Towards more efficient molecular simulations

Gabriel STOLTZ, Tony LELIEVRE

(CERMICS, Ecole des Ponts)

Presentation of the institutions at play

- **CERMICS: Applied mathematics laboratory of Ecole des Ponts**
  - 18 permanent members (16 HdR)
  - about 30 PhD students and 10 postdocs
  - research directions:
    - applied probability
    - modeling, analysis and simulation
    - optimization and operations research

- **MATHERIALS: project-team of Inria Paris**
  - 8 permanent researchers
  - strong overlap with CERMICS
  - analysis and development of simulation methods for multiscale models (incl. stochastic homogenization) and molecular simulations (quantum/classical)

What can (and cannot) applied mathematics do?

We, applied mathematicians...

- work on simplified models (one dimensional reaction coordinates, overdamped Langevin dynamics, etc) $\rightarrow$ this make us look like fools...
- exaggerate sources of errors (for instance by considering situations which rarely happen in practice) $\rightarrow$ this makes us insufferable

...but the aim is to

- rigorously understand why some methods work and some don't (mathematical proofs)
- devise new numerical strategies based on this theoretical understanding
- validate them on toy examples
- transfer the knowledge to practitionners by helping them to implement the methods into their own codes

Free energy computations (1)

- Canonical distribution $\mu(dq\,dp) = Z^{-1}\mathrm{e}^{-\beta H(q,p)}\,dq\,dp$

- For a given reaction coordinate $\xi(q)$, compute the function

$$F(z) = \int_{\xi^{-1}\{z\}} \mathrm{e}^{-\beta H(q,p)}\,\delta_{\xi(q)-z}(dq)$$

Thermodynamic integration, free energy perturbation, nonequilibrium techniques, adaptive methods (ABF, $\star$-metadynamics, ...)

- ABF in the simplest case ($\xi(q) = q_1$)

$$\left\{\begin{array}{c} dq_t = -\nabla\Big(V - F_t \circ \xi\Big)(q_t)\,dt + \sqrt{2\beta^{-1}}\,dW_t, \\ F_t'(z) = \mathbb{E}\left(f(q_t)\,\middle|\,\xi(q_t) = z\right) \end{array}\right\}$$

Free energy computations (2)

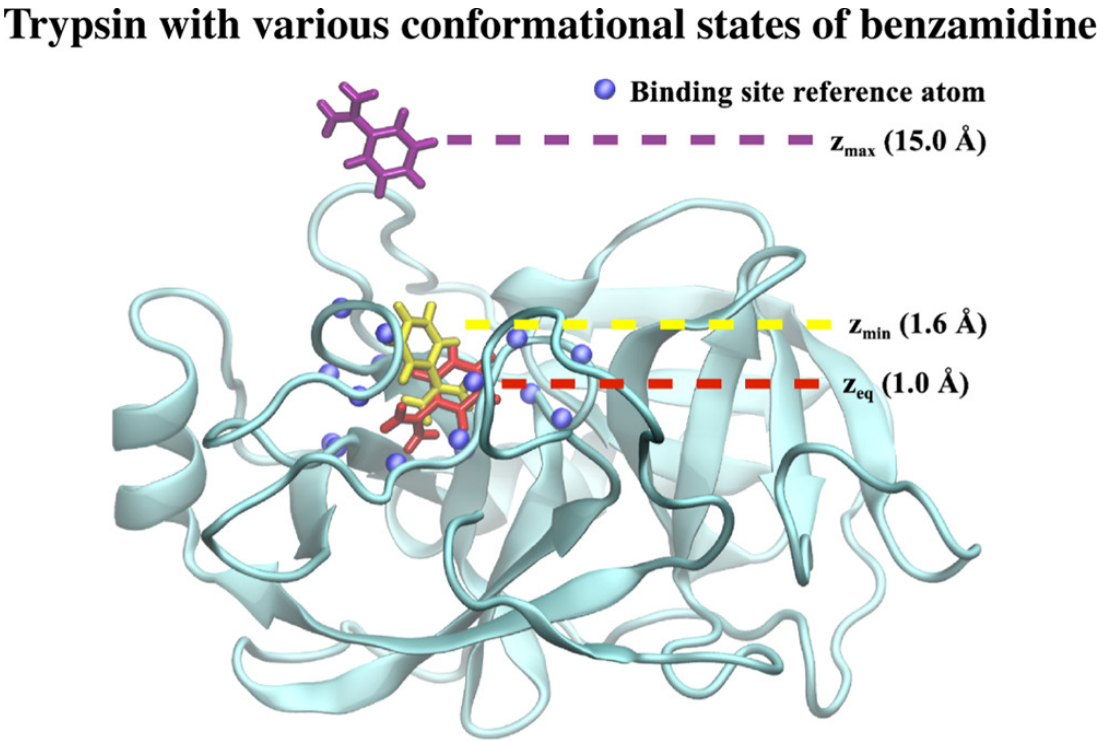
- **Our contributions for ABF include:**
  - a theoretical understanding of the improved convergence rate compared to unbiased dynamics (entropy methods)
  - improvements of the simulation methods: use of replicas and selection (implemented in NAMD)
  - multidimensional reaction coordinate case: projecting the current estimated mean force onto a gradient

- **Other results for adaptive dynamics:**
  - convergence of Wang-Landau type dynamics (including Self-Healing Umbrella Sampling and Well-Tempered metadynamics)
  - suggestion of modifications in the methods to improve convergence rates (e.g. allow for larger steps in the free energy update for WTM)

- **Also results for other types of dynamics/methods:**
  - constraints (overdamped/underdamped, TI, Jarzynski–Crooks, etc)
  - temperature accelerated molecular dynamics

Sampling reactive trajectories (1)

- Motivation: Unbinding of a ligand from a protein



- Challenge: bridge the gap between timescales
  - Elementary time-step for the molecular dynamics $= 10^{-15}\,\mathrm{s}$
  - Dissociation time $\simeq 0.02\,\mathrm{s}$

Sampling reactive trajectories (2)

- **Mathematical setting: rare event computation**
  - Stochastic process $(X_t)_{t\geq 0}$, stopping times τ_A and τ_B
  - Aim: simulate and compute the very small probability $\mathbb{P}(\tau_B < \tau_A)$
  - Here sets A and B defined as metastable states (bound/unbound)

- Splitting technique: find intermediate events easier to simulate

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

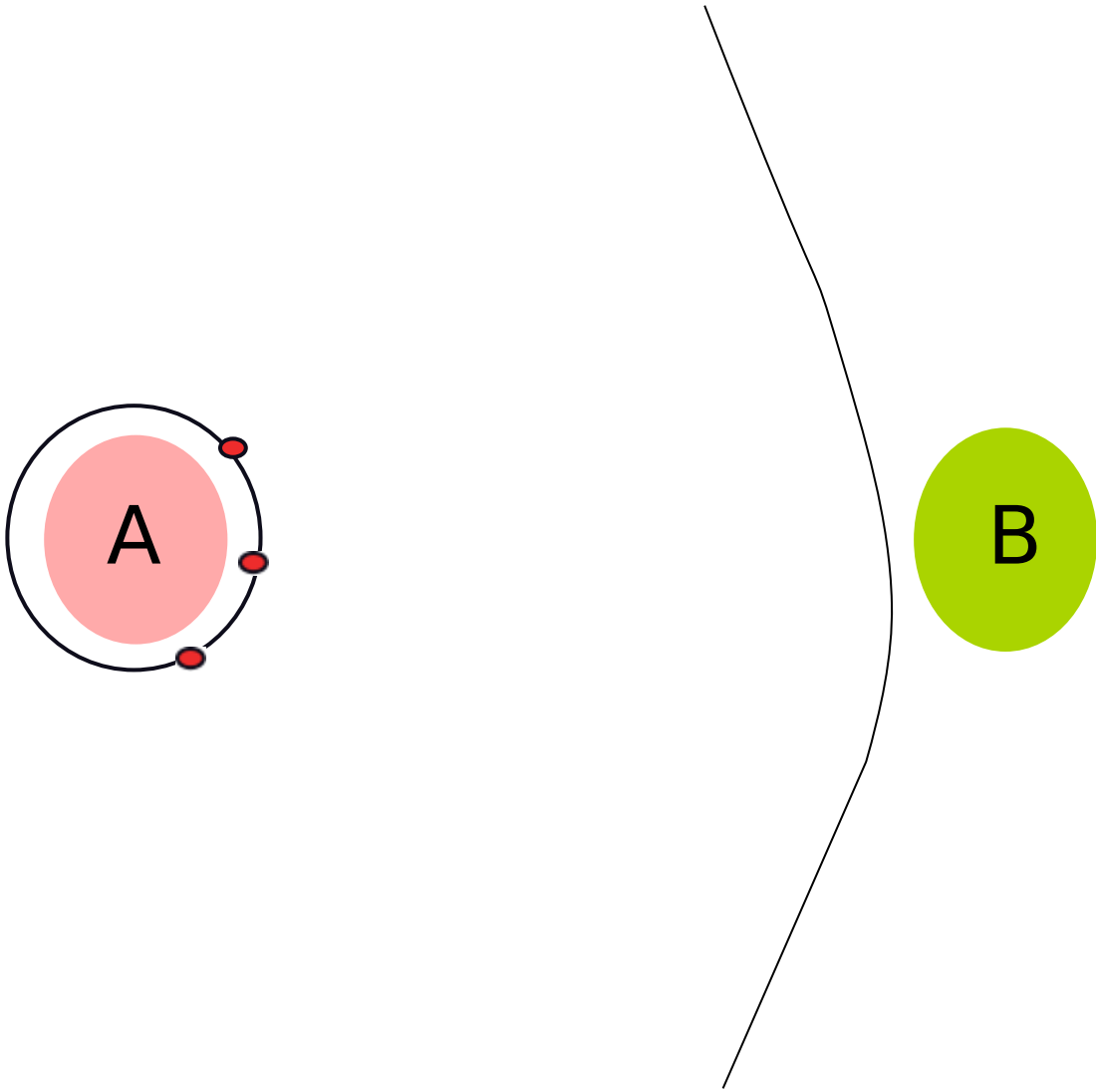
and simulate the successive conditional events: for $k = 1, 2, \ldots,$

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

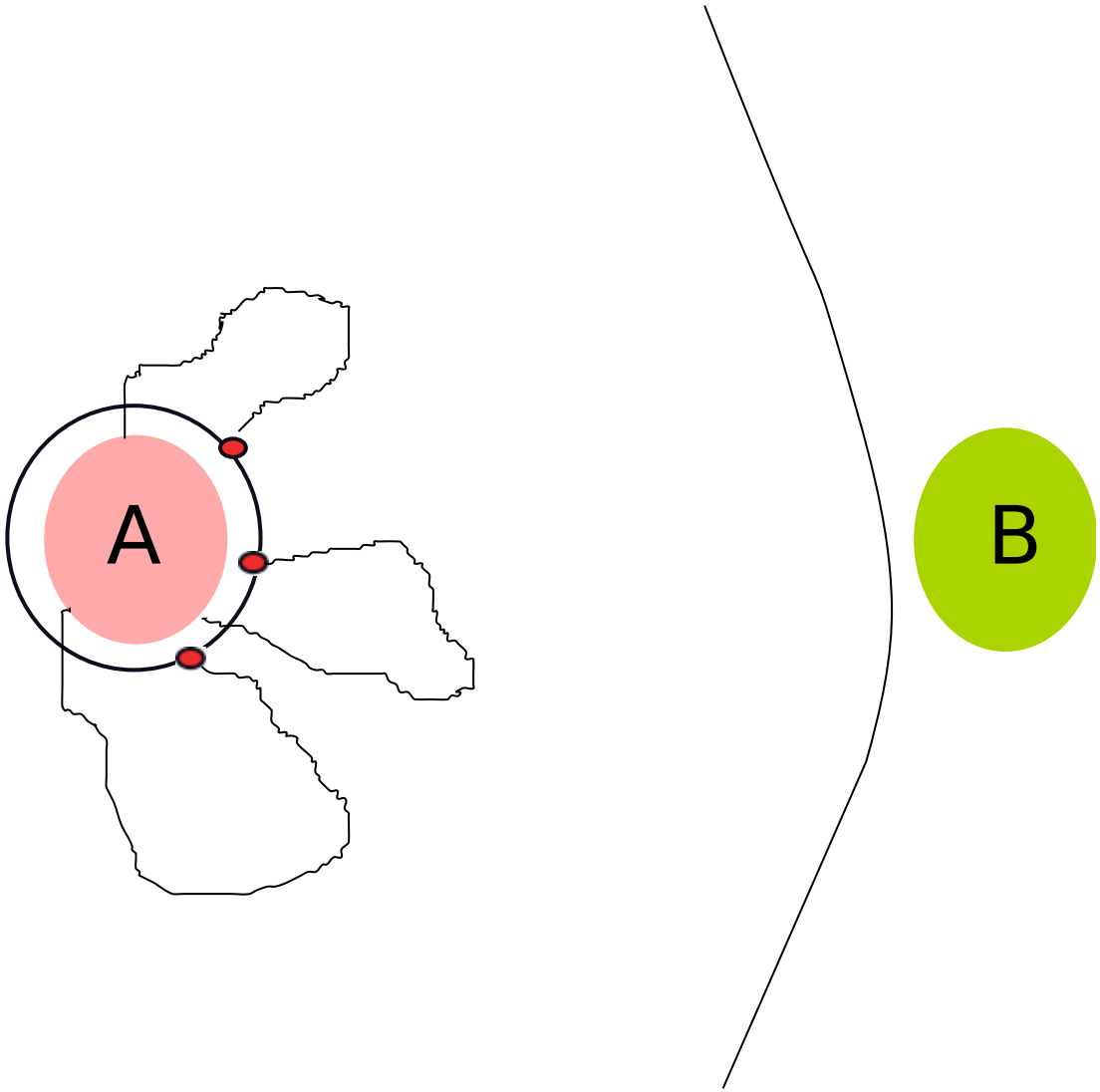
where $\tau_z = \inf\{t, \xi(X_t) > z\}$ for good importance function $\xi \in \mathbb{R}$.

- Adaptive feature: build the intermediate levels $(z_i)_{i\geq 1}$ on the fly $\rightarrow$ Adaptive Multilevel Splitting algorithm (Cérou/Guyader, 2007)

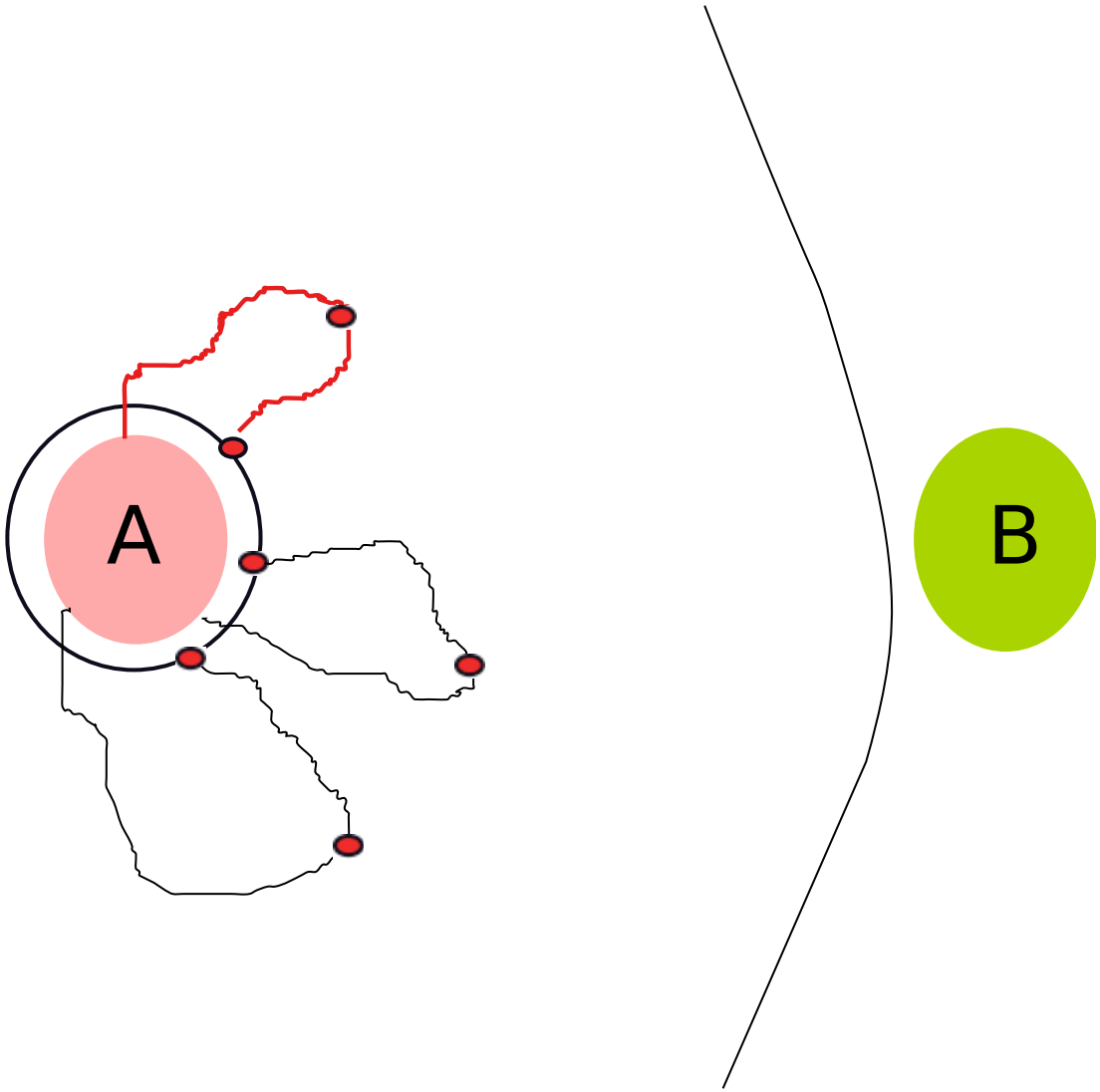
Schematic illusration of the AMS Algorithm



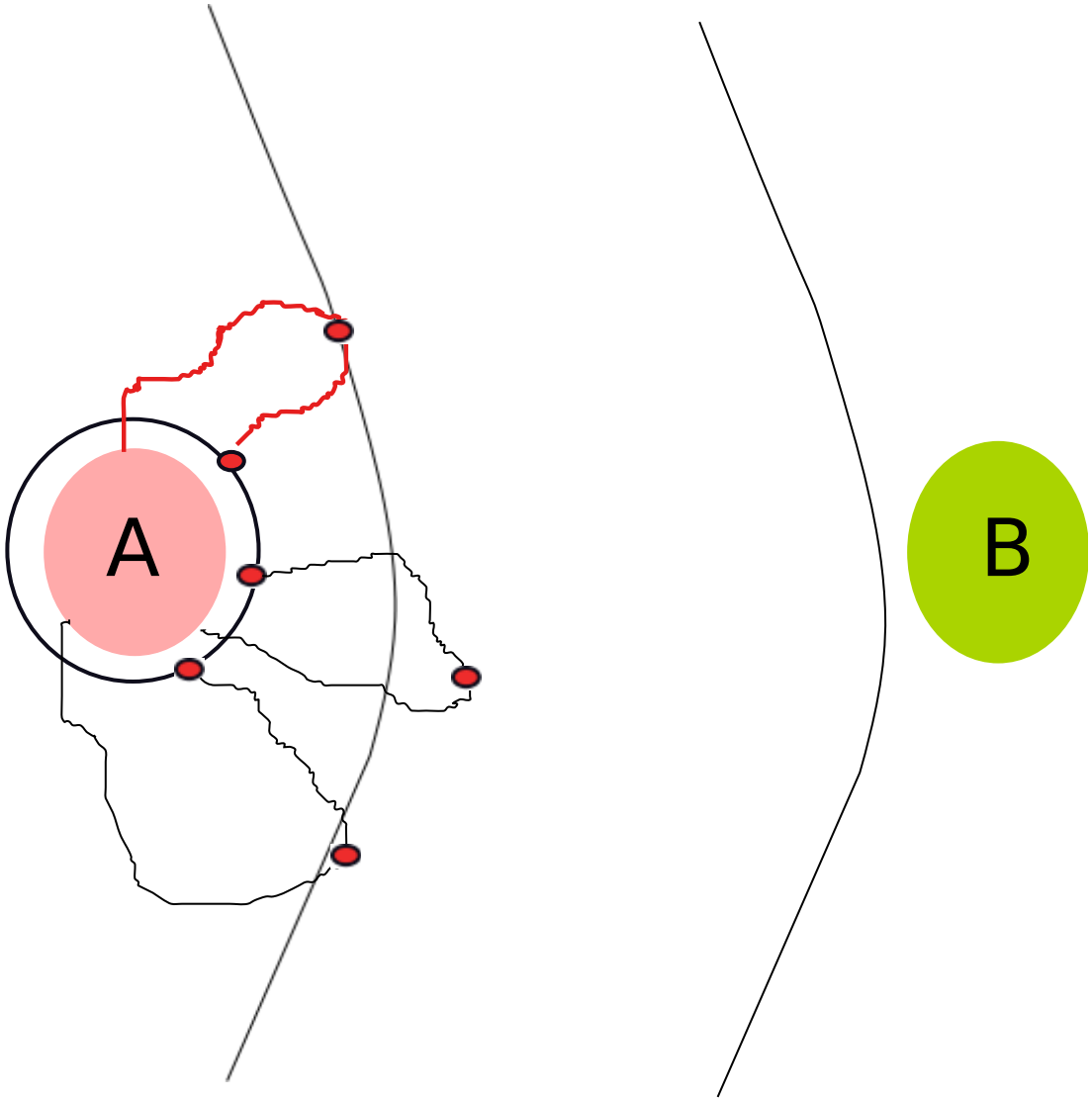
Schematic illusration of the AMS Algorithm



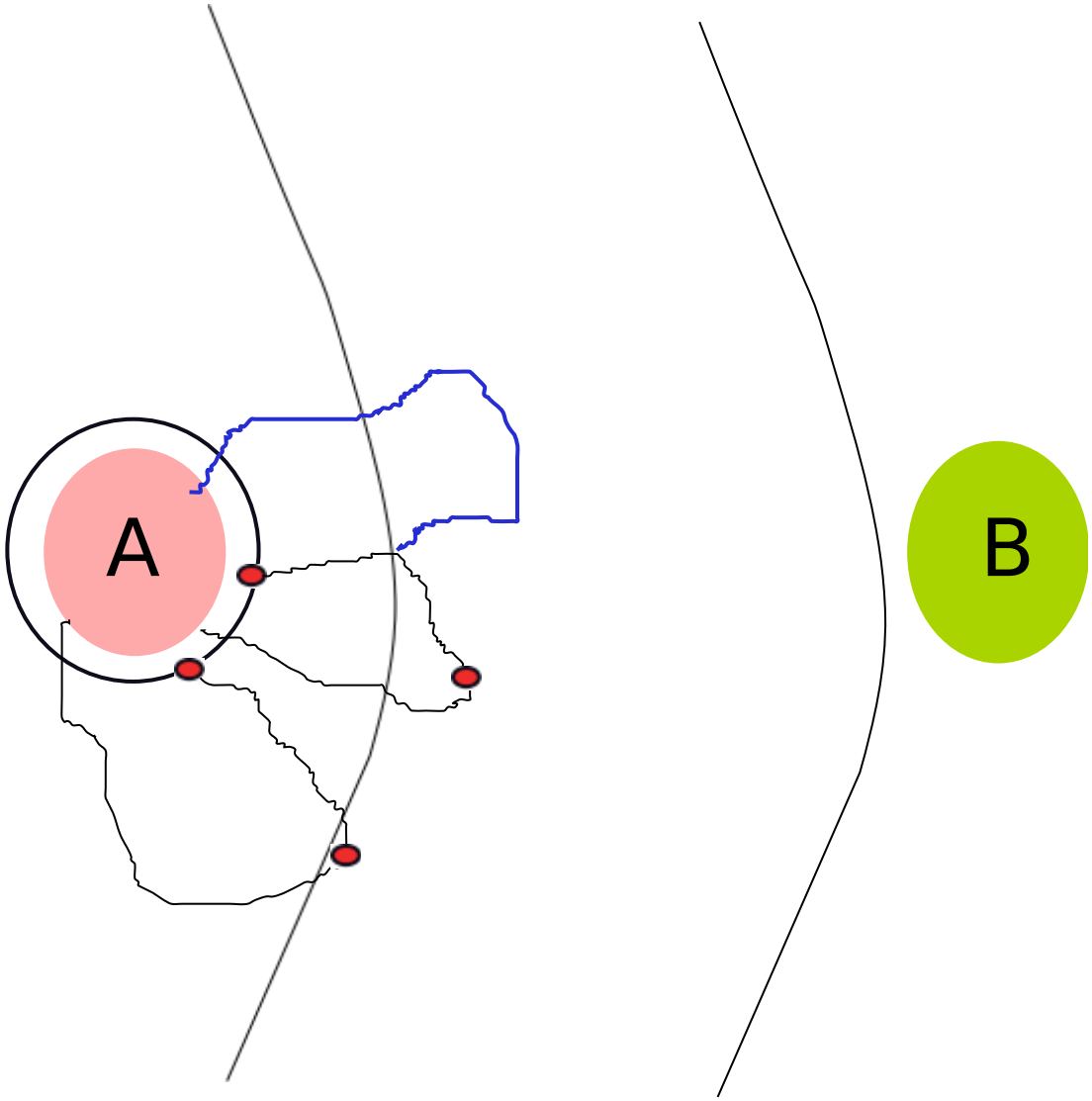
Schematic illusration of the AMS Algorithm



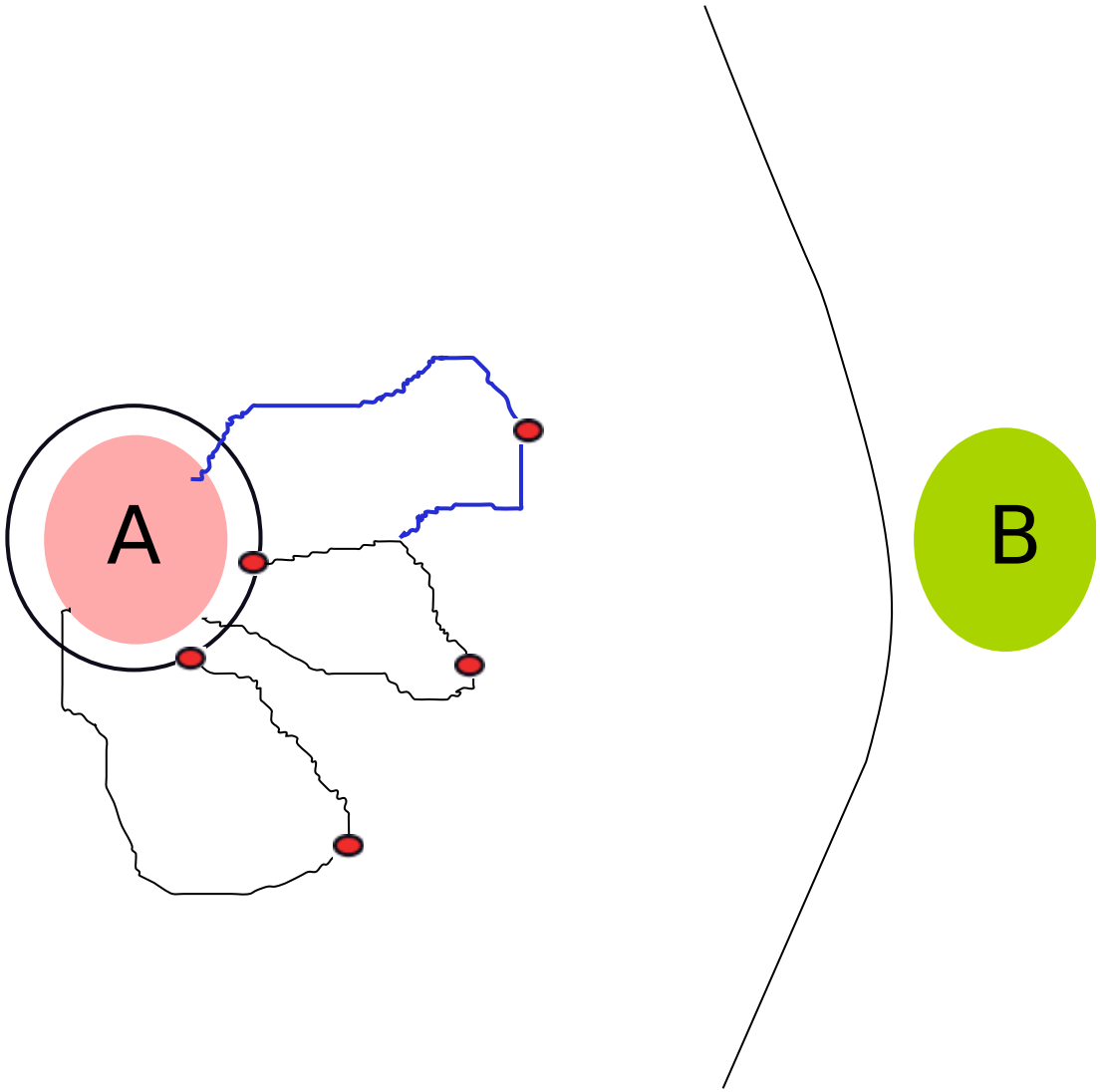
Schematic illusration of the AMS Algorithm



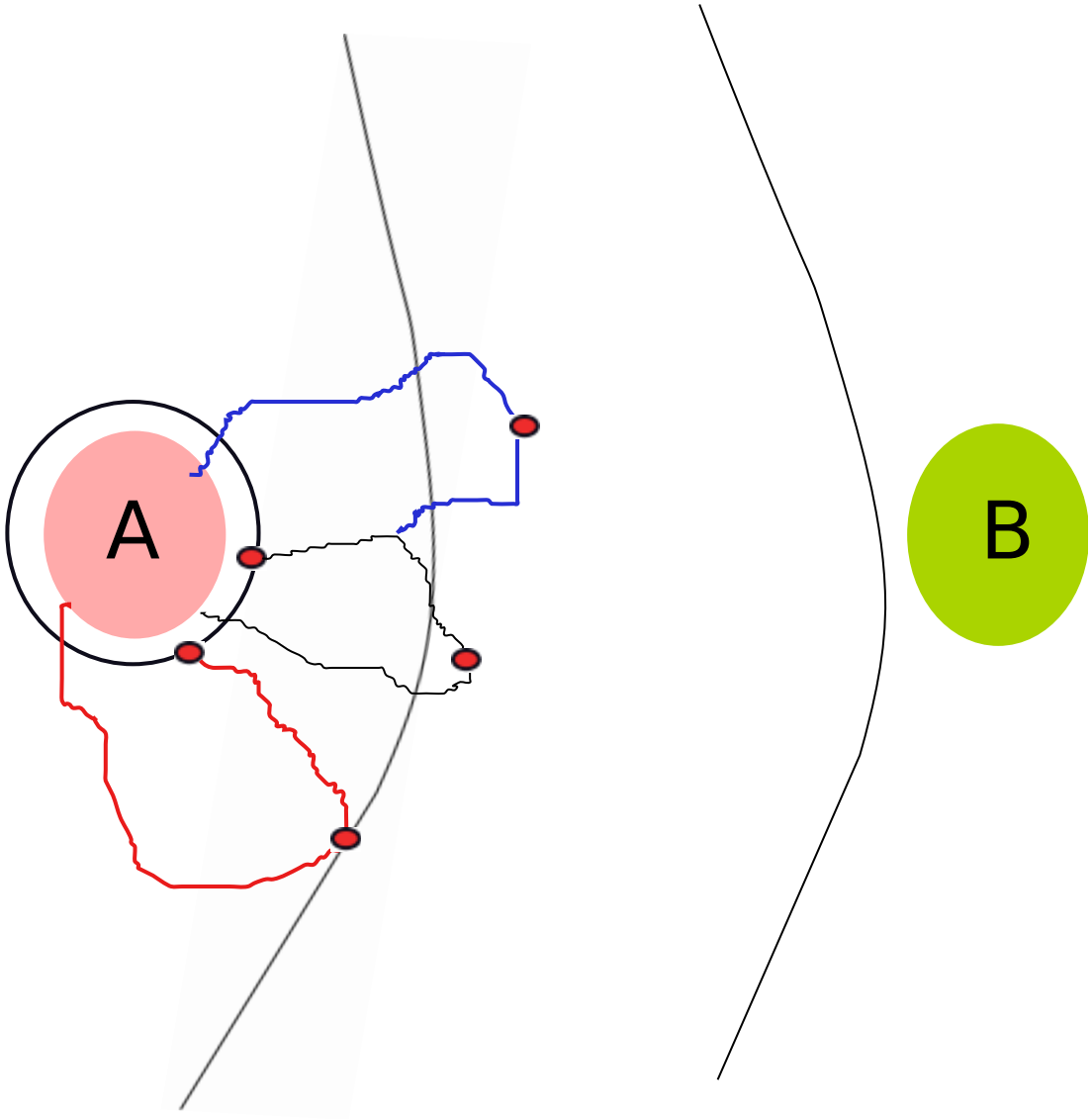
Schematic illusration of the AMS Algorithm



Schematic illusration of the AMS Algorithm



Schematic illusration of the AMS Algorithm



Sampling reactive trajectories (3)

- In collaboration with the group of K. Schulten (C. Mayne and I. Teo), AMS is currently implemented in the NAMD code.

- **Unbinding event of benzamidine from trypsin:**
  - MD setup: about 70 000 atoms, CHARMM36 force field, NPT conditions (298 K)
  - Estimated dissociation rate: $k_{\text{off}} = (260 \pm 240) s^{-1}$ which is in the same order of magnitude as the experimental rate $(600 \pm 300) s^{-1}$
  - Overall simulation time: $2.3\ \mu s$ which is 4 orders of magnitude shorter than than the estimated dissociation time

Computation of transport coefficients

- **Green-Kubo formulas:** integrated correlation functions

Effective diffusion at equilibrium

Unperiodized displacement $Q_t - Q_0 = \int_0^t M^{-1} p_s \, ds$

$$D = \lim_{t \to +\infty} \frac{\mathbb{E}\left[(Q_t - Q_0) \otimes (Q_t - Q_0)\right]}{2t} = \int_0^{+\infty} \mathbb{E}_0 \left[M^{-1} p_t \otimes M^{-1} p_0\right] dt$$

- Alternatively: linear response of steady-state nonequilibrium dynamics

- **Issues/questions :**
  - bias due to discretization in time

$$\int_0^{+\infty} \mathbb{E}\Big(\psi(x_t)\varphi(x_0)\Big) dt = \Delta t \sum_{n=0}^{+\infty} \mathbb{E}_{\Delta t} \left(\widetilde{\psi}_{\Delta t, \alpha}\left(x^n\right) \varphi\left(x^0\right)\right) + \mathrm{O}(\Delta t^\alpha)$$

  - variance reduction (in progress)

Results in 1D for $\varphi = \psi = V'$ and cosine potential