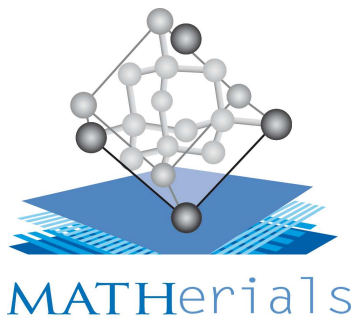
Molecular dynamics algorithms: numerical and mathematical analysis

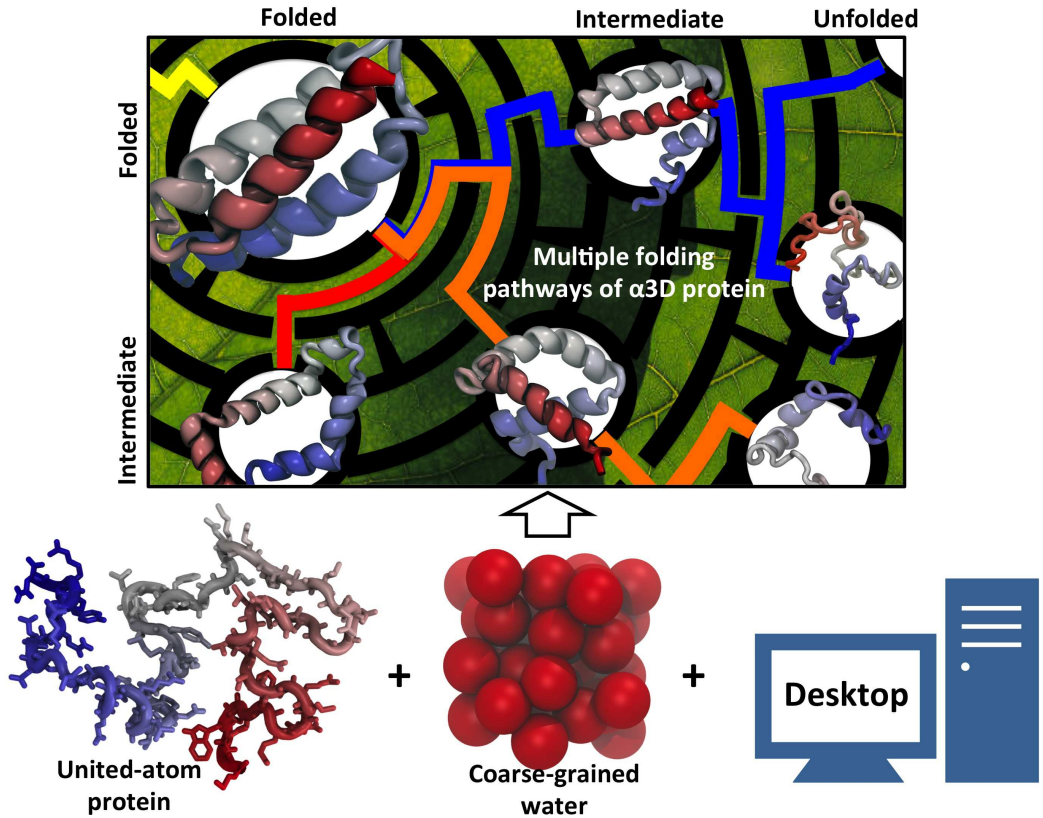
T. Lelièvre

CERMICS - Ecole des Ponts ParisTech & Equipe Matherials - INRIA



On Future Synergies for Stochastic and Learning Algorithms
CIRM, 26th Sept. 2021

Molecular dynamics



Simulation of protein folding (Courtesy of K. Schulten's group)

Molecular dynamics

Molecular dynamics consists in simulating on the computer the evolution of atomistic systems, as a numerical microscope:

- Understand the link bewteen macroscopic properties and microscopic ingredients
- Explore matter at the atomistic scale
- Simulate new materials, new molecules
- Interpret experimental results

Applications: biology, chemistry, materials science

Molecular dynamics comes of age:

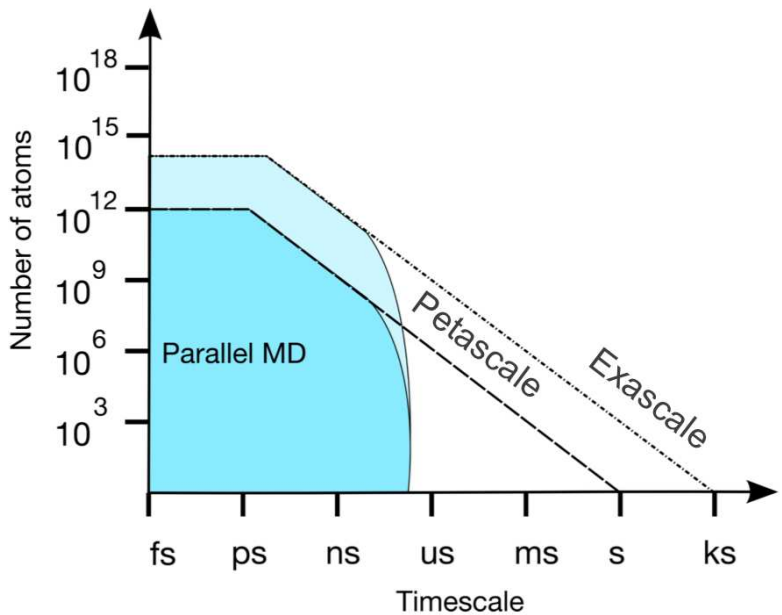
- 1/4 of CPU time worldwide is devoted to computations at the molecular scale
- 2013 Chemistry Nobel prize: Arieh Warshel, Martin Karplus and Michael Levitt. "Today the computer is just as important a tool for chemists as the test tube. Simulations are so realistic that they predict the outcome of traditional experiments."

Challenges

Main challenges:

- Improved models (force fields, coarse-grained force fields): polarisability, water, chemical reactions
- Improved sampling methods (access long time scales): thermodynamic quantities, and dynamical properties
- Incorporate data: Bayesian approaches, data sciences

Spatial parallelism is very effective, but temporal reach of heroic brute force MD is limited to $1\mu s$ or less.



Courtesy of Danny Perez (LANL)

Langevin and over-damped Langevin dynamics

The basic modeling ingredient in molecular dynamics: a potential function V which associates to a configuration $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_{N_{atom}})$ in $\mathbb{R}^d$ ($d = 3N_{atom}$) an energy $V(\boldsymbol{x}) \in \mathbb{R}$. Let us also introduce the inverse temperature: $\beta^{-1} = k_B T$.

The *Langevin dynamics* writes:

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

The *over-damped Langevin* dynamics writes:

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

These dynamics are both ergodic wrt the canonical measure: $\lim_{t\to\infty} \frac{1}{t}\int_0^t \varphi(\boldsymbol{X}_s)ds = \int_{\mathbb{R}^d} \varphi d\mu$ where

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

Thermodynamic and dynamical quantities

These dynamics are used to compute macroscopic quantities:

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

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

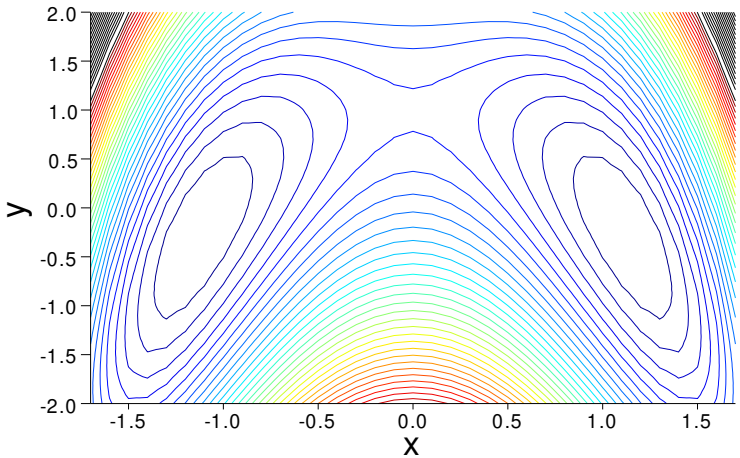
(ii) Dynamical quantities (averages over trajectories): diffusion coefficients, viscosity, transition rates,...

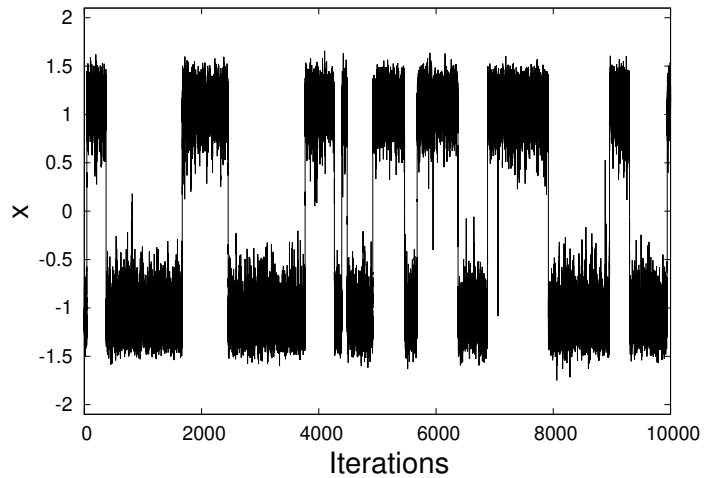
$$\mathbb{E}(\mathcal{F}((\boldsymbol{X}_t)_{t\geq 0})) \simeq \frac{1}{M}\sum_{m=1}^{M} \mathcal{F}((\boldsymbol{X}_t^m)_{t\geq 0}).$$

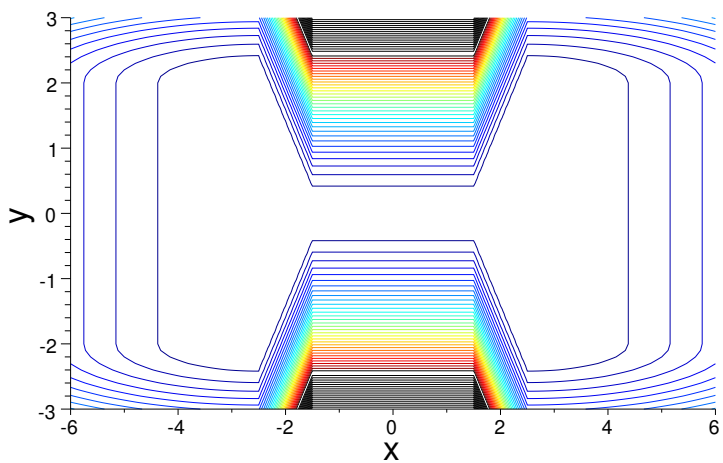
Difficulties: (i) high-dimensional problem ($N \gg 1$); (ii) $\boldsymbol{X}_t$ is a metastable process and μ is a multimodal measure.

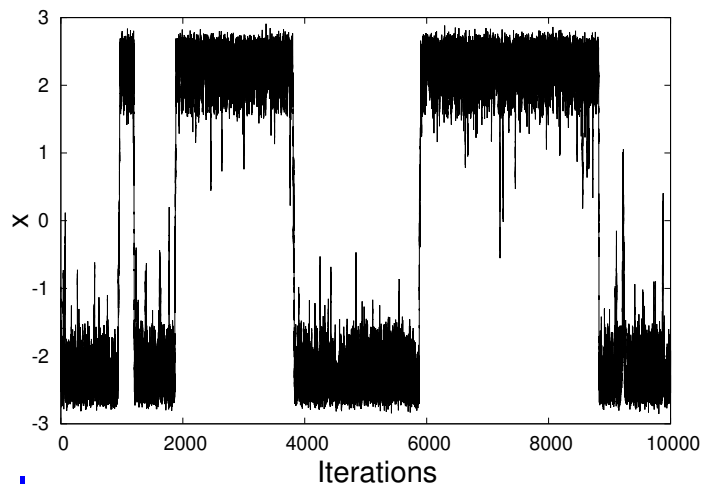
Metastability: energetic and entropic barriers

A two-dimensional schematic picture









→ • Slow convergence of trajectorial averages

• Transitions between metastable states are rare events

Challenges

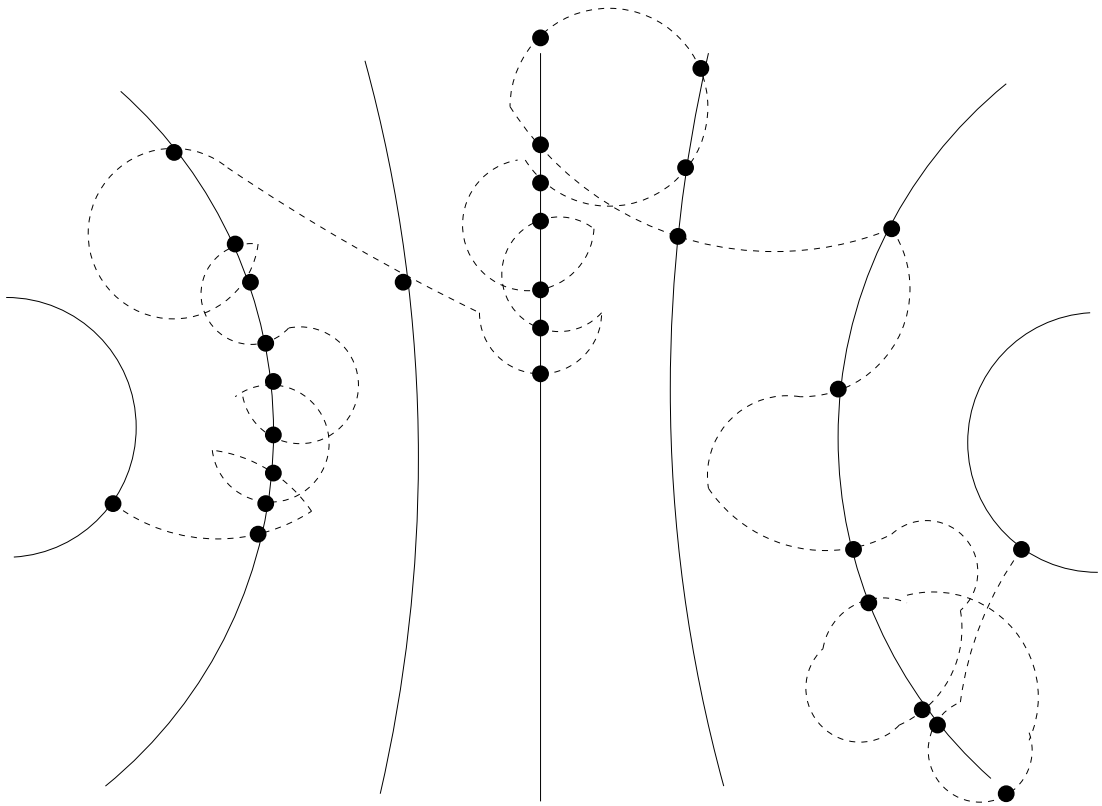
Examples of hot topics in mathematics for MD:

- Sampling of reactive trajectories, rare event sampling (A. Guyader, C. Hartmann, TL, E. Vanden Eijnden, J. Weare, ...)
- Sampling of probability measures on manifolds, constrained MD (P. Breiding, P. Diaconis, J. Goodman, TL, ...)
- Sampling of non equilibrium stationary state, non-reversible dynamics (J. Bierkens, G. Stoltz, ...)
- Towards better force fields (G. Csanyi, C. Ortner, A.V. Shapeev, ...)
- Effective dynamics, Mori-Zwanzig (T. Hudson, F. Legoll, TL, W. Zhang, ...)

Today:

- Free energy adaptive biasing algorithms
- Sampling of metastable dynamics

Free energy and adaptive biasing techniques

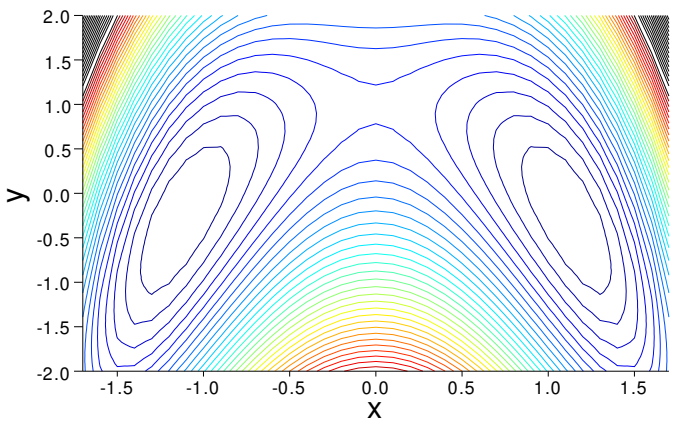


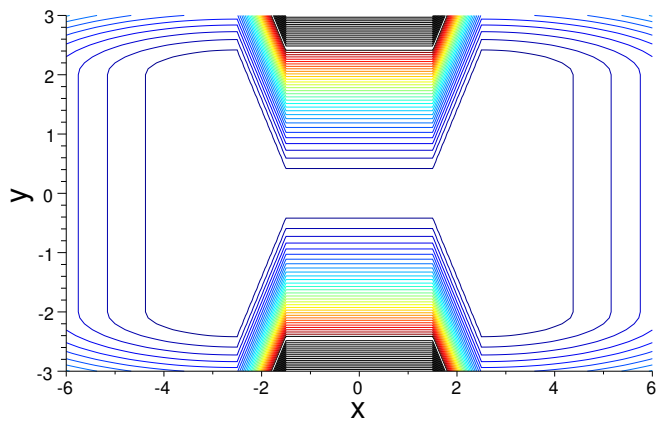
Adaptive biasing techniques

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

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

For example, in the 2D simple examples: $\xi(x, y) = x$.





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:

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

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

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

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

Adaptive biasing techniques

In the simple case $\xi(x, y) = x$, we have:

- The image of the measure μ by ξ:

$$\xi_* \mu\,(dx) = \exp(-\beta A(x))\, dx$$

where the free energy A is defined by:

$$A(x) = -\beta^{-1} \ln \left(\int_{\Sigma(x)} e^{-\beta V(x,y)} dy \right),$$

and $\Sigma(x) = \{(x, y),\ y \in \mathbb{R}\}$.

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

$$\mu_{\Sigma(x)}(dy) = \frac{\exp(-\beta V(x, y))\, dy}{\exp(-\beta A(x))}.$$

Adaptive biasing techniques

The bottom line of adaptive methods is the following: for "well chosen" ξ the potential $V - A \circ \xi$ is "less rugged" than V. Indeed, by construction $\xi_* \exp(-\beta(V - A \circ \xi)) = 1_{\mathbb{T}}$.

Problem: A is unknown ! Idea: use a time dependent potential of the form

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

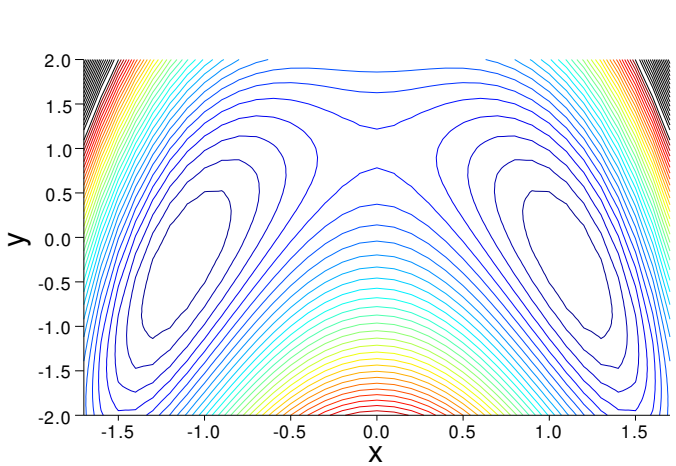
where A_t is an approximation at time t of A, given the configurations visited so far.

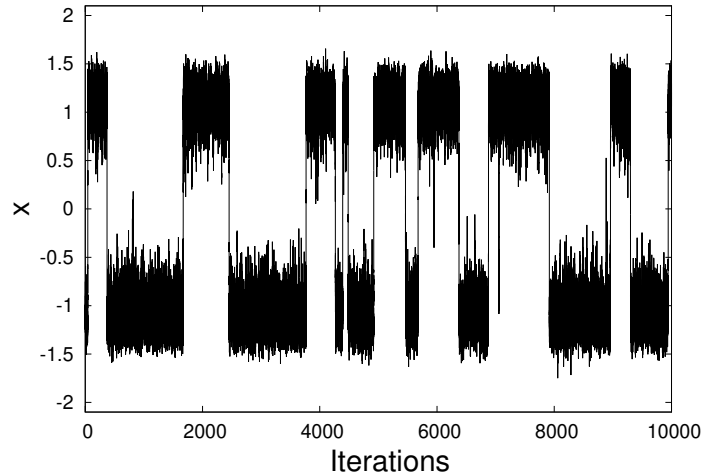
Hopes:

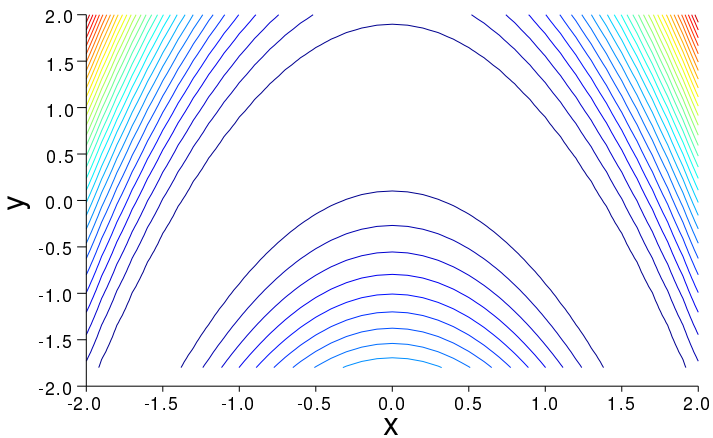
- build a dynamics which goes quickly to equilibrium,
- compute free energy profiles.

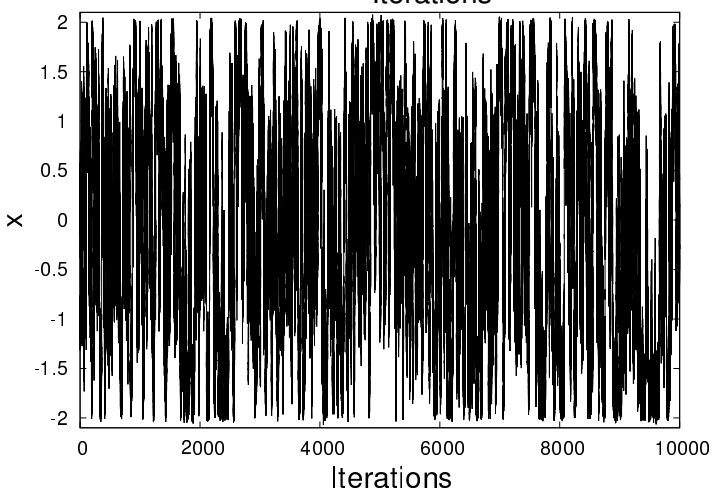
Wang-Landau, ABF, metadynamics: *Darve, Pohorille, Hénin, Chipot, Laio, Parrinello, Wang, Landau,...*

Free energy biased dynamics (1/2)



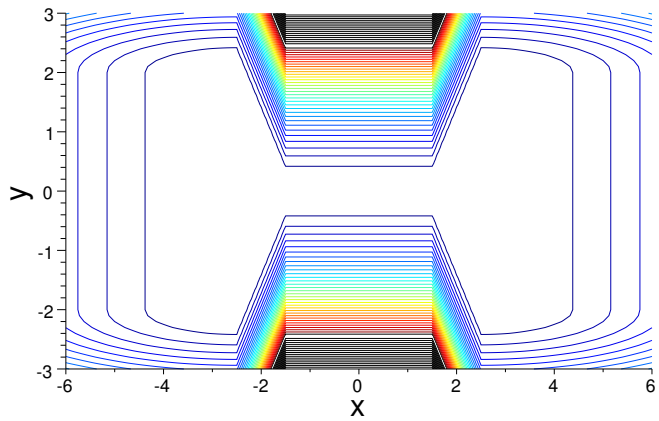


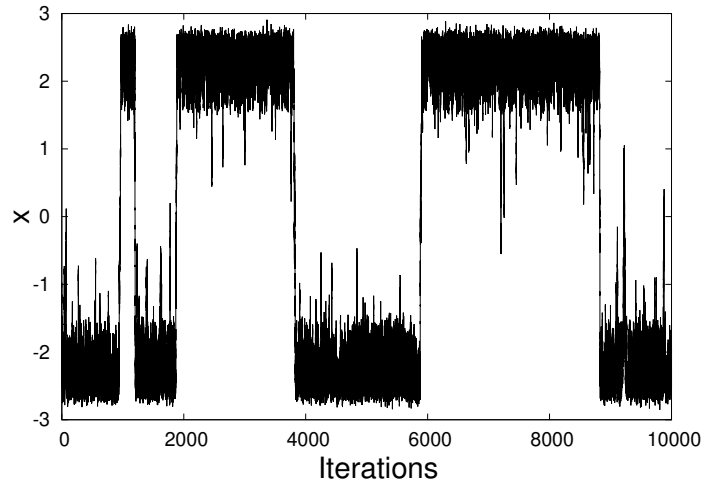


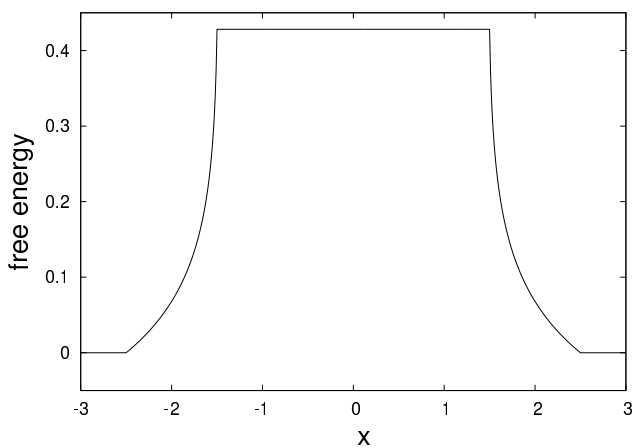


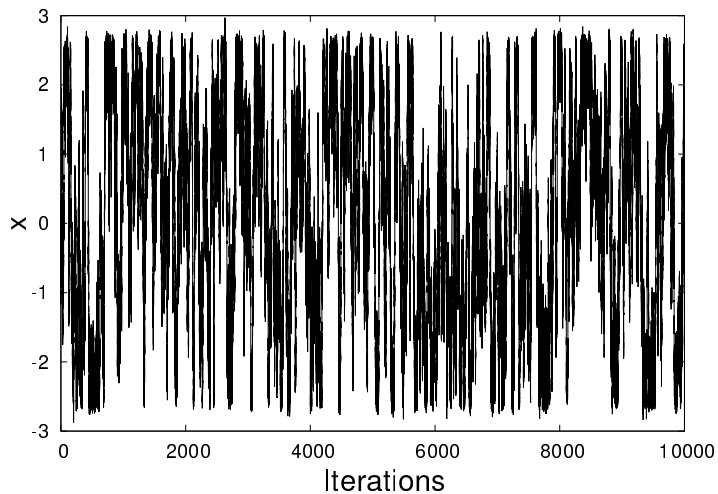
A 2D example of a free energy biased trajectory: energetic barrier.

Free energy biased dynamics (2/2)









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

Updating strategies

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 (ABF) method, the idea is to use the formula

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

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

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

The ABF method

In the simple case $\xi(x,y) = x$, remember that

$$A(x) = -\beta^{-1} \ln \left(\int_{\Sigma(x)} e^{-\beta V(x,y)} dy \right),$$

so that

$$A'(x) = \frac{\displaystyle\int_{\Sigma(x)} \partial_x V \, e^{-\beta V(x,y)} \, dy}{\displaystyle\int_{\Sigma(x)} e^{-\beta V(x,y)} \, dy} = \int_{\Sigma(x)} \partial_x V \, d\mu_{\Sigma(x)}$$

where $\mu_{\Sigma(x)}$ is the prob meas μ conditioned to $\xi(x,y) = x$.
Notice that actually, whatever A_t is,

$$A'(z) = \frac{\displaystyle\int_{\Sigma(z)} f e^{-\beta(V - A_t \circ \xi)} \, \delta_{\xi(x)-z}(dx)}{\displaystyle\int_{\Sigma(z)} e^{-\beta(V - A_t \circ \xi)} \, \delta_{\xi(x)-z}(dx)}.$$

The ABF method

Thus, we would like to simulate:

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

but A is unknown...

The ABF method

The 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 ABF method

The 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:

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

where $\boldsymbol{X}_t \sim \psi(t, \boldsymbol{x})\, d\boldsymbol{x}$.

The ABF method

The ABF dynamics is then:

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

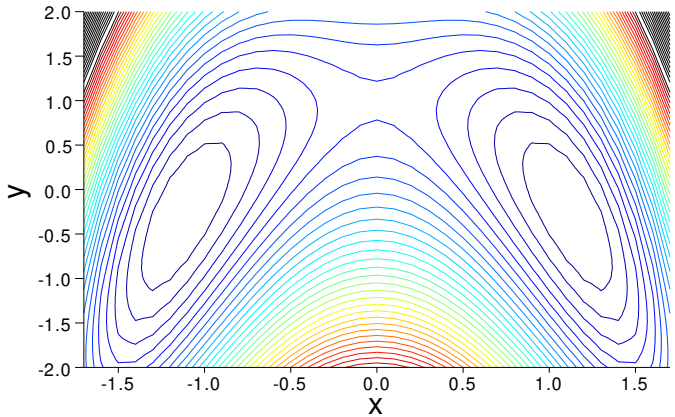
The associated (nonlinear) Fokker-Planck equation writes:

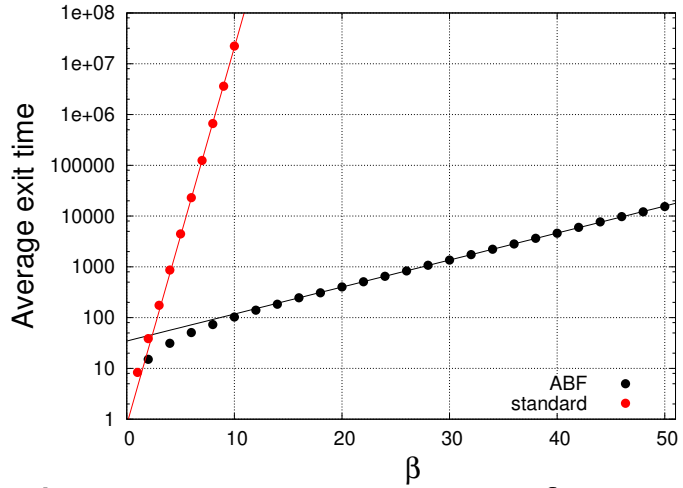
$$\begin{cases} \partial_t \psi = \operatorname{div}\left(\nabla(V - A_t \circ \xi)\psi + \beta^{-1}\nabla\psi\right), \\ A_t'(z) = \dfrac{\displaystyle\int_{\Sigma(z)} f\, \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})}{\displaystyle\int_{\Sigma(z)} \psi\, \delta_{\xi(\boldsymbol{x})-z}(d\boldsymbol{x})},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

where $\boldsymbol{X}_t \sim \psi(t, \boldsymbol{x})\, d\boldsymbol{x}$.

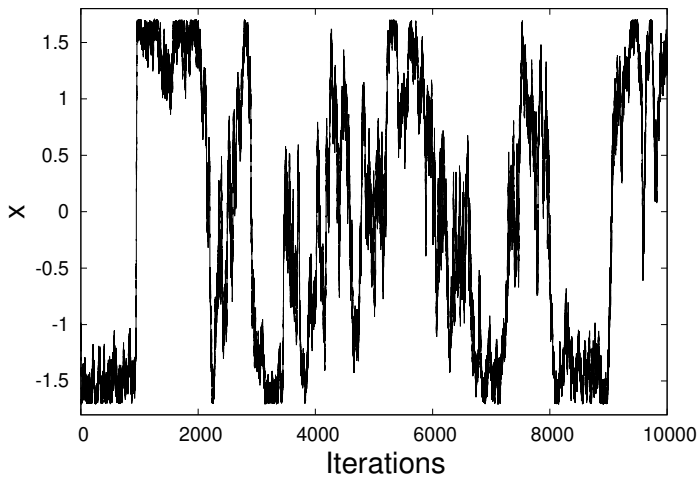
Questions: Does A_t' converge to A' ? What did we gain compared to the original gradient dynamics ?

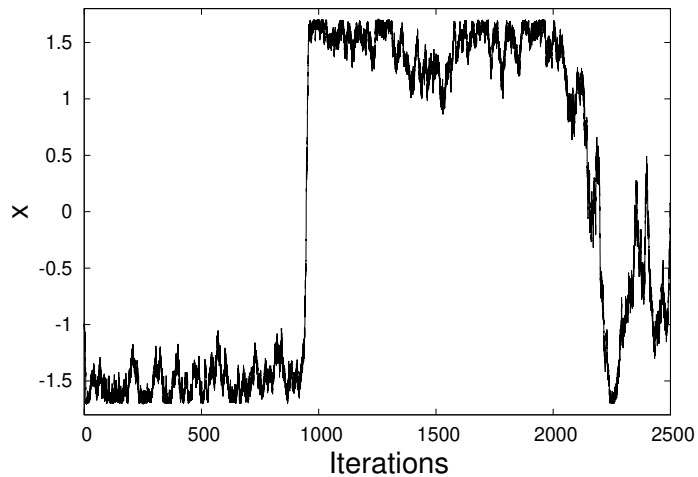
Back to the 2D example





Left: the 2D potential – energetic barrier; Right: average exit time from the left well

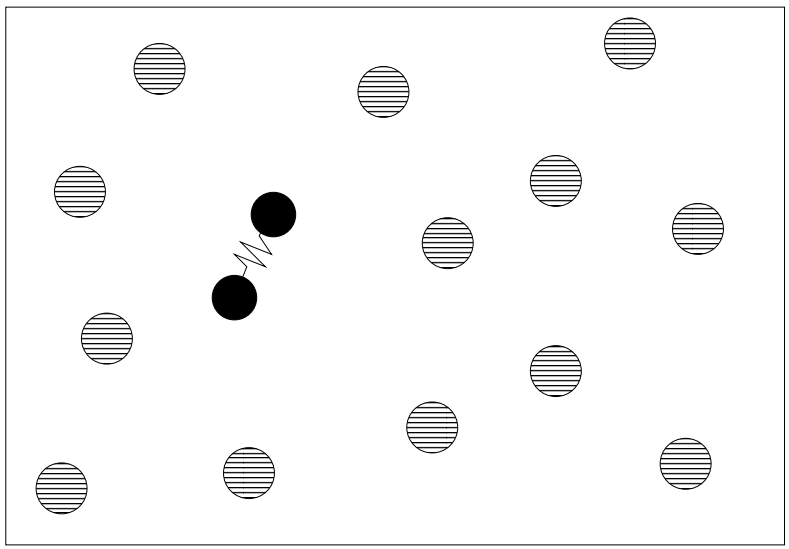




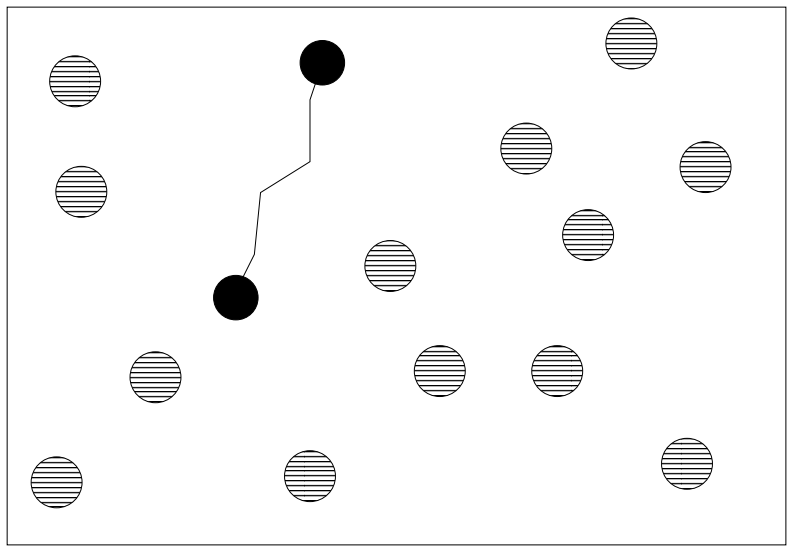
The ABF trajectory (right: zoom on the first 2500 iterations)

A toy example

Influence of the solvation on a dimer conformation [Dellago, Geissler]



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. $\longrightarrow$ simulation

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 approach for partial differential equations (PDEs): 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,

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

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

Metastability $\iff$ 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, $\rho \gg R$: the free energy can be computed very efficiently.

Two ingredients of the proof:

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

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

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

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

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

Practical implementation of ABF

The estimate of the conditional average $\mathbb{E}(f(\boldsymbol{X}_t)|\xi(\boldsymbol{X}_t)=z)$ can be done using two (complementary) approaches:

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

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

  This approach is easy to parallelize, flexible (selection mechanisms) and efficient in cases with multiple reactive paths.

- Use trajectorial averages along a single path:

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

  The longtime behavior is much more difficult to analyze.

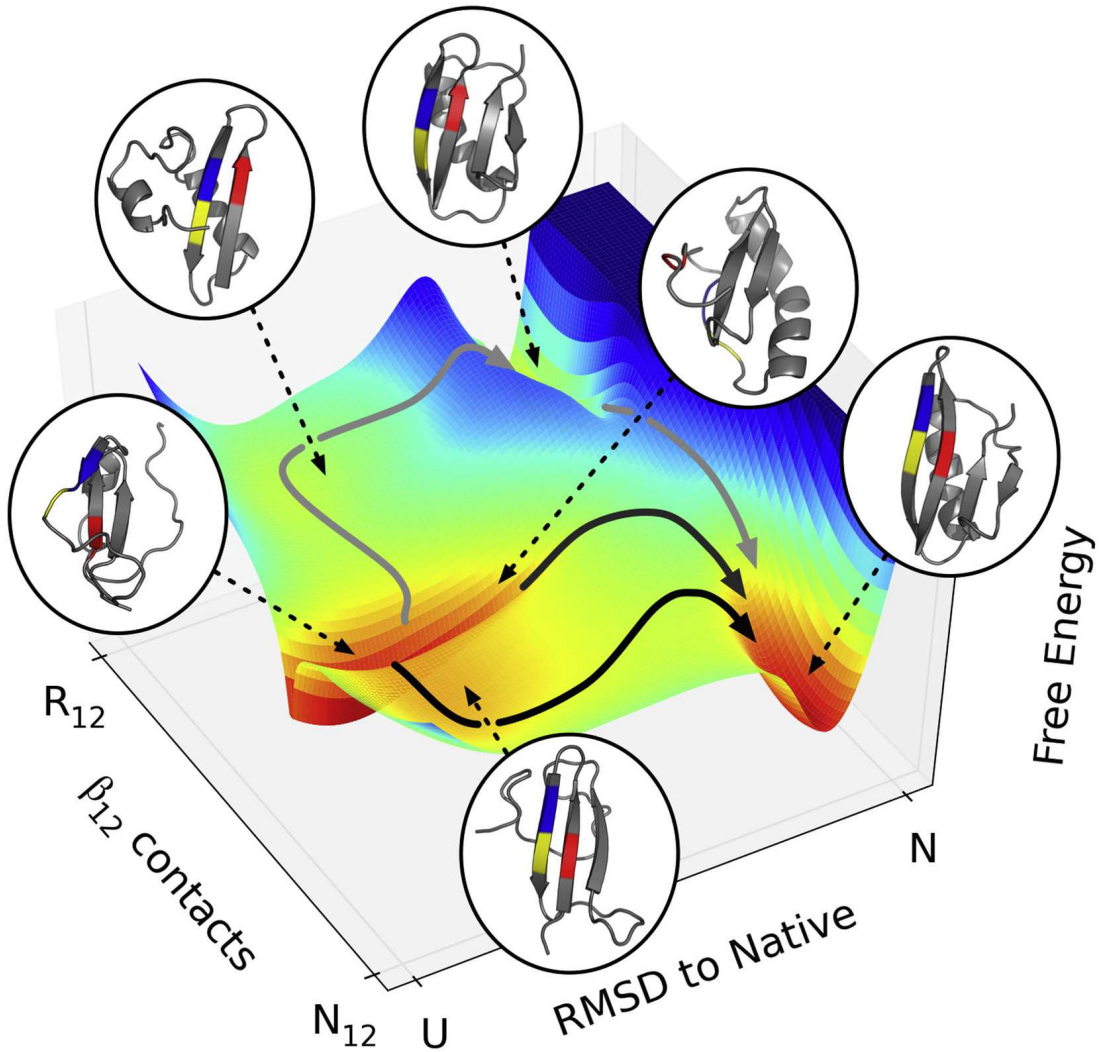
Other results and recent developments

Adaptive biasing 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].

Recent works:

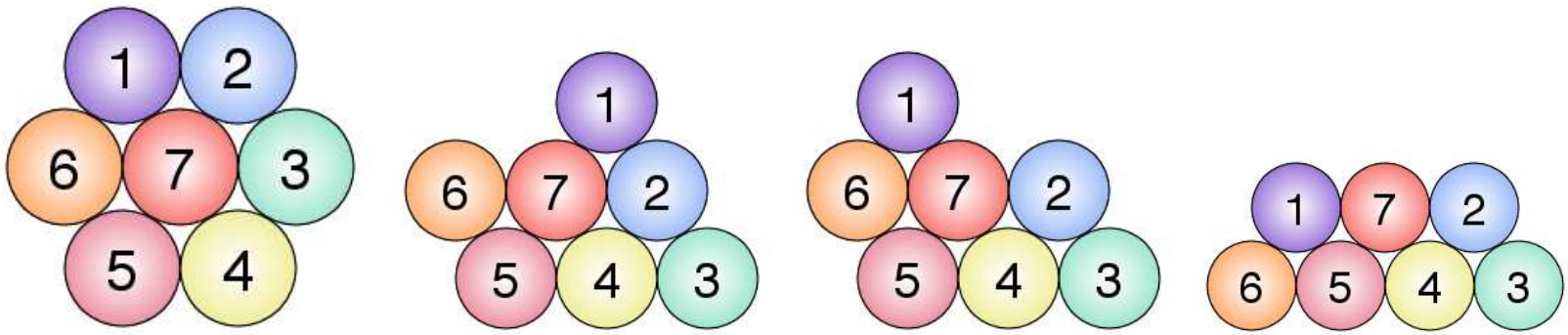
- Non-gradient force fields [L. Maurin, P. Monmarché, TL, 2021]
- Efficiency of techniques which use trajectorial averages to learn the bias (Wang Landau, metadynamics, SHUS) [M. Benaïm, G. Fort, B. Jourdain, TL, P. Monmarché, G. Stoltz, P.A. Zitt, 2014-2021]
- Extension to Langevin dynamics [M. Benaïm, P. Monmarché, 2018]
- Computation of the collective variable using auto-encoders [Z. Belkacemi, E. Gkeka, TL, G. Stoltz]

Metastability: the quasi-stationary distribution approach



C.R. Schwantes, D. Shukla, V.S.Pande, Biophysical Journal, vol. 110, 2016

Metastability: a toy example



(a) $V = -12.53$ (b) $V = -11.50$ (c) $V = -11.48$ (d) $V = -11.40$

Figure: Low energy conformations of the 7 atoms Lennard-Jones cluster.

$\longrightarrow$ simulation

Two models for dynamics

The basic modeling ingredient in molecular dynamics: a potential function V which associates to a configuration $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_{N_{atom}})$ in $\mathbb{R}^d$ $(d = 3N_{atom})$ an energy $V(\boldsymbol{x}) \in \mathbb{R}$.

From V, two kinds of dynamics are considered:

- Langevin and over-damped Langevin dynamics: Markov processes with values in continuous state space
- kinetic Monte Carlo model or Markov state model (first order kinetics): Markov processes with values in discrete state space (jump Markov process)

Question: Can a mathematically rigorous link be made between these two kinds of models ?

The exit event

Let us focus on 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 an ensemble of subsets of $\mathbb{R}^d$ (states). Let us consider one of them: $\mathcal{S} \subset \mathbb{R}^d$. The exit event from $\mathcal{S}$ is given by

$$(\tau_{\mathcal{S}}, \boldsymbol{X}_{\tau_{\mathcal{S}}})$$

where $\tau_{\mathcal{S}} = \inf\{t > 0,\, \boldsymbol{X}_t \notin \mathcal{S}\}$.

Objective: build a jump Markov model to simulate the exit event $(\tau_{\mathcal{S}}, \boldsymbol{X}_{\tau_{\mathcal{S}}})$.

This is useful theoretically (justification of Markov state models and Eyring-Kramers laws) and numerically (accelerated dynamics *à la* Voter).

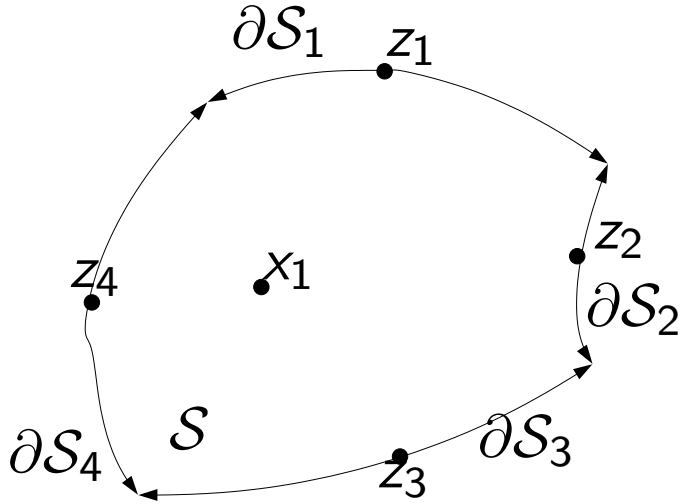
Kinetic Monte Carlo

Kinetic Monte Carlo (or Markov state) models are built as follows:

- define exit regions from $\mathcal{S}$: $\partial\mathcal{S} = \cup_{j=1}^{J} \partial\mathcal{S}_j$
- associate a rate k_j with an exit through $\partial\mathcal{S}_j$

and then (jump Markov model)

- the exit time $\tau_{\mathcal{S}}^{kMC}$ is exponentially distributed with parameter $\sum_{j=1}^{J} k_j$
- the exit region is $I_{\mathcal{S}}^{kMC}$ with law $\mathbb{P}(I_{\mathcal{S}}^{kMC} = i) = \frac{k_i}{\sum_{j=1}^{J} k_j}$
- $I_{\mathcal{S}}^{kMC}$ and $\tau_{\mathcal{S}}^{kMC}$ are independent random variables

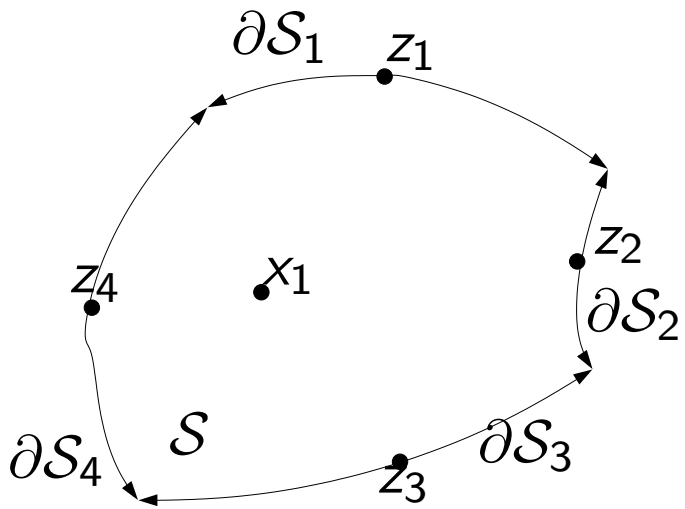


Eyring-Kramers laws

Formulas for transition rates. Let us introduce the local minima $(z_j)_{j=1,\dots,J}$ of V on ∂S, and associated exit regions ∂S_i. The parameters k_j are computed using the Eyring-Kramers formula (Harmonic Transition State Theory):

$$k_j^{HTST} = \nu_j \, \mathrm{e}^{-\beta[V(z_j)-V(x_1)]}$$

where ν_j is an explicit prefactor and $x_1 = \arg\min_{\mathcal{S}} V$.



A theoretical question

Question: can we relate the exit event $(\tau_{\mathcal{S}}, \boldsymbol{X}_{\tau_{\mathcal{S}}})$ for the original dynamics with the exit event $(\tau_{\mathcal{S}}^{kMC}, I_{\mathcal{S}}^{kMC})$ for the jump Markov process?

Two steps:

- Introduce the Quasi-Stationary Distribution
- Consider the small temperature regime $\beta \to \infty$ (semi-classical limit)

Step 1: The Quasi-Stationary Distribution

Definition of the QSD: A probability measure ν with support $\mathcal{S}$ is a QSD for the Markov process $(\boldsymbol{X}_t)_{t\geq 0}$ iff for all $t > 0$,

$$\boldsymbol{X}_0 \sim \nu \Longrightarrow \mathcal{L}(\boldsymbol{X}_t | \tau_{\mathcal{S}} > t) = \nu$$

Existence, uniqueness, convergence: Assume the state is bounded in positions. For the Langevin and the overdamped Langevin dynamics, there exists a unique QSD ν in $\mathcal{S}$. Moreover, for any $\boldsymbol{X}_0$ in $\mathcal{S}$,

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

Remark: Quantitative definition of a metastable exit:

exit time$\gg$ local equilibration time

Fundamental properties of the QSD

Assume $\boldsymbol{X}_0 \sim \nu$, then:

- the first exit time $\tau_{\mathcal{S}}$ is exponentially distributed since:

$$\begin{aligned}\mathbb{P}^{\nu}(\tau_{\mathcal{S}} > s+t) &= \mathbb{P}^{\nu}(\tau_{\mathcal{S}} > s+t | \tau_{\mathcal{S}} > s)\, \mathbb{P}^{\nu}(\tau_{\mathcal{S}} > s) \\ &= \mathbb{P}^{\nu}(\tau_{\mathcal{S}} > t)\, \mathbb{P}^{\nu}(\tau_{\mathcal{S}} > s)\end{aligned}$$

- and $\tau_{\mathcal{S}}$ is independent of the first hitting point $\boldsymbol{X}_{\tau_{\mathcal{S}}}$ since:

$$\begin{aligned}\mathbb{P}^{\nu}(\boldsymbol{X}_{\tau_{\mathcal{S}}} \in A, \tau_{\mathcal{S}} > s) &= \mathbb{P}^{\nu}(\boldsymbol{X}_{\tau_{\mathcal{S}}} \in A | \tau_{\mathcal{S}} > s)\, \mathbb{P}^{\nu}(\tau_{\mathcal{S}} > s) \\ &= \mathbb{P}^{\nu}(\boldsymbol{X}_{\tau_{\mathcal{S}}} \in A)\, \mathbb{P}^{\nu}(\tau_{\mathcal{S}} > s)\end{aligned}$$

Consequence: Starting from ν, the exit event from $\mathcal{S}$ can be exactly written as one jump of a kinetic Monte Carlo model with rates

$$k_j = \frac{\mathbb{P}^{\nu}(\boldsymbol{X}_{\tau_{\mathcal{S}}} \in \partial\mathcal{S}_j)}{\mathbb{E}^{\nu}(\tau_{\mathcal{S}})}.$$

Step 2: The small temperature regime

One has explicit formulas for $\mathbb{E}(\tau_{\mathcal{S}})$ and the distribution of $\boldsymbol{X}_{\tau_{\mathcal{S}}}$.
The first eigenstate (λ_1, u_1) of the Fokker-Planck operator with Dirichlet boundary conditions on $\partial\mathcal{S}$ satisfies (for the overdamped Langevin dynamics):

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

Then, $\nu = \frac{u_1(\boldsymbol{x})d\boldsymbol{x}}{\int_{\mathcal{S}} u_1}$,

$$\mathbb{E}^{\nu}(\tau_{\mathcal{S}}) = \frac{1}{\lambda_1}$$

and

$$\mathbb{P}^{\nu}(\boldsymbol{X}_{\tau_{\mathcal{S}}} \in \partial\mathcal{S}_i) = -\frac{\int_{\partial\mathcal{S}_i} \partial_n u_1 \, d\sigma}{\beta\lambda_1 \int_{\mathcal{S}} u_1(x)\, dx}.$$

Thus, $k_i = -\frac{\int_{\partial\mathcal{S}_i} \partial_n u_1 \, d\sigma}{\beta \int_{\mathcal{S}} u_1(x)\, dx}$. Can we then show that $k_i \simeq k_i^{HTST}$?

Justifying Eyring-Kramers laws

Theorem [Di Gesu, TL, Le Peutrec, Nectoux, 2019]
Under some geometric assumptions, starting from the QSD, in the limit $\beta \to \infty$, the exit rates for the overdamped Lang dyn are

$$k_i = \widetilde{\nu}_i^{OL} \, \mathrm{e}^{-\beta[V(z_i)-V(x_1)]} \, (1 + O(\beta^{-1}))$$

where

$$\widetilde{\nu}_i^{OL} = \sqrt{\frac{\beta}{2\pi}} \partial_n V(z_i) \frac{\sqrt{\det(\nabla^2 V)(x_1)}}{\sqrt{\det(\nabla^2 V_{|\partial S})(z_i)}}.$$

The proof is based on tools from semi-classical analysis and properties of the low-lying spectrum of the Witten Laplacians on 0-forms and 1-forms.

Among the geometric assumptions, one imposes some kind of "separation" between the saddle points in terms of Agmon distances, which appears to be necessary numerically.

Generalizations and perspectives

If the state is metastable, the QSD is a good intermediate between continuous-state space dynamics and jump Markov models.

The mathematical analysis gives the proper geometric setting under which the kinetic Monte Carlo model can be built and the Eyring-Kramers formulas can be used to parameterize it.

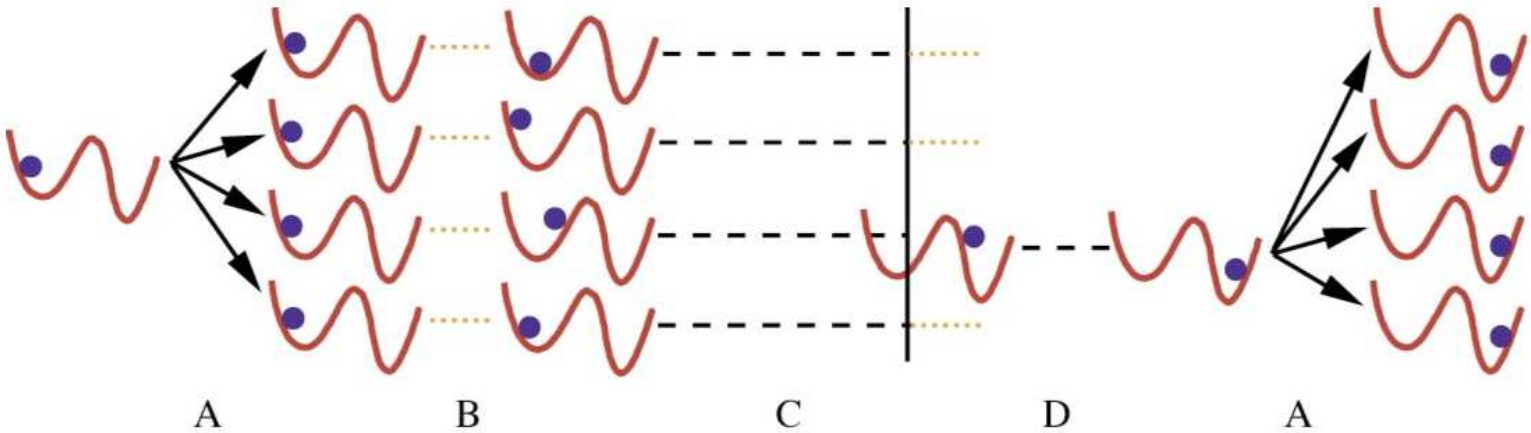
Recent works:

- QSD for Langevin dynamics [Guillin, TL, Ramil, Reygner, Wu]
- Saddle points on $\partial\mathcal{S}$ [TL, Le Peutrec, Nectoux]

Open questions:

- Small temperature regime on the Langevin dynamics
- Non-reversible dynamics

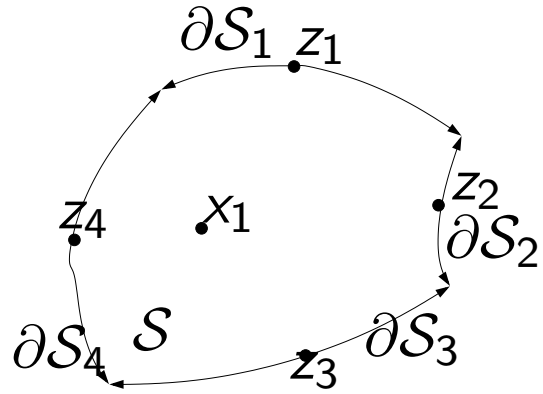
From theory to algorithms



A.F. Voter, Annu. Rev. Mater. Res., vol. 32, 2002.

How to sample efficiently the exit event?

If the process remains sufficiently long in a state, the exit event can be modeled by one jump of a Markov state model. This can be used to simulate efficiently the exit event: accelerated dynamics *à la* A.F. Voter.



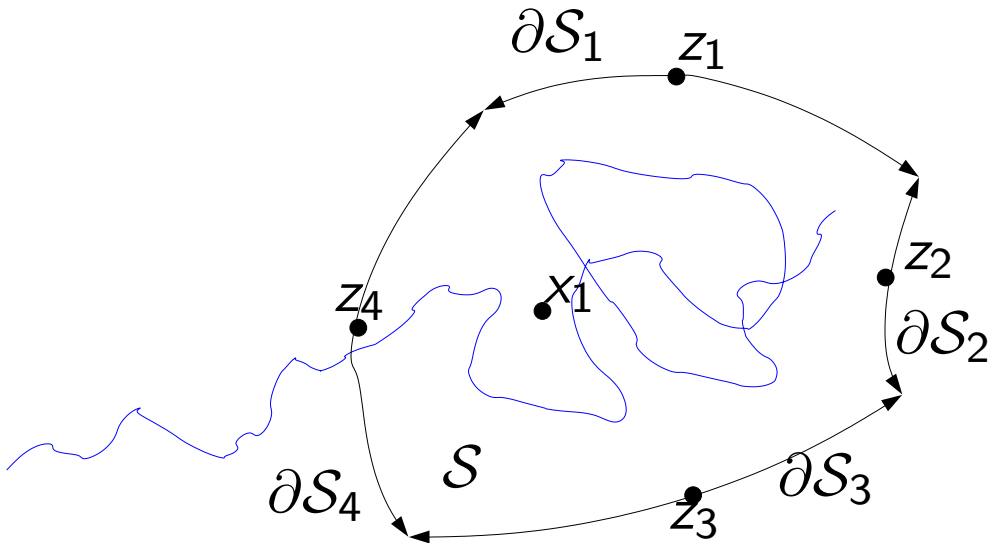
Two steps:

- Estimate the decorrelation time, namely the time to reach the QSD
- Use the underlying jump Markov process to efficiently sample the exit event

Decorrelation time

How long should we wait in practice so that $\mathcal{L}(\boldsymbol{X}_t|\tau_{\mathcal{S}} > t)$ is close to the QSD ν?

- Theoretically: exponential decay
  $\|\mathcal{L}(\boldsymbol{X}_t|\tau_{\mathcal{S}} > t) - \nu\|_{TV} \leq C(\mathcal{L}(\boldsymbol{X}_0)) \exp\left(-(\lambda_2 - \lambda_1)t\right)$;
- Numerically: simulate $\mathcal{L}(\boldsymbol{X}_t|\tau_{\mathcal{S}} > t)$ via interacting particle system (Fleming-Viot particle system), and test stationarity to estimate the convergence time to the QSD (Gelman-Rubin convergence diagnostic).



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 [Villemonais]

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

satisfies:

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

Accelerated dyamics

Once the QSD has been reached, there are three ideas to efficiently sample $(\tau_{\mathcal{S}}, \boldsymbol{X}_{\tau_{\mathcal{S}}})$:

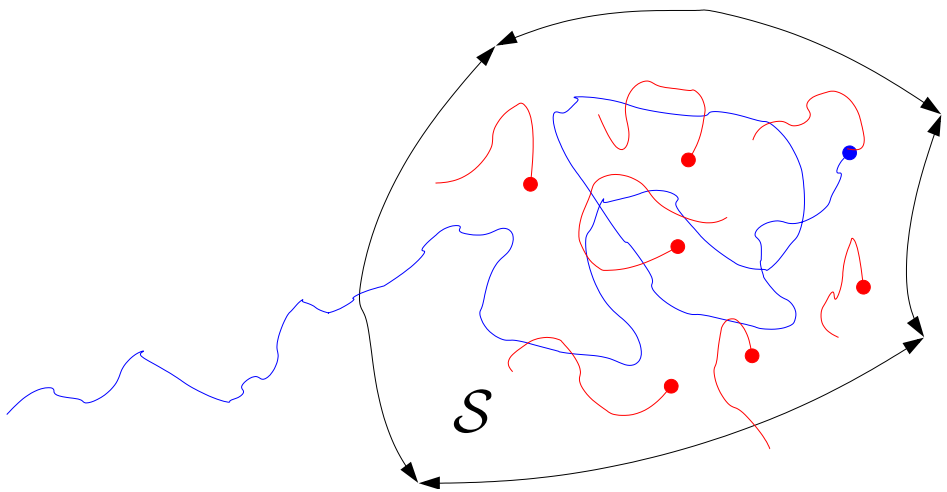
- use parallel architectures to accelerate the sampling: parallel replica, parsplicing
- raise the minimum of the potential inside the state $\mathcal{S}$ (but not on $\partial\mathcal{S}$): hyperdynamics
- raise the temperature: temperature accelerated dynamics

The Parallel Replica Algorithm

Perform many independent exit events in parallel [Voter, 1998]

Two steps:

- Distribute N independent initial conditions in $\mathcal{S}$ according to the QSD ν ;
- Evolve N replicas from these initial conditions, consider the first exiting replica, and multiply the first exit time by the number of replicas.



The Parallel Replica Algorithm

Why is it consistent?

- Exit time is independent of exit point so that

$$\boldsymbol{X}^{I_0}_{\tau^{I_0}_{\mathcal{S}}} \stackrel{\mathcal{L}}{=} \boldsymbol{X}^{1}_{\tau^{1}_{\mathcal{S}}}$$

  where $I_0 = \arg\min_i(\tau^i_{\mathcal{S}})$.

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

$$N \min(\tau^1_{\mathcal{S}}, \ldots, \tau^N_{\mathcal{S}}) \stackrel{\mathcal{L}}{=} \tau^1_{\mathcal{S}}.$$

Remark: For this algorithm, one just needs two properties: $\tau_{\mathcal{S}}$ is exponentially distributed, and independent of the exit point $\boldsymbol{X}_{\tau_{\mathcal{S}}}$. The Eyring-Kramers formulas are not used.

The generalized Parallel Replica algorithm

[Binder, Hédin, TL, Simpson]

1. Run a reference walker, using standard MD.
2. Each time the reference walker enters a state, start a Fleming-Viot particle process (with N replicas simulated in parallel) with initial condition the entering point.
3. If the reference walker exits before the Fleming Viot particle process reaches stationarity go back to 1. Else go to the parallel step.
4. Parallel step: Starting from the end points of the Fleming-Viot particle process (approximately i.i.d. with law the QSD), run independent MD and consider the first exit event. Multiply the first exit time by N and go back to 1, using the first exit point as initial condition.

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

The generalized Parallel Replica algorithm

- The algorithm does not require a partition of the state space but only an ensemble of states.
- The time to reach the QSD is estimated each time the process enters a new state (it depends on the state and on the initial condition within the state).

Numerical results

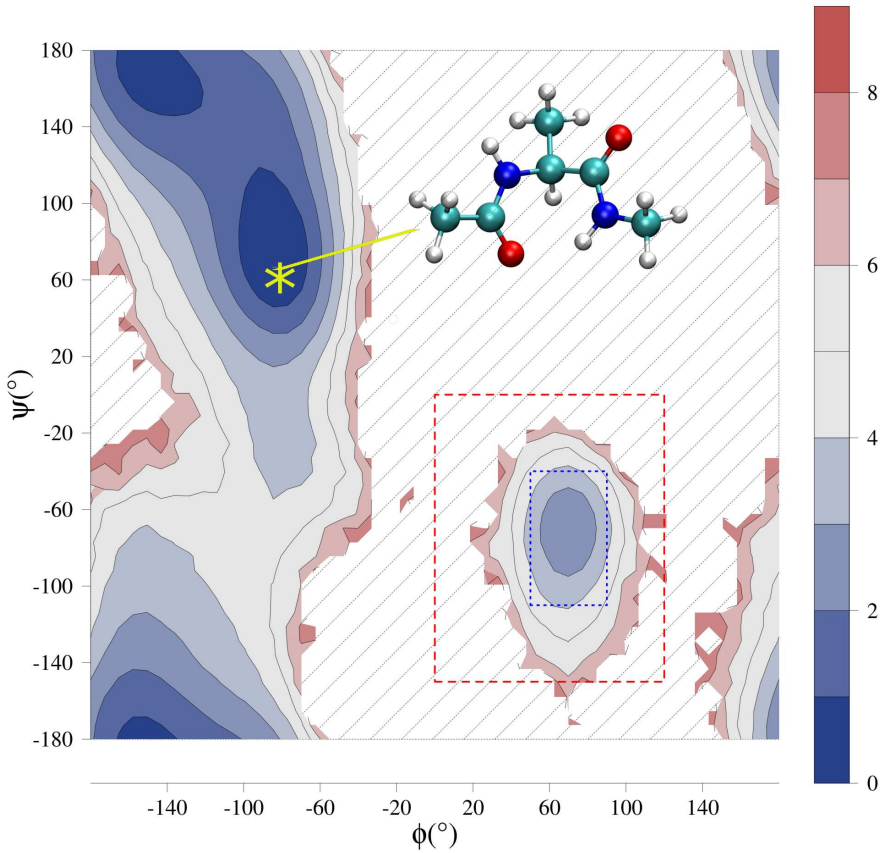
We tested the generalized Parallel Replica algorithm on biological systems [Hédin, TL]:

- Conformational equilibrium of the alanine dipeptide
- Dissociation of the FKBP-DMSO protein-ligand system

Main differences with materials science: definition of the states using collective variables, the states do not define a partition, much more rugged landscapes.

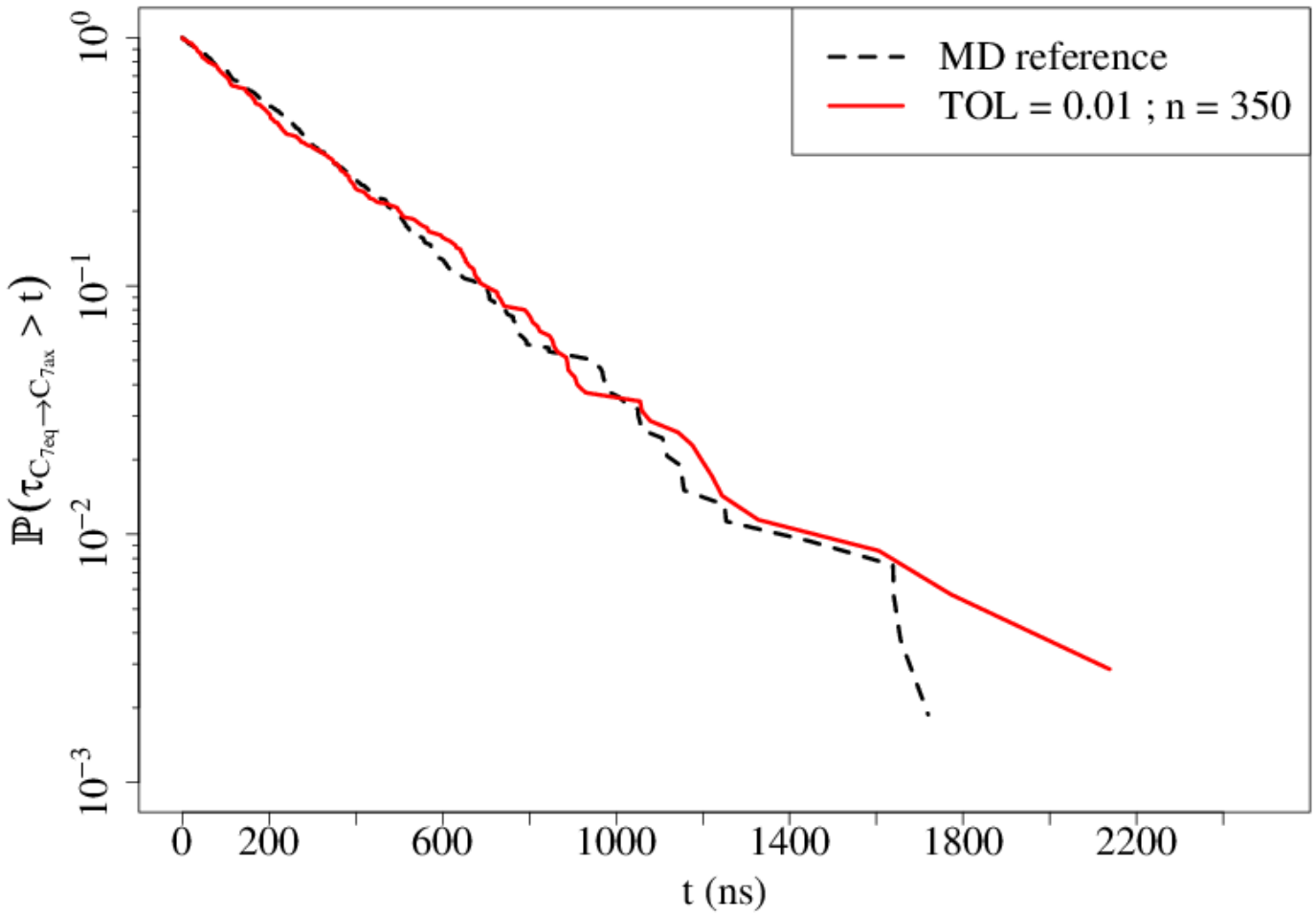
Current implementation within OpenMM, see
`https://gitlab.inria.fr/parallel-replica`

Alanine dipeptide (1/5)



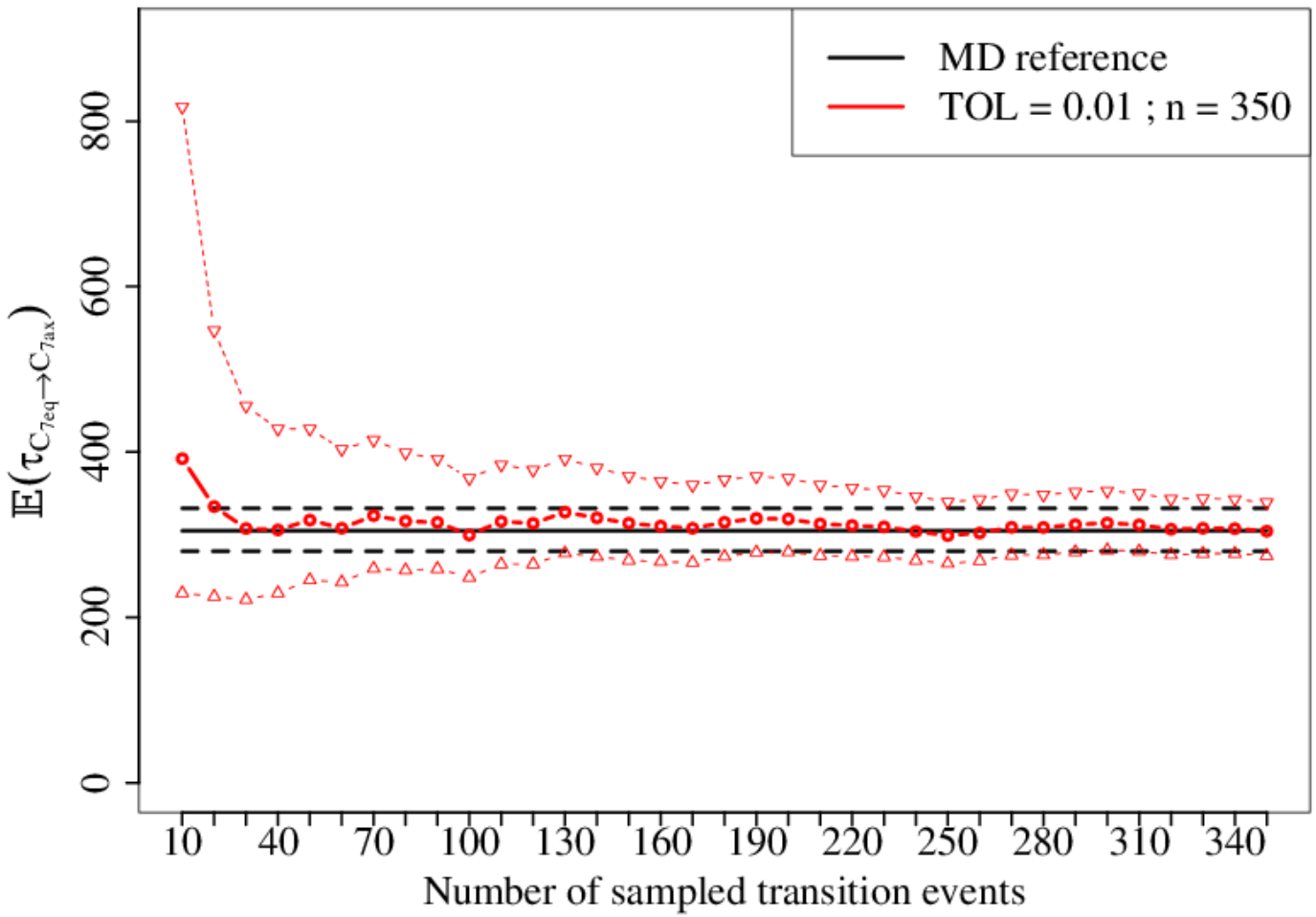
Definition of ParRep domains based on a free energy surface: we study the transition time from C_{7eq} (outside the red rectangle) to C_{7ax} (inside the red rectangle).

Alanine dipeptide (2/5)



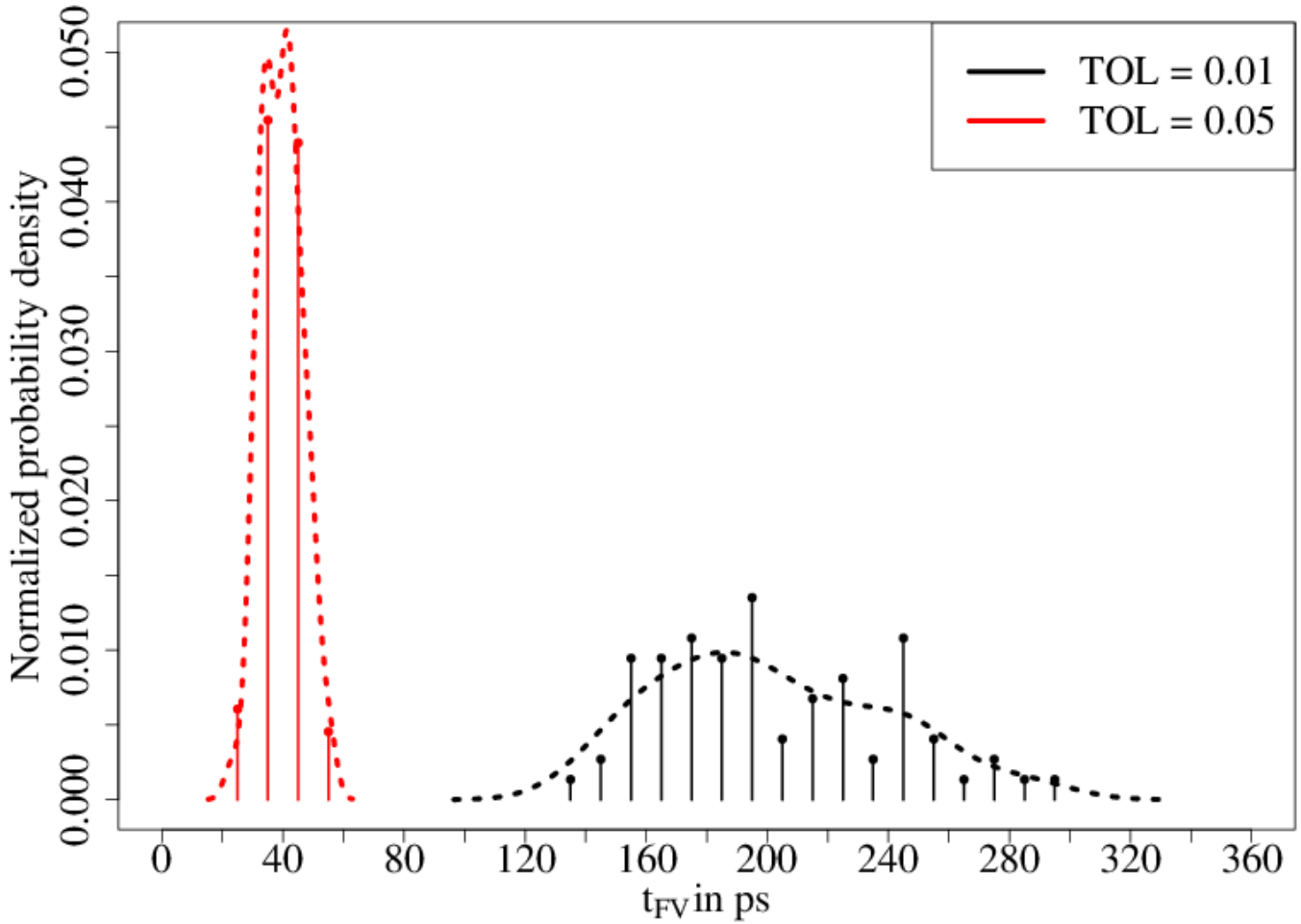
Cumulative distribution function of the transition time.

Alanine dipeptide (3/5)



Convergence of the mean transition time.

Alanine dipeptide (4/5)



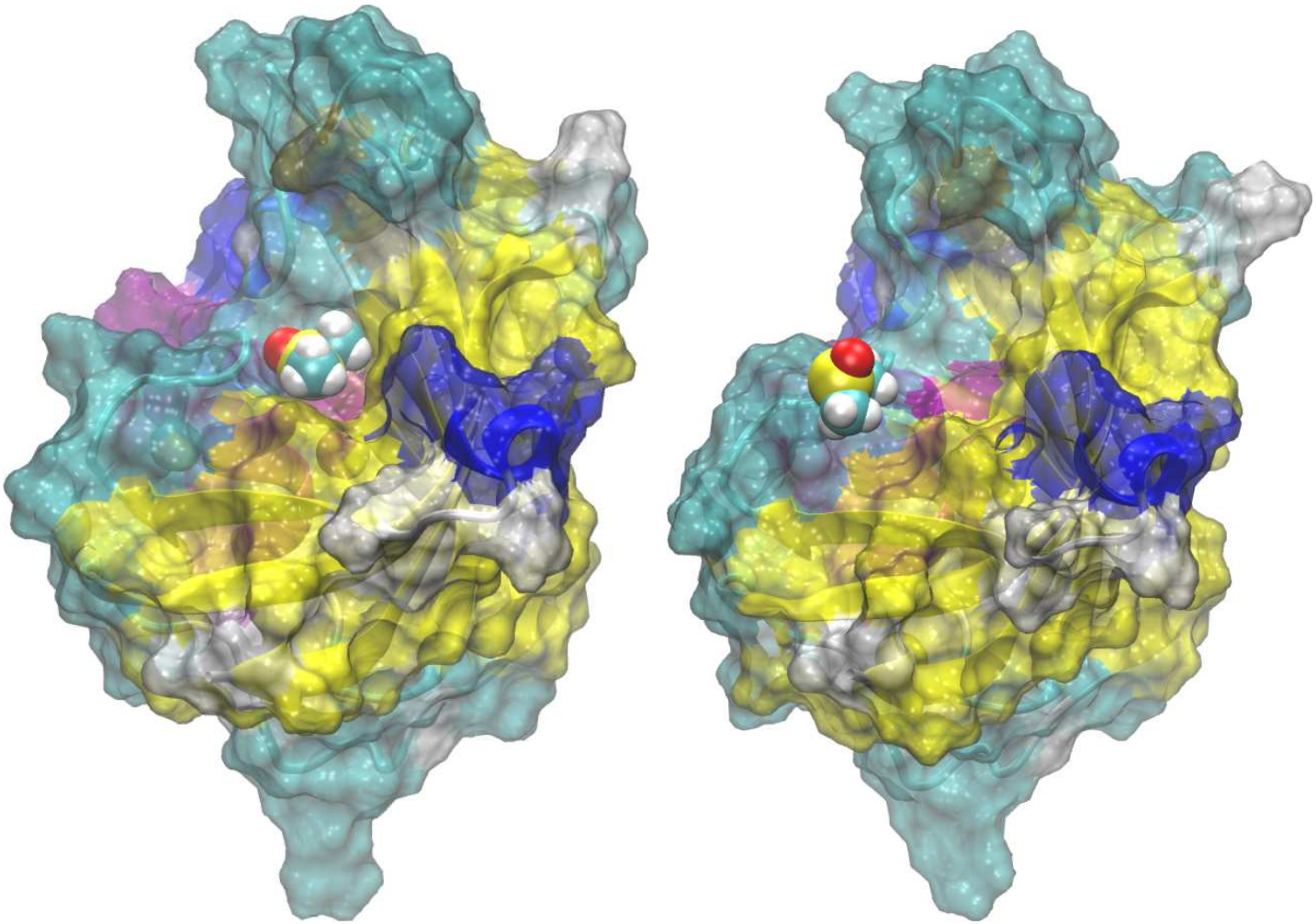
Distribution of the correlation times computed by FV.

Alanine dipeptide (5/5)

tol	WT(s)	t_{sim}(ns)	Speed(ns/day)	Eff. speedup	(Eff./Max)
0.01	6015	10008	143752	156	70%
0.025	5239	10103	166609	181	80%
0.05	4973	10032	174296	189	84%

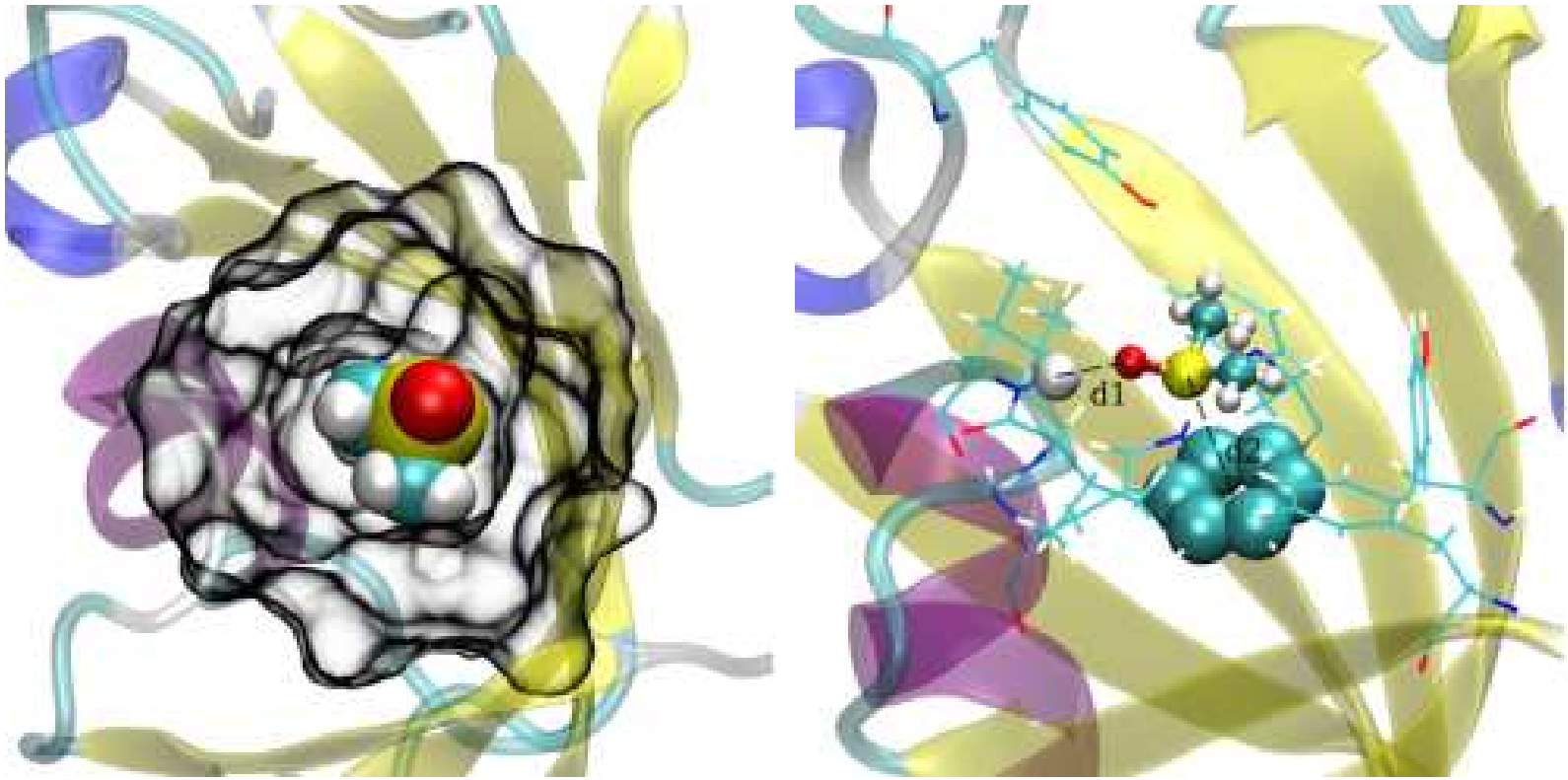
Effective speed-up as a function of the tolerance, for $N = 224$ replicas run in parallel (speed of a reference Langevin dynamics is 921 ns/day).

FKBP-DMSO (1/4)



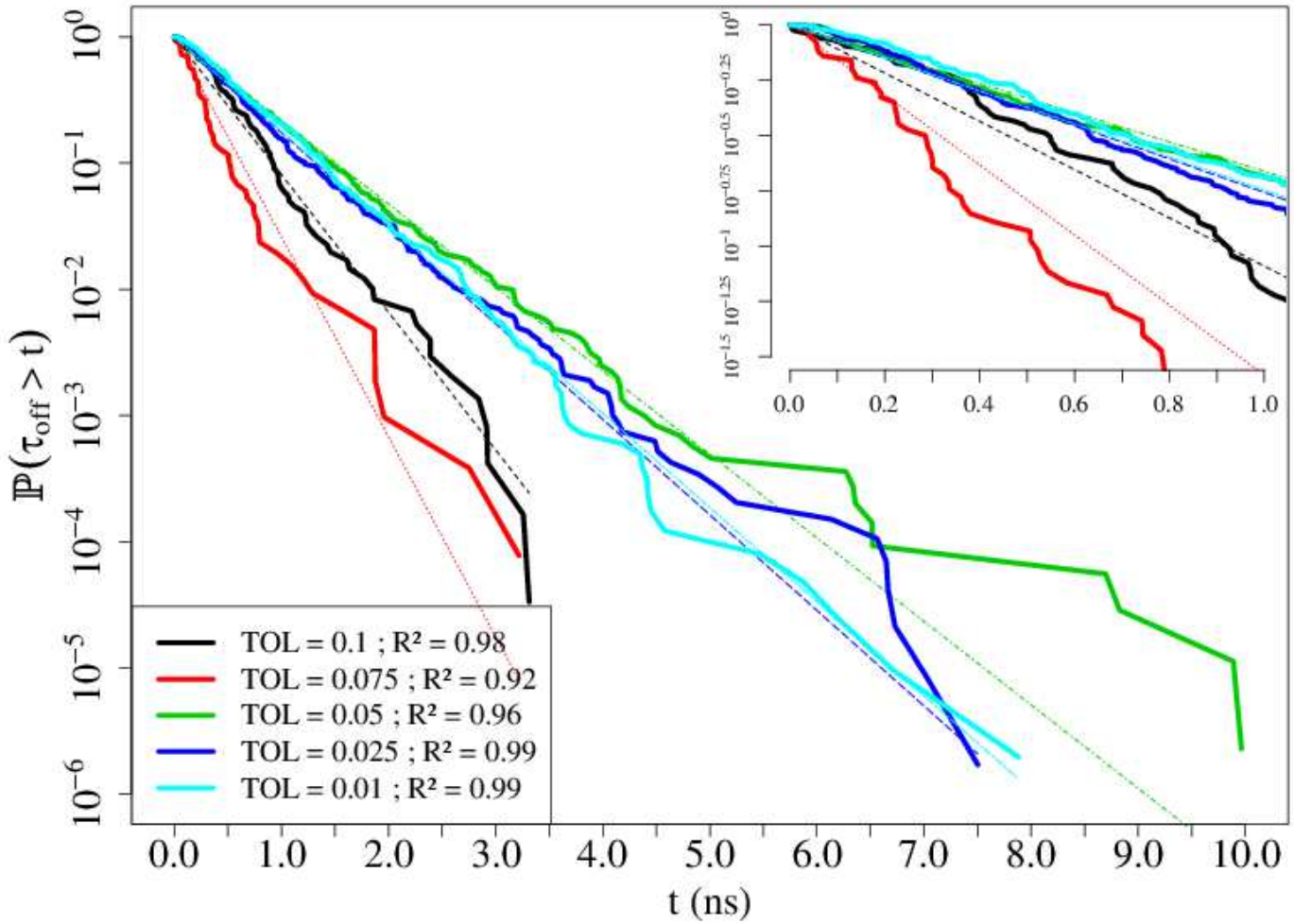
FKBP-DMSO complex,
corresponding to the RCSB-PDB entry “1D7H”.

FKBP-DMSO (2/4)



DMSO in its binding cavity ; distances used to define the cavity.

FKBP-DMSO (3/4)



Cumulative distribution function of the dissociation times.

FKBP-DMSO (4/4)

TOL	WT(s)	t_{sim}(ns)	Speed (ns/day)	Eff. speedup	(Eff./Max)
0.01	85142	403.5	409.4	79.5	56.8%
0.025	79574	457.6	496.8	96.5	68.9%
0.05	84455	482.2	493.4	95.8	68.4%

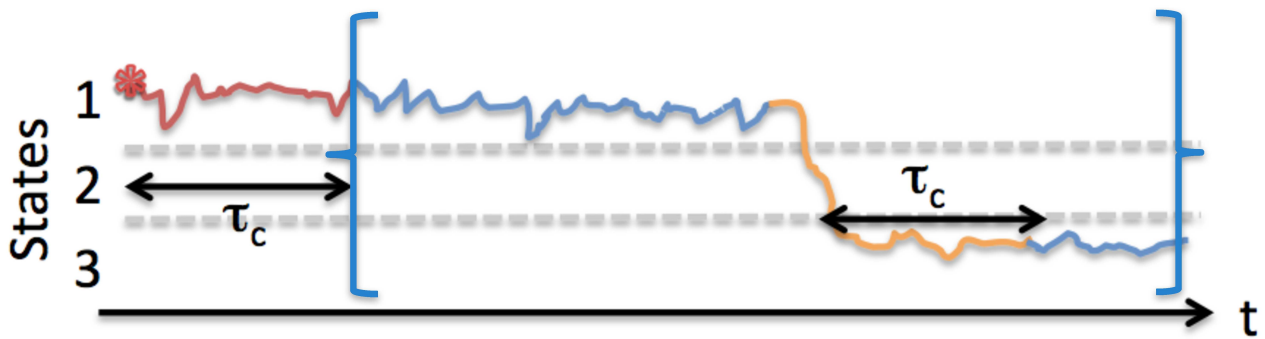
Effective speed-up as a function of the tolerance, for $N = 140$ replicas run in parallel (speed of a reference Langevin dynamics is 5.15 ns/day).

The Parallel Trajectory Splicing algorithm

Precompute the exit events [Perez, Cubuk, Waterland, Kaxiras, Voter, 2015]

Algorithm:

- Simulate in parallel short trajectories which start from the QSD in a state, and end at the QSD in a state.
- Glue together these short trajectories to build the full dynamics.

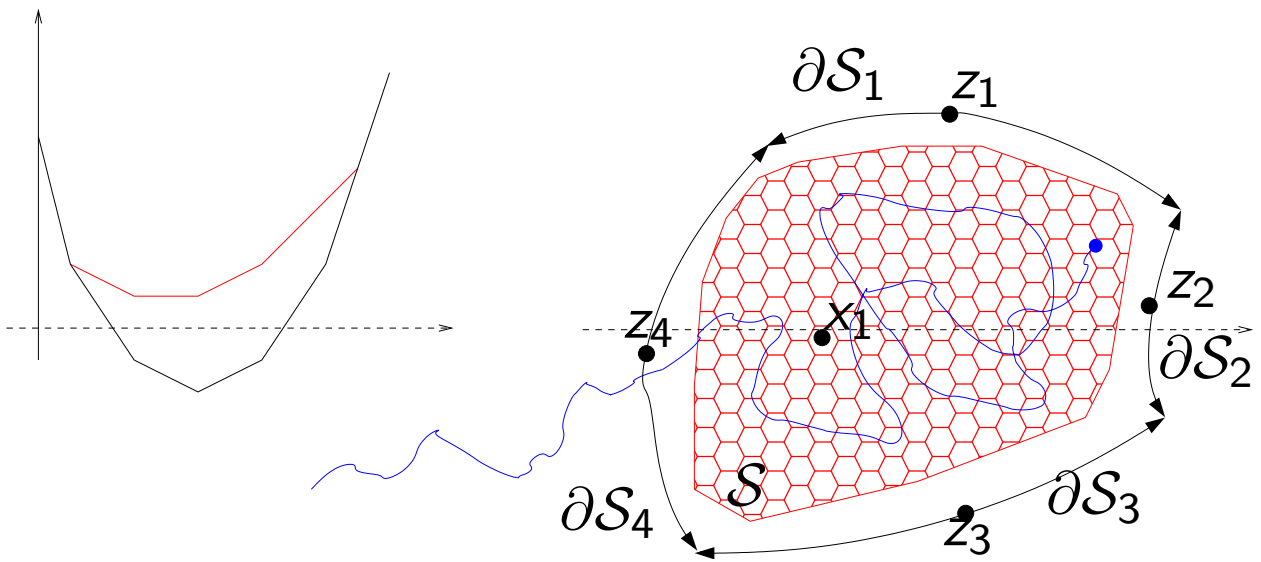


Hyperdynamics (1/2)

Raise the potential in $\mathcal{S}$ to reduce the exit time [Voter, 1997]

Two steps:

- Equilibrate on the biased potential $V + \delta V$;
- Wait for an exit and multiply the exit time $\tau_{\mathcal{S}}^{\delta V}$ by the boost factor $B = \frac{1}{\tau_{\mathcal{S}}^{\delta V}} \int_0^{\tau_{\mathcal{S}}^{\delta V}} \exp(\beta\, \delta V(\boldsymbol{X}_t))\, dt$.



Hyperdynamics (2/2)

Why is it consistent ?

Assumptions on δV: (i) $\delta V = 0$ on ∂S and (ii) δV is sufficiently small so that the Theorem above applies.

Recall the formula for the exit rates:

$$k_i = \widetilde{\nu}_i^{OL} \, \mathrm{e}^{-\beta[V(z_i)-V(x_1)]} \, (1 + O(\beta^{-1}))$$

where $\widetilde{\nu}_i^{OL} = \sqrt{\frac{\beta}{2\pi}} \partial_n V(z_i) \frac{\sqrt{\det(\nabla^2 V)(x_1)}}{\sqrt{\det(\nabla^2 V_{|\partial S})(z_i)}}$.

One easily check that $k_i / \sum_{j=1}^J k_j$ is independent of δV and

$$\frac{\sum_{j=1}^J k_j(V+\delta V)}{\sum_{j=1}^J k_j(V)} = \sqrt{\frac{\det(\nabla^2(V+\delta V))(x_1)}{\det(\nabla^2(V))(x_1)}} \mathrm{e}^{\beta \delta V(x_1)} (1 + O(\beta^{-1}))$$

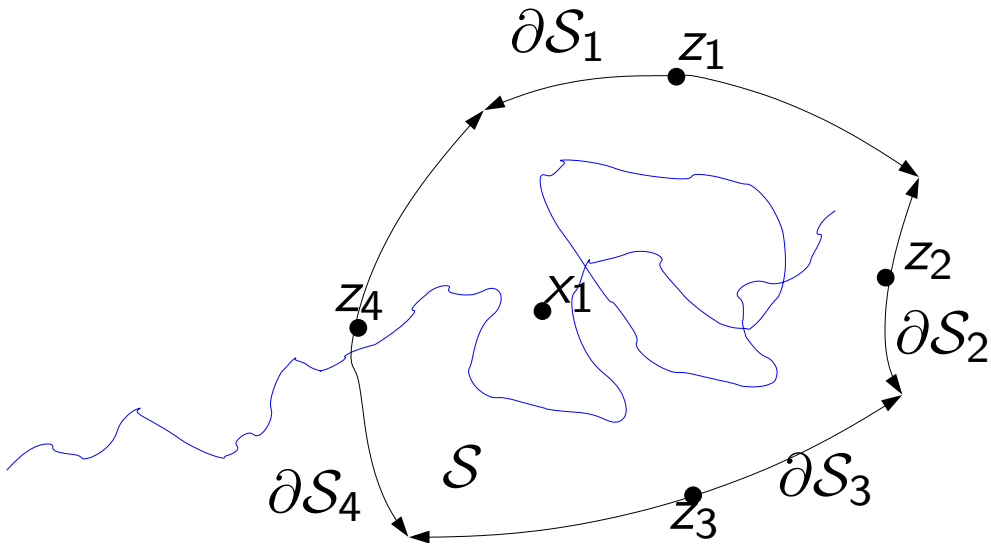
$$= \frac{\int_S \exp(-\beta V)}{\int_S \exp(-\beta(V+\delta V))} (1 + O(\beta^{-1})) \simeq B$$

Temperature Accelerated Dynamics (1/2)

Increase the temperature to reduce the exit time [Sorensen, Voter, 2000]

Algorithm:

- Observe the exit events from $\mathcal{S}$ at high temperature ;
- Extrapolate the high temperature exit events to low temperature exit events.



Temperature Accelerated Dynamics (2/2)

Recall that, starting from the QSD, the exit event from a given state $\mathcal{S}$ can exactly be modelled using a kinetic Monte Carlo model with rates

$$k_i = \widetilde{\nu}_i^{OL} \, \mathrm{e}^{-\beta[V(z_i)-V(x_1)]} \, (1 + O(\beta^{-1}))$$

where $\widetilde{\nu}_i^{OL} = \sqrt{\frac{\beta}{2\pi}} \partial_n V(z_i) \frac{\sqrt{\det(\nabla^2 V)(x_1)}}{\sqrt{\det(\nabla^2 V_{|\partial S})(z_i)}}$.

Thus,

$$\frac{k_i^{lo}}{k_i^{hi}} \simeq \sqrt{\frac{\beta^{lo}}{\beta^{hi}}} \exp(-(\beta^{lo} - \beta^{hi})(V(z_i) - V(x_1))).$$

Algorithm: observe exit events at high temperature, extrapolate the rates to low temperature, stop when the extrapolated event will not modify anymore the low temperature exit event.

Remark: TAD can be seen as a smart saddle point search method.

Generalizations and perspectives

- The parallel replica is a very versatile algorithm: it applies *e.g.* to non reversible dynamics, discrete-in-time dynamics, continuous-time Markov Chain [Aristoff, Plechac, Wang]. It does not require estimates of the exit rates.
- Hyper and TAD are more efficient, but require the temperature to be sufficiently small so that estimates of the rates by the Eyring-Kramers formulas hold true.

All these techniques require "good" metastable states:

exit time $>$ convergence time to the QSD.

Conclusion

There are mathematical characterizations of good coarse-graining representations (spectral gaps, convergence times *vs* exit times).

Could we use those characterizations together with advanced learning techniques (auto-encoder, sparse methods) to get better coarse-grained descriptions?

- Identify slow variables
- Sparse representation of the committor function
- Identify metastable states

References

Some papers I mentioned:

- A. Binder, TL and G. Simpson, *A Generalized Parallel Replica Dynamics*, Journal of Computational Physics, 284, 2015.
- G. Di Gesù, TL, D. Le Peutrec and B. Nectoux, *Sharp asymptotics of the first exit point density*, Annals of PDE, 5, 2019.
- F. Hédin and TL, *gen.parRep: a first implementation of the Generalized Parallel Replica dynamics for the long time simulation of metastable biochemical systems*, Computer Physics Communications, 239, 2019.
- C. Le Bris, TL, M. Luskin and D. Perez, *A mathematical formalization of the parallel replica dynamics*, Monte Carlo Methods and Applications, 18(2), 2012.
- TL, M. Rousset and G. Stoltz, *Long-time convergence of an Adaptive Biasing Force method*, Nonlinearity, 21, 2008.

References

Recent review papers:

- TL and Gabriel Stoltz, *Partial differential equations and stochastic methods in molecular dynamics*. Acta Numerica, 25, 2016.
- TL, *Mathematical foundations of Accelerated Molecular Dynamics methods*, In: W. Andreoni and S. Yip (eds) Handbook of Materials Modeling, Springer, 2018.