Stochastic dynamics and numerical methods in molecular dynamics

T. Lelièvre

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

Joint work with C. Chipot, C. Le Bris, M. Luskin, K. Minoukadeh, D. Perez, M. Rousset and G. Stoltz.

Oberwolfach
August 2011

Introduction

The aim of molecular dynamics simulations is to understand the relationships between the macroscopic properties of a molecular system and its atomistic features. In particular, one would like to to evaluate numerically macroscopic quantities from models at the microscopic scale.

Some examples of macroscopic quantities:

(i) Thermodynamics quantities: stress, heat capacity, free energy (average of some observable wrt an equilibrium measure)

$$\mathbb{E}(\varphi(\boldsymbol{X})) = \int_{\mathbb{R}^d} \varphi(\boldsymbol{x})\, \mu(d\boldsymbol{x}).$$

(ii) Dynamical quantities: diffusion coefficients, viscosity, transition rates (average over trajectories at equilibrium)

$$\mathbb{E}(\mathcal{F}((\boldsymbol{X}_t)_{t\geq 0})) = \int_{\mathcal{C}^0(\mathbb{R}_+,\mathbb{R}^d)} \mathcal{F}((\boldsymbol{x}_t)_{t\geq 0}))\, \mathcal{W}(d((\boldsymbol{x}_t)_{t\geq 0})).$$

Introduction

Many applications in various fields: biology, physics, chemistry, materials science. Molecular dynamics computations consume today a lot of CPU time.

A molecular dynamics model amounts essentially in choosing a potential V which associates to a configuration $(x_1, ..., x_N) = x \in \mathbb{R}^{3N}$ an energy $V(x_1, ..., x_N)$.

In the canonical (NVT) ensemble, configurations are distributed according to the Boltzmann-Gibbs probability measure:

$$d\mu(x) = Z^{-1} \exp(-\beta V(x))\, dx,$$

where $Z = \int \exp(-\beta V(x))\, dx$ is the partition function and $\beta = (k_B T)^{-1}$ is proportional to the inverse of the temperature.

Introduction

Typically, V is a sum of potentials modelling interaction between two particles, three particles and four particles:

$$V = \sum_{i<j} V_1(x_i, x_j) + \sum_{i<j<k} V_2(x_i, x_j, x_k) + \sum_{i<j<k<l} V_3(x_i, x_j, x_k, x_l).$$

For example, $V_1(x_i, x_j) = V_{LJ}(|x_i - x_j|)$ where $V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)$ is the Lennard-Jones potential.

Difficulties: (i) high-dimensional problem ($N \gg 1$) ; (ii) μ is a multimodal measure.

Introduction

To sample μ, ergodic dynamics wrt to μ are used. A typical example is the *over-damped Langevin* (or gradient) dynamics:

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

which is a limit (when the mass goes to zero or the damping parameter to infinity) of the *Langevin dynamics*:

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

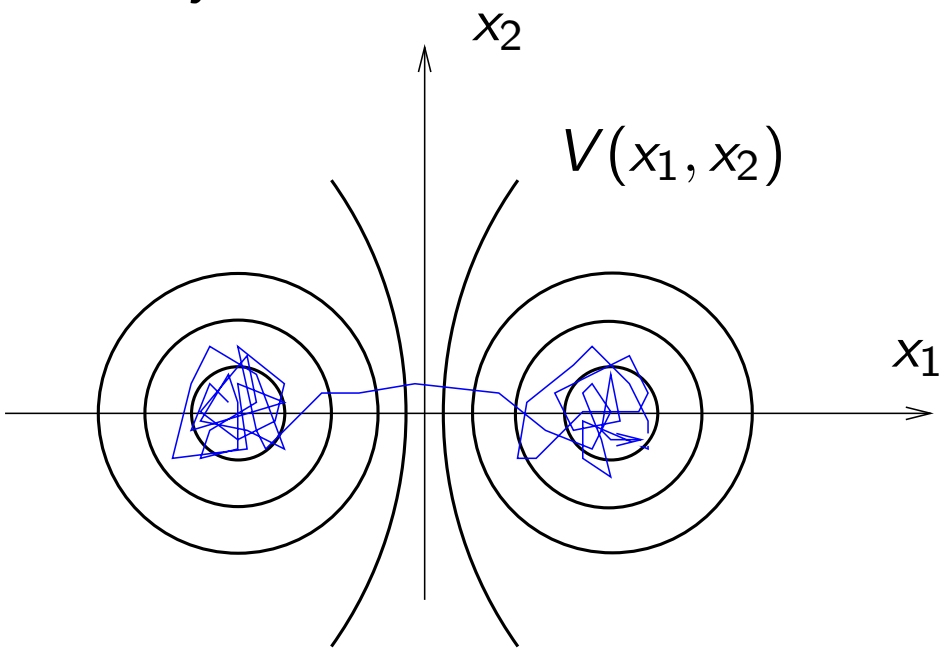
where M is the mass tensor and γ is the friction coefficient.

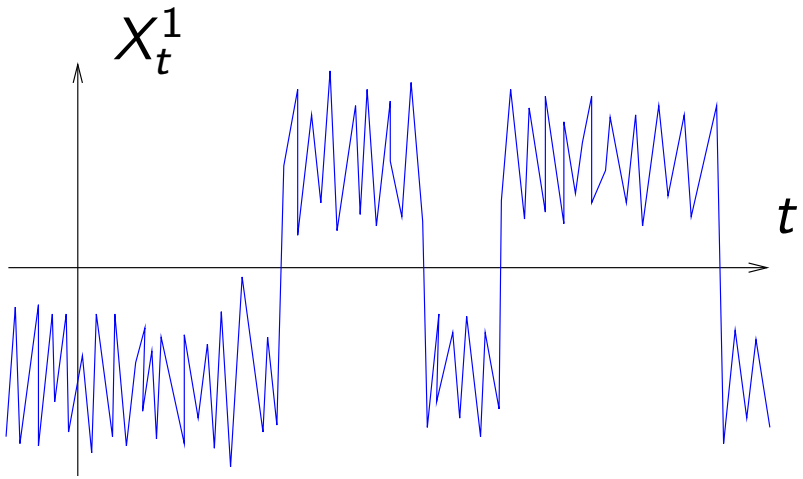
In the following, we mainly consider the over-damped Langevin dynamics.

Introduction

Difficulty: In practice, $\boldsymbol{X}_t$ is a metastable process, so that the convergence to equilibrium is very slow.

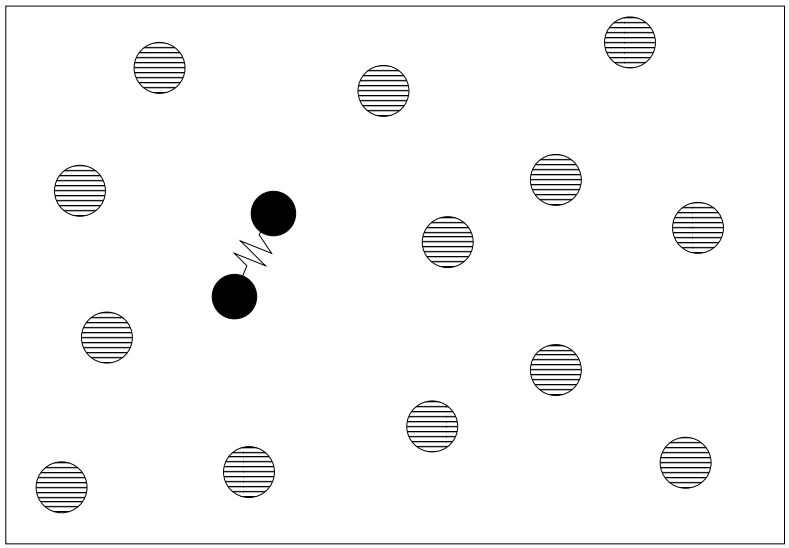
A 2d schematic picture: X_t^1 is a slow variable (a metastable dof) of the system.



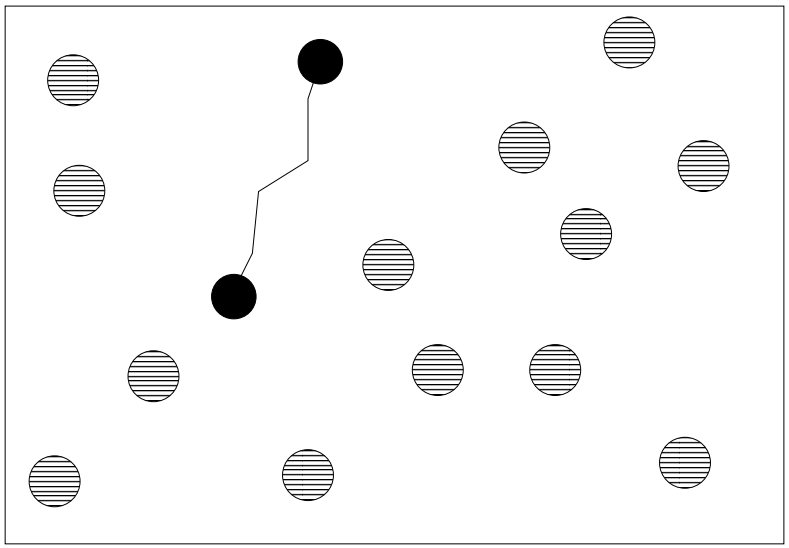


Introduction

A more realistic example (Dellago, Geissler): Influence of the solvation on a dimer conformation.



Compact state.

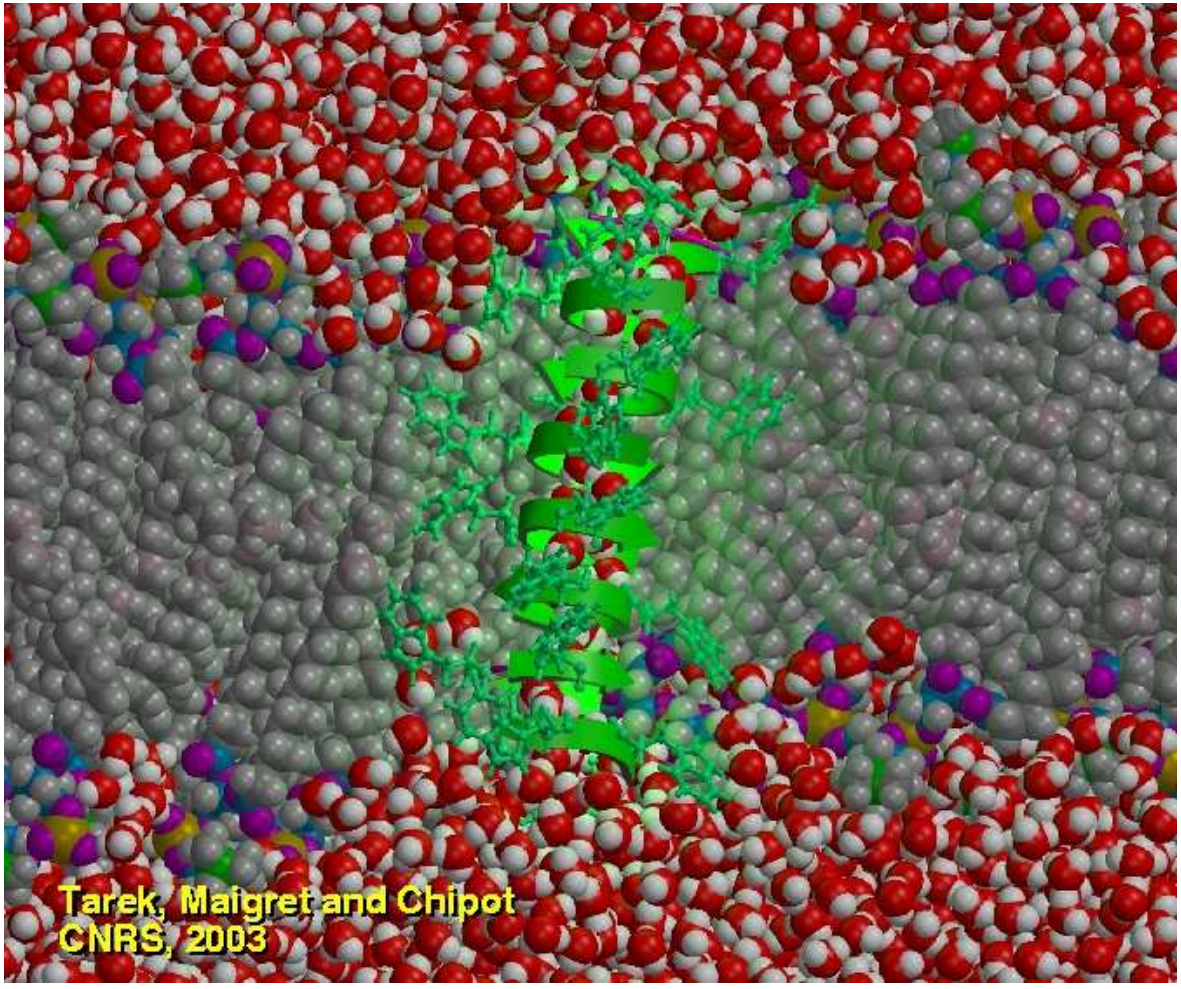


Stretched state.

The particles interact through a pair potential: truncated LJ for all particles except the two monomers (black particles) which interact through a double-well potential. A slow variable is the distance between the two monomers.

Introduction

A "real" example: ions canal in a cell membrane. *(C. Chipot)*.



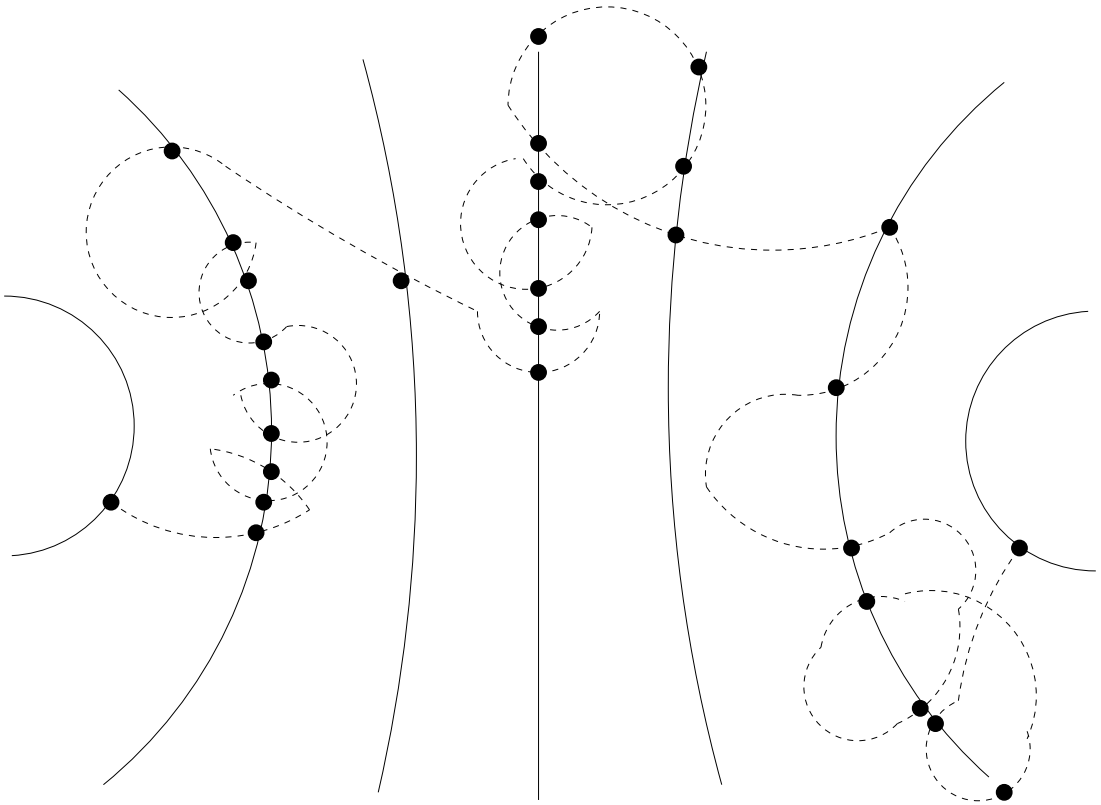
Introduction

One contral numerical difficulty is thus metastability.

Outline of the talk:

1. Adaptive biasing techniques: These belong to one class of numerical methods to compute thermodynamic quantities, and in particular free energy differences.
2. The Parallel Replica dynamics: This is one instance of an algorithm to generate efficiently metastable dynamics.

Adaptive biasing techniques



Adaptive biasing techniques

We suppose in this part that we know a slow variable of dimension 1: $\xi(\boldsymbol{X}_t)$, where $\xi : \mathbb{R}^d \longrightarrow \mathbb{T}$ is a so-called reaction coordinate.

This reaction coordinate will be used to bias the dynamics (adaptive importance sampling technique).

Adaptive biasing techniques

Let us introduce two probability measures associated to μ and ξ:

- The image of the measure μ by ξ:

$$\xi * \mu\,(dz) = \exp(-\beta A(z))\,dz$$

where the free energy A is defined by:

$$A(z) = -\beta^{-1} \ln \left(\int_{\Sigma(z)} e^{-\beta V} \delta_{\xi(x)-z}(dx) \right),$$

with $\Sigma(z) = \{x,\, \xi(x) = z\}$ is a (smooth) submanifold of $\mathbb{R}^d$, and $\delta_{\xi(x)-z}(dx)\,dz = dx$.

- The probability measure μ conditioned to $\xi(x) = z$:

$$\mu_{\Sigma(z)}(dx) = \frac{\exp(-\beta V(x))\,\delta_{\xi(x)-z}(dx)}{\exp(-\beta A(z))}.$$

Adaptive biasing techniques

The bottom line of adaptive methods is the following: for "well chosen" ξ the potential $V - A \circ \xi$ is less metastable than V. But A is unknown !

Principle: use a time dependent potential of the form

$$\mathcal{V}_t(x) = V(x) - A_t(\xi(x))$$

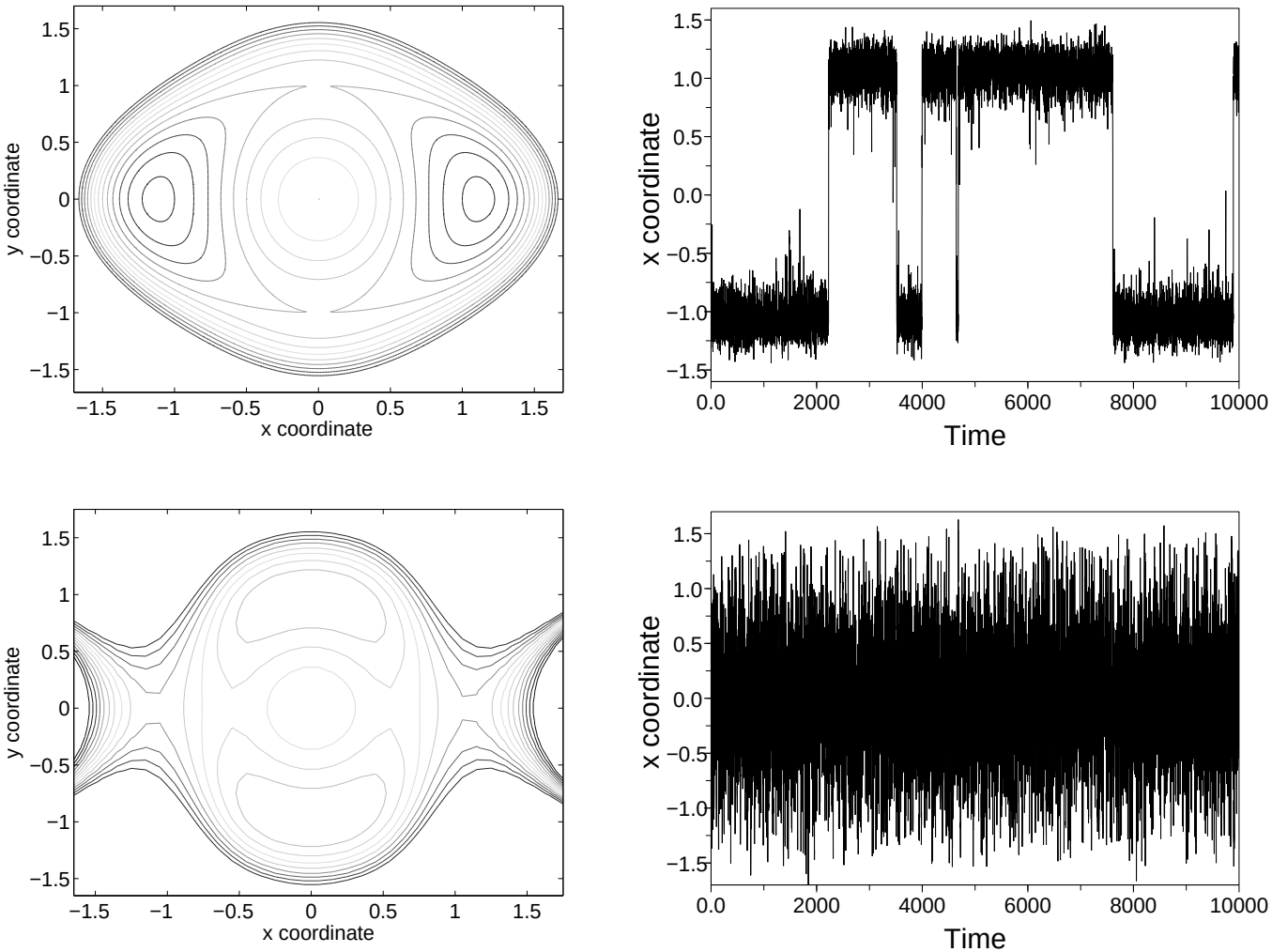
where A_t is an approximation at time t of A, given the configurations visited so far.

Hopes:

- build a dynamics which goes quickly to equilibrium,
- compute free energy profiles.

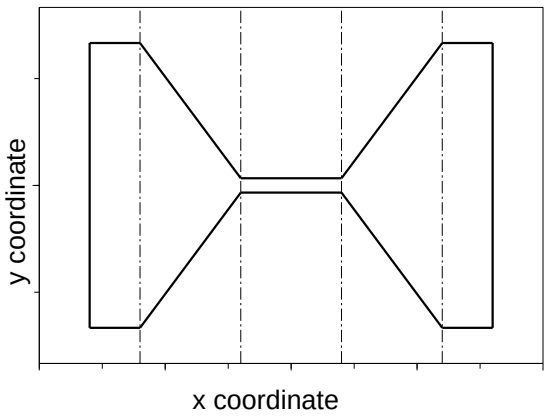
Wang-Landau, ABF, metadynamics: *Darve, Pohorille, Hénin, Chipot, Laio, Parrinello, Wang, Landau,...*

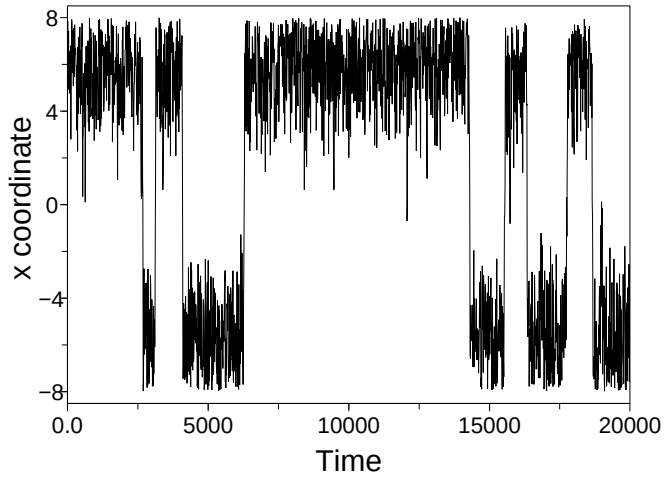
Adaptive biasing techniques

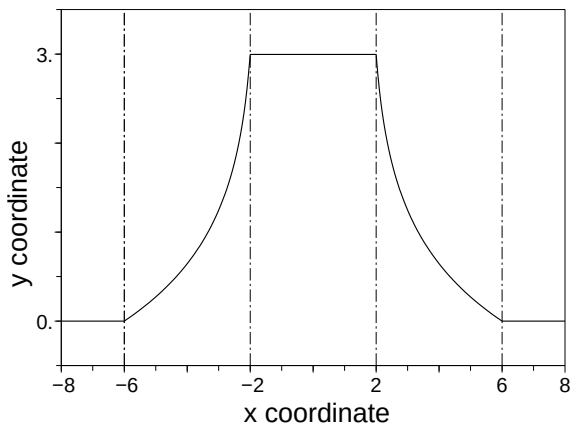


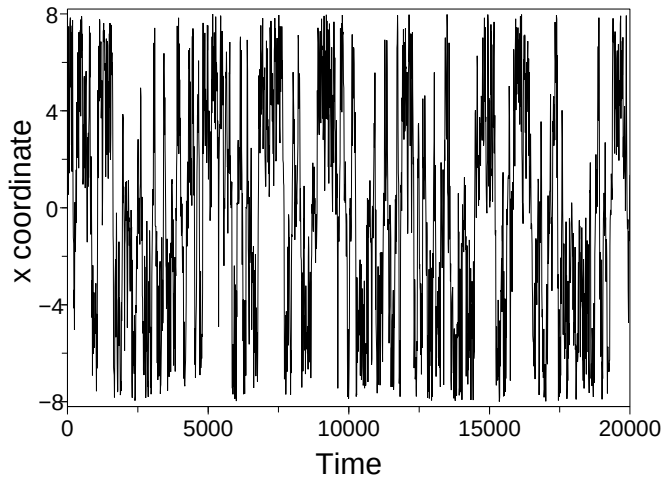
A 2d example of a free energy biased trajectory: energetic barrier.

Adaptive biasing techniques









A 2d example of a free energy biased trajectory: entropic barrier.

The ABF method

How to update A_t ? Two methods depending on wether A'_t (Adaptive Biasing Force) or A_t (Adaptive Biasing Potential) is approximated.

For the Adaptive Biasing Force method, the idea is to use the formula

$$A'(z) = \frac{\int \left(\frac{\nabla V \cdot \nabla \xi}{|\nabla \xi|^2} - \beta^{-1} \mathrm{div} \left(\frac{\nabla \xi}{|\nabla \xi|^2} \right) \right) e^{-\beta V} \, \delta_{\xi(x)-z}(dx)}{\int e^{-\beta V} \, \delta_{\xi(x)-z}(dx)}$$

$$= \int f \, d\mu_{\Sigma(z)} = \mathbb{E}_\mu(f(\boldsymbol{X})|\xi(\boldsymbol{X}) = z).$$

The mean force $A'(z)$ is the mean of f with respect to $\mu_{\Sigma(z)}$.

The ABF method

Important remark: whatever A_t, the mean force associated with the Gibbs distribution

$$\psi^{\mathrm{eq}} \propto \exp(-\beta\, \mathcal{V}_t)(\boldsymbol{x})\, d\boldsymbol{x} = \exp(-\beta(V - A_t \circ \xi))(\boldsymbol{x})\, d\boldsymbol{x}$$

is the original mean force A':

$$\frac{\int f\, \psi^{\mathrm{eq}}\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}{\int \psi^{\mathrm{eq}} \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})} = A'(z).$$

Thus, use as an approximation of $A'(z)$:

$$A'_t(z) = \mathbb{E}(f(\boldsymbol{X}_t)|\xi(\boldsymbol{X}_t) = z).$$

The ABF method

A typical ABF dynamics is then:

$$\begin{cases} d\boldsymbol{X}_t = -\nabla(V - A_t \circ \xi)(\boldsymbol{X}_t)\, dt + \sqrt{2\beta^{-1}} d\boldsymbol{W}_t, \\ A_t'(z) = \mathbb{E}\left(f(\boldsymbol{X}_t) | \xi(\boldsymbol{X}_t) = z\right). \end{cases}$$

The associated (nonlinear) Fokker-Planck equation writes:

$$\begin{cases} \partial_t \psi = \operatorname{div}\left(\nabla(V - A_t \circ \xi)\psi + \beta^{-1}\nabla\psi\right), \\ A_t'(z) = \dfrac{\displaystyle\int f\, \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}{\displaystyle\int \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}, \end{cases}$$

where $\boldsymbol{X}_t \sim \psi(t, \boldsymbol{x})\, d\boldsymbol{x}$. A numerical illustration.
Questions: Does A_t' converge to A' ? What did we gain compared to the original gradient dynamics ?

Longtime convergence and entropy (1)

Recall the original gradient dynamics:

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

The associated (linear) Fokker-Planck equation writes:

$$\partial_t \phi = \mathrm{div}\, \left(\nabla V \phi + \beta^{-1} \nabla \phi\right).$$

where $\boldsymbol{Q}_t \sim \phi(t, \boldsymbol{q})\, d\boldsymbol{q}$.

The metastable behaviour of $\boldsymbol{Q}_t$ is related to the multimodality of μ, which can be quantified through the rate of convergence of ϕ to $\phi_\infty = Z^{-1} \exp(-\beta V)$.

A classical PDE approach: use entropy techniques.

Longtime convergence and entropy (2)

Notice that the Fokker-Planck equation rewrites

$$\partial_t \phi = \beta^{-1} \mathrm{div} \left(\phi_\infty \nabla \left(\frac{\phi}{\phi_\infty} \right) \right).$$

Let us introduce the entropy:

$$E(t) = H(\phi(t,\cdot)|\phi_\infty) = \int \ln \left(\frac{\phi}{\phi_\infty} \right) \phi.$$

We have (Csiszár-Kullback inequality):

$$\|\phi(t,\cdot) - \phi_\infty\|_{L^1} \leq \sqrt{2E(t)}.$$

Longtime convergence and entropy (3)

$$\begin{aligned}\frac{dE}{dt} &= \int \ln\left(\frac{\phi}{\phi_\infty}\right)\partial_t\phi \\ &= \beta^{-1}\int \ln\left(\frac{\phi}{\phi_\infty}\right)\operatorname{div}\left(\phi_\infty\nabla\left(\frac{\phi}{\phi_\infty}\right)\right) \\ &= -\beta^{-1}\int\left|\nabla\ln\left(\frac{\phi}{\phi_\infty}\right)\right|^2\phi =: -\beta^{-1}I(\phi(t,\cdot)|\phi_\infty).\end{aligned}$$

If V is such that the following Logarithmic Sobolev inequality (LSI(R)) holds: $\forall\phi$ pdf,

$$H(\phi|\phi_\infty) \leq \frac{1}{2R}I(\phi|\phi_\infty)$$

then $E(t) \leq E(0)\exp(-2\beta^{-1}Rt)$ and thus ϕ converges to ϕ_∞ exponentially fast with rate $\beta^{-1}R$.

Metastability $\Longleftrightarrow$ small R

Convergence of ABF (1)

A convergence result [TL, M. Rousset, G. Stoltz, Nonlinearity 2008] : Recall the ABF Fokker-Planck equation:

$$\begin{cases} \partial_t \psi = \mathrm{div}\, \left(\nabla (V - A_t \circ \xi) \psi + \beta^{-1} \nabla \psi \right), \\ A_t'(z) = \dfrac{\int f\, \psi\, \delta_{\xi(x)-z}(dx)}{\int \psi\, \delta_{\xi(x)-z}(dx)}. \end{cases}$$

Suppose:

(H1) "Ergodicity" of the microscopic variables: the conditional probability measures $\mu_{\Sigma(z)}$ satisfy a LSI(ρ),

(H2) Bounded coupling: $\left\| \nabla_{\Sigma(z)} f \right\|_{L^\infty} < \infty$,

then

$$\| A_t' - A' \|_{L^2} \leq C \exp(-\beta^{-1} \min(\rho, r) t).$$

The rate of convergence is limited by:

- the rate r of convergence of $\overline{\psi} = \int \psi\, \delta_{\xi(x)-z}(dx)$ to $\overline{\psi_\infty}$,
- the LSI constant ρ (the real limitation).

Convergence of ABF (2)

In summary:

- Original gradient dynamics: $\exp(-\beta^{-1} R t)$ where R is the LSI constant for μ ;
- ABF dynamics: $\exp(-\beta^{-1} \rho t)$ where ρ is the LSI constant for the conditioned probability measures $\mu_{\Sigma(z)}$.

If ξ is well chosen, $\rho \gg R$.

Two ingredients of the proof:

(1) The marginal $\overline{\psi}(t,z) = \int \psi(t,x)\, \delta_{\xi(x)-z}(dx)$ satisfies a closed PDE:

$$\partial_t \overline{\psi} = \beta^{-1} \partial_{z,z} \overline{\psi} \text{ on } \mathbb{T},$$

and thus, $\overline{\psi}$ converges towards $\overline{\psi_\infty} \equiv 1$, with exponential speed $C \exp(-4\pi^2 \beta^{-1} t)$. (Here, $r = 4\pi^2$).

Convergence of ABF (3)

(2) The total entropy can be decomposed as [N. Grunewald, F. Otto, C. Villani, M. Westdickenberg, Ann. IHP, 2009]:

$$E = E_M + E_m$$

where

The total entropy is $E = H(\psi|\psi_\infty)$,

The macroscopic entropy is $E_M = H(\overline{\psi}|\overline{\psi_\infty})$,

The microscopic entropy is

$$E_m = \int H\Big(\psi(\cdot|\xi(x) = z)\Big|\psi_\infty(\cdot|\xi(x) = z)\Big)\,\overline{\psi}(z)\,dz.$$

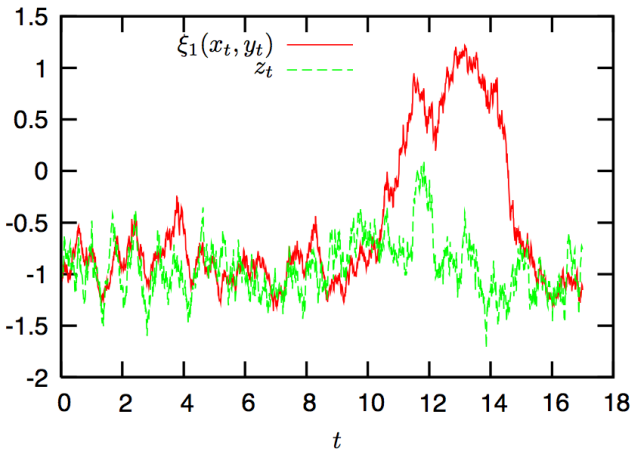
We already know that E_M goes to zero: it remains only to consider E_m...

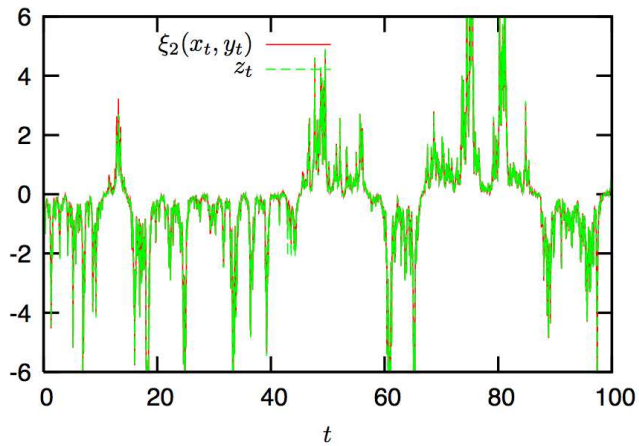
Convergence of ABF (4)

Other results based on this set of assumptions:

- [TL, JFA 2008] LSI for the cond. meas. $\mu_{\Sigma(z)}$
  + LSI for the marginal $\overline{\mu}(dz) = \xi * \mu(dz)$
  + bdd coupling ($\|\nabla_{\Sigma(z)} f\|_{L^\infty} < \infty$) $\Longrightarrow$ LSI for μ.
- [F. Legoll, TL, Nonlinearity, 2010] Effective dynamics for $\xi(\boldsymbol{Q}_t)$. Uniform control in time:

$$H(\mathcal{L}(\xi(\boldsymbol{Q}_t))|\mathcal{L}(z_t)) \leq C \left(\frac{\|\nabla_{\Sigma(z)} f\|_{L^\infty}}{\rho} \right)^2 H(\mathcal{L}(\boldsymbol{Q}_0)|\mu).$$





Discretization of ABF (1)

Discretization of adaptive methods can be done using two (complementary) approaches:

- Use trajectorial averages along a single path:

$$\mathbb{E}(f(\boldsymbol{X}_t)|\xi(\boldsymbol{X}_t)=z) \simeq \frac{\int_0^t f(\boldsymbol{X}_s)\,\delta^\alpha(\xi(\boldsymbol{X}_s)-z)\,ds}{\int_0^t \delta^\alpha(\xi(\boldsymbol{X}_s)-z)\,ds}.$$

- Use empirical means over many replicas (interacting particle system):

$$\mathbb{E}(f(\boldsymbol{X}_t)|\xi(\boldsymbol{X}_t)=z) \simeq \frac{\sum_{m=1}^N f(\boldsymbol{X}_t^{m,N})\,\delta^\alpha(\xi(\boldsymbol{X}_t^{m,N})-z)}{\sum_{m=1}^N \delta^\alpha(\xi(\boldsymbol{X}_t^{m,N})-z)}.$$

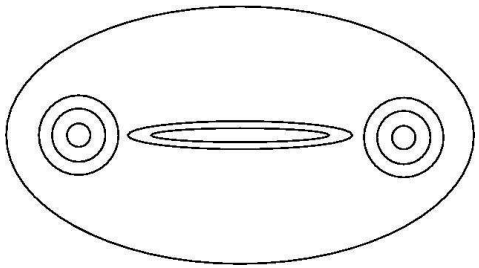
Discretization of ABF (2)

Interest of a discretization using an interacting particle system:

- Convergence can be more easily analyzed !
- Very efficient parallelization.
- A selection mechanism may be added to duplicate "innovative particles" and kill "redundant particles". [TL, M. Rousset, G. Stoltz, J Chem Phys 2007].
- Better sampling of all reactive paths.

Multiple channel cases

In some practical cases (multi-channel case), ρ may be small... Are these convergence results optimal ?



Numerically, it is observed [C. Chipot, TL, K. Minoukadeh, 2010] that in such a situation, the ABF method actually converges rapidly, in particular when using implementations using many replicas.

Theoretically, it can be shown [TL, K. Minoukadeh, 2010] that, in a bi-channel situation, the ABF method actually converges with a rate limited by the LSI constants of the conditional measures in each channel.

Adaptive biasing techniques

Coming back to the original aim: how to use free energy to compute canonical averages $\int \varphi d\mu = \int \varphi Z^{-1} \exp(-\beta V)$?

- Importance sampling:

$$\int \varphi \, d\mu = \frac{\int \varphi \exp(-\beta A \circ \xi) \, Z_A^{-1} \exp(-\beta(V - A \circ \xi)}{\int \exp(-\beta A \circ \xi) \, Z_A^{-1} \exp(-\beta(V - A \circ \xi))}.$$

- Conditioning:

$$\int \varphi d\mu = \frac{\int_z \left(\int_{\Sigma(z)} \varphi \, d\mu_{\Sigma(z)} \right) e^{-\beta A(z)} \, dz}{\int_z e^{-\beta A(z)} \, dz}.$$

This requires the sampling of the conditional probability measure $\mu_{\Sigma(z)}$ which can be done using projected Langevin dynamics [TL, M. Rousset, G. Stoltz, 2011].

Adaptive biasing techniques: conclusions

Interesting features of the algorithm: parallelization and adaptivity.

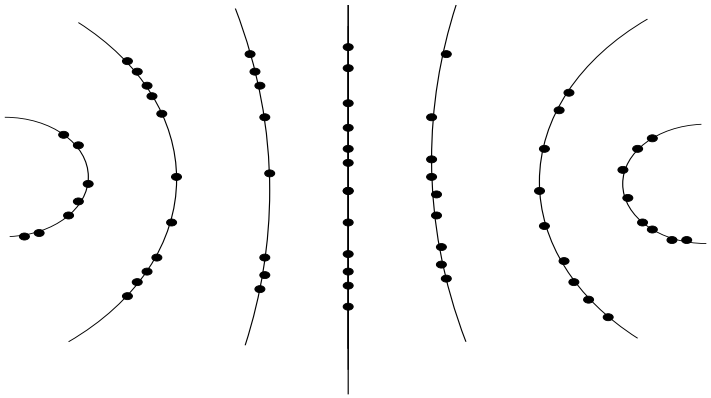
Entropy approaches are powerful techniques to investigate multimodal measures, metastable dynamics and analyze sampling algorithms.

These techniques can be used whenever the sampling of a multimodal measure is involved, for example for statistical inference in Bayesian statistics [N. Chopin, TL, G. Stoltz, 2011].

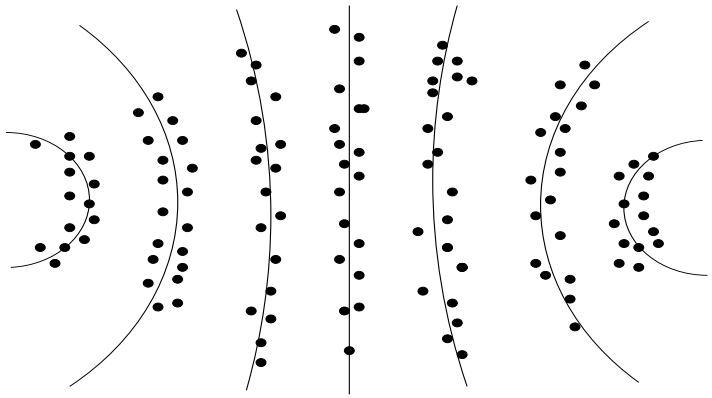
Challenges: (i) Extensions to high-dimensional RCs ; (ii) Extensions to cases with no detailed balance.

Free energy calculation methods

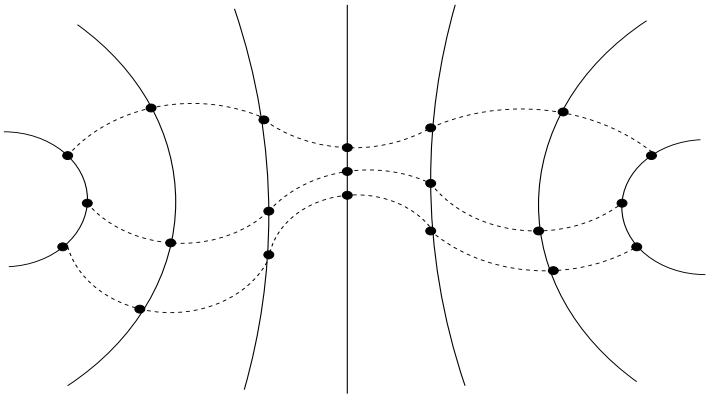
There many other numerical methods to compute free energies.



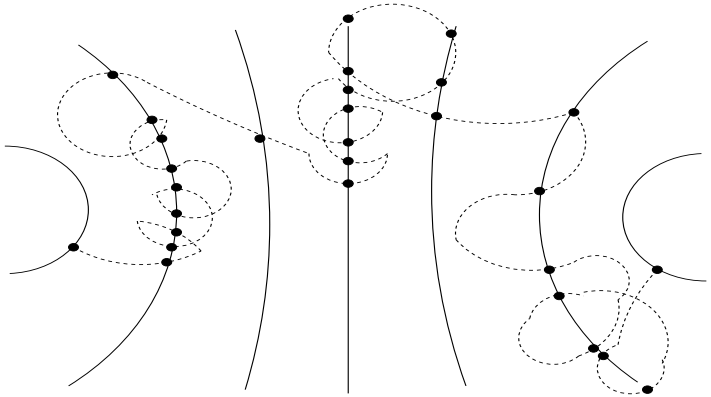
(a) Thermodynamic integration.



(b) Histogram method.

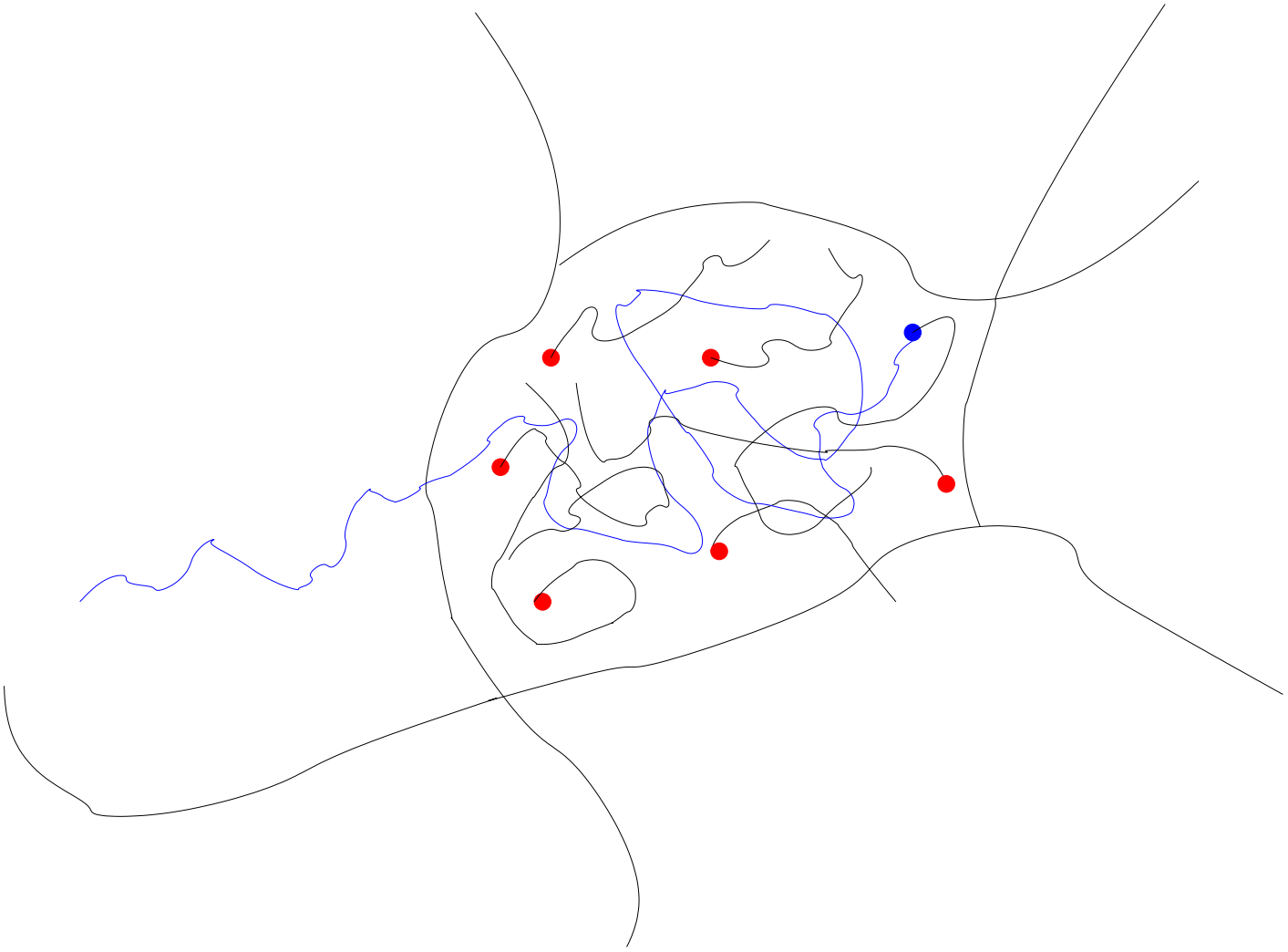


(c) Out of equilibrium dynamics.



(d) Adaptive dynamics.

The Parallel Replica Algorithm



The Parallel Replica Algorithm

The Parallel Replica Algorithm, proposed by A.F. Voter in 1998, is a method to get efficiently a "coarse-grained projection" of a dynamics.
Let us consider again the overdamped Langevin dyanmics:

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

and let assume that we are given a smooth 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.

The aim of the parallel replica dynamics is to 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 Parallel Replica Algorithm

Initialization: Consider an initial condition $\boldsymbol{X}_0^{ref}$ for a reference walker, the associated initial condition $S_0 = \mathcal{S}(\boldsymbol{X}_0^{ref})$, and a simulation time counter $T_{simu} = 0$.

Then, one iteration of the algorithm goes through three steps.

- The decorrelation step: Let the reference walker $(\boldsymbol{X}^{ref}_{T_{simu}+t})_{t\geq 0}$ evolve over a time interval $t \in [0, \tau_{corr}]$. Then,
  - If the process leaves the well during the time interval (*i.e.* $\exists t \leq \tau_{corr}$ such that $\mathcal{S}\left(\boldsymbol{X}^{ref}_{T_{simu}+t}\right) \neq \mathcal{S}\left(\boldsymbol{X}^{ref}_{T_{simu}}\right)$) advance the simulation clock by τ_{corr} and restart the decorrelation step ;
  - otherwise, advance the simulation clock by τ_{corr} and proceed to the dephasing step.

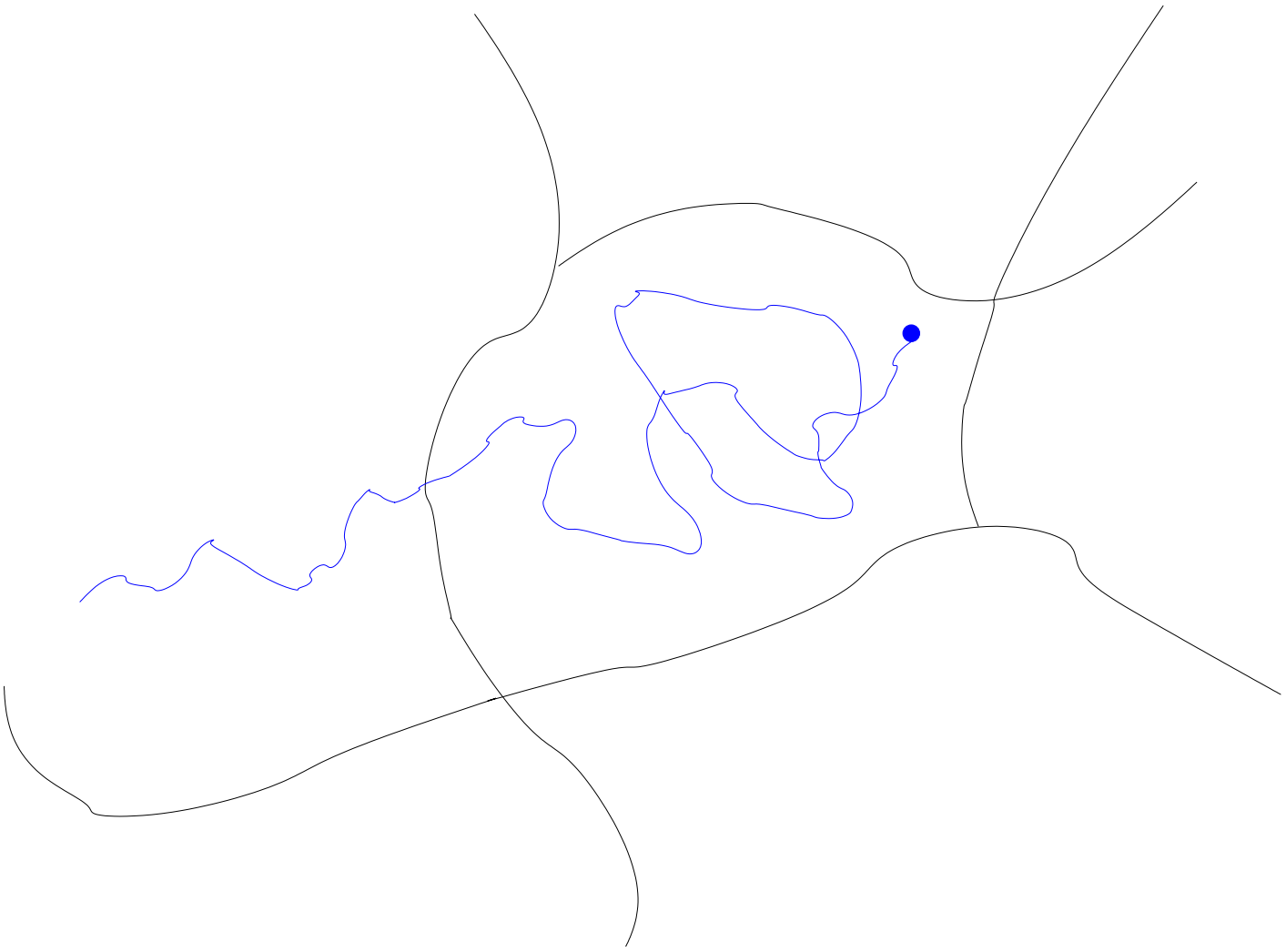
  During all this step, $S_{T_{simu}+t} := \mathcal{S}\left(\boldsymbol{X}^{ref}_{T_{simu}+t}\right)$.

The Parallel Replica Algorithm

The reference walker enters a new state

The Parallel Replica Algorithm

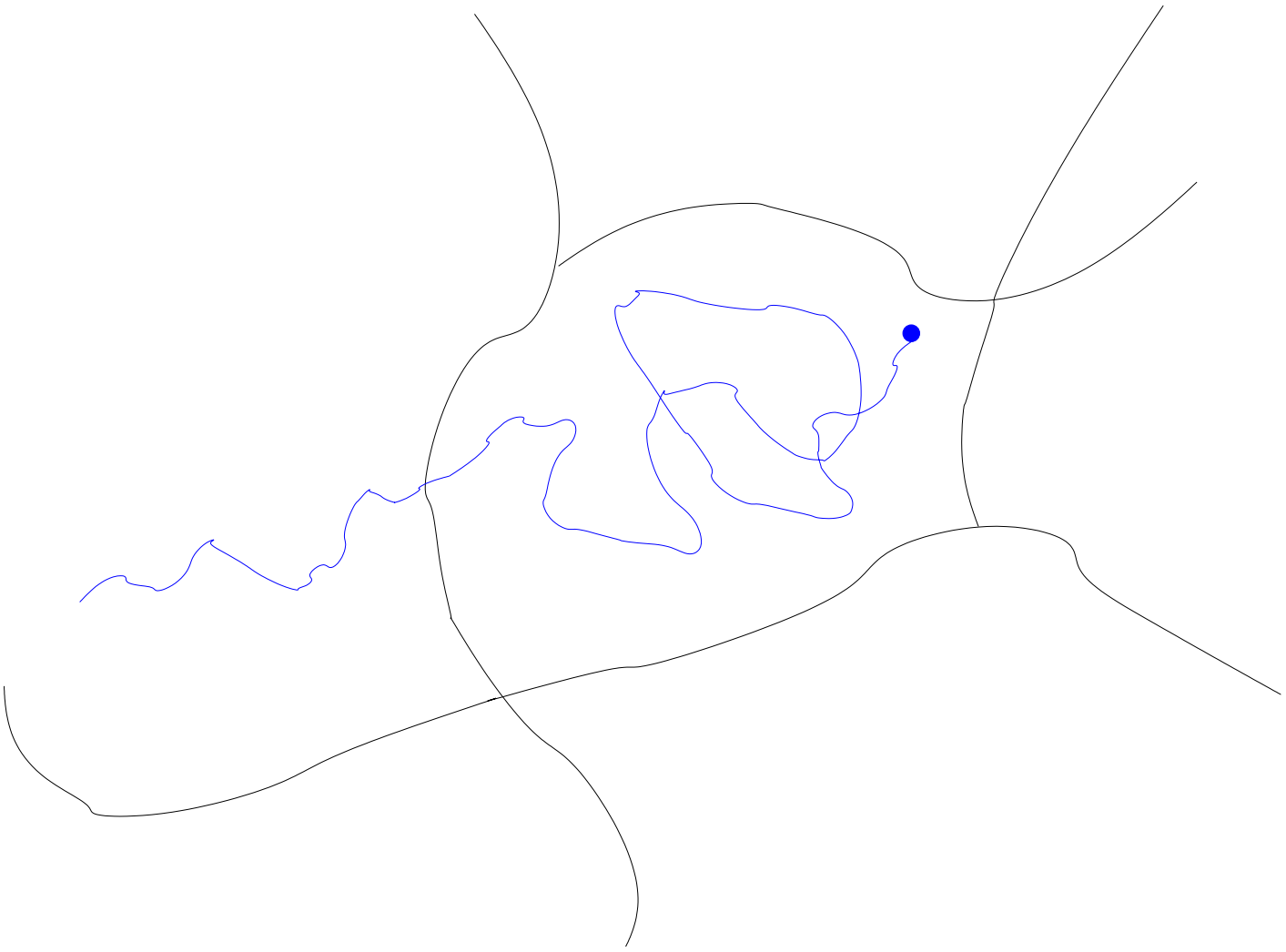
Decorrelation step: wait for a time τ_{corr}.



- The dephasing step: Duplicate the walker $\boldsymbol{X}_{T_{simu}}^{ref}$ into N replicas. Let these replicas evolve independently and in parallel over a time interval of length $\tau_{dephase}$. If a replica leaves the well during this time interval, restart the dephasing step for this replica. Throughout this step, the simulation counter is stopped.

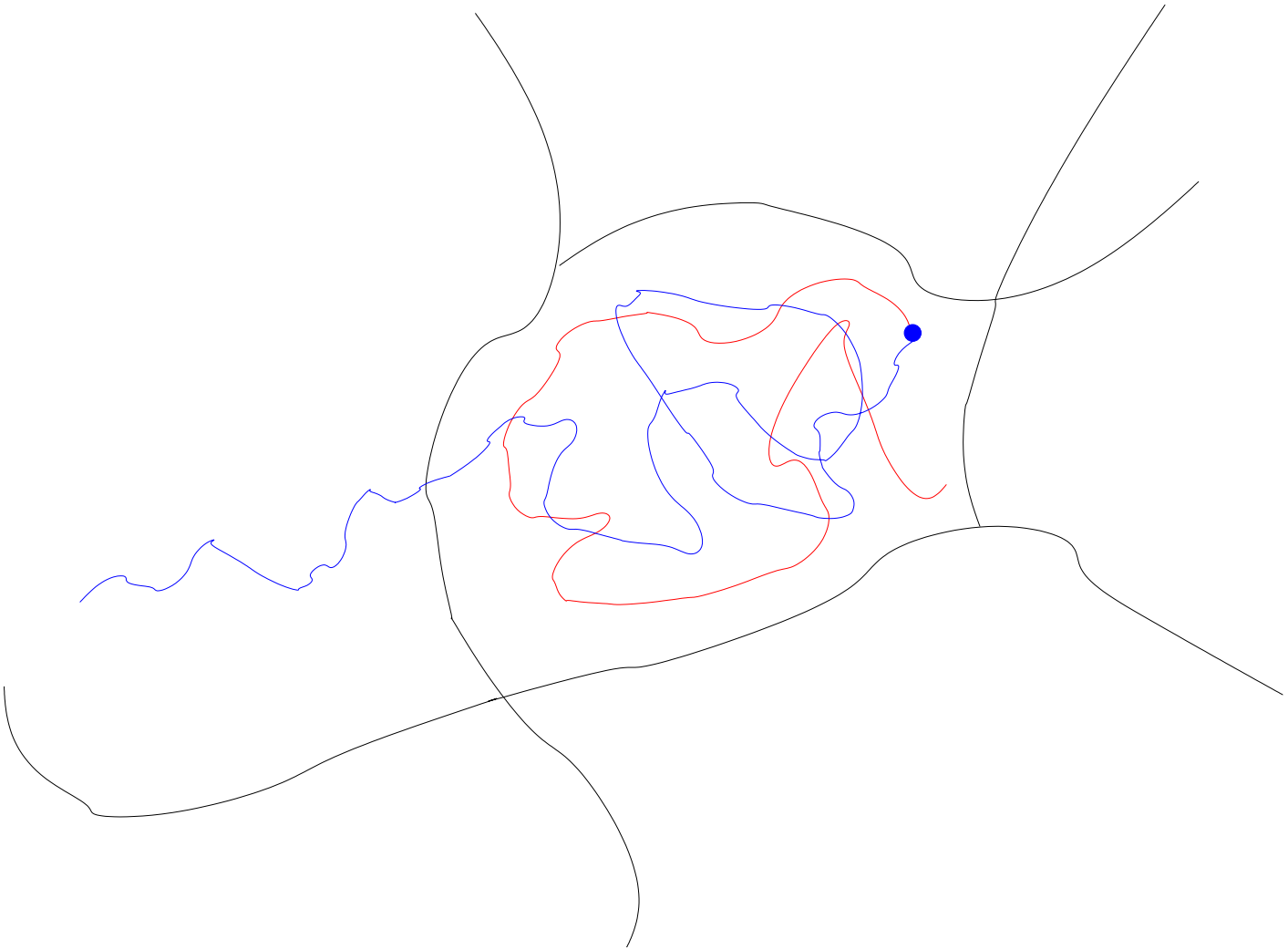
The Parallel Replica Algorithm

Dephasing step.



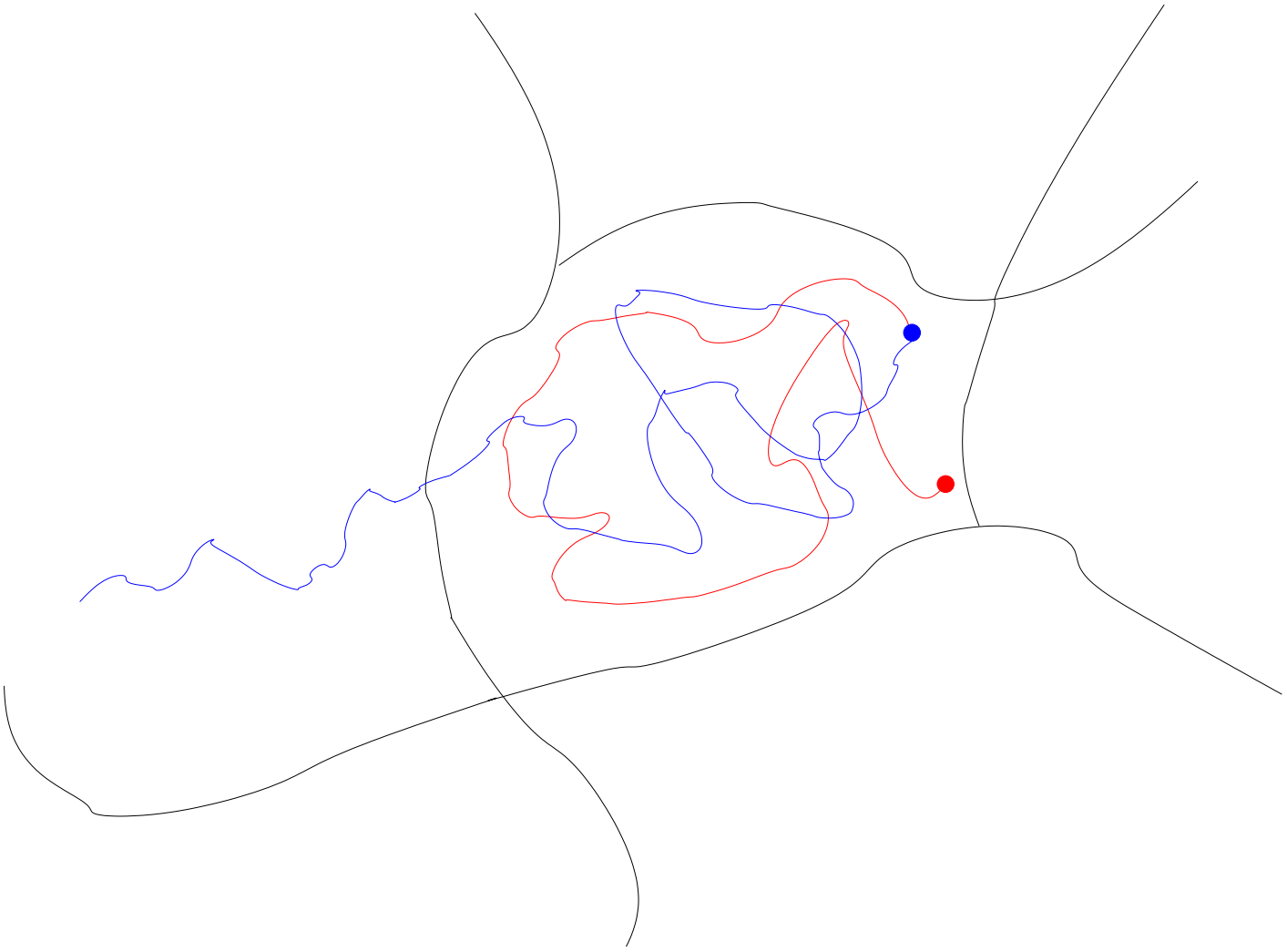
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



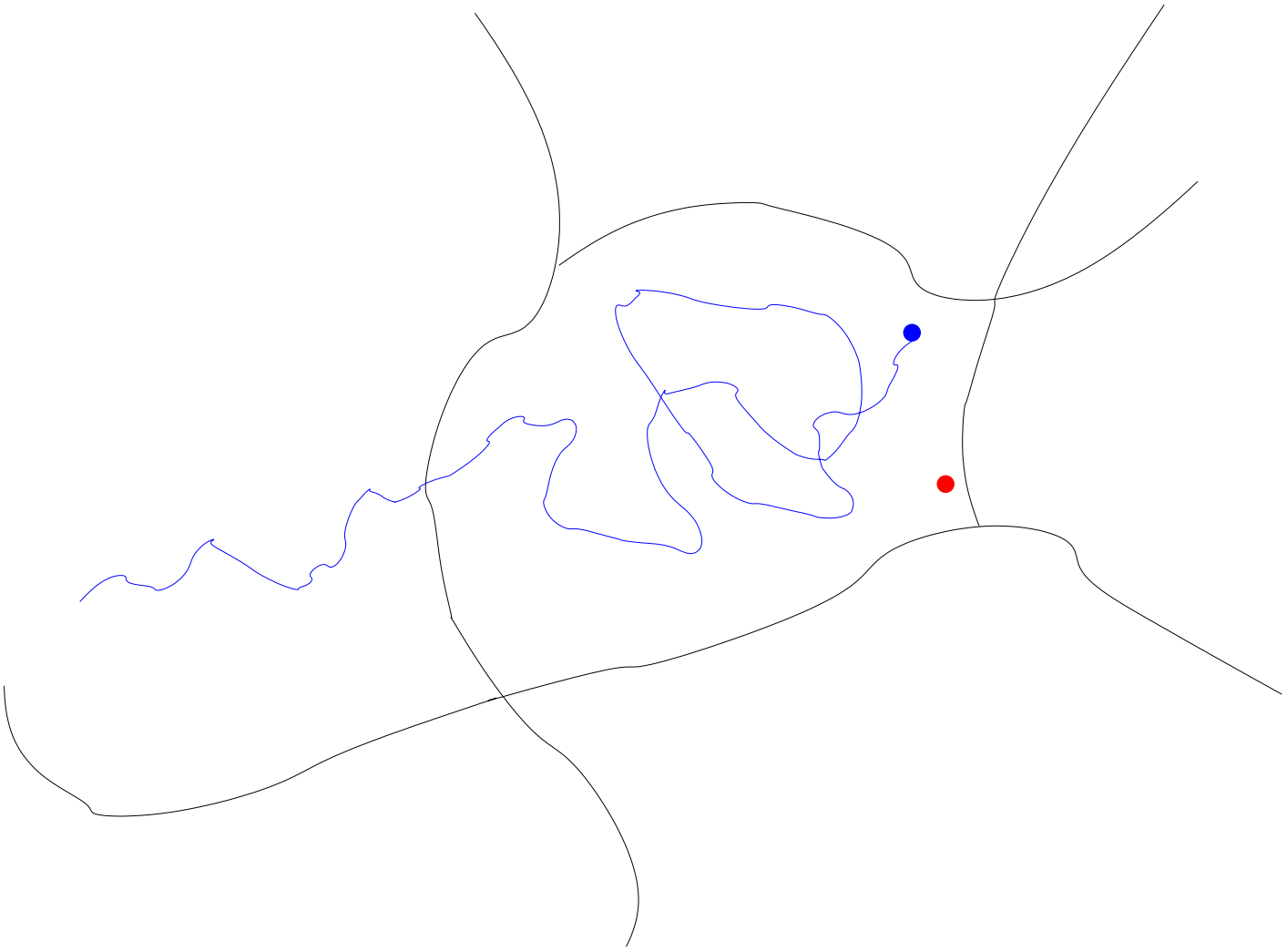
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



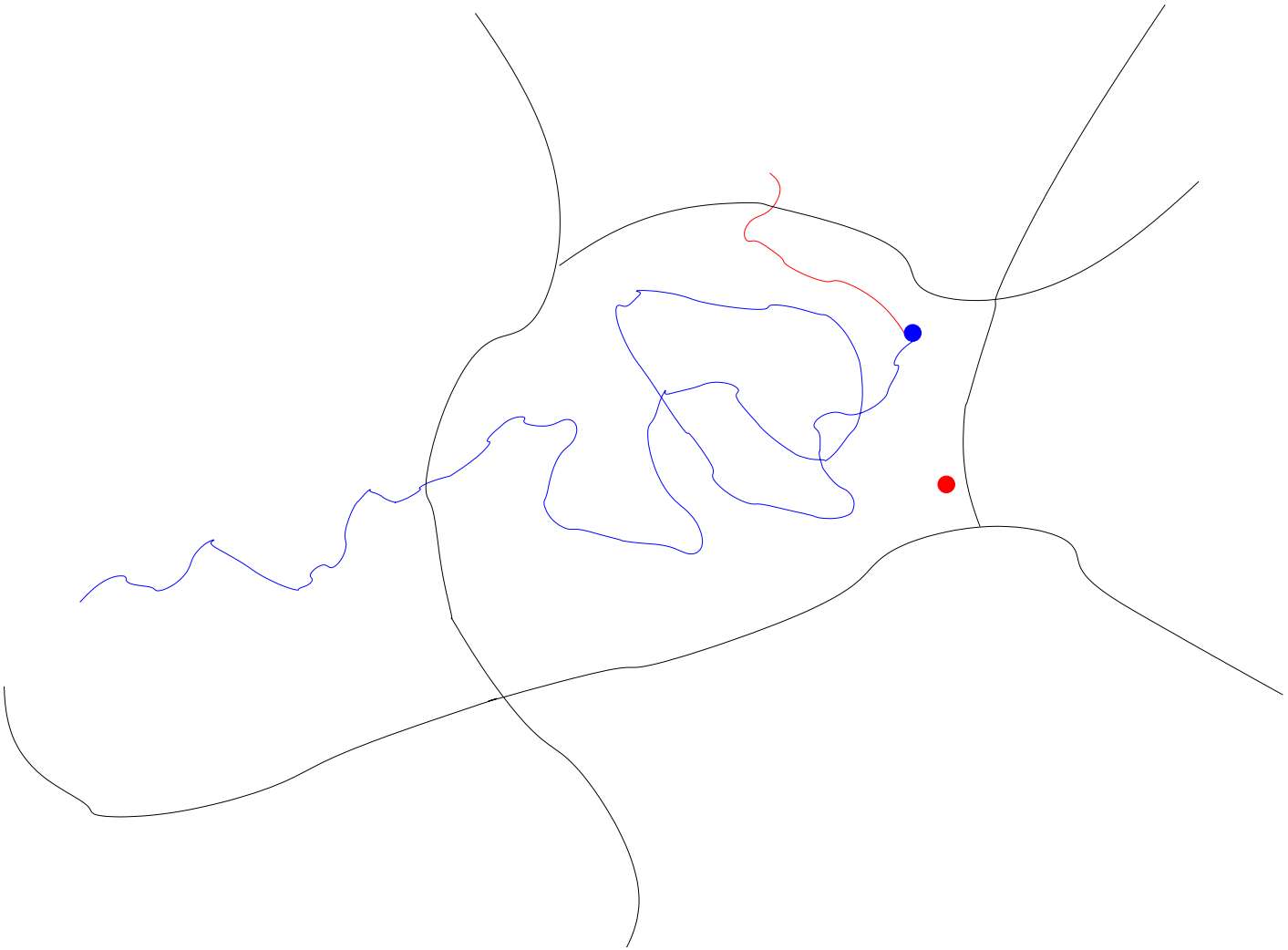
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



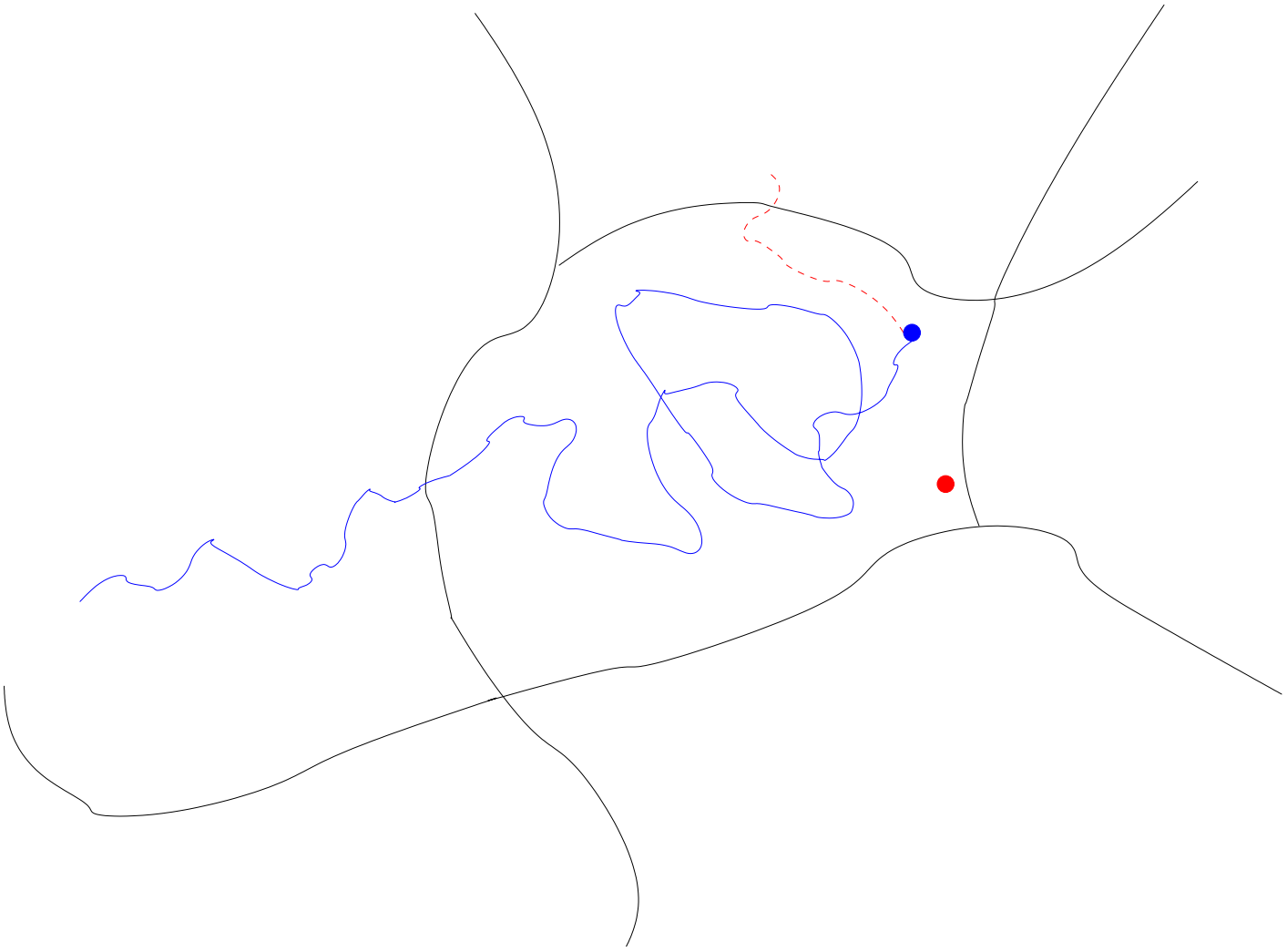
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



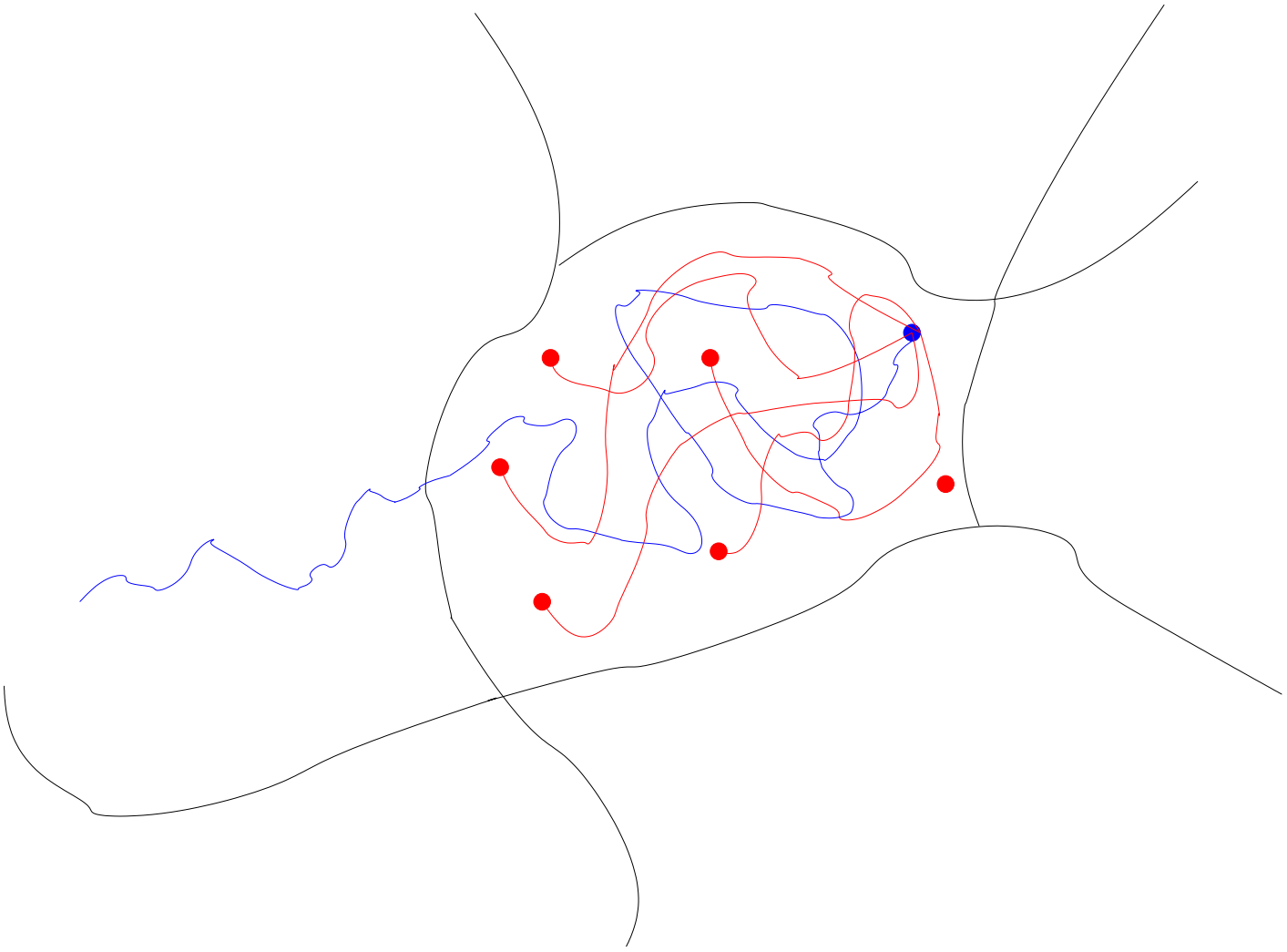
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



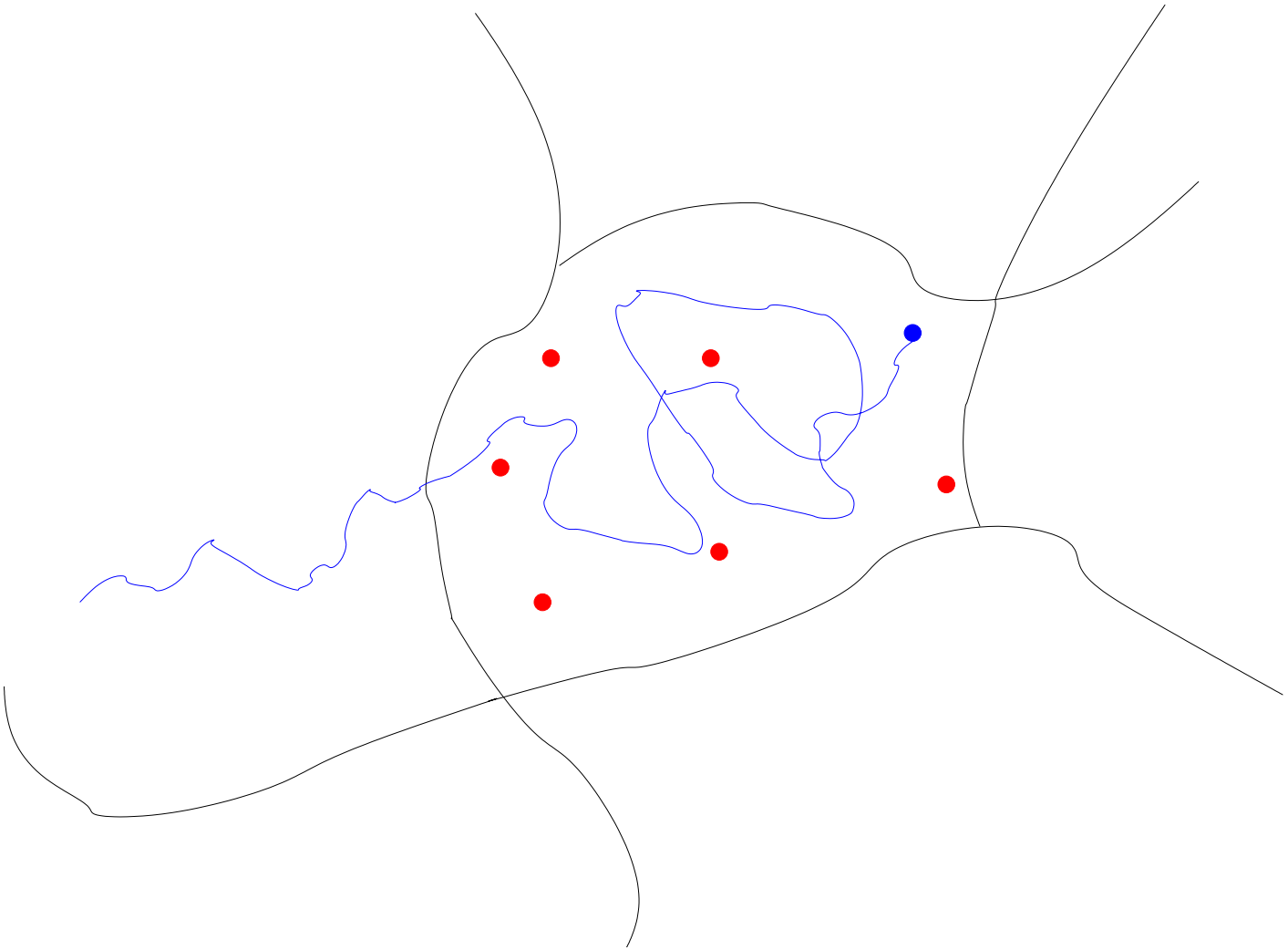
The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



The Parallel Replica Algorithm

Dephasing step: generate new initial conditions in the state.



- The parallel step: Let all the replicas evolve independently and track the first escape event:

$$T = \inf_k T_W^k = T_W^{K_0}$$

where $K_0 = \arg\inf_k T_W^k$ and

$$T_W^k = \inf\{t \geq 0,\ \mathcal{S}(\boldsymbol{X}^k_{T_{simu}+t}) \neq \mathcal{S}(\boldsymbol{X}^k_{T_{simu}})\}$$

is the first time the k-th replica leaves the well. Then:

$$T_{simu} = T_{simu} + NT \text{ and } \boldsymbol{X}^{ref}_{T_{simu}+NT} = \boldsymbol{X}^{K_0}_{T_{simu}+T}.$$

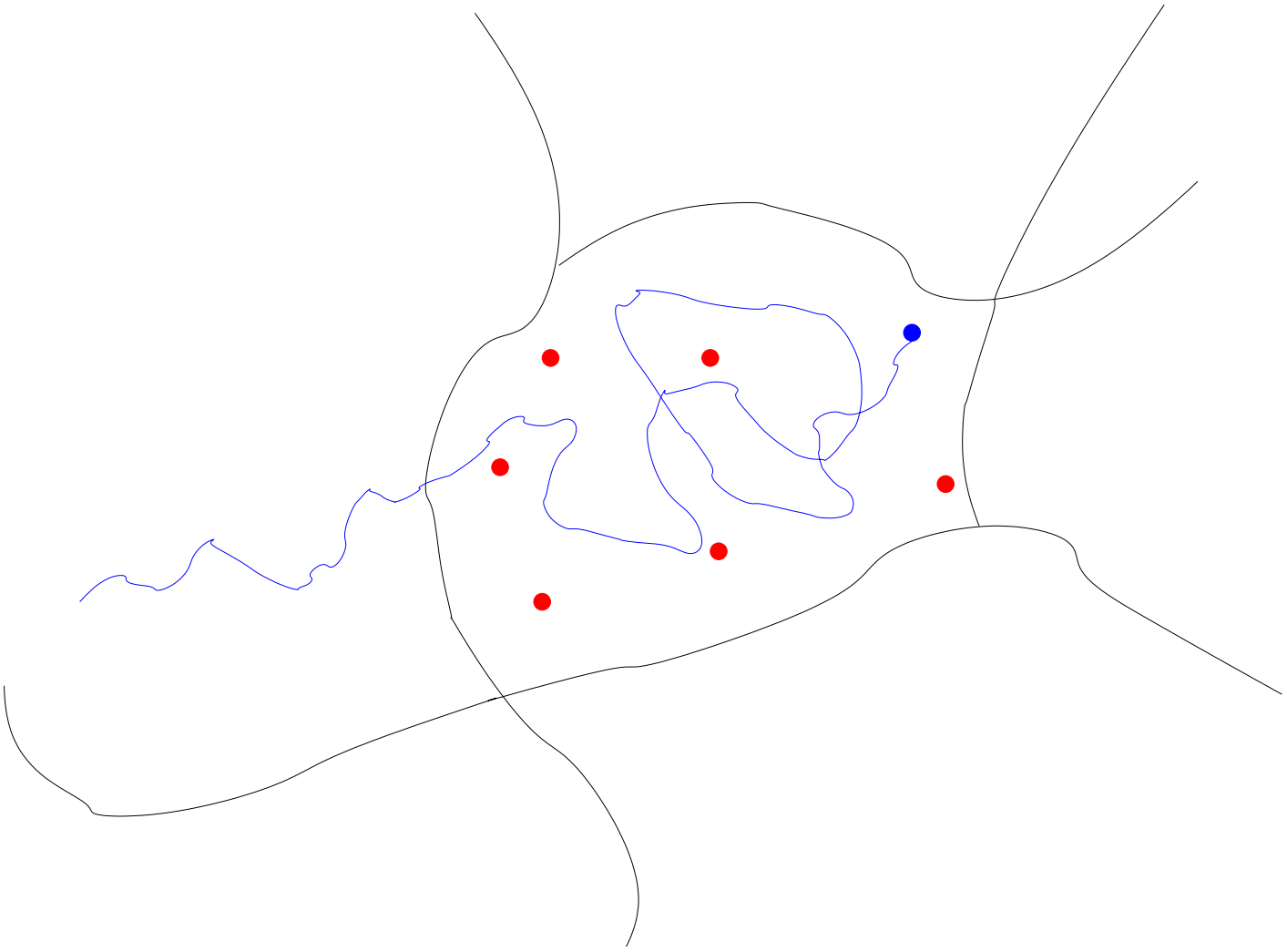
Moreover, over $[T_{simu}, T_{simu} + NT]$, the state dynamics S_t is constant and defined as:

$$S_t = \mathcal{S}(\boldsymbol{X}^1_{T_{simu}}).$$

Then, go back to the decorrelation step...

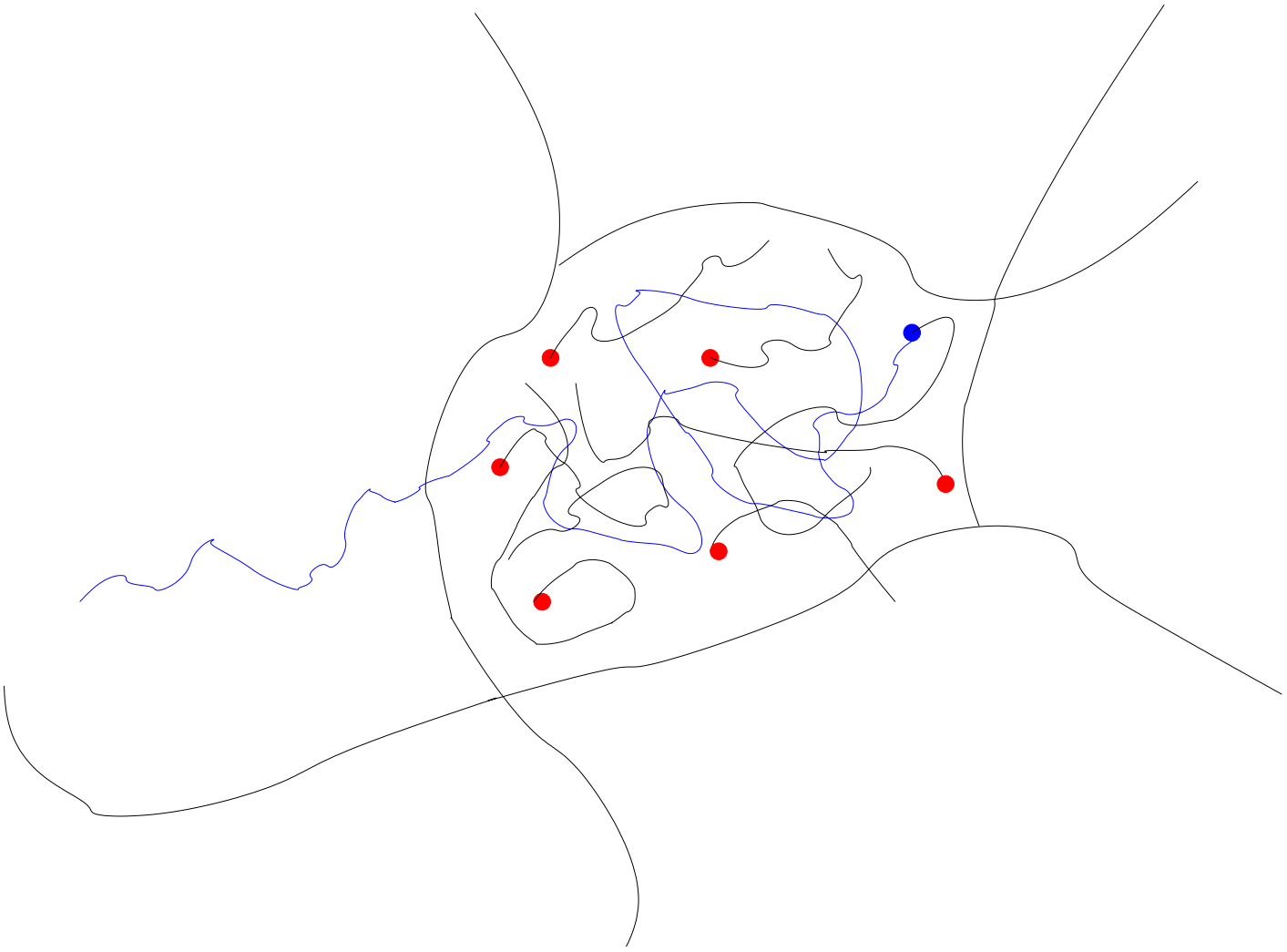
The Parallel Replica Algorithm

Parallel step.



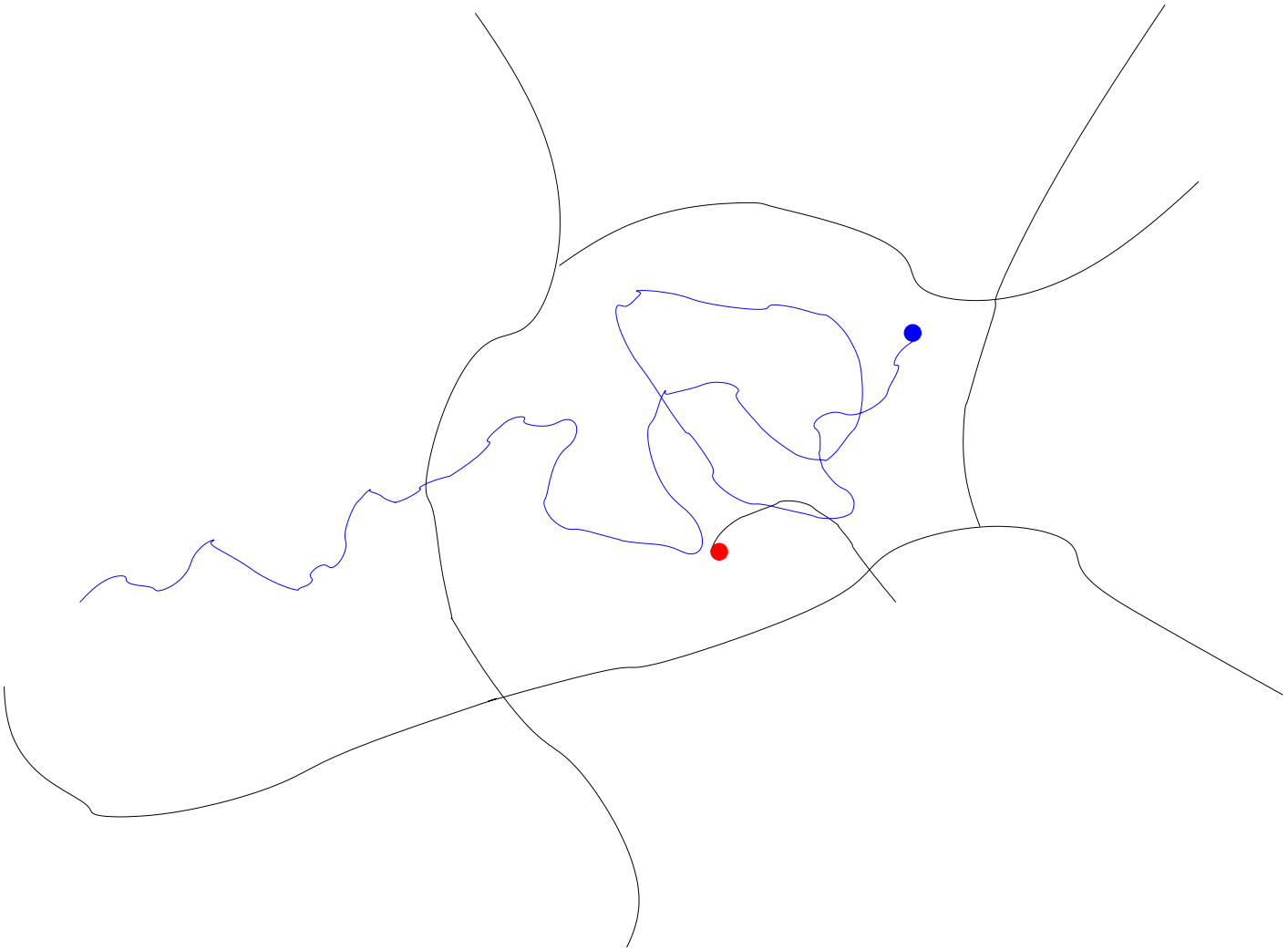
The Parallel Replica Algorithm

Parallel step: run independent trajectories in parallel...



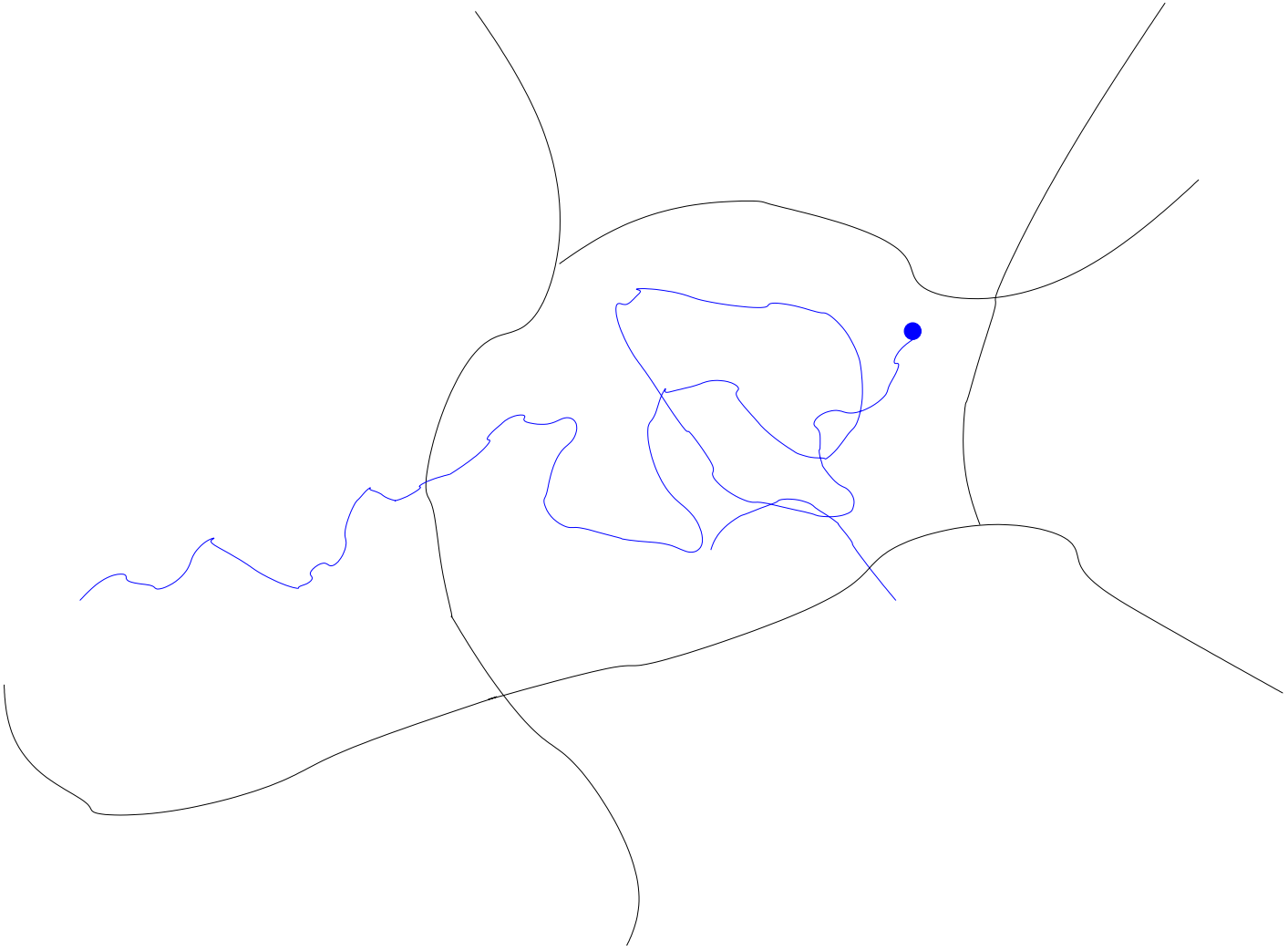
The Parallel Replica Algorithm

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



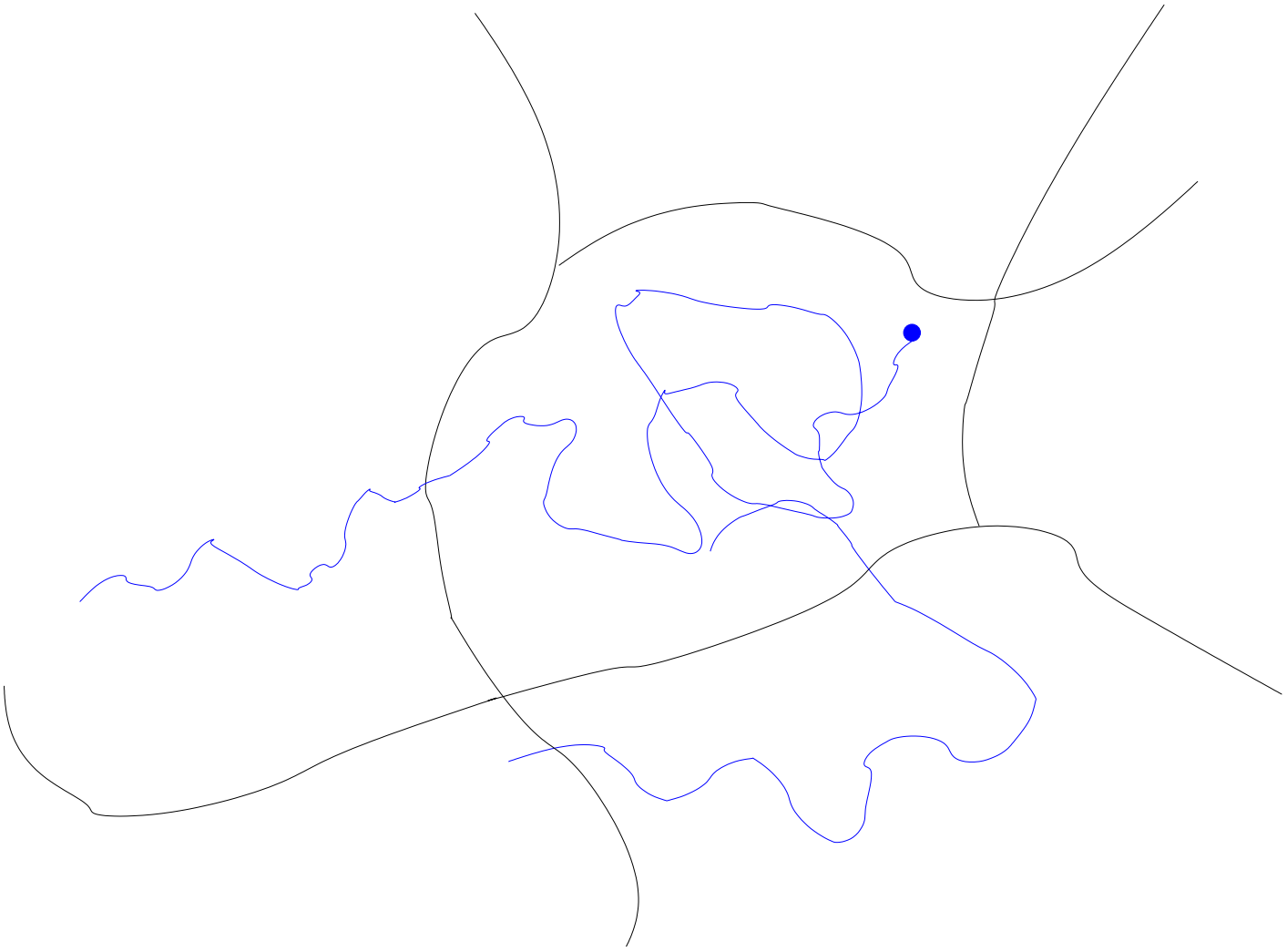
The Parallel Replica Algorithm

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



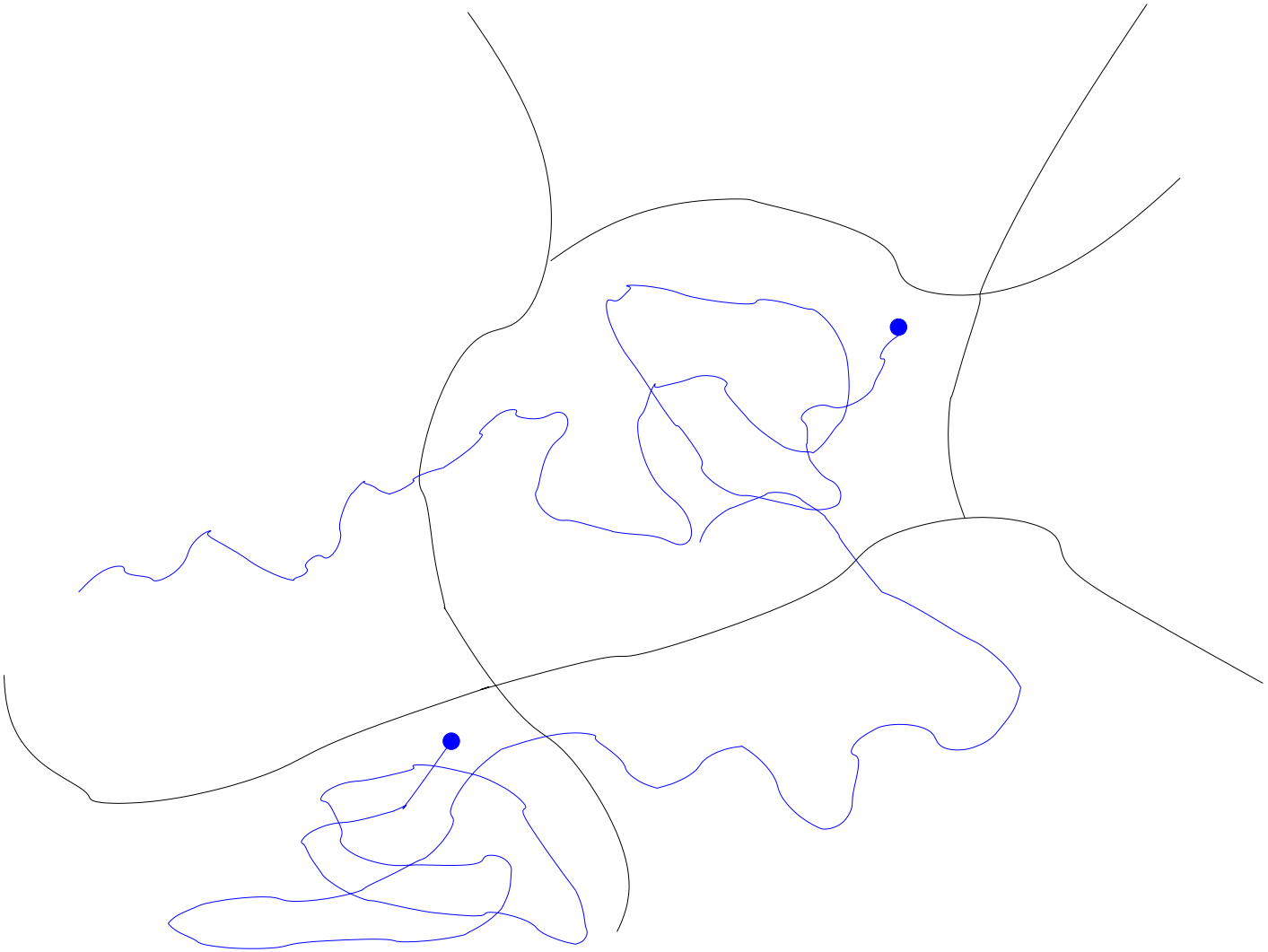
The Parallel Replica Algorithm

A new decorrelation step starts...



The Parallel Replica Algorithm

New decorrelation step



The Parallel Replica Algorithm

Analysis of the algorithm: the parallel step would introduce no error if

- the escape time T_W^1 was exponentially distributed
- and independent of the next visited state.

This essentially amounts to assuming that $\mathcal{S}(\boldsymbol{X}_t)$ is a Markov chain...

How to analyze the error introduced by the algorithm ?

This is related to the general question: how to relate a continuous state space Markov dynamics to a discrete state space Markov dynamics ? Pitfalls: (i) the temperature is not necessarily small (ii) the partition of the state space may be anything (iii) no thermodynamic limit in general (non-homogeneous systems).

The quasi-stationary distribution

The quasi-stationary distribution (QSD) ν for $\boldsymbol{X}_t$ and associated to the actual well W is a probability measure which is (i) supported by W and such that (ii): $\forall t > 0$, $\forall A \subset W$,

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

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

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

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

The quasi-stationary distribution and the dephasing step

Property of the QSD: If $\boldsymbol{X}_0 \sim \nu$ then, the first exit time T_W from W is exponentially distributed with parameter λ_1 and is a random variable independent of the first hitting point $\boldsymbol{X}_{T_W}$ on ∂W.

The dephasing step is very much related to the so-called Fleming-Viot process and may be seen as a way to get N i.i.d. random variables distributed according to the QSD.

Remark: In general, T_W exponentially distribtued is *not* sufficient for $\boldsymbol{X}_0$ to be distributed according to ν.

The parallel step

As announced above, starting from the QSD, the parallel step is exact. This is stated precisely here.

Let us start from N initial conditions $\boldsymbol{X}_0^k$ i.i.d. in the well W and let the processes evolve independently. Let us denote

$$T_W^k = \inf\{t > 0, \boldsymbol{X}_t^k \notin W\}$$

the escape time for the k-th replica, and

$$T = T_W^{K_0} \text{ where } K_0 = \arg \min_{k \in \{1,\dots,N\}} T_W^k$$

the *first* escape time over all processes.

- Assume that T_W^1 is exponentially distributed [OK starting from QSD.] Then NT has the same law as T_W^1.
- Assume that T_W^1 is independent of $\boldsymbol{X}_{T_W^1}^1$ [OK starting from QSD.] Then $\boldsymbol{X}_{T_W^{K_0}}^{K_0}$ has the same distribution as $\boldsymbol{X}_{T_W^1}^1$ and is independent of $T_W^{K_0}$.

The parallel step

Remark: If the CPUs are not synchronized, everything works likewise by changing

$$T_{simu} = T_{simu} + NT_W^{K_0}$$

to $T_{simu} = T_{simu}+$ the sum of the elapsed times on each CPU when the first escape event occurs.

The decorrelation step

We would like to quantify the error introduced by the dephasing and parallel steps, when the decorrelation step is successful.

As shown above, 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: conclusions

This can be generalized to other dynamics (coarse-graining of kMC).

Main results:

- The QSD is a good intermediate between continuous state dynamics and kMC-like approximations.
- The error analysis holds whatever the partition. But the method requires metastability between the states to be computationally efficient.
- The parameter τ_{corr} should be adjusted in terms of the two first eigenvalues of the Fokker-Planck operator with absorbing boundary conditions.

Conclusion

Many numerical methods in MD are based on some dimension reduction and coarse-graining techniques.

Main open problems in MD:

- How to generate efficiently metastable dynamics ?
- Out-of-equilibrium systems: models, analysis, sampling methods ?

Conclusion

A few references:

- T. Lelièvre, M. Rousset and G. Stoltz, *Long-time convergence of an Adaptive Biasing Force method*, Nonlinearity, 21, 1155-1181, 2008.
- T. Lelièvre, M. Rousset and G. Stoltz, *Free energy computations, a mathematical perspective*, Imperial College Press, 2010.
- C. Le Bris, T. Lelièvre, M. Luskin and D. Perez, *A mathematical formalization of the parallel replica dynamics*, `http://arxiv.org/abs/1105.4636`.