

Exploring machine learned reaction coordinates in conjunction with rare events sampling methods in ab-initio molecular dynamics for catalytic reactions

T. Pigeon*^{1,2}, M. Corral-Valero², A. Anciaux-Sedrakian³, G. Stoltz¹, P. Raybaud², T. Lelièvre¹,

¹ CERMICS, Ecole des Ponts ParisTech, 6/8 av. Blaise Pascal - Citée Descartes, 77455 Marne-La-Vallée, FRANCE and Equipe-projet MATERIALS, Inria Paris, 2 rue Simone Iff, CS 42112 75589 Paris Cedex 12

² IFP Energies nouvelles, Rond-point de l'échangeur de Solaize, BP3, 69360 Solaize, France

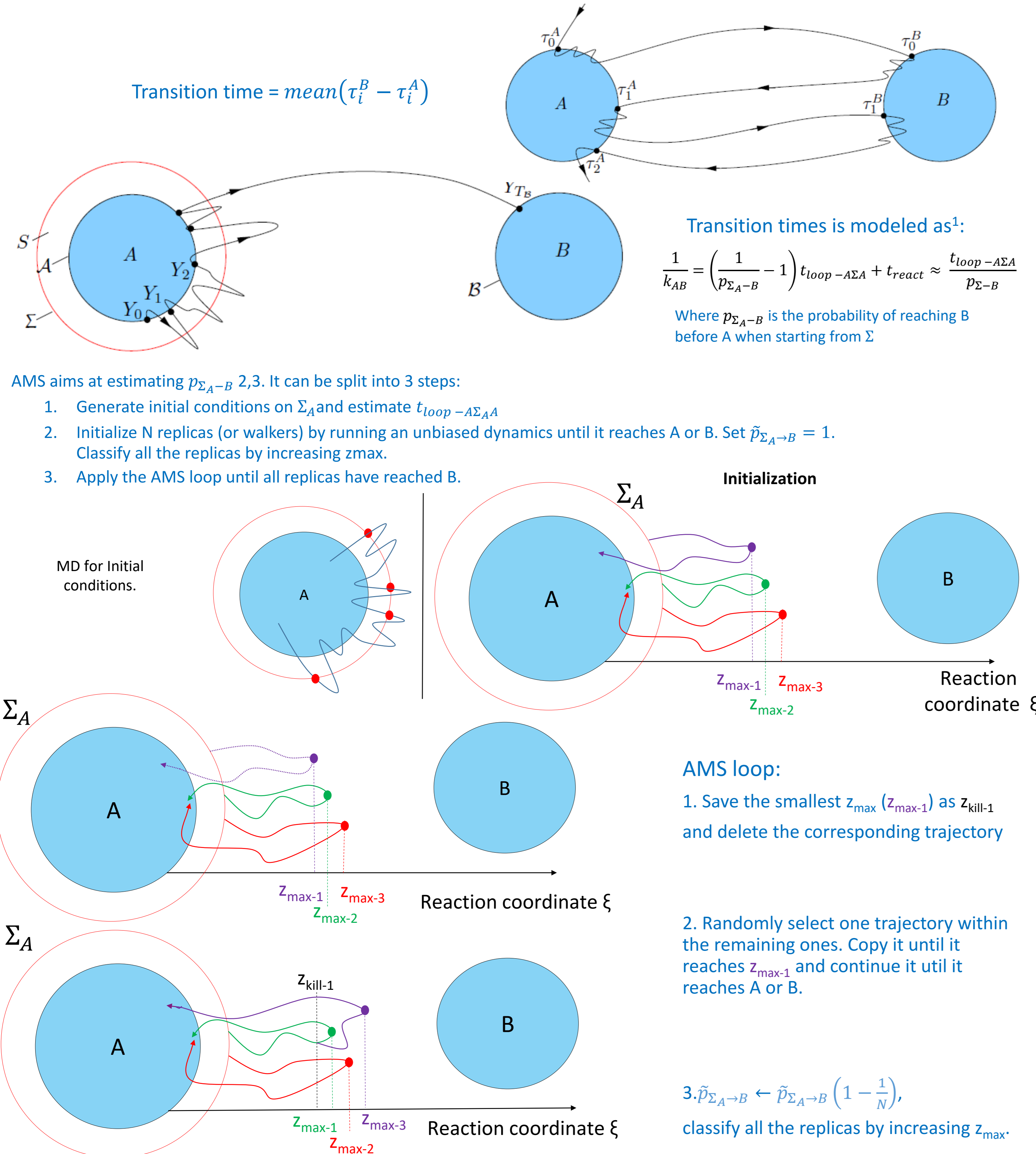
³ IFP Energies nouvelles, 1-4 Avenue du Bois Préau, 92852 Reuil-Malmaison, France

* thomas.pigeon@inria.fr

Introduction

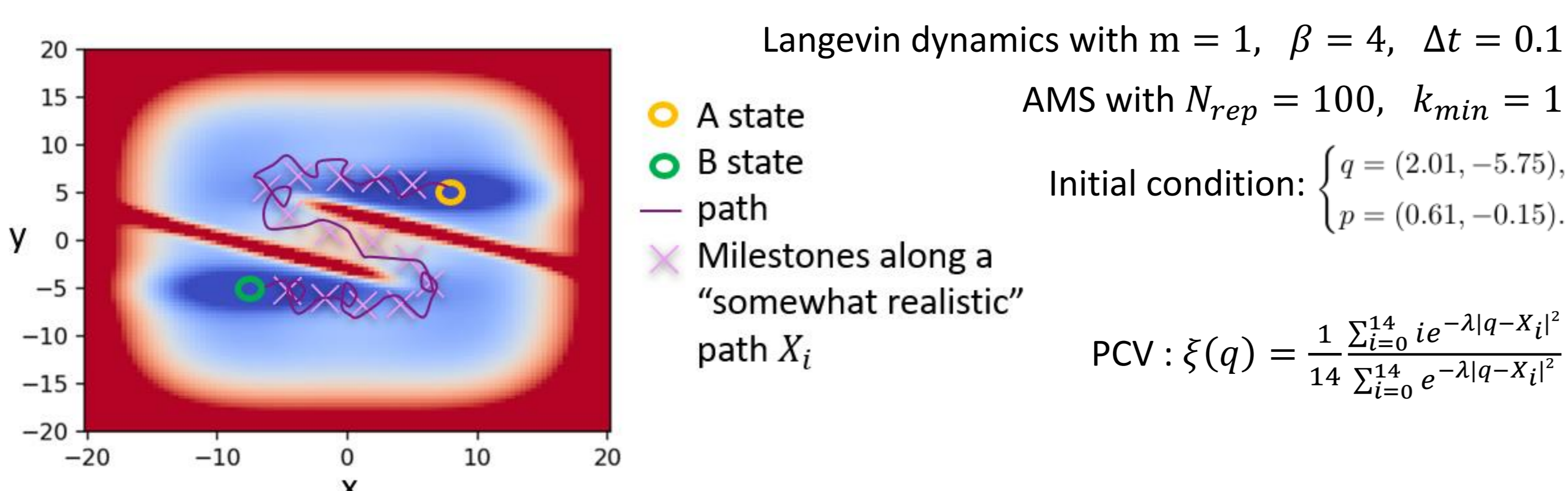
An elementary step of a chemical reaction links two metastable states. The transition mechanism and the time to perform the transition for each of these elementary chemical step is required to better understand the catalytic processes. Direct numerical simulations (DNS) are generally out of reach when using *ab initio* molecular dynamics (AIMD) because of their computational cost. Dedicated rare event simulation methods have therefore been devised to overcome this limitation. In this work, one of these methods called Adaptive Multilevel Splitting (AMS)^{1,2} is studied, combined with *ab initio* molecular dynamics (AIMD). As many other rare event methods, it requires some information on the progress of the transition, quantified here using a one-dimensional Reaction Coordinate (RC). Determining such a RC challenges chemical intuition for complex systems, which is why we turn instead to machine learning approaches. This approach is applied to explore the reactivity of water on alumina surface as a case study.

Adaptive Multi-level Splitting (AMS) algorithm



AMS behavior on a two-dimensional case

Estimation of the probability of reaching B before A given a fixed initial condition given that $\tilde{p}_{\Sigma \rightarrow B} \approx 10^{-5}$.



	DNS	AMS $\xi(x, y) = x$	AMS $\xi(x, y) = y$	AMS PCV
FM	3.14×10^1	1.57×10^0	6.41×10^1	6.45×10^2

Figure of merit: $FM(X) = \frac{1}{c \text{Var}(X)}$
 c: computational cost per evaluation

AMS algorithm is designed to estimate low probabilities and in the given case, the target probability is such that the efficiency of AMS compared to DNS strongly depends on the reaction coordinate.

References:

- 1 M. Baudel, A. Guyader, T. Lelièvre, *arXiv preprint*, (2020) arXiv:2008.0979
- 2 F. Cérou, A. Guyader, *Stoch. Anal. Appl.* **25**, (2007) 417.
- 3 L. J. S. Lopes, T. Lelièvre, *J. Comput. Chem.* **40**, (2019) 1198.
- 4 M. Digne, P. Sautet, P. Raybaud, P. Euzen, H. Toulhoat, *J. Catal.*, **211**(1), (2002) 1.
- 5 G. Kresse, J. Hafner, *Phys. Rev. B*, **47**, (1993) 558.
- 6 S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **132**, (2010) 154104.

Learning a CV for a more realistic case: dissociation of water on γ -alumina

Illustration with our "toy model": Water dissociation on a (100) γ -Al₂O₃ surface.⁴ This material is a widely used catalytic support. It is active in presence of water through its Brønsted and Lewis acidities.

The molecule can be either dissociated (D_i states) or associated (A_i). Rotation (r) and transitions between different A_i ↔ D_j states are possible.

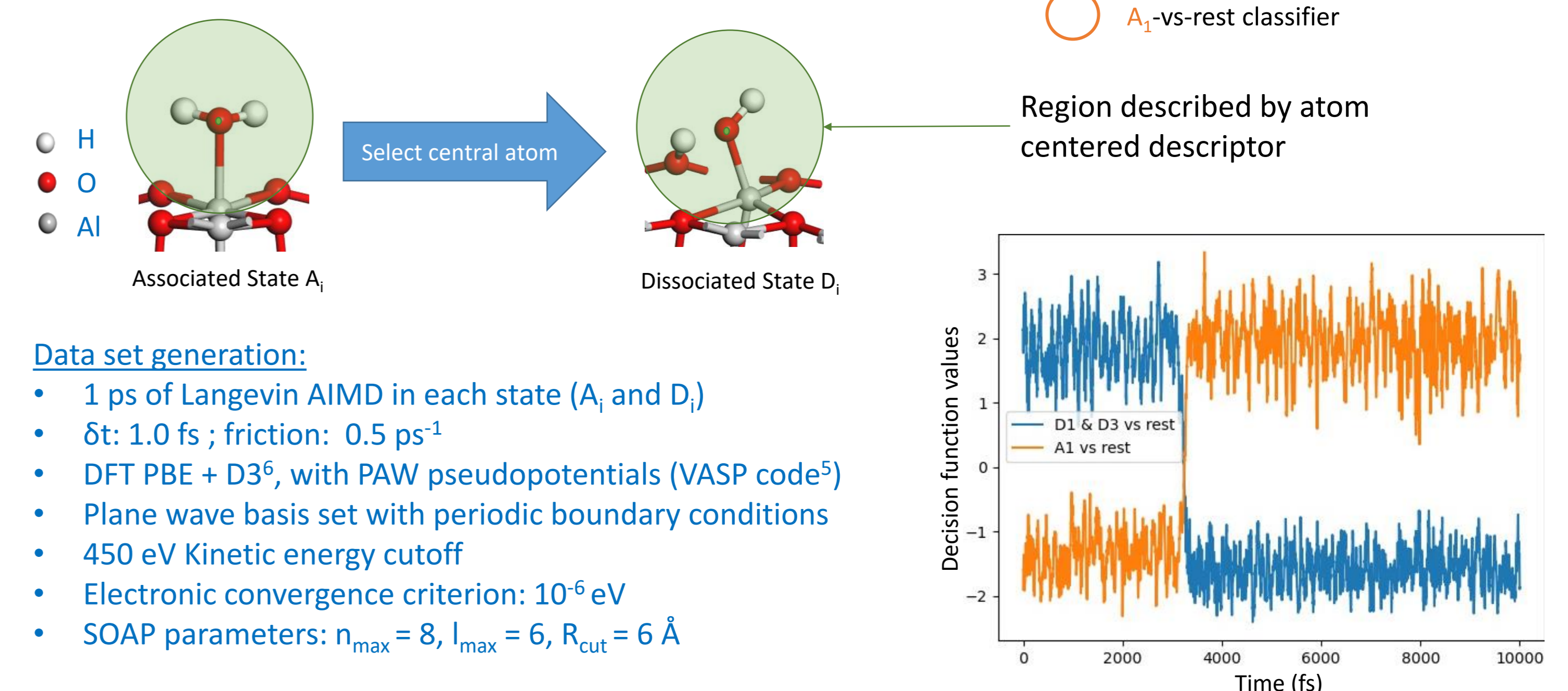
Method:

1. Identify the different metastable states (intermediates)
2. Run short dynamics in these states to gather structures
3. Train machine learning model with the proper labelling

Focus on 1D Reaction Coordinates which requires to be invariant under:

- System symmetries
- Rotation
- Atom indices permutation

This is enforced by using the Smooth Overlap of Atomic Positions (SOAP) atom centered descriptor.



Data set generation:

- 1 ps of Langevin AIMD in each state (A_i and D_j)
- δt : 1.0 fs ; friction: 0.5 ps⁻¹
- DFT PBE + D3⁶, with PAW pseudopotentials (VASP code⁵)
- Plane wave basis set with periodic boundary conditions
- 450 eV Kinetic energy cutoff
- Electronic convergence criterion: 10⁻⁶ eV
- SOAP parameters: $n_{max} = 8$, $l_{max} = 6$, $R_{cut} = 6 \text{ \AA}$

Reaction Coordinate : 1-vs-all Support Vector Machines (SVM) classifiers decision function. Σ defined through these RCs.

Using AMS and these RCs allows to estimate the probabilities:

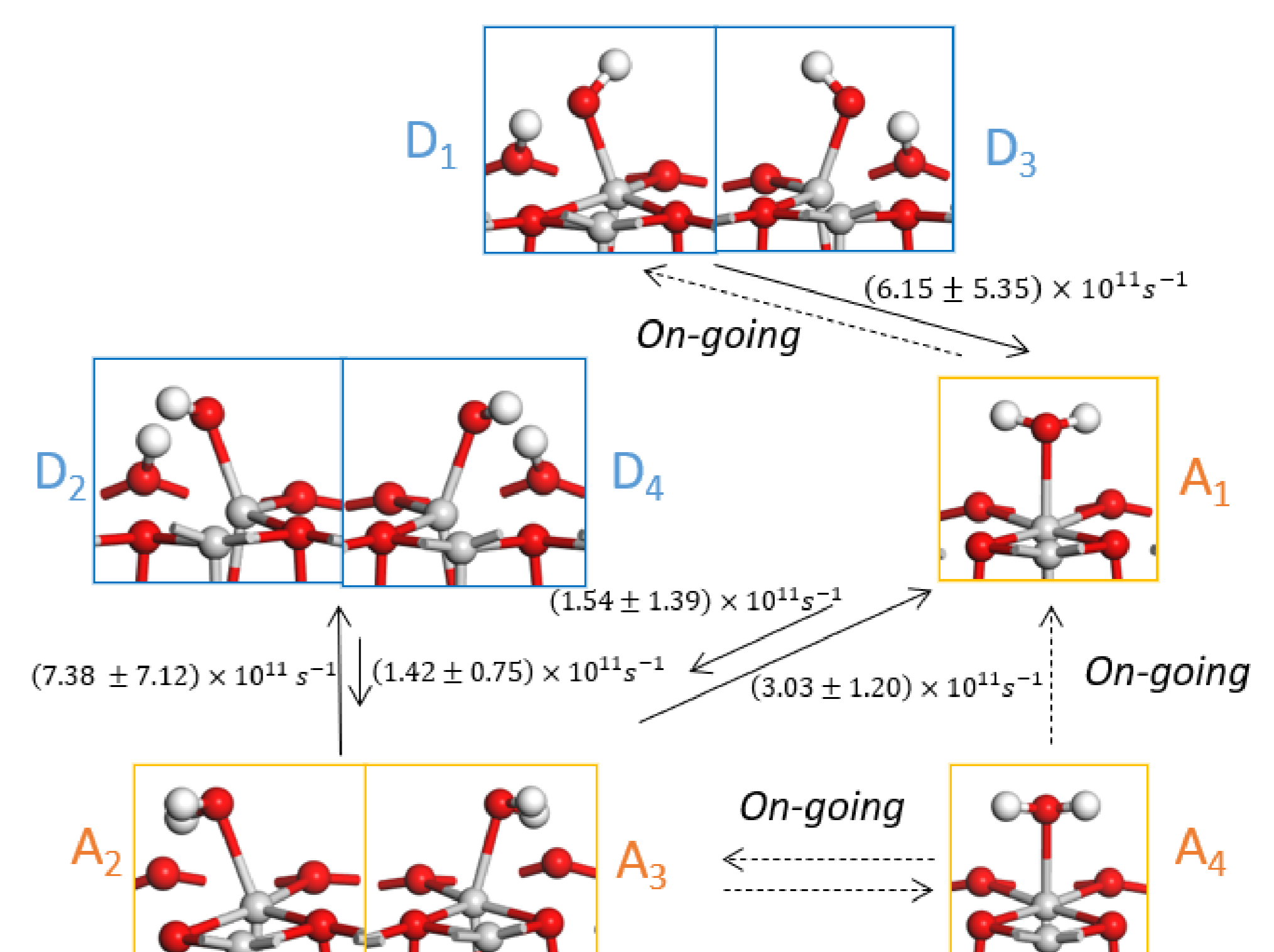
$\tilde{p}_{\Sigma A_1 \rightarrow any}$, $\tilde{p}_{\Sigma A_2 A_3 \rightarrow any}$, $\tilde{p}_{\Sigma A_4 \rightarrow any}$, $\tilde{p}_{\Sigma D_1 D_3 \rightarrow any}$ and $\tilde{p}_{\Sigma D_2 D_4 \rightarrow any}$

State to state probability are computed given the number of replicas n_X^{in} finishing in the state and:

$$\tilde{p}_{\Sigma X \rightarrow Y} = \frac{n_X^{in}}{N} \tilde{p}_{\Sigma X \rightarrow any}$$

Transition rates are finally computed:

$$\tilde{k}_{X \rightarrow Y} \approx \frac{\tilde{p}_{\Sigma X \rightarrow Y}}{t_{loop - X \Sigma X}}$$



Conclusion and perspectives

The SOAP-SVM Reaction Coordinates used in this case study has proven to be able to successfully estimate rates with AMS. A transition which was not anticipated intuitively was observed (A₄ → A₁). The paths observed will be used to build path collective variables to better characterize the rates and obtain a better sampling of the transition path. Another perspective is to use the reactive trajectories to build some reaction coordinate with some auto-encoders models.