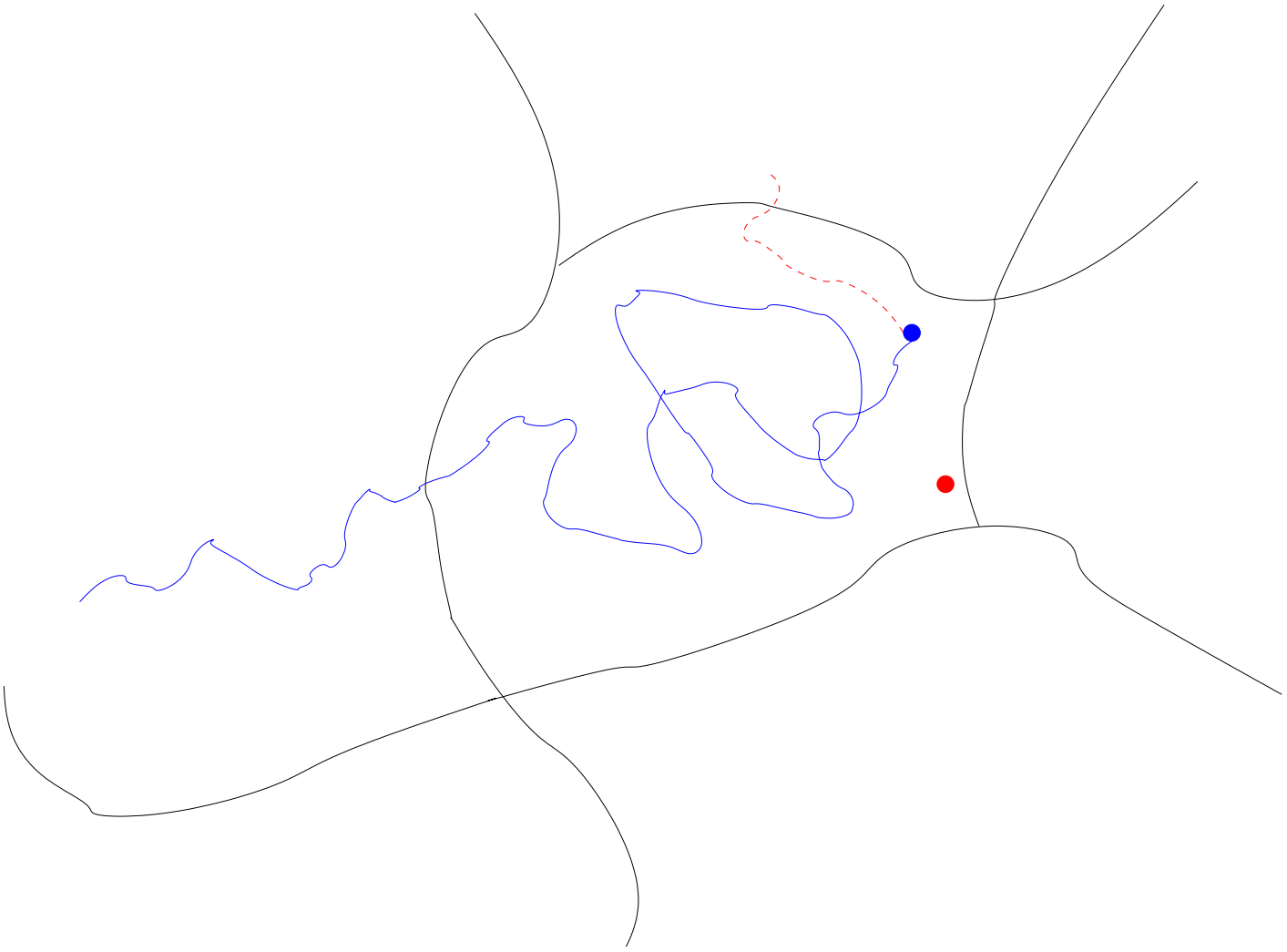
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



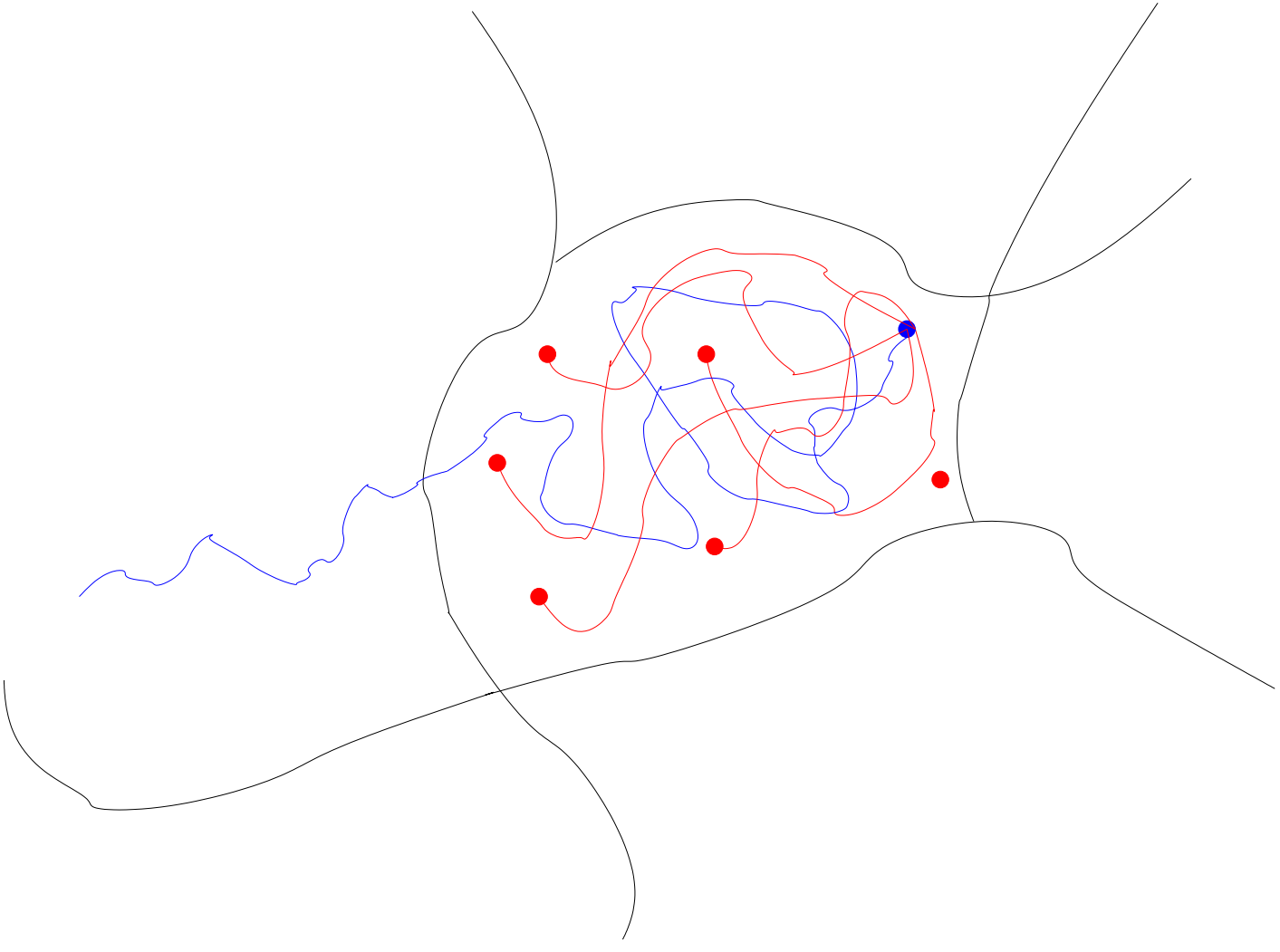
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



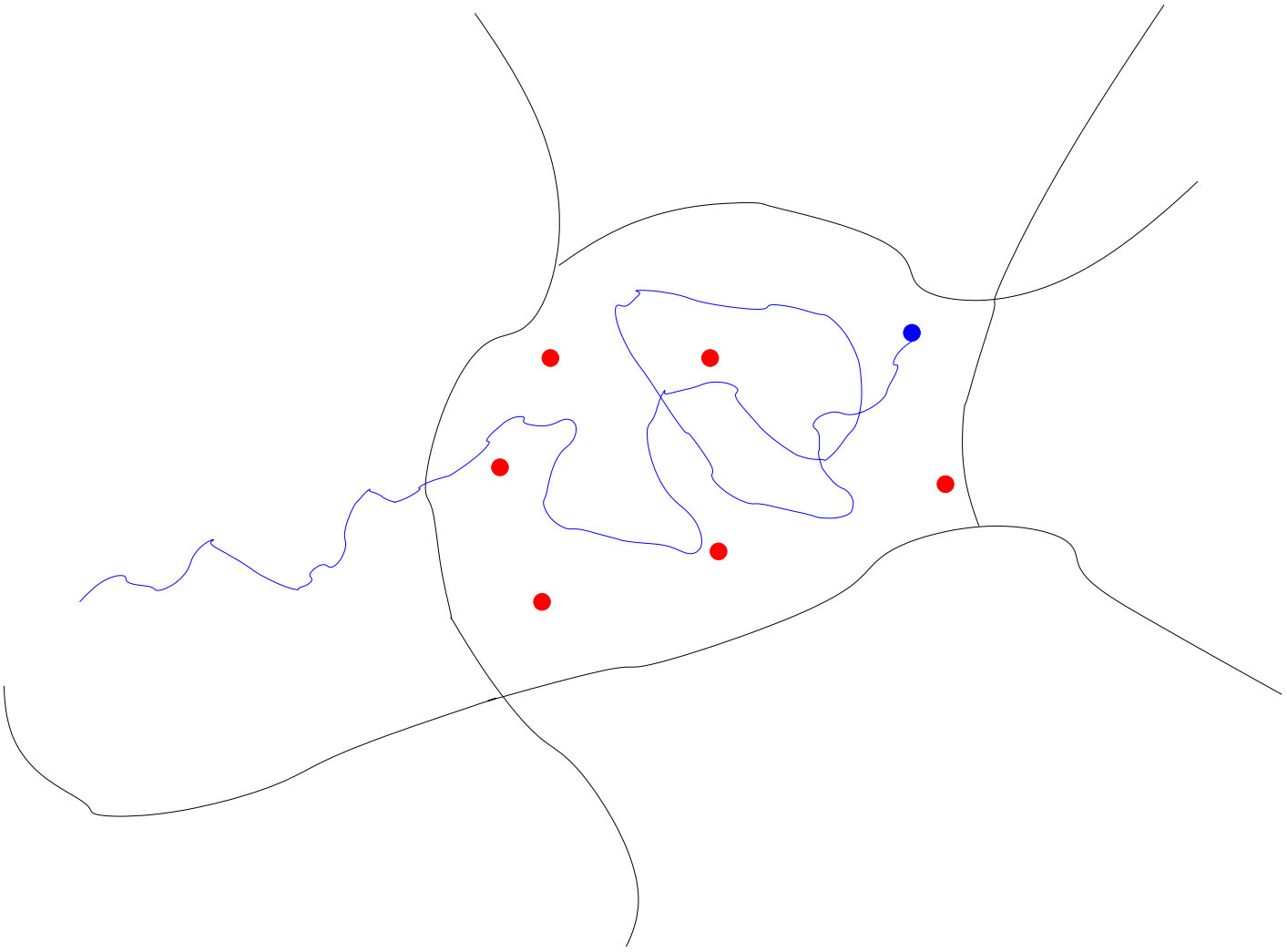
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



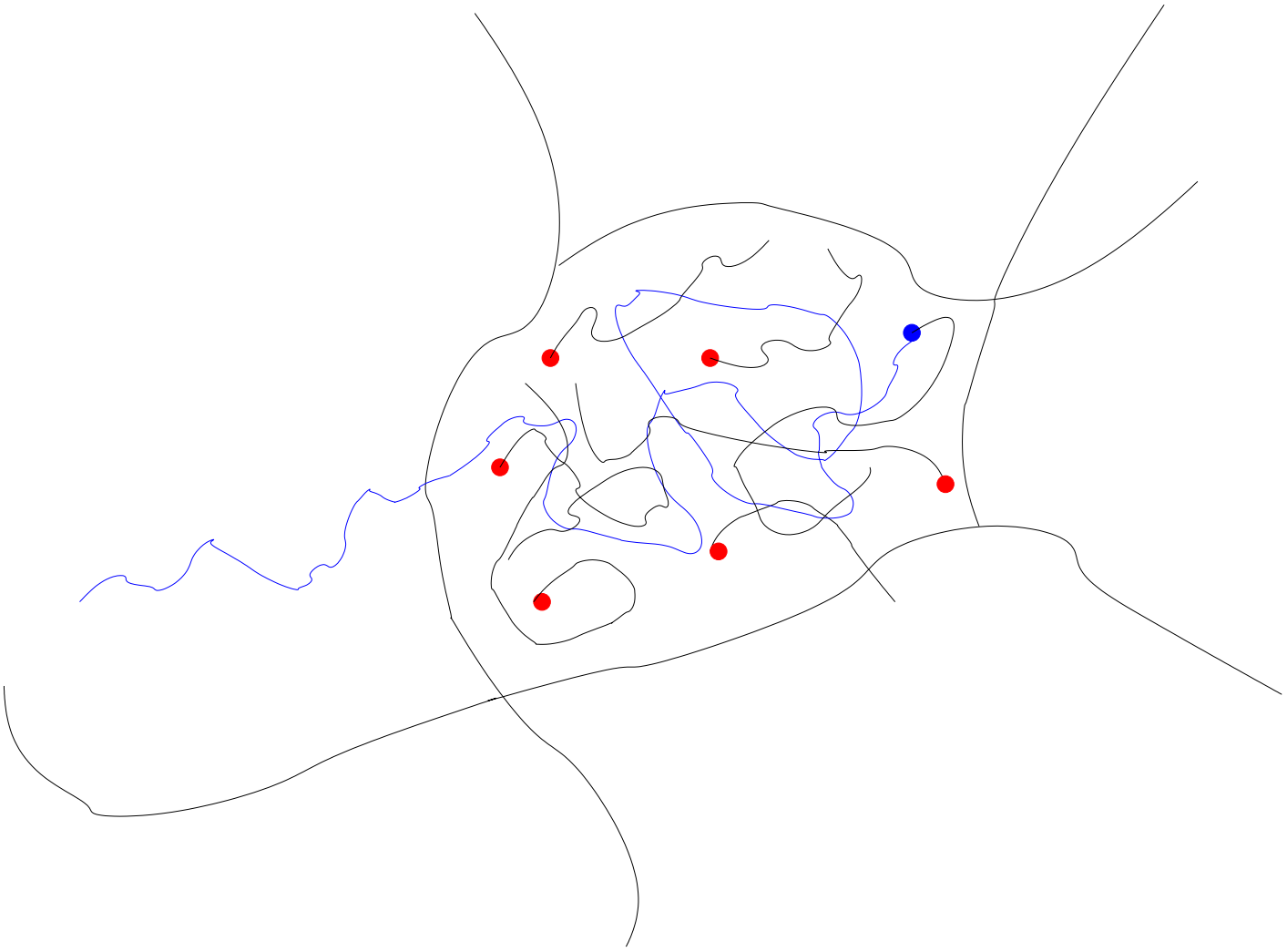
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



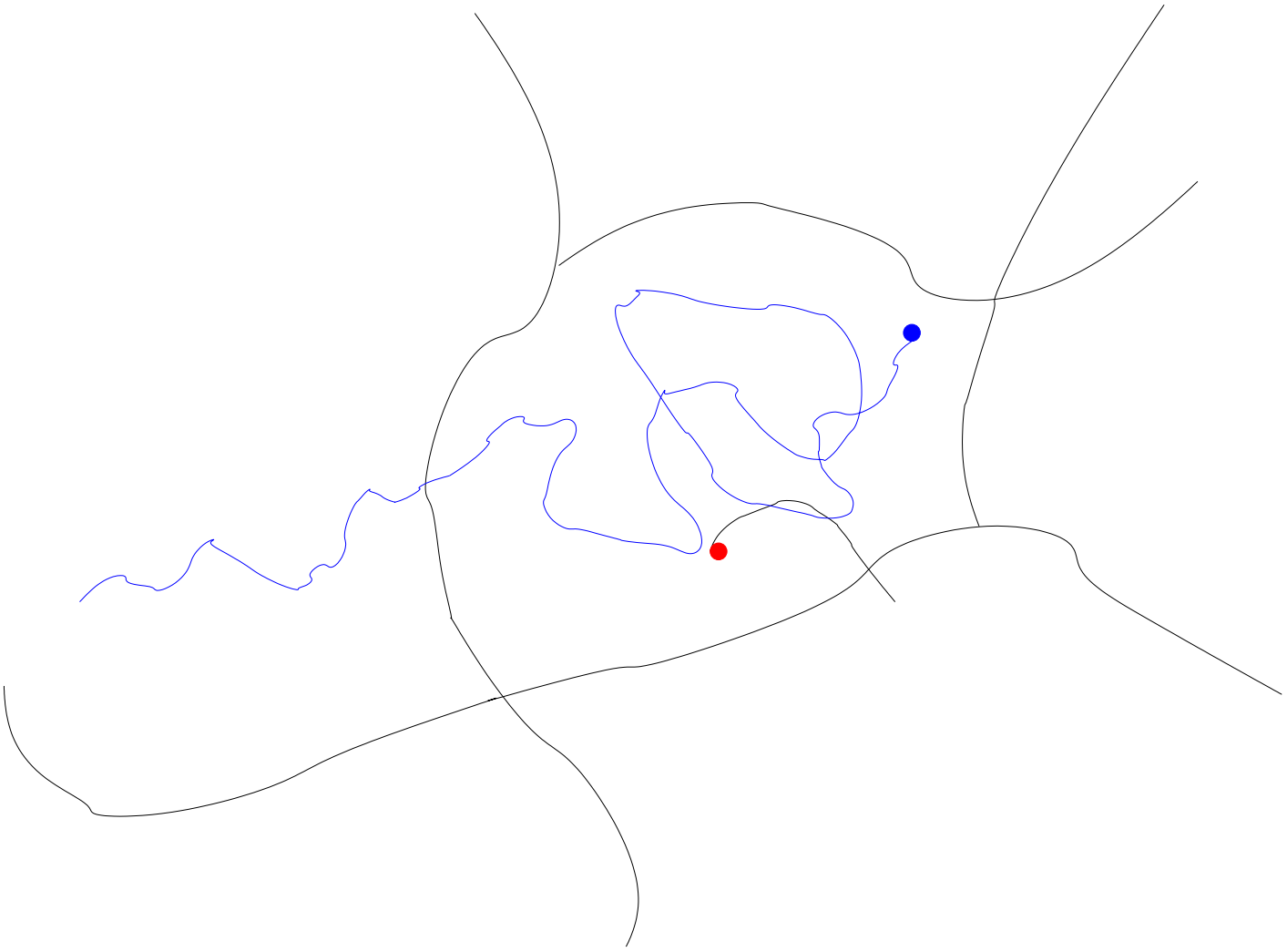
The Parallel Replica Algorithm

Parallel step: run independent trajectories in parallel...



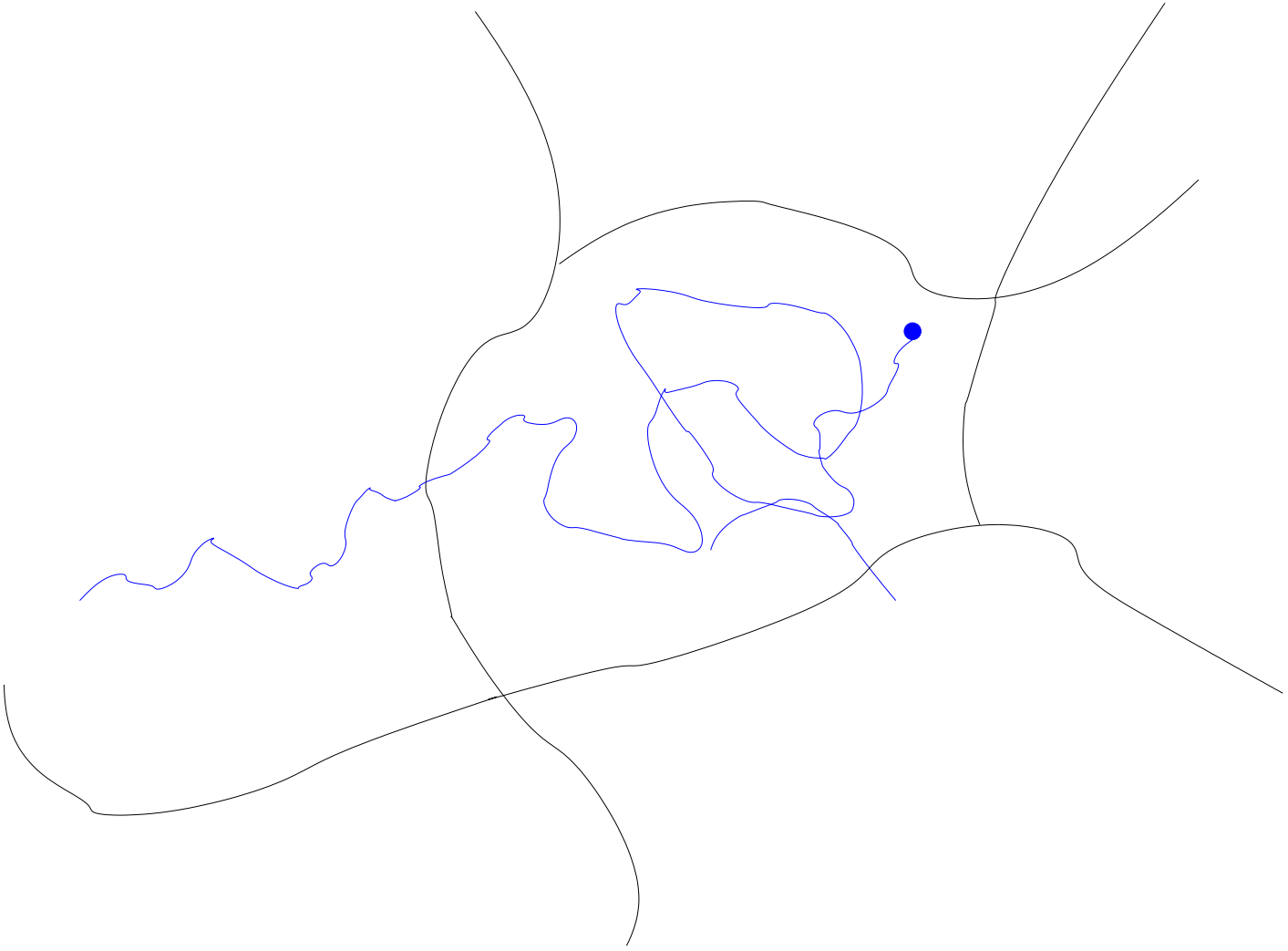
The Parallel Replica Algorithm

Parallel step: ... and detect the first transition event.



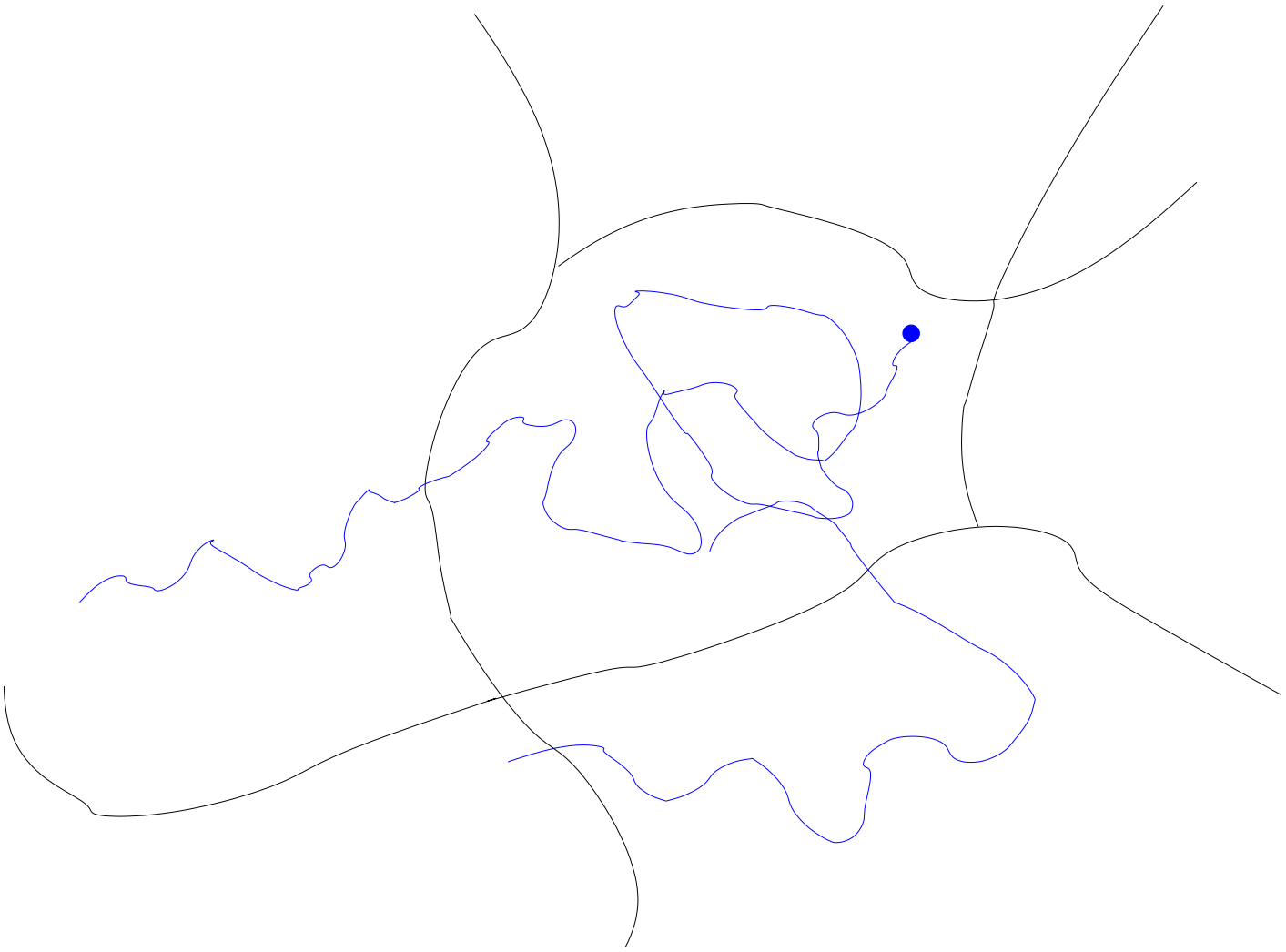
The Parallel Replica Algorithm

Parallel step: update the time clock: $T_{simu} = T_{simu} + NT$.



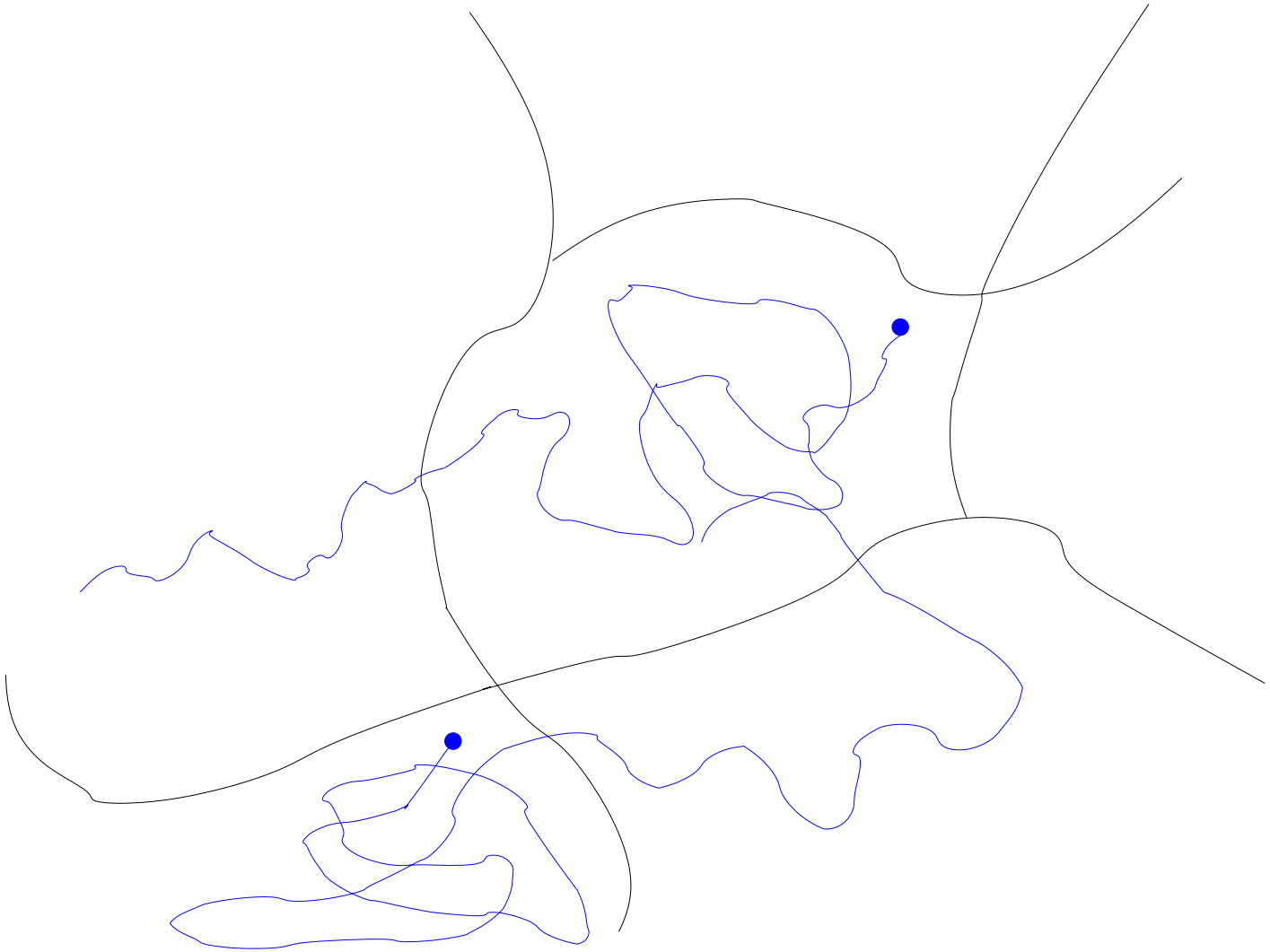
The Parallel Replica Algorithm

A new decorrelation step starts...



The Parallel Replica Algorithm

New decorrelation step



The Parallel Replica Algorithm

The three steps of ParRep:

- Decorrelation step: does the reference walker remain trapped in a set ?
- Dephasing step: prepare many initial conditions in this trapping set.
- Parallel step: detect the first escaping event.

The decorrelation step

How to quantify the error introduced by the dephasing and parallel steps, when the decorrelation step is successful ?

When the decorrelation step is successful, it is assumed that the reference walker is distributed according to the QSD : if it was indeed the case, the algorithm would be exact. The decorrelation step can be seen as a way to probe this assumption. What is the error introduced there ?

The decorrelation step

We have the following error estimate in total variation norm: for $t \geq \frac{C}{\lambda_2 - \lambda_1}$,

$$\sup_{f, \|f\|_{L^\infty} \leq 1} \left| \mathbb{E}(f(T_W - t, \boldsymbol{X}_{T_W}) | T_W \geq t) - \mathbb{E}^\nu(f(T_W, \boldsymbol{X}_{T_W})) \right| \leq C \exp(-(\lambda_2 - \lambda_1)t),$$

where $-\lambda_2 < -\lambda_1 < 0$ are the two first eigenvalues of L^* with absorbing boundary conditions on ∂W.

This shows that τ_{corr} should be chosen such that:

$$\tau_{corr} \geq \frac{\overline{C}}{\lambda_2 - \lambda_1}.$$

On the other hand, it should be smaller than the typical time to leave the well, $\mathbb{E}(T_W)$. Since $\mathbb{E}^\nu(T_W) = 1/\lambda_1$, this typically implies the spectral gap requirement,

$$\frac{\overline{C}}{\lambda_2 - \lambda_1} \leq \frac{1}{\lambda_1}.$$

The Parallel Replica Algorithm

This algorithm is very versatile: it works for entropic barriers, and for any partition of the state space into states. But it requires some a priori knowledge on the system: the equilibration time τ_{corr} attached to each state S.

Two questions: How to choose τ_{corr} ? How to sample the QSD ?

We recently proposed a generalized Parallel Replica algorithm [Binder, TL, Simpson, 2014] to solve these issues. It is based on two ingredients:

- the Fleming-Viot particle process
- the Gelman-Rubin statistical test

The Fleming-Viot particle process

Start N processes i.i.d. from μ_0, and iterate the following steps:

1. Integrate (in parallel) N realizations ($k = 1, \ldots, N$)
$$d\boldsymbol{X}_t^k = -\nabla V(\boldsymbol{X}_t^k)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t^k$$
until one of them, say $\boldsymbol{X}_t^1$, exits;
2. Kill the process that exits;
3. With uniform probability $1/(N-1)$, randomly choose one of the survivors, $\boldsymbol{X}_t^2, \ldots, \boldsymbol{X}_t^N$, say $\boldsymbol{X}_t^2$;
4. Branch $\boldsymbol{X}_t^2$, with one copy persisting as $\boldsymbol{X}_t^2$, and the other becoming the new $\boldsymbol{X}_t^1$.

It is known that the empirical distribution

$$\mu_{t,N} \equiv \frac{1}{N} \sum_{k=1}^{N} \delta_{\boldsymbol{X}_t^k}$$

satisfies:

$$\lim_{N\to\infty} \mu_{t,N} = \mathcal{L}(\boldsymbol{X}_t | t < T_W).$$

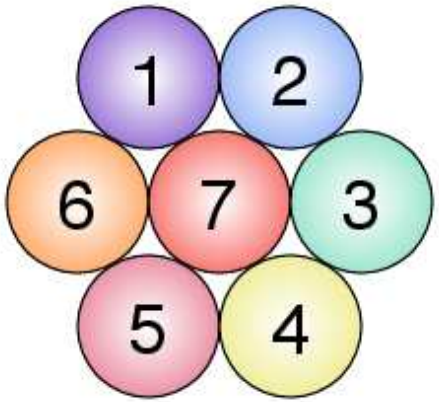
The generalized Parallel Replica algorithm

The generalized Parallel Replica algorithm consists in using a Fleming-Viot particle process for the dephasing step and running in parallel the decorrelation and the dephasing steps.

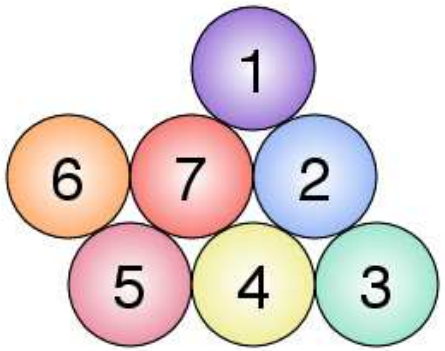
If the Fleming Viot particle process reaches stationarity before the reference walker, go to the parallel step. Otherwise, restart a new decorrelation / dephasing step.

The time at which the Fleming-Viot particle process becomes stationary is determined using the Gelman-Rubin statistical test.

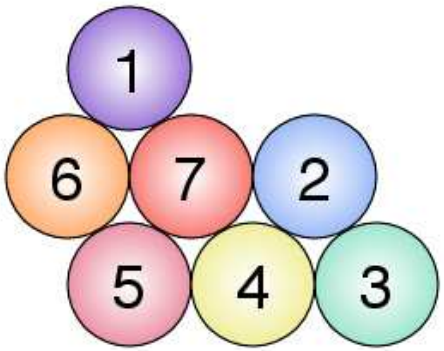
Numerical test case: the 7 atoms LJ cluster



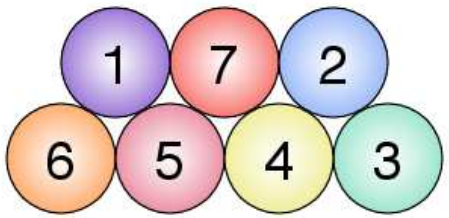
(a) C_0, $V = -12.53$



(b) C_1, $V = -11.50$



(c) C_2, $V = -11.48$



(d) C_3, $V = -11.40$

We study the escape from the configuration C_0 using overdamped Langevin dynamics with $\beta = 6$. The next visited states are C_1 or C_2.

Numerical test case: the 7 atoms LJ cluster

Method	TOL	$\langle T \rangle$	$\mathbb{P}[C_1]$	$\mathbb{P}[C_2]$
Serial	–	17.0	(0.502, 0.508)	(0.491, 0.498)
ParRep	0.2	19.1	(0.508, 0.514)	(0.485, 0.492)
ParRep	0.1	18.0	(0.506, 0.512)	(0.488, 0.494)
ParRep	0.05	17.6	(0.505, 0.512)	(0.488, 0.495)
ParRep	0 .01	17.0	(0.504, 0.510)	(0.490, 0.496)

Method	TOL	$\langle t_{\mathrm{corr}} \rangle$	$\langle$Speedup$\rangle$	% Dephased
Serial	–	–	–	–
ParRep	0.2	0.41	29.3	98.5%
ParRep	0.1	.98	14.9	95.3%
ParRep	0.05	2.1	7.83	90.0%
ParRep	0 .01	11	1.82	52.1%

Numerical test case: the 7 atoms LJ cluster

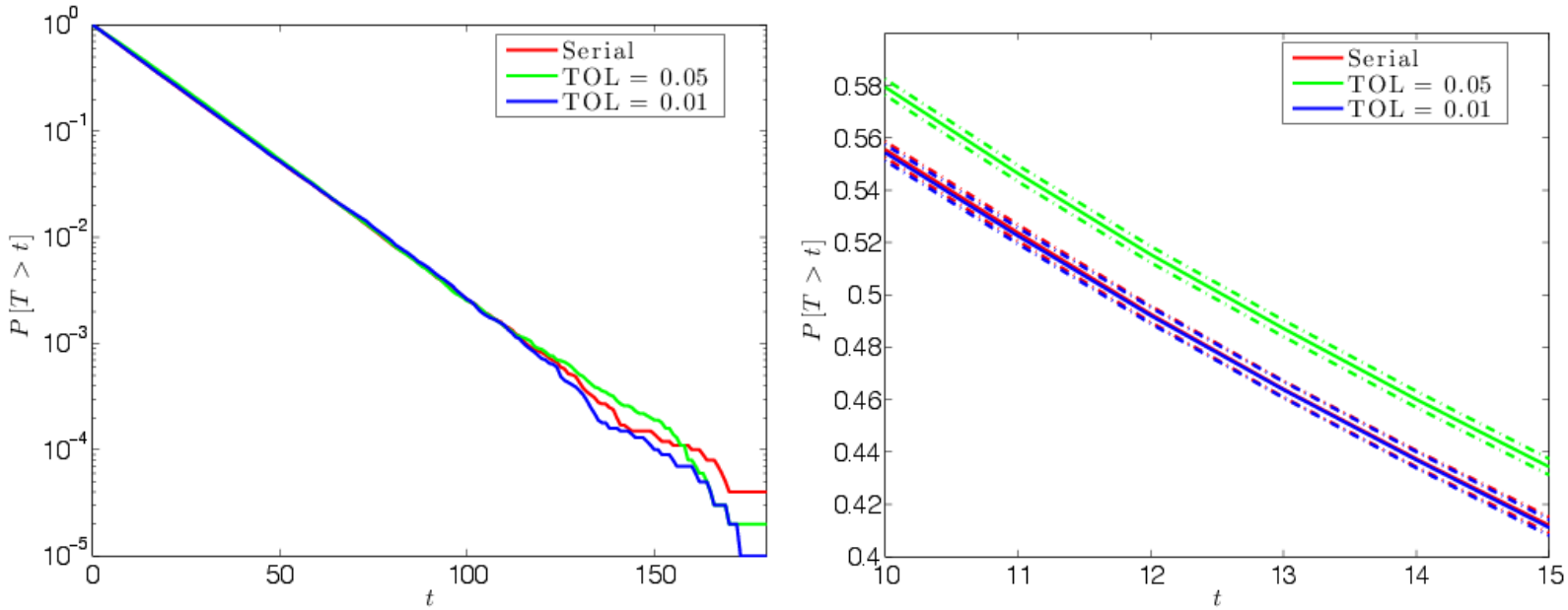
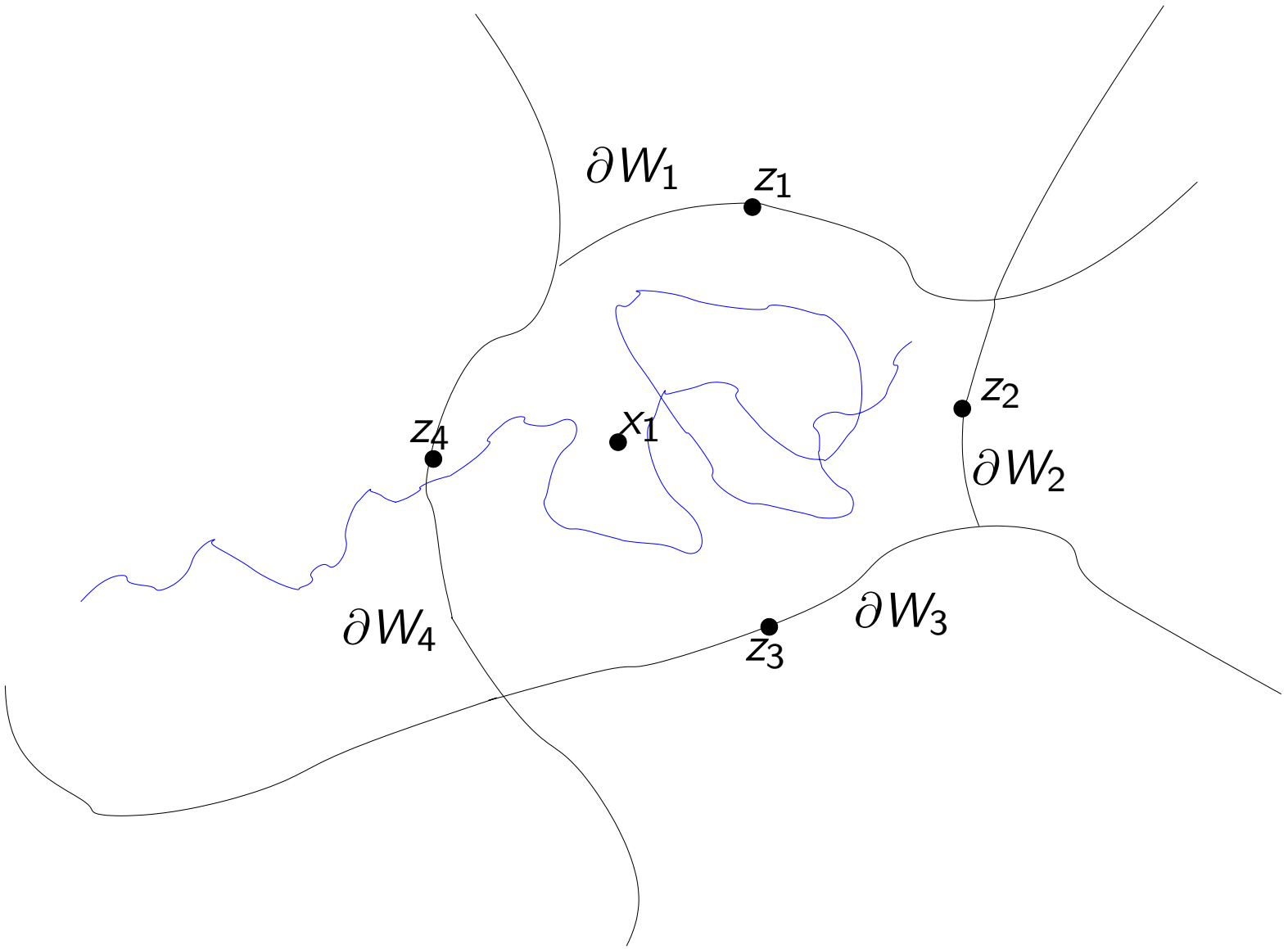


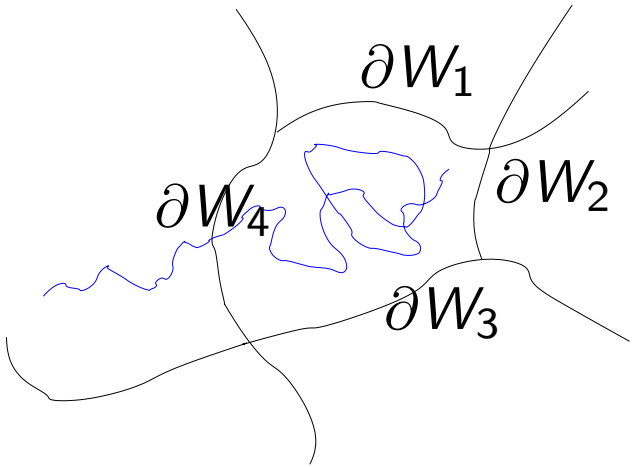
Figure: $\mathrm{LJ}_7^{2\mathrm{D}}$: Cumulative distribution function of the escape time from C_0.

kinetic Monte Carlo and Harmonic Transition State Theory



kMC models

Let us go back to the kinetic Monte Carlo model.



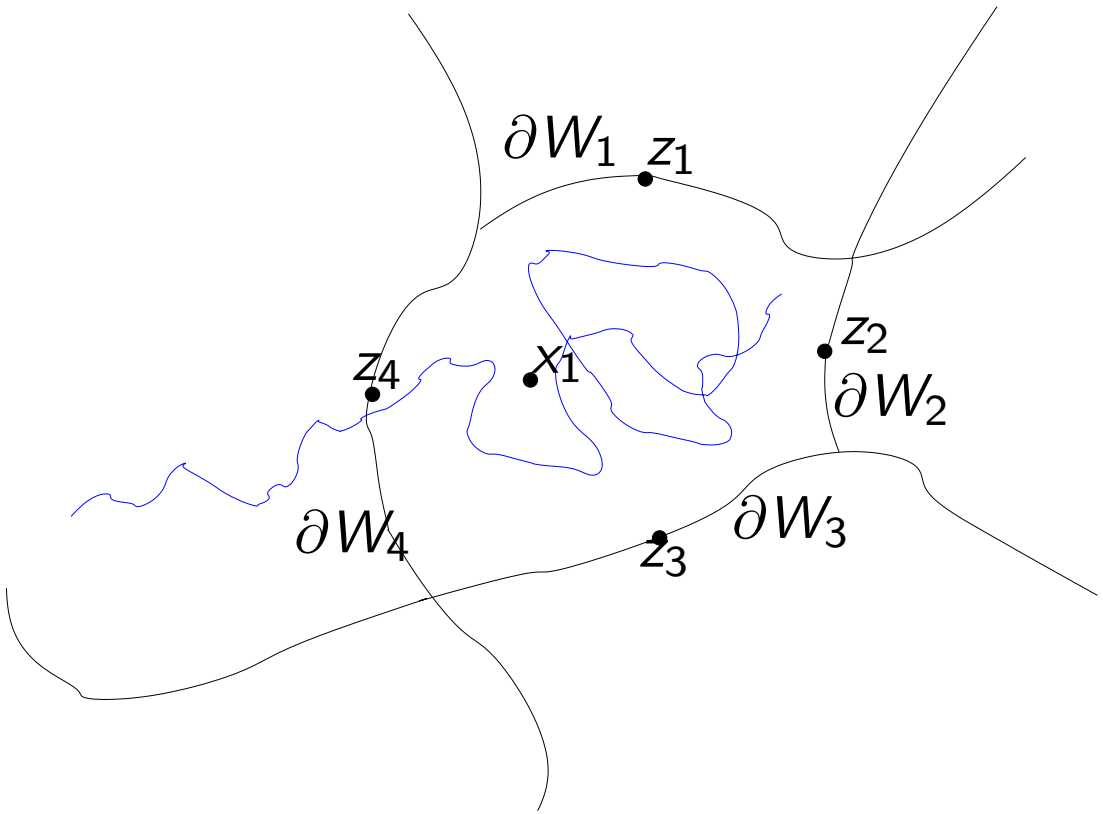
To each exit region ∂W_i is associated a rate $k(i)$. Let $\tau_i \sim \mathcal{E}(k(i))$ be independent exponential random variables. And then,

- The exit time is $\min(\tau_1, \ldots, \tau_I)$;
- The exit region is $\arg\min(\tau_1, \ldots, \tau_I)$.

Thus, (i) exit time and exit region are independent r.v. ; (ii) exit time is $\mathcal{E}(k(1) + \ldots + k(I))$; (iii) exit region is i with prob. $\frac{k(i)}{k(1)+\ldots+k(I)}$.

The Eyring Kramers law and HTST

In practice, kMC models are parameterized using HTST.



We assume in the following $V(z_1) < V(z_2) < \ldots < V(z_I)$.

Eyring Kramers law (HTST): $k(i) = A_i \exp\left(-\beta(V(z_i) - V(x_1))\right)$
where A_i is a prefactor depending on V at z_i and x_1.

kMC and HTST

Thus, one obtains the following law for the exit event:

- exit time and exit region are independent r.v.
- exit time is $\mathcal{E}(k(1)+\ldots+k(I))$ and, when β is large

$$k(1)+\ldots+k(I) \simeq k(1) = A_1 \exp\left(-\beta(V(z_1)-V(x_1))\right)$$

- exit region is i with probability $\frac{k(i)}{k(1)+\ldots+k(I)}$ and, when β is large,

$$\frac{k(i)}{k(1)+\ldots+k(I)} \simeq \frac{k(i)}{k(1)} = \frac{A_i}{A_1} \exp\left(-\beta(V(z_i)-V(z_1))\right)$$

Our aim: justify and analyze this method.

Back to overdamped Langevin and the QSD

Starting from the QSD $d\nu = u_1(x)dx$, we already know that

- the exit time T_W and the exit point X_{T_W} are independent r.v.
- the exit time is exponentially distributed with parameter λ_1
- the exit region is ∂W_i with probability

$$p(i) = \mathbb{P}(\boldsymbol{X}_{T_W} \in \partial W_i) = -\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda \displaystyle\int_W u_1(x)\, dx}.$$

We thus need to prove that

$$\lambda_1 \simeq A_1 \exp\left(-\beta(V(z_1) - V(x_1))\right)$$

and

$$-\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx} \simeq \frac{A_i}{A_1} \exp\left(-\beta(V(z_i) - V(z_1))\right).$$

Small temperature regime

The question is thus: consider (λ_1, u_1) such that (first eigenvalue eigenfunction pair)

$$\begin{cases} \operatorname{div}\left(\nabla V u_1 + \beta^{-1}\nabla u_1\right) = -\lambda_1 u_1 \text{ on } W, \\ u_1 = 0 \text{ on } \partial W. \end{cases}$$

We assume wlg $u_1 > 0$ and $\int u_1^2 e^{\beta V} = 1$.
In the small temperature regime ($\beta \to \infty$), prove that

$$\lambda_1 \simeq A_1 \exp\left(-\beta(V(z_1) - V(x_1))\right)$$

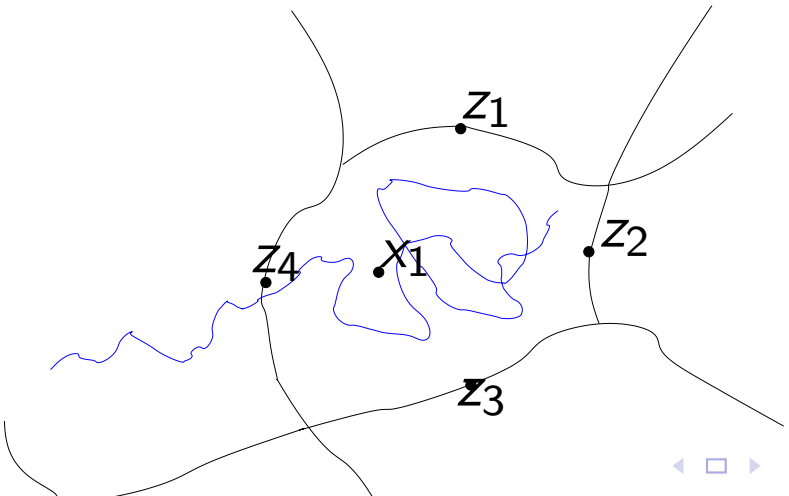
and

$$-\frac{\displaystyle\int_{\partial W_i} \partial_n u_1 \, d\sigma}{\beta\lambda_1 \displaystyle\int_W u_1(x)\, dx} \simeq \frac{A_i}{A_1} \exp\left(-\beta(V(z_i) - V(z_1))\right).$$

Assumptions

- W is an open bounded smooth domain in $\mathbb{R}^d$.
- $V : \overline{W} \to \mathbb{R}$ is a Morse function with a single critical point x_1. Moreover, $x_1 \in W$ and $V(x_1) = \min_{\overline{W}} V$.
- $\partial_n V > 0$ on ∂W and $V|_{\partial W}$ is a Morse function with local minima reached at $z_1, \ldots, z_I$ with $V(z_1) < \ldots < V(z_I)$.
- $V(z_1) - V(x_1) > V(z_I) - V(z_1)$
- $\forall i \in \{1, \ldots I\}$, consider B_{z_i} the basin of attraction for the dynamics $\dot{x} = -\nabla_T V(x)$ and assume that

$$\inf_{z \in B_{z_i}^c} d_a(z, z_i) > V(z_I) - V(z_1)$$



Agmon distance

Here, d_a is the Agmon distance:

$$d_a(x,y) = \inf_{\gamma} \int_0^1 g(\gamma(t))|\gamma'(t)|\,dt$$

where $g = \begin{cases} |\nabla V| \text{ in } W \\ |\nabla_T V| \text{ in } \partial W \end{cases}$, and the infimum is over all Lipschitz paths $\gamma : [0,1] \to \overline{W}$ such that $\gamma(0) = x$ and $\gamma(1) = y$. A few properties:

- One has $\forall x, y \in \overline{W}$, $|V(x) - V(y)| \leq d_a(x,y) \leq C|x-y|$
- On a neighborhood $\mathcal{V}$ of a local minima z_i, the function $x \mapsto d_a(x, z_i)$ satisfies the eikonal equation: $|\nabla \Phi|^2 = |\nabla V|^2$ on $\mathcal{V}$ with boundary conditions $\Phi = V$ on $\mathcal{V} \cap \partial W$, and $\Phi \geq V(z_i)$.

Results

[G. Di Gesu, TL, D. Le Peutrec and B. Nectoux] In the limit $\beta \to \infty$, the exit rate is

$$\lambda_1 = \sqrt{\frac{\beta}{2\pi}} \partial_n V(z_1) \frac{\sqrt{\det(Hess V)(x_1)}}{\sqrt{\det(Hess V_{|\partial W})(z_1)}} \mathrm{e}^{-\beta(V(z_1)-V(x_1))}(1+O(\beta^{-1})).$$

Moreover, for all open set Σ_i containing z_i such that $\overline{\Sigma}_i \subset B_{z_i}$,

$$\frac{\int_{\Sigma_i} \partial_n u_1 \, d\sigma}{\int_W u_1} = -C_i(\beta) e^{-\beta(V(z_i)-V(x_1))}(1 + O(\beta^{-1})),$$

where $C_i(\beta) = \frac{\beta^{3/2}}{\sqrt{2\pi}} \partial_n V(z_i) \frac{\sqrt{\det(Hess V)(x_1)}}{\sqrt{\det(Hess V|_{\partial W})(z_i)}}$. Therefore,

$$\mathbb{P}^\nu(X_{T_W} \in \Sigma_i) = \frac{\partial_n V(z_i) \sqrt{\det \mathrm{Hess}(V|_{\partial W})(z_1)}}{\partial_n V(z_1) \sqrt{\det \mathrm{Hess}(V|_{\partial W})(z_i)}} e^{-\beta(V(z_i)-V(z_1))}(1+O(\beta^{-1})).$$

Related results in the literature (1/3)

The result on λ_1 is well known and actually holds under weaker assumptions. See for example [Helffer Nier] [Le Peutrec].

Similar formulas are obtained concerning the problem on the whole domain to compute the cascade of timescales down to the global minimum.

- Potential theoretic approaches [Bovier, Schuette, Hartmann,...]
- Spectral analysis of the Fokker Planck operator on the whole space and semi-classical analysis [Holley, Kusuoka, Stroock, Miclo, Sjöstrand, Helffer, Nier, Pavliotis, Schuette]

Warning: The exit rate is (1/2) times the transition rate !

Related results in the literature (2/3)

Another approach to study the exit problem from a domain: Large deviation techniques [Freidlin, Wentzell, Day, Vanden Eijnden, Weare, Touchette,...].

Compared to our approach, the assumptions in LD are much less stringent but LD only provides the exponential rates (not the prefactors) and LD does not provide error bounds. (Moreover the fact that the exit time is exponentially distributed and the independance property between exit time and exit point are only obtained when $\beta = \infty$.)

There are also PDE versions of these results see [Matkowsky, Schuss, Maier, Stein] for formal expansions, and [Holley, Kusuoka, Stroock, Kamin, Friedman, Mathieu, Perthame] for precise results.

Typical result [Freidlin, Wentzell, Theorem 5.1]: for all $W' \subset\subset W$, for any $\gamma > 0$, for any $\delta > 0$, there exists $\delta_0 \in (0, \delta]$ and $\beta_0 > 0$ such that for all $\beta \geq \beta_0$, for all $x \in W'$ such that $f(x) < \min_{\partial W} f$ and for all $y \in \partial W$,

$$\exp(-\beta(V(y) - V(z_1) + \gamma)) \leq \mathbb{P}^x(X_{T_W} \in \mathcal{V}_{\delta_0}(y)) \leq \exp(-\beta(V(y) - V(z_1) - \gamma))$$

Related results in the literature (3/3)

Why do we care about prefactors ?

Consider a situation with two local minima on the boundary ($V(z_1) < V(z_2)$). Compare

- the probability to leave through Σ_2 such that $z_2 \in \Sigma_2$, $\overline{\Sigma_2} \subset B_{z_2}$ and
- the probability to leave through Σ such that $\overline{\Sigma} \subset B_{z_1}$ and $\inf_\Sigma V = V(z_2)$.

Then, in the limit $\beta \to \infty$,

$$\frac{\mathbb{P}^\nu(X_{T_W} \in \Sigma)}{\mathbb{P}^\nu(X_{T_W} \in \Sigma_2)} = O(\beta^{-1/2}).$$

Discussion on the assumptions (1/5)

The assumption

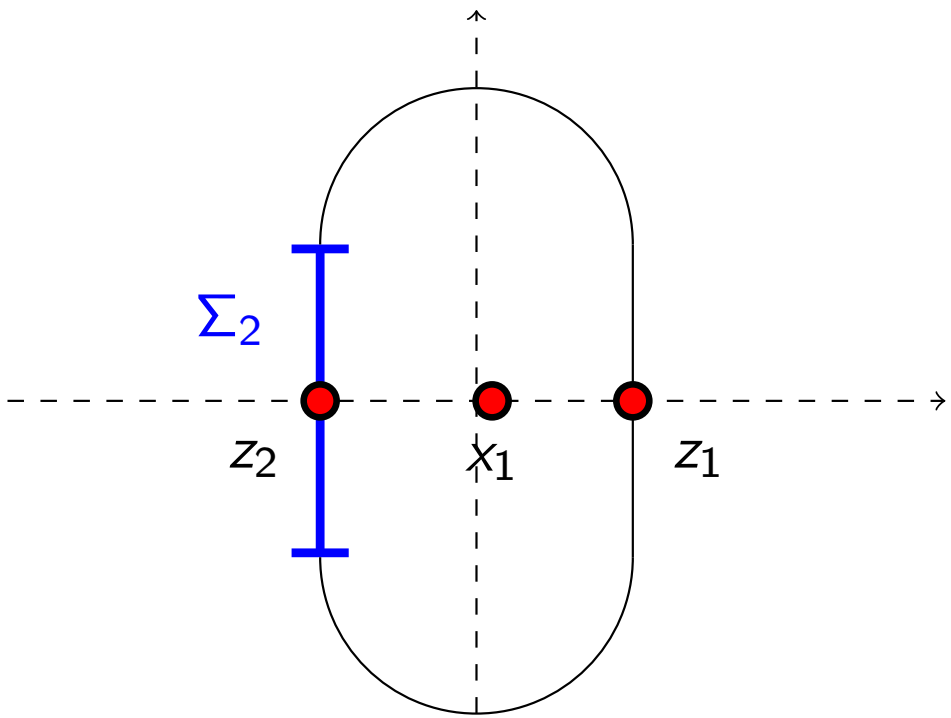
$$\forall i \in \{1, \dots I\}, \inf_{z \in B_{z_i}^c} d_a(z, z_i) > V(z_I) - V(z_1)$$

seems indeed important to get the expected results.

Discussion on the assumptions (2/5)

Let us consider the potential function $V(x, y) = x^2 + y^2 - ax$ with $a \in (0, 1/9)$ on the domain W. Two saddle points: $z_1 = (1, 0)$ and $z_2 = (-1, 0)$ (and $V(z_2) - V(z_1) = 2a$). One can check that the above assumptions are satisfied.

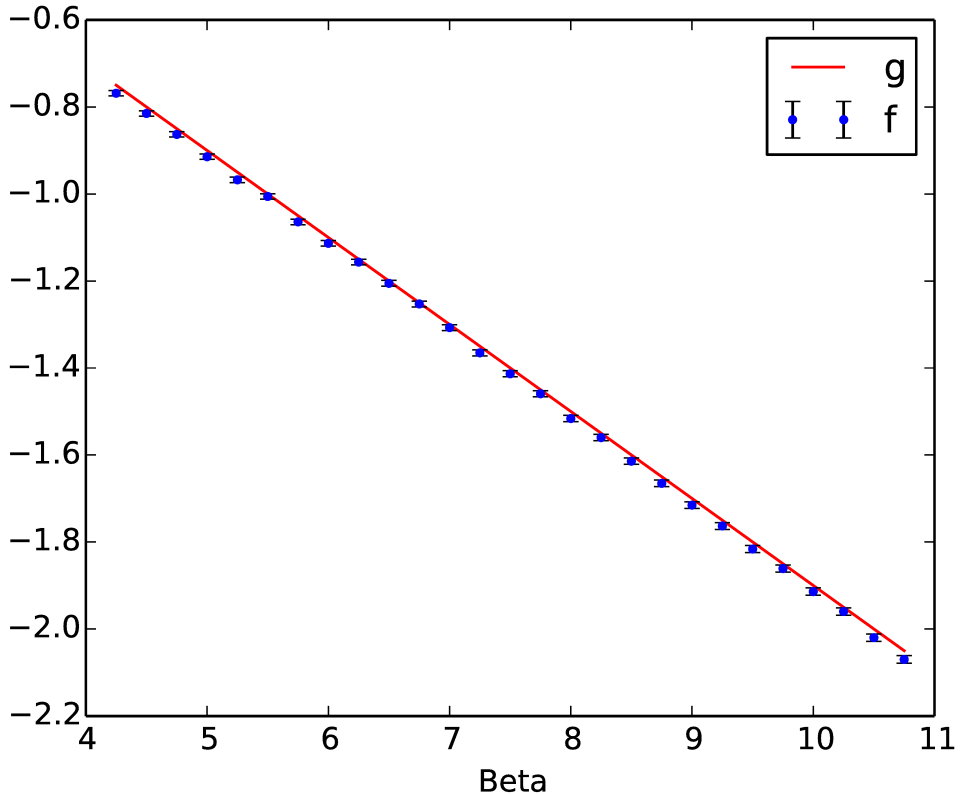
The domain W



Discussion on the assumptions (3/5)

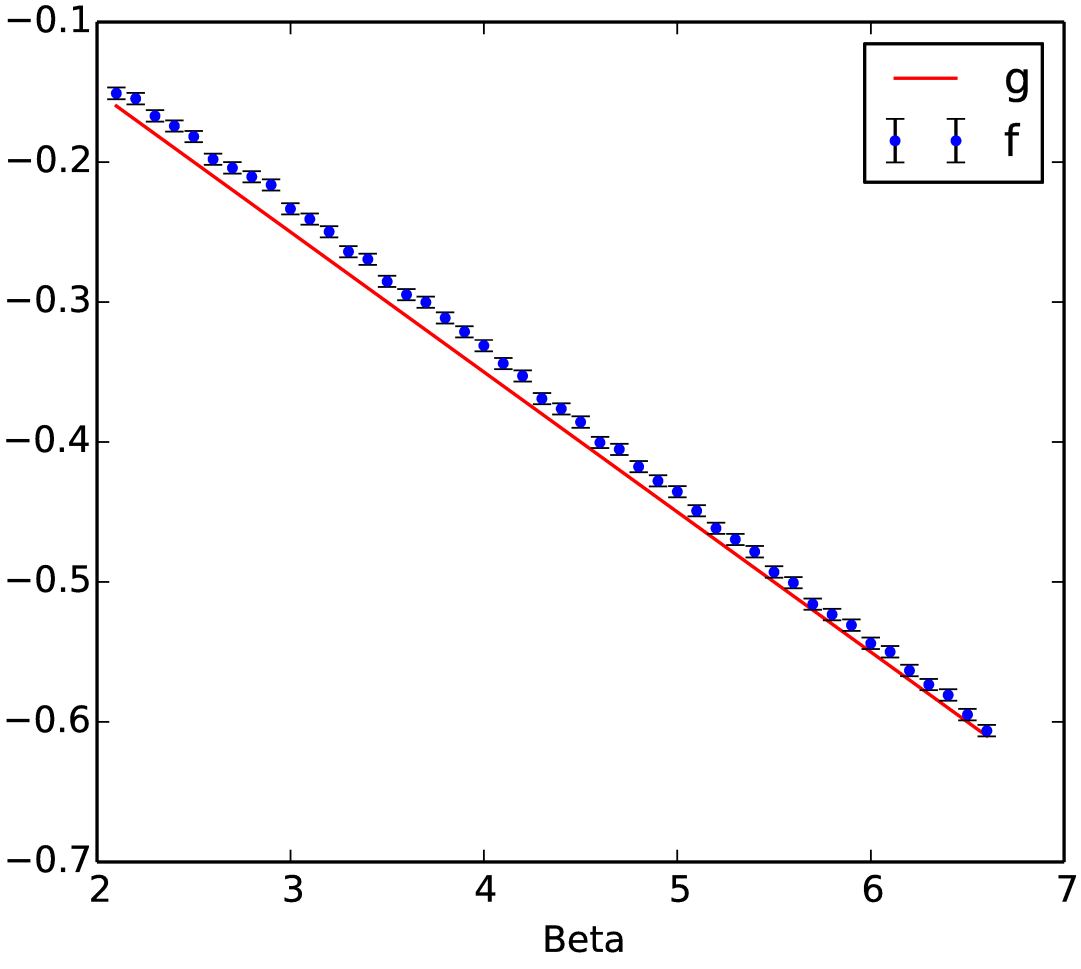
With $a = 1/10$, let us plot

- the numerical results $f : \beta \mapsto \ln\ \mathbb{P}^{\nu}(X_{T_W} \in \Sigma_2)$
- the theoretical result $g : \beta \mapsto \ln B_2 - \beta(V(z_2) - V(z_1))$, where $B_2 = \frac{\partial_n V(z_2)\sqrt{\det\ \mathrm{Hess}(V|_{\partial W})(z_1)}}{\partial_n V(z_1)\sqrt{\det\ \mathrm{Hess}(V|_{\partial W})(z_2)}}$ is the expected prefactor.



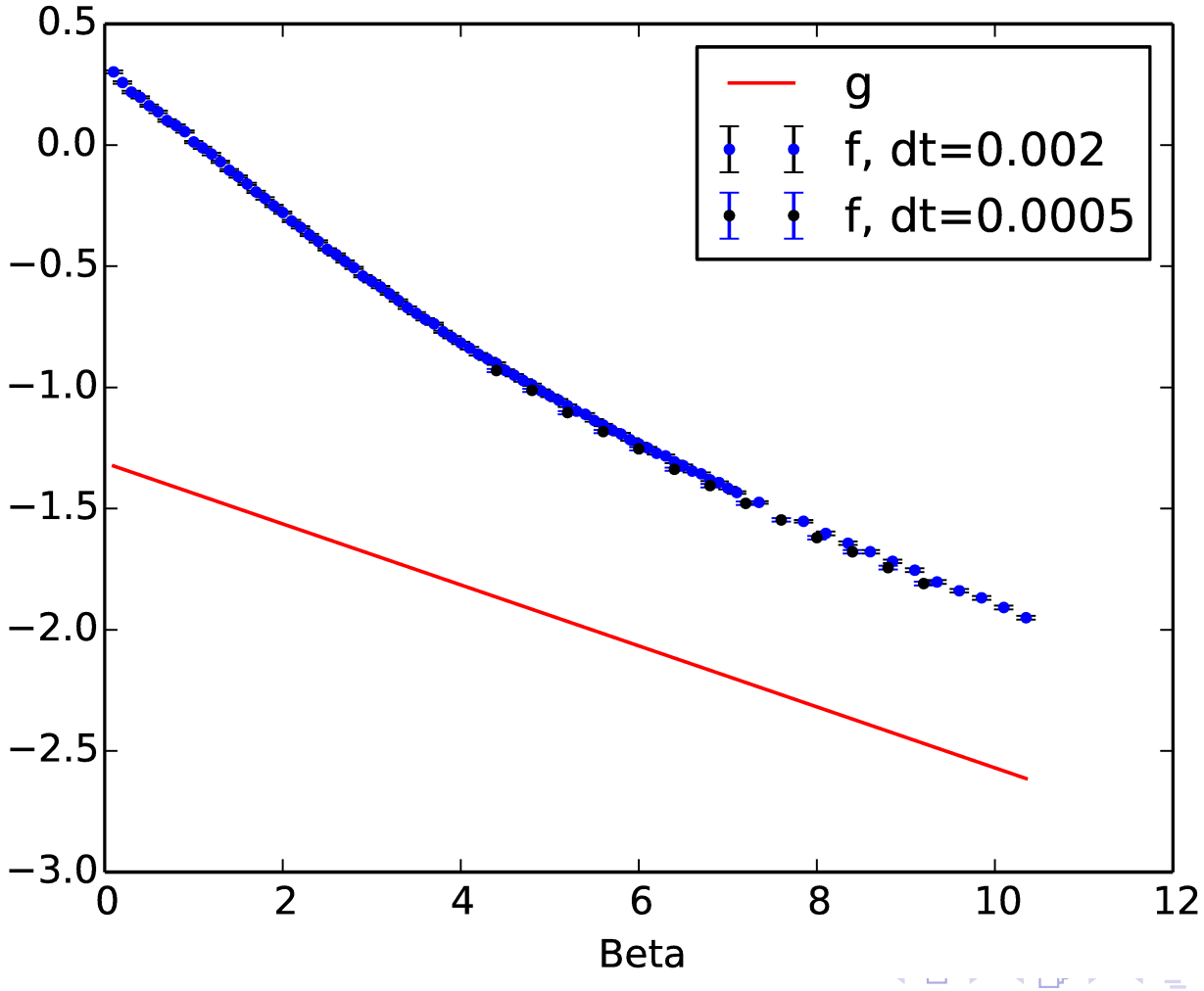
Discussion on the assumptions (4/5)

Same result with $a = 1/20$.



Discussion on the assumptions (5/5)

We now modify the potential such that the assumption on the Agmon distance is not satisfied anymore.



Sketch of the proof

The difficult part is to find an approximation for $\int_{\Sigma_i} \partial_n u_1 d\sigma = \int_{\Sigma_i} \partial_n v_1 e^{-\beta V}$, where $v_1 = u_1 e^{\beta V}$.
We have

$$\begin{cases} L^{(0)} v_1 = -\lambda_1 v_1 \text{ on } W, \\ v_1 = 0 \text{ on } \partial W, \end{cases}$$

where $L^{(0)} = \beta^{-1}\Delta - \nabla V \cdot \nabla$ is a self adjoint operator on $L^2(e^{-\beta V})$. We are interested in $\nabla v_1 \cdot n$, and ∇v_1 satisfies

$$\begin{cases} L^{(1)} \nabla v_1 = -\lambda_1 \nabla v_1 \text{ on } W, \\ \nabla_T v_1 = 0 \text{ on } \partial W, \\ (\beta^{-1}\mathrm{div} \ - \nabla V \cdot) \nabla v_1 = 0 \text{ on } \partial W, \end{cases}$$

where

$$L^{(1)} = \beta^{-1}\Delta - \nabla V \cdot \nabla - \mathrm{Hess}(V).$$

Therefore ∇v_1 is an eigenvector (eigen-1-form) of $-L^{(1)}$ associated with the small eigenvalue λ_1.

Sketch of the proof

Let $\Pi^{(p)} = 1_{[0,\beta^{-3/2}]}(-L^{(p)})$ be the spectral projection operator on small eigenvalues. We know that, for β large, $\dim(\mathrm{Ran}\Pi^{(0)}) = 1$ and $\dim(\mathrm{Ran}\Pi^{(1)}) = I$ [Helffer,Sjöstrand]:

$$\mathrm{Ran}\Pi^{(0)} = \mathrm{Span}(v_1)$$

$$\mathrm{Ran}\Pi^{(1)} = \mathrm{Span}(\psi_1, \ldots, \psi_I).$$

Since $\nabla v_1 \in \mathrm{Ran}\Pi^{(1)}$,

$$\int_{\Sigma_i} \partial_n v_1 \exp(-\beta V) = \sum_{j=1}^{I} \langle \nabla v_1, \psi_j \rangle_{L^2(e^{-\beta V})} \int_{\Sigma_i} \psi_j \cdot n e^{-\beta V}.$$

The idea is now to build so-called quasi-modes which approximate the eigenvectors of $L^{(0)}$ and $L^{(1)}$ associated with small eigenvalues in the regime $\beta \to \infty$, in order to approximate the terms in the sum.

Sketch of the proof

- $\mathrm{Ran}\Pi^{(0)}$: an approximation of v_1 is given by

$$\tilde{v} = Z^{-1} 1_{W'}$$

  where $W' \subset\subset W$.
- $\mathrm{Ran}\Pi^{(1)}$: an approximation of $\mathrm{Ran}\Pi^{(1)}$ is $\mathrm{Span}(\tilde{\psi}_1, \ldots, \tilde{\psi}_I)$ where $(\tilde{\psi}_i)_{1 \leq i \leq I}$ are solutions to auxiliary eigenvalue problems, attached to the local minima $(z_i)_{1 \leq i \leq I}$.

Two tools:

- Agmon estimates (the support of $\tilde{\psi}_i$ is essentially in a neighborhood of z_i):

$$\exists N > 0, \ \| e^{\beta d_a(z_i, \cdot)/2} \tilde{\psi}_i \|_{H^1(e^{-\beta V})} = O(\beta^N).$$

- WKB approximations:

$$\exists N > 0, \ \tilde{\psi}_i \simeq Z_i^{-1} d(e^{\beta V/2} e^{-\beta d_a(z_i, \cdot)/2}) \beta^p.$$

Sketch of the proof

The last step consists in projecting the approximation of ∇v_1 on the approximation of $\mathrm{Ran}\Pi^{(1)}$.

Using the assumptions $V(z_1) - V(x_1) > V(z_I) - V(z_1)$ and $\inf_{z \in B_{z_i}^c} d_a(z, z_i) > V(z_I) - V(z_i)$, one can check that $\tilde{v}$ and $(\tilde{\psi}_i)_{i=1\dots I}$ are such that

- [Normalization] $\tilde{v} \in H_0^1(e^{-\beta V})$ and $\|\tilde{v}\|_{L^2(e^{-\beta V})} = 1$. $\forall i$, $\tilde{\psi}_i \in H_T^1(e^{-\beta V})$ and $\|\tilde{\psi}_i\|_{L^2(e^{-\beta V})} = 1$.
- [Good quasimodes]
  - $\forall \delta > 0$

$$\|(1 - \Pi^{(0)})\tilde{v}\|^2_{L^2(e^{-\beta V})} = O(e^{-\beta(V(z_1) - V(x_1) - \delta)}),$$

  - $\exists \varepsilon > 0$, $\forall i$,

$$\|(1 - \Pi^{(1)})\tilde{\psi}_i\|^2_{H^1(e^{-\beta V})} = O(e^{-\beta(V(z_I) - V(z_1) + \varepsilon)})$$

- [Orthonomality of quasimodes] $\exists \varepsilon_0 > 0$, $\forall i < j$

$$\langle \tilde{\psi}_i, \tilde{\psi}_j \rangle_{L^2(e^{-\beta V})} = O\left(e^{-\frac{\beta}{2}(V(z_j) - V(z_i) + \varepsilon_0)} \right).$$

Sketch of the proof

- [Decomposition of $\nabla\tilde{v}$] $\exists C_i, p, \forall i,$

$$\langle \nabla\tilde{v}, \tilde{\psi}_i \rangle_{L^2(e^{-\beta V})} = C_i \ \beta^{-p} e^{-\frac{\beta}{2}(V(z_i)-V(x_1))} \ (1 + O(\beta^{-1}) \ ).$$

- [Normal components of the quasimodes] $\exists B_i, m, \forall i, j$

$$\int_{\Sigma_i} \tilde{\psi}_j \cdot n \, e^{-\beta V} d\sigma = \begin{cases} B_i \ \beta^{-m} \ e^{-\frac{\beta}{2}V(z_i)} \ (\ 1 \ + O(\beta^{-1}) \ ) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Then for $i = 1, ..., n$, when $\beta \to \infty$

$$\int_{\Sigma_i} \partial_n v_1 \ e^{-\beta V} d\sigma = C_i B_i \ \beta^{-(p+m)} \ e^{-\frac{\beta}{2}(2V(z_i)-V(x_1))} \ (1 + O(\beta^{-1}))$$

Sketch of the proof

The proof is based on the formula:

$$\int_{\Sigma_i} \partial_n v_1 \exp(-\beta V) = \sum_{j=1}^{I} \langle \nabla v_1, \psi_j \rangle_{L^2(e^{-\beta V})} \int_{\Sigma_i} \psi_j \cdot n e^{-\beta V}.$$

Using the fact that $v_1 = \Pi^{(0)} \tilde{v}$ and that $(\psi_1, \ldots, \psi_I)$ can be obtained by a Gram-Schmidt procedure on $(\Pi^{(1)} \tilde{\psi}_1, \ldots, \Pi^{(1)} \tilde{\psi}_I)$, one can rewrite this formula in terms of $\tilde{v}$ and $(\tilde{\psi}_i)_{1 \leq i \leq n}$. Injecting the estimates then yields the result.

Conclusions

- There are two other accelerated dynamics methods: Hyperdynamics and Temperature Accelerated Dynamics. From ParRep to Hyper to TAD, the underlying assumptions for the algorithms to be correct are more and more stringent. In particular, TAD is based on the fact that the Eyring-Kramers formula yield a correct approximation of the exit event.
- The QSD is a good intermediate between continuous state dynamics and kMC-like approximations (Markov state models). Transition rates could be defined starting from the QSD.
- The QSD can be used to analyze the validity of kMC models and the Eyring-Kramers law, in the small temperature regime.

Simulating dynamics: conclusions (1/2)

There are other mathematical settings to characterize / quantify metastability:

- Large deviation techniques [Freidlin, Wentzell, Vanden Eijnden, Weare, Touchette,...] and Onsager-Machlup functionals [Stuart, Pinsky, Theil]
- Potential theoretic approaches [Bovier, Schuette, Hartmann,...]
- Spectral analysis of the Fokker Planck operator on the whole space and semi-classical analysis [Schuette, Helffer, Nier, Pavliotis]

Simulating dynamics: conclusions (2/2)

There are many other numerical techniques:

- Going from state A to state B:
  - *Local search*: the string method [E, Ren, Vanden-Eijnden], max flux [Skeel], transition path sampling methods [Chandler, Bolhuis, Dellago],
  - *Global search, ensemble of trajectories*: AMS, transition interface sampling [Bolhuis, van Erp], forward flux sampling [Allen, Valeriani, ten Wolde], milestoning techniques [Elber, Schuette, Vanden-Eijnden]
- Importance sampling approaches on paths, reweighting [Dupuis, Vanden-Einjden, Weare, Schuette, Hartmann]
- Saddle point search techniques [Mousseau, Henkelman] and graph exploration
- Starting from a long trajectory, extract states: clustering, Hidden Markov chain [Schuette]

References

Adaptive Multilevel Splitting algorithm:

- C.-E. Bréhier, M. Gazeau, L. Goudenège , TL and M. Rousset, *Unbiasedness of some generalized Adaptive Multilevel Splitting algorithms*, Annals of Applied Probability, 2016.
- F. Cérou, A. Guyader, TL and D. Pommier, *A multiple replica approach to simulate reactive trajectories*, Journal of Chemical Physics, 2011.
- I. Teo, C.G. Mayne, K. Schulten and TL, *Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time*, Journal of Chemical Theory and Computation, 2016.

References

Accelerated dynamics:

- A. Binder, TL and G. Simpson, *A Generalized Parallel Replica Dynamics*, Journal of Computational Physics, 2015.
- G. Di Gesù, TL, D. Le Peutrec and B. Nectoux, *Jump Markov models and transition state theory: the Quasi-Stationary Distribution approach*, Faraday Discussion, 2016.
- G. Di Gesù, TL, D. Le Peutrec and B. Nectoux, *Sharp asymptotics of the first exit point density*, `arXiv:1706.08728`, 2017.
- C. Le Bris, TL, M. Luskin and D. Perez, *A mathematical formalization of the parallel replica dynamics*, Monte Carlo Methods and Applications, 2012.
- TL and F. Nier, *Low temperature asymptotics for Quasi-Stationary Distributions in a bounded domain*, Analysis & PDE, 2015.
- TL and G. Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*, Acta Numerica, 2016.