Rare events sampling methods and machine learning techniques for the simulation of catalytic reactions

Tony Lelièvre

CERMICS Ecole des Ponts, Institut Polytechnique de Paris & MATHERIALS team INRIA

Co-authors: T. Pigeon (IFPEN), P. Raybaud (IFPEN) and G. Stoltz (ENPC, INRIA)

In collaboration with M. Corral Valero, A. Anciaux-Sedrakian, and M. Moreaud

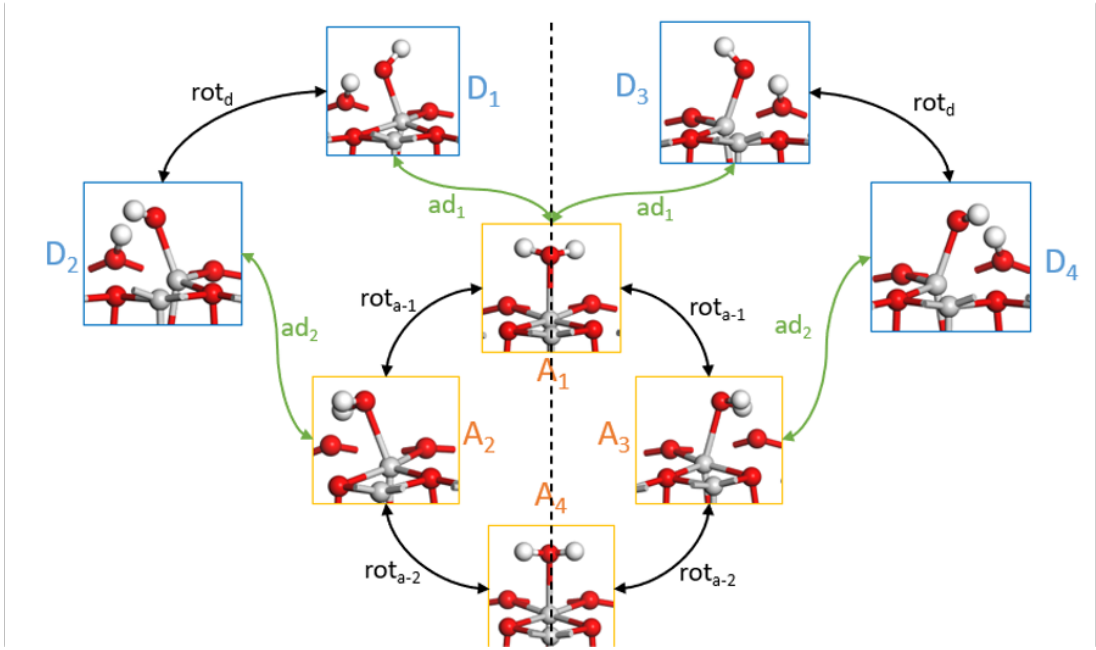
Journée IFPEN-Inria, Paris, 9 décembre 2024

Catalytic reactions

Motivation: simulation of catalytic reactions involved in the conversion of biomass, in order to uncover the molecular mechanisms and quantify the reaction rates

Objective: optimization and control of catalytic reactions:

- Ranking of existing catalysts for a given target
- Design of new efficient catalysts (without experiments)



Example: dissociation of water molecules on the (100)-surface of γ alumina

Modelling

System of N atoms with:

- Positions: $\mathbf{q} \in \Omega$ where $\Omega = \mathbb{R}^{3N}$ or $\mathbb{T}^{3N}$
- Momenta: $\mathbf{p} \in \mathbb{R}^{3N}$
- Hamiltonian: $H(\mathbf{q}, \mathbf{p}) = \frac{1}{2}\mathbf{p}^T M^{-1}\mathbf{p} + V(\mathbf{q})$

The function V models the interactions between atoms. For catalytic processes (chemical reactions), computationally expensive methods must be used to evaluate V: ab initio molecular dynamics (AIMD)

At temperature $\beta^{-1} = k_B T$, positions and momenta are distributed according to the Boltzmann–Gibbs measure:

$$\rho(d\mathbf{q}, d\mathbf{p}) = Z^{-1} \mathrm{e}^{-\beta H(\mathbf{q},\mathbf{p})} d\mathbf{q} d\mathbf{p}$$

Modelling

Evolution in time of the system modelled by Langevin dynamics (friction $\gamma > 0$)

$$\begin{cases} d\mathbf{q}_t = M^{-1}\mathbf{p}_t dt \\ d\mathbf{p}_t = -\nabla V(\mathbf{q}_t)dt - \gamma \mathbf{p}_t dt + \sqrt{\frac{2\gamma}{\beta}} M^{\frac{1}{2}} d\mathbf{W}_t \end{cases}$$

This dynamics is ergodic w.r.t. ρ

Questions we would like to address:

- What are the main modes (reactants, intermediates, products) of the Boltzmann–Gibbs measure?
- How much time does it take for the Langevin dynamics to go from one mode (metastable state) to another?
- What is (are) the path(s) followed during these transitions?

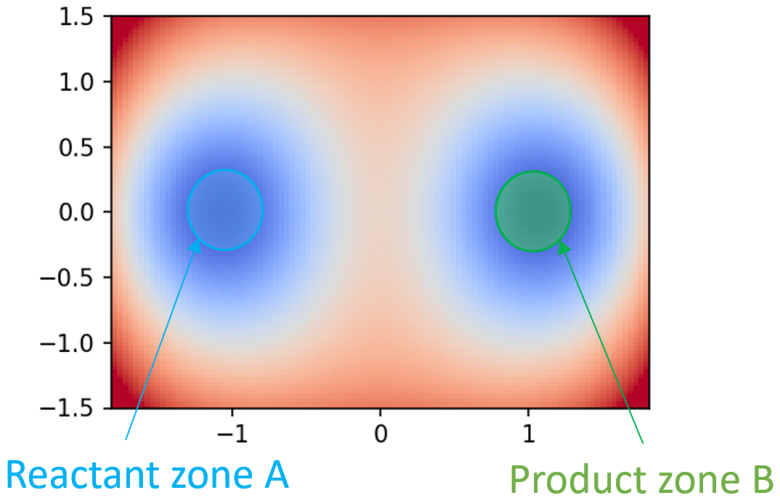
Fundamental problem: a naive simulation is too costly
(discretization timestep $\simeq 10^{-15} s \ll$ timescale of transitions $\simeq 10^{-6} - 10^{3} s$)
$\longrightarrow$ The dynamics are metastable, and observing a transition is a rare event

Table of Contents

Table of Contents

Equilibrium transition time

Two metastable states A and B



The transition time from A to B is:

$$T_{A\to B} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\tau_B^k - \tau_A^k)$$

with:

$$\tau_A^k = \inf \left\{ t > \tau_B^{k-1} \;\middle|\; \mathbf{q}_t \in A \right\}$$

$$\tau_B^k = \inf \left\{ t > \tau_A^k \;\middle|\; \mathbf{q}_t \in B \right\}$$

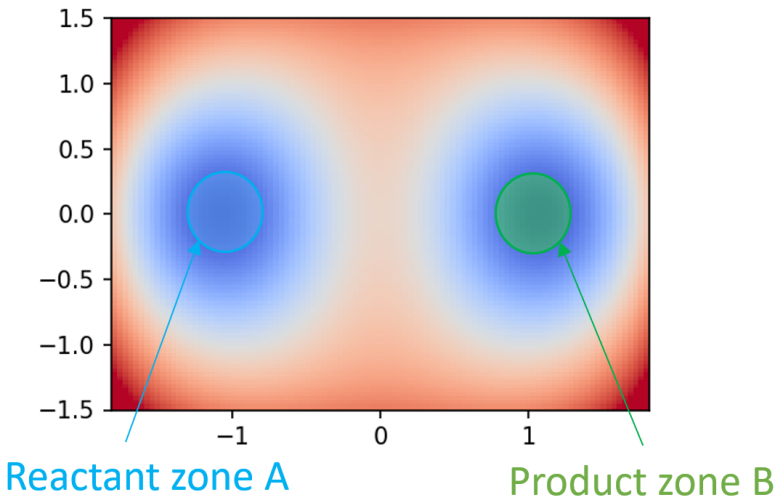
The reaction rate is defined by $k_{A\to B} = T_{A\to B}^{-1}$

Computing the reaction rate is difficult:

(i) One needs to sample the equilibrium first entrance distribution in A

(ii) The typical transition time $(\tau_B^k - \tau_A^k)$ is large

Hill relation



The reaction rate satisfies the Hill relation:

$$k_{A\to B} = \Phi_A \, \mathbb{P}_{\partial A}(\tau_B < \tau_A)$$

where

$$\tau_X = \inf\left\{ t \in (0;+\infty) \;\middle|\; \mathbf{q}_t \in X \right\}$$

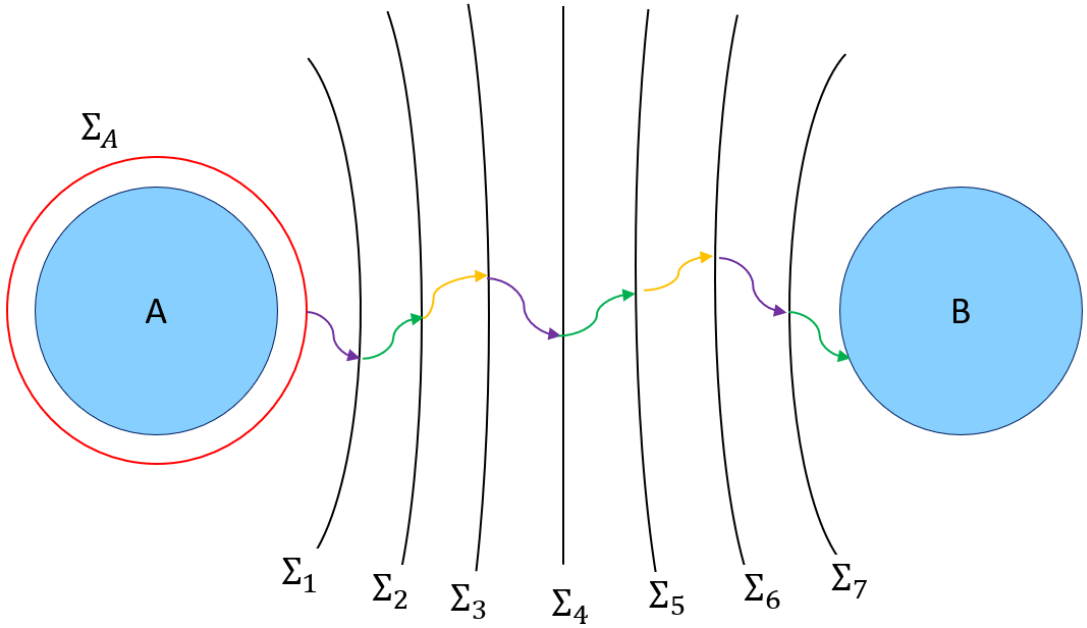
- Φ_A is the frequency of exits from A (easy to compute)
- $\mathbb{P}_{\partial A}(\tau_B < \tau_A)$ is the probability of reaching B before A starting at the equilibrium distribution ρ restricted to ∂A

The Hill relation is exact for the overdamped Langevin and Langevin dynamics [Hill, 2005] [Baudel, Guyader, TL, Stoch. Proc. Appl. 2020] [TL, Ramil, Reygner, Ann. IHP 2022]

Problem (i) is solved, and problem (ii) has been replaced by the estimation of a rare event probability

Sampling a rare event: splitting algorithms

Initialization: a surface $\Sigma_A = \Sigma_0$ (such that $\mathbb{P}(\tau_{\Sigma_A} < \tau_A)$ is not too small) is introduced to capture "actual exits" from A



$$\mathbb{P}(\tau_B < \tau_A) = \mathbb{P}(\tau_{\Sigma_A} < \tau_A) \left(\prod_{i=0}^{6} \mathbb{P}(\tau_{\Sigma_{i+1}} < \tau_A | \tau_{\Sigma_i} < \tau_A) \right) \mathbb{P}(\tau_B < \tau_A | \tau_{\Sigma_7} < \tau_A)$$

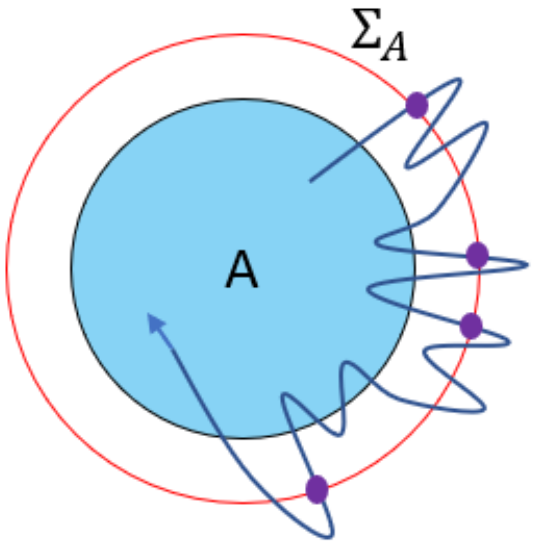
By the Markov property, given an ensemble of trajectories such that $\tau_{\Sigma_i} < \tau_A$, one can estimate the probability $\mathbb{P}(\tau_{\Sigma_{i+1}} < \tau_A | \tau_{\Sigma_i} < \tau_A)$ by spawning trajectories from the first hitting points of Σ_i

How to choose the surfaces Σ_i ?

Adaptive multilevel splitting (AMS)

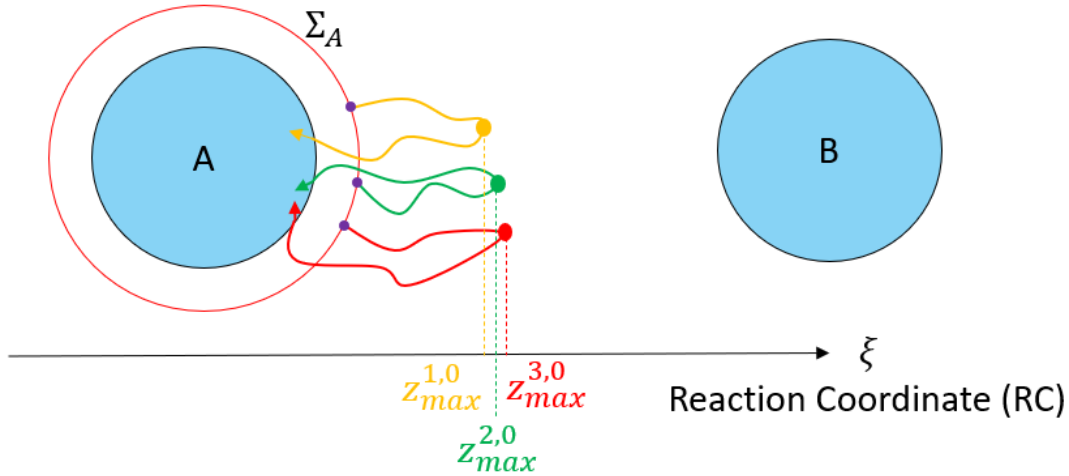
AMS is a multiple replicas approach designed to place the Σ_i's automatically so that $\mathbb{P}(\tau_{\Sigma_{i+1}} < \tau_A | \tau_{\Sigma_i} < \tau_A)$ is independent of i (to get minimal variance)
[Cérou, Guyader, Stoch. Anal. Appl. 2007]

1. Initial conditions for replicas are sampled on Σ_A by running MD (discretized Langevin dynamics)



Adaptive multilevel splitting (AMS)

2. Run N_{rep} replicas starting from the hitting points on Σ_A until A or B is reached

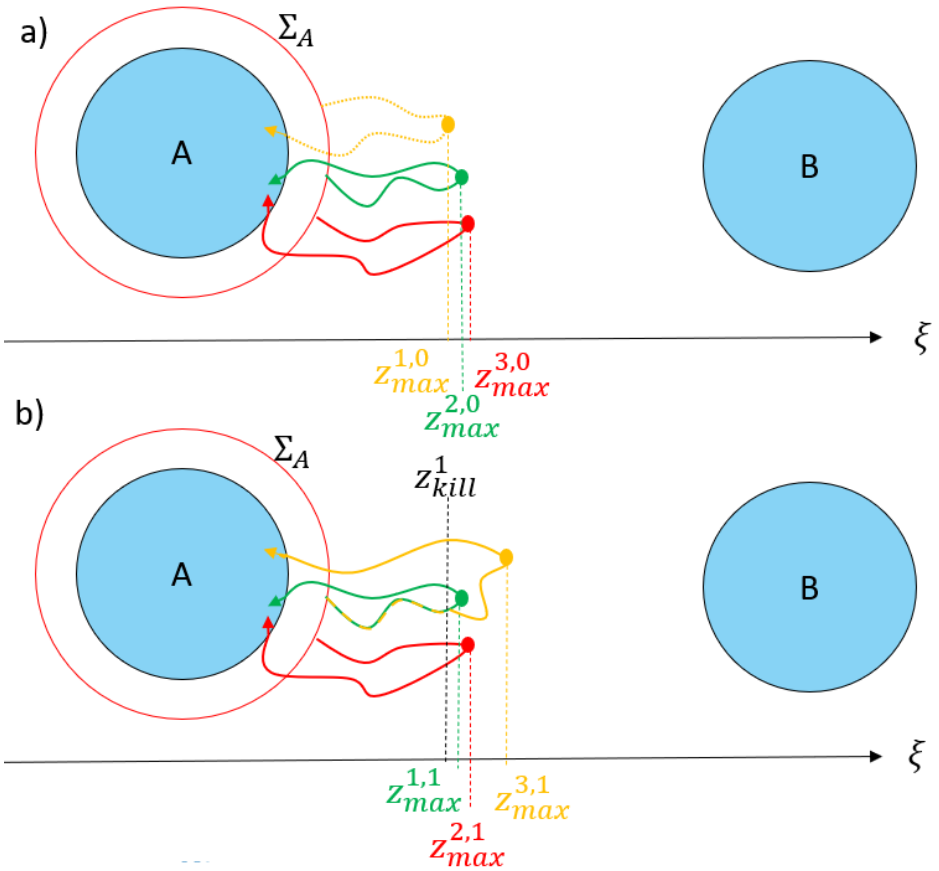


Rank the replicas using a 1D collective variable (CV) $\xi : \Omega \to \mathbb{R}$
The surfaces Σ_i will be level sets of ξ

Defining $\Sigma_1 = \left\{ \mathbf{q} \in \Omega | \xi(\mathbf{q}) = z_{\max}^{1,0} \right\}$, we have:

$$\mathbb{P}\left(\tau_{\Sigma_1} < \tau_A | \tau_{\Sigma_0} < \tau_A\right) \simeq \frac{N_{\mathrm{rep}} - 1}{N_{\mathrm{rep}}}$$

Adaptive multilevel splitting (AMS)



3. Delete the first replica, and replace it by branching randomly any of the remaining replicas from the killing level $z_{\max}^{1,0}$

Repeat step 3 until all replicas reach a level
$\{q, \xi(q) > \xi_{max}\} \supset B$

Adaptive multilevel splitting (AMS)

The estimator of

$$\mathbb{P}(\tau_B < \tau_A) = \mathbb{P}(\tau_{\Sigma_A} < \tau_A) \left(\prod_{i=0}^{j-1} \mathbb{P}(\tau_{\Sigma_{i+1}} < \tau_A | \tau_{\Sigma_i} < \tau_A) \right) \mathbb{P}(\tau_B < \tau_A | \tau_{\Sigma_j} < \tau_A)$$

is given by:

$$\hat{p}_{A \to \Sigma_A} \left(\frac{N_{\mathrm{rep}} - 1}{N_{\mathrm{rep}}} \right)^J \hat{p}_{\xi_{max} \to B}$$

where

- J is the (random) number of iterations to reach $\{q, \xi(q) > \xi_{max}\}$
- $\hat{p}_{A \to \Sigma_A}$ (resp. $\hat{p}_{\xi_{max} \to B}$) is a naive Monte Carlo estimator of $\mathbb{P}(\tau_{\Sigma_A} < \tau_A)$ (resp. $\mathbb{P}(\tau_B < \tau_A | \tau_{\xi_{max}} < \tau_A)$)

Adaptive multilevel splitting (AMS)

Unbiasedness

Theorem [Brehier, Gazeau, Goudenège, TL, Rousset, J. Appl. Probab. 2016]
The AMS estimator is unbiased whatever the choices of ξ and N_{rep}

Practical counterparts: the algorithm is easy to parallelize; one can check the results by changing ξ

Variance

The variance of the estimator depends on N_{rep} (asymptotically, $\propto N_{\mathrm{rep}}^{-1}$) and on ξ.
The optimal ξ is the committor function $p_{A\to B}$ defined by:

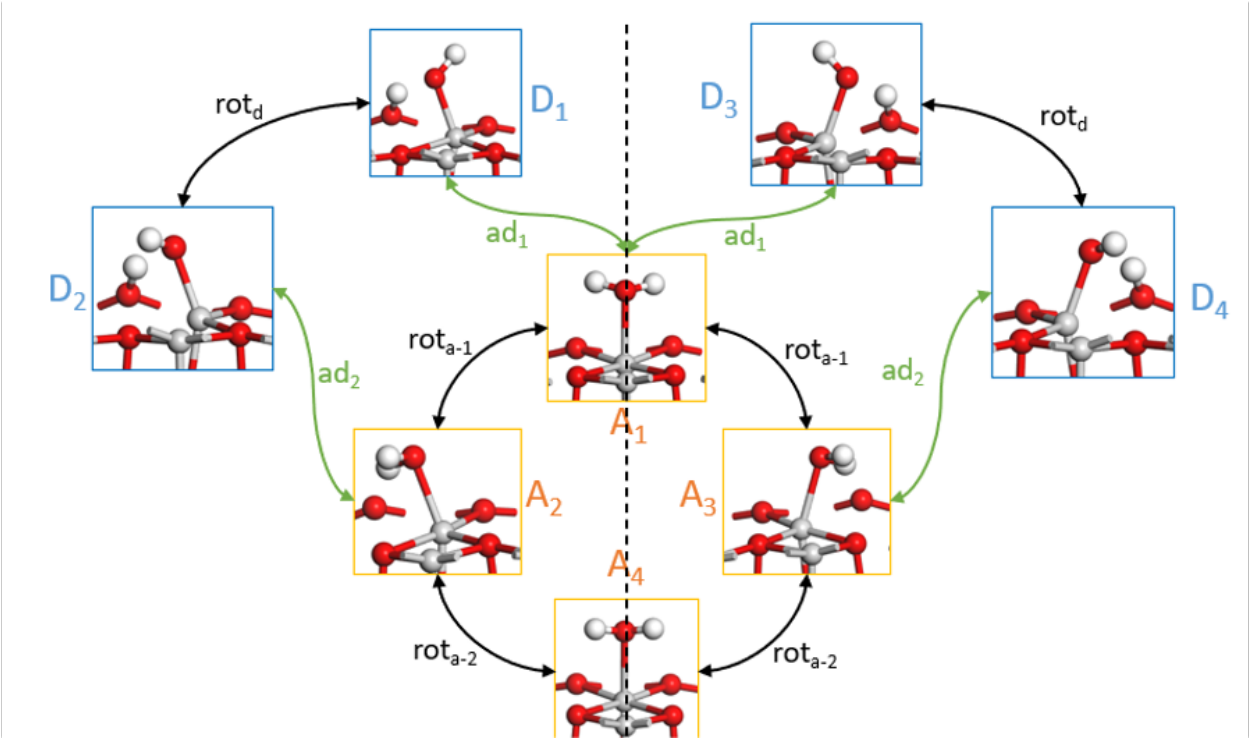
$$\forall \mathbf{q} \in \Omega, p_{A\to B}(\mathbf{q}) = \mathbb{P}_{\mathbf{q}}(\tau_B < \tau_A)$$

How to choose the collective variable ξ in practice? Can it be learnt "on the fly"?

Table of Contents

Dissociation of water on (100) surface

Metastable states of H_2O on the (100) surface of γ-alumina



This is a multistate problem

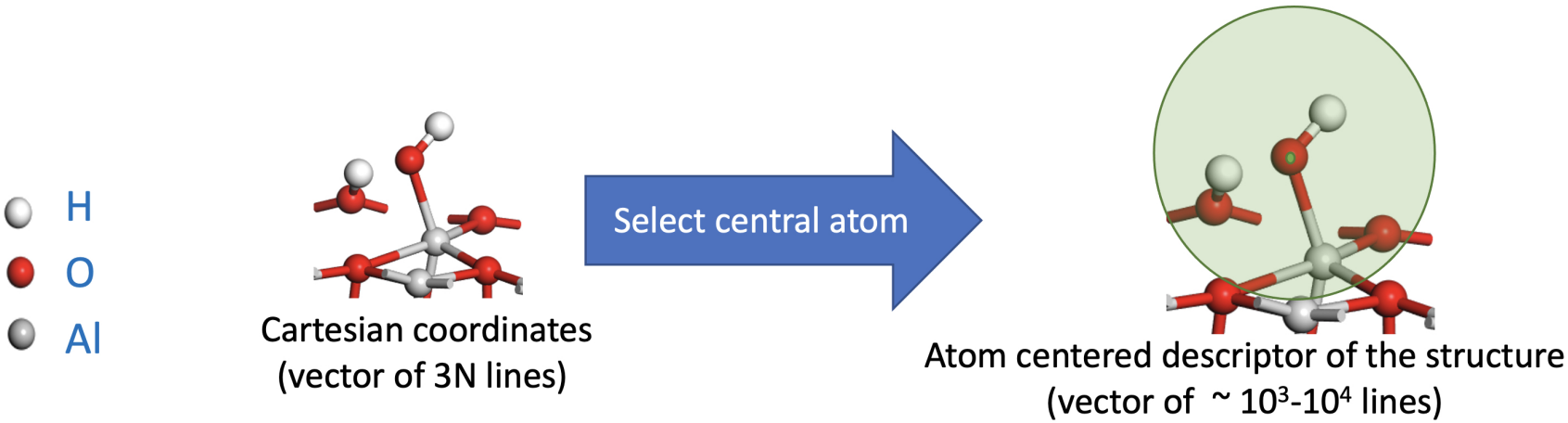
AMS can be used to analyze the transitions:

- Start from A_1, choose $R = A_1$ and $P = A_2A_3 \cup A_4 \cup D_1D_3 \cup D_2D_4$: What is the most probable exit from A_1?
- Start from A_1, choose $R = A_1 \cup A_2A_3 \cup A_4 \cup D_2D_4$ and $P = D_1D_3$: Focus on the $A_1 \to D_1D_3$ transition

Construction of the SOAP-SVM collective variable

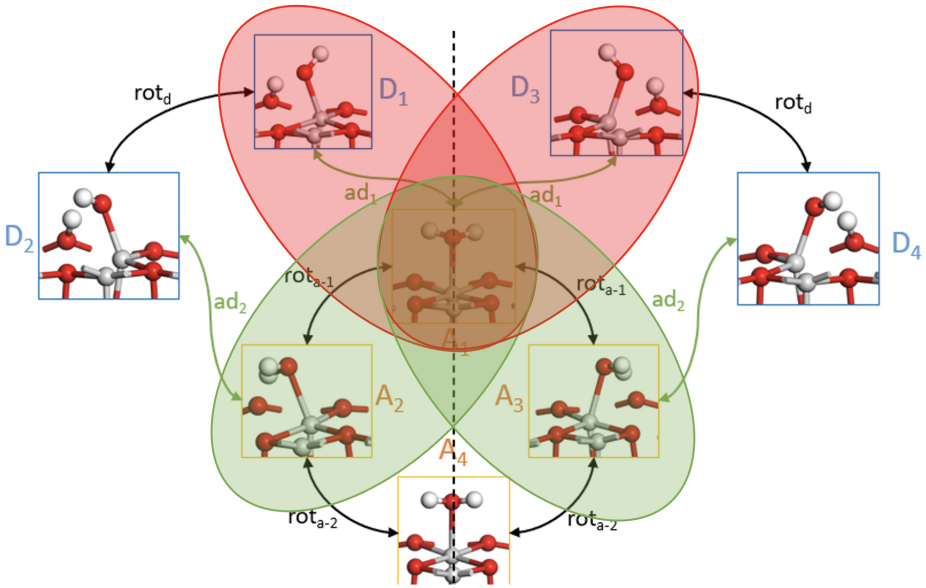
Two ingredients to build the CV:

- Smooth Overlap of Atomic Positions (SOAP) atom centered descriptors [Bartok, Kondor, Csanyi, Phys. Rev. B, 2013]
- Support-Vector Machine (SVM) to build a classifier decision function to separate the states (using short dynamics in each state)



Comparison with harmonic Transition State Theory

Here are the results obtained on two specific transitions (dissociation and rotation) [$\sim 2.\,10^6$ CPU hours for the whole reaction network]



Dissociation	Hill	hTST
$k_{A_1 \to D_1 D_3}$	$1.6\,10^9 s^{-1}$	$3.4\,10^{11} s^{-1}$
$k_{D_1 D_3 \to A_1}$	$2.3\,10^{10} s^{-1}$	$1.1\,10^{12} s^{-1}$

Rotation	Hill	hTST
$k_{A_1 \to A_2 A_3}$	$3.8\,10^{10} s^{-1}$	$7.6\,10^{10} s^{-1}$
$k_{A_2 A_3 \to A_1}$	$1.5\,10^{11} s^{-1}$	$2.1\,10^{12} s^{-1}$

hTST rates are larger (entropy estimation, recrossing)

[Pigeon, Stoltz, Corral-Valero, Anciaux-Sedrakian, Moreaud, TL, Raybaud, J. Chem. Theory Comput., 2023]

Table of Contents

Committor function

We consider for simplicity the overdamped Langevin dynamics ($\gamma \to +\infty$)

$$d\mathbf{q}_t = -\nabla V(\mathbf{q}_t)dt + \sqrt{\frac{2}{\beta}}d\mathbf{W}t$$

with infinitesimal generator:

$$\mathcal{L} = -\nabla V \cdot \nabla + \frac{1}{\beta}\Delta$$

The committor function

$$p_{A\to B}(\mathbf{q}) = \mathbb{P}_{\mathbf{q}}\left(\tau_B < \tau_A\right)$$

satisfies the PDE:

$$\begin{aligned}
&\forall \mathbf{q} \in \Omega\backslash(\overline{A} \cup \overline{B}), \quad \mathcal{L}p_{A\to B}(\mathbf{q}) = 0\\
&\forall \mathbf{q} \in \overline{A},\ p_{A\to B}(\mathbf{q}) = 0\\
&\forall \mathbf{q} \in \overline{B},\ p_{A\to B}(\mathbf{q}) = 1
\end{aligned}$$

Methods to learn the committor with neural networks

- Pointwise approximations [Frassek, Arjun, Bolhuis, J. Chem. Phys. 2021] [Lopes, TL, J. Comput. Chem., 2019]
- Variational formulation [Khoo, Lu, Ying, arXiv:1802.10275 2018] [Li, Lin, Ren, J. Chem. Phys. 2019] [Rotskoff, Vanden-Eijnden, arXiv:2008.06334, 2020]
- Fixed point on the transition operator [Strahan, Finkel, Dinner, Weare, J. Comput. Phys. 2023] [Li, Khoo, Ren, Ying, MSML 2022] [He, Chipot, Roux, J. Phys. Chem. Lett. 2022]

A new loss function

Ito formula leads to:

$$dp_{A\to B}(\mathbf{q}_t) = \mathcal{L}p_{A\to B}(\mathbf{q}_t)dt + \sqrt{\frac{2}{\beta}}\nabla p_{A\to B}(\mathbf{q}_t)\cdot d\mathbf{W}_t$$

Then, $\forall \mathbf{q}_0 \in \Omega\backslash(\overline{A}\cup\overline{B})$:

$$p_{A\to B}(\mathbf{q}_t)\mathbb{1}_{t<\tau_{\overline{A}\cup\overline{B}}} + \mathbb{1}_{\overline{B}}(\mathbf{q}_{\tau_{\overline{A}\cup\overline{B}}})\mathbb{1}_{t\geqslant\tau_{\overline{A}\cup\overline{B}}} - p_{A\to B}(\mathbf{q}_0) = \int_0^{t\wedge\tau_{\overline{A}\cup\overline{B}}} \sqrt{\frac{2}{\beta}}\nabla p_{A\to B}(\mathbf{q}_s)\cdot d\mathbf{W}_s$$

Thus the committor is solution to (for all $t>0$ and all μ):

$$\operatorname*{arginf}_f \int_{\Omega\backslash(\overline{A}\cup\overline{B})} \mathbb{E}\left[\left(f(\mathbf{q}_t)\mathbb{1}_{t<\tau_{\overline{A}\cup\overline{B}}} + \mathbb{1}_{\overline{B}}(\mathbf{q}_{\tau_{\overline{A}\cup\overline{B}}})\mathbb{1}_{t\geqslant\tau_{\overline{A}\cup\overline{B}}} - f(\mathbf{q}_0) - \int_0^{t\wedge\tau_{\overline{A}\cup\overline{B}}} \sqrt{\frac{2}{\beta}}\nabla f(\mathbf{q}_s)\cdot d\mathbf{W}_s\right)^2\right]\mu(d\mathbf{q}_0)$$

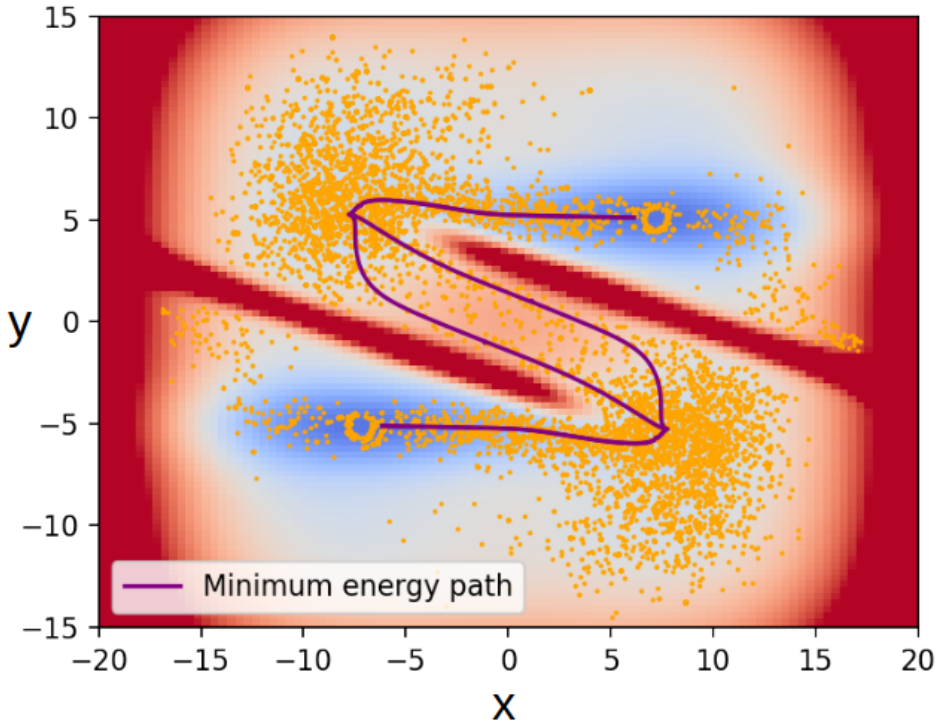
Iterative learning

The method consists in:

- Defining a first putative CV (approximation of the committor function)
- Running AMS (forward $A \to B$ and backward $B \to A$) with this CV ($p_{B \to A} = 1 - p_{A \to B}$)
- Updating the CV by minimizing the loss: the function f is approximated by a neural network
- Iterating, and enriching the training dataset, until the CV converges

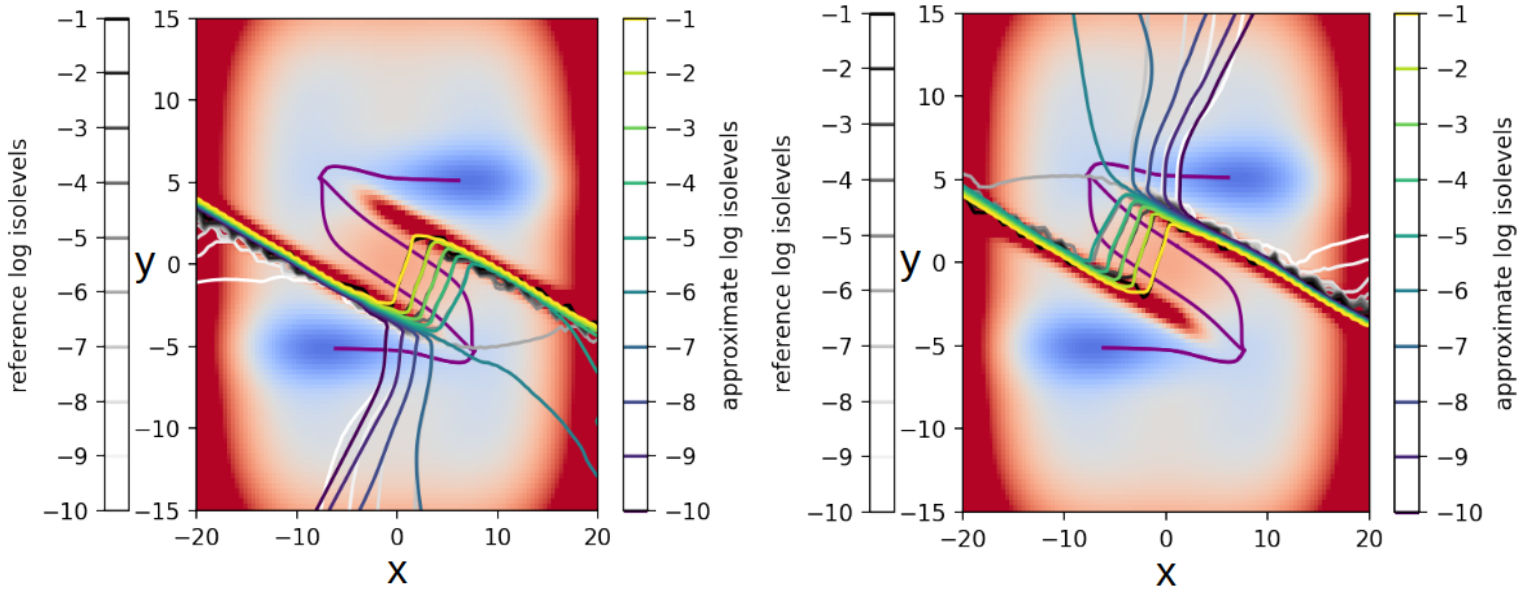
Illustration of the method

The Z potential



Feedforward model with architecture: (2, 20, 20, 1)
Optimizer: Adam, learning rate .001

Illustration of the method



Isolevels of the logarithm of the obtained approximation of the committor function
Results obtained after 5 iterations
(left: forward, right: backward)

Illustration of the method

CV	linear interpolant	NN committor	FE committor
Forward $p \pm \frac{1.96}{\sqrt{10}} \sigma_p$	$(3.39 \pm 6.58) \times 10^{-8}$	$(4.41 \pm 1.07) \times 10^{-7}$	$(5.58 \pm 1.02) \times 10^{-7}$
Backward $p \pm \frac{1.96}{\sqrt{10}} \sigma_p$	$(0.89 \pm 1.52) \times 10^{-8}$	$(6.57 \pm 1.48) \times 10^{-7}$	$(5.09 \pm 1.) \times 10^{-7}$

95% confidence interval of the transition probability obtained after 10 forward and backward AMS runs ($N_{\mathrm{rep}} = 100$) using various reaction coordinates.

NN committor = Neural network committor

FE committor = committor obtained by a finite element approximation

[Pigeon, Stoltz, TL, in preparation]

Conclusion & perspectives

- Promising approaches to use ML built CVs for rare event sampling methods
- On-going works to couple AMS sampling with active learning methods for ML built force-fields [T. Pigeon]
- Perspectives (PhD of P. Marmey, MAMABIO project of the PEPR B-BEST):
  (i) Applications to the dehydration of alcohols
  (ii) New algorithms combining importance sampling with AMS