Computing Surface Reaction Rates by Adaptive Multilevel Splitting Combined with Machine Learning and Ab Initio Molecular Dynamics

Thomas Pigeon, Gabriel Stoltz, Manuel Corral-Valero, Ani Anciaux-Sedrakian, Maxime Moreaud, Tony Lelièvre, Pascal Raybaud

Computing Surface Reaction Rates by Adaptive Multilevel Splitting Combined with Machine Learning and Ab Initio Molecular Dynamics

Thomas Pigeon,[*,†,‡,¶] Gabriel Stoltz,[‡,†] Manuel Corral-Valero,[¶] Ani Anciaux-Sedrakian,[§] Maxime Moreaud,[¶] Tony Lelièvre,[*,‡,†] and Pascal Raybaud[*,¶]

†*MATHERIALS team-project, Inria Paris, 2 Rue Simone Iff, 75012 Paris, France*

‡*CERMICS, École des Ponts ParisTech, 6-8 Avenue Blaise Pascal, 77455,Marne-la-Vallée, France*

¶*IFP Energies Nouvelles, Rond-Point de l'Echangeur de Solaize, BP 3, 69360 Solaize, France*

§*IFP Energies Nouvelles, 1 et 4 avenue de Bois-Préau, F-92852 Rueil-Malmaison Cedex, France*

E-mail: thomas.pigeon@inria.fr; tony.lelievre@enpc.fr; raybaud@ifpen.fr

Abstract

Computing accurate rate constants for catalytic events occurring at the surface of a given material represents a challenging task with multiple potential applications in chemistry. To address this question, we propose an approach based on a combination of the rare event sampling method called Adaptive Multilevel Splitting (AMS) and ab initio molecular dynamics (AIMD). The AMS method requires a one dimensional reaction coordinate to index the progress of the transition. Identifying a good reaction coordinate is difficult, especially for high dimensional problems such a those encountered in catalysis. We probe various approaches to build reaction coordinates such as Support Vector Machine and path collective variables. The AMS is implemented so as to communicate with a DFT-plane wave code. A relevant case study in catalysis: the change of conformation and the dissociation of a water molecule chemisorbed on the (100) γ-alumina surface is used to evaluate our approach. The calculated rate constants and transition mechanisms are discussed and compared to those obtained by a conventional static approach based on the Eyring-Polanyi equation with harmonic approximation. It is revealed that the AMS method may provide rate constants which are smaller than the static approach by up to two orders of magnitude due to entropic effects involved in the chemisorbed water.

1 Introduction

The determination of chemical reaction rate constants is of tremendous importance to better understand and quantify the kinetics of molecular transformations. This can be a challenging task, especially in catalysis where multiple elementary steps are involved for one targeted reaction. Evaluating each of them by experimental methods being often out of reach, an alternative lies in the theoretical modeling of each of them. Thanks to the significant increase of computational resources, quantum simulation approaches are widely used nowadays to address numerous catalytic systems involved in petrochemistry, fine chemistry and biomass conversion.[1–4]

However, at the simulation time scale, such chemical transformations are rare events. The typical time step for the integration of stochastic dynamics modeling the evolution of the system is of the order of 10^{-15} s, while the frequency of most of the chemical reactions of interest is, at least, several order of magnitude higher. Moreover, to accurately simulate catalytic activation of chemical bond breaking and formation, the simulation must include the explicit treatment of valence electrons and the quantum chemical calculation of the Hellmann-Feynman forces for each step of the dynamics.[5] Such an *ab initio* molecular dynamics (AIMD) approach becomes so computationally demanding that it is generally impossible to simulate a trajectory that is long enough to observe multiple reaction events, allowing the accurate quantification of rate constants.

Theoretical approaches most commonly used to explore chemical transformations are based on transition state theory (TST).[6] Within this formalism, the reactant and product are considered to be separated in phase space by a dynamical bottleneck,[7] which can be characterized as a surface in the configuration space. For a reaction with only one reactive path and only one energy barrier to cross, assuming momenta are not relevant for the transition process, this surface should contain the first order saddle points. The term transition state (TS) is versatile as sometimes it refers to a first order saddle point and sometimes to a isocommittor surface, as defined by IUPAC.[8] Considering TS to be surfaces, the reaction rate

can be approximated as the frequency at which this surface is crossed. The most common approach to compute the reaction rate constant is called harmonic TST (hTST) as it allows to reduce general TST expression into the "generalized" Eyring–Polanyi equation thanks to harmonic approximation of the potential energy surface.[6,9–11]

$$k_{\mathrm{hTST}} = \kappa(T)\frac{k_{\mathrm{B}}T}{h}\mathrm{e}^{-\frac{\Delta G^{\ddagger}}{k_{\mathrm{B}}T}}, \tag{1}$$

where $\Delta G^{\ddagger}$ is the free energy of activation computed as the difference of the free energy of the metastable basin and of the transition state, k_{B} the Boltzmann constant, h the Planck constant, T the temperature and $\kappa(T)$ the transmission coefficient. This last quantity has to be between 0 and 1 and accounts for the recrossing of the surface, as discussed later on. The free energy of activation is approximated via an harmonic approximation around the saddle point and the minima. Although hTST is one of the most widely used method to determine activation free energies and the rate constants of chemical events, particularly catalytic ones, it suffers from some weaknesses. Among them, the harmonic approximation of the potential energy surfaces as well as the determination of the prefactor κ in (1) might be questionable. In general when the entropy of the metastable state and the transition state differ by a non negligible amount, the harmonic approximation can lead to significant errors. This can occur in various systems of interest to catalysis such as solid–liquid interfaces, zeolites, porous solids and supported nano-particles.[12] More general expression for the TST rate, using a one dimensional reaction coordinate[7,13] and relying on sampling methods to estimate free energies,[14–16] were proposed to overcome some limitations of (1). However, TST reaction rates contain a transmission coefficient $\kappa \in (0, 1]$, accounting for the recrossing of the transition state surface, which is rather difficult to evaluate and which explains why bare TST overestimates the transition rate.[7,11,13,17–19]

There are of course alternative approaches to TST. A first one is based on the evolution of a time correlation function[13,20] which found applications in Transition Path Sampling

(TPS)[21] or other approaches such as the recent work relying on Onsager–Machlup path probability distribution of Ref. 22. Another alternative, that we will use in the present work, is provided by approaches based on the Hill relation:[23]

$$k_{\text{Hill}} = \Phi_R p_{R\to P}(\partial R), \tag{2}$$

where Φ_R is the flux of trajectories leaving the reactant state R and the committor probability at the boundary $p_{R\to P}(\partial R)$, the probability of reaching the product state P before returning to R starting from the boundary ∂R. In other words, this relation states that the rate constant is the average rate at which the system attempts to leave the initial state times the probability of success. Relation (2) has been proven correct assuming that the reactant state R is metastable for systems evolving according to the overdamped Langevin dynamics[24] or Langevin dynamics.[25] The Hill relation is used in various approaches corresponding to so-called path sampling methods such as Transition Interface Sampling (TIS),[26] Forward Flux Sampling (FFS),[27] Weighted Ensembles (WE)[28] and Adaptive Multi-level Splitting (AMS).[29] All these methods are designed to compute the probability $p_{R\to P}(\partial R)$, which is the most difficult object to evaluate in (2). As a side product, these methods sample some reactive trajectories.

All the methods described previously have different precisions and computational efficiencies. On the one hand, the hTST approach is, by far, the most inexpensive methodology in terms of computational resources but, as mentioned above, it may lead to significant errors. On the other hand, the computational cost of enhanced sampling methods to estimate free energies is not negligible. The Hill relation has the advantage of being exact compared to TST approach but the required computational cost can be high depending on the method to compute the probability $p_{R\to P}(\partial R)$. Moreover, numerical methods sampling reactive trajectories offer the possibility of performing a more detailed analysis of reaction mechanisms.

Most methods to compute reaction rate constants require the definition of a Collective Variable (CV), either to define the states of the system, its free energy, or to use it as a one dimensional Reaction Coordinate (RC) indexing the progress of the transition. In many situations, reactions go through one or a few channels in phase space. CVs should describe these channels with a minimal number of dimensions. Usually, CVs are defined thanks to chemical intuition or through the expert knowledge of the chemical system. They are typically based on key distances or angles associated with atoms central to the reaction mechanism. Nonetheless, this kind of heuristic approach can have some limitations especially when the studied mechanism is *a priori* unknown. Automatic and data based approaches using various Machine Learning (ML) methods currently offer very appealing perspectives in this context. Recent reviews[30–33] provide an overview of current options to propose CVs and discuss their advantages and drawbacks. These methods bear the promise of more systematic and efficient ways to define CVs, albeit at the expense of interpretability compared to intuitive CVs such as angles or distances. Nonetheless, machine-learned CVs are becoming common practice in the field. For example, Support Vector Machine (SVM) models trained on a set of data generated by molecular dynamics were used for exploring the configurational transitions of model protein molecules.[34] In material sciences, the combination of SVM and AIMD was used for the mechanistic study of the diffusion of Al atoms on Al (100) surface.[35] To the best of our knowledge, SVM has not been used to explore more complex reactive events, such as chemical bond breaking/formation catalyzed by an oxide material's surface such as proposed in the present work.

To benchmark an innovative methodology based on the Hill relation for exploring reaction mechanisms occurring on catalytic materials, we chose in this work a relevant case study: the reactivity of water on the (100) orientation of $\gamma-$alumina, a widely used support in heterogeneous catalysis applied to biomass conversion.[36,37] Comprehensive DFT based studies have revealed the versatile nature of active sites (Lewis Al and Bronsted Al-OH), their thermodynamic properties[38–42] and their kinetic ones (TS and activation barriers) by us-

ing predominantly hTST calculations.[36,37,43,44] As for the study of many chemical reactions, especially in catalysis, most of the reaction rate constants are computed within the TST framework.[4] Unbiased AIMD simulations have been also applied to decipher the gamma-alumina's reactivity, its local structure and spectroscopic features, in the presence of liquid water in order to obtain a better understanding of phenomena occurring during the catalyst preparation or catalytic reaction.[45,46] TPS was used in particular for studying the catalytic reactivity of other oxide materials,[47,48] also in combination with the blue-moon ensemble formalism.[49]

Methods based on the Hill relation and rare event simulation methods are rarely used for studying chemical reactions[50] and to the best of our knowledge, they have never been used to describe reactions in heterogeneous catalysis. In particular, the AMS method has only been used for molecular dynamics applications to study the isomerization of small biomolecules[51] or a protein-ligand dissociation,[52] up to now.

Hence, the aim of the present work is to highlight how AMS applied to AIMD rare event sampling, combined with ML approach, is able to compute reaction rate constants via the Hill relation in a relevant case study for heterogeneous catalysis. The CVs and RCs are built using SVM or Path Collective Variables and well-chosen chemical descriptors.[53]

Considering the challenge of the chemical reactivity of the alumina catalysts highlighted before, we will aim at determining rate constants for a reaction network involving various water rotation, dissociation and association events on the (100) γ-alumina surface.

This article is organized as follows. In the methods section, the general computational approach following a flowchart leading to the determination of rate constants is described. First, it is presented how the implementation of AMS coupled to a reference plane wave-DFT software enables the determination of rate constants. In a second part, numerical tools such as SVM and path collective variables (PCV) used to define CVs and RCs are presented. The results section first describes the catalytic model system of water activated on the γ-alumina surface, used to probe the theoretical approach. The constructions of CVs and RCs

corresponding to the water molecule transformation path are then explained. Finally, the numerical values of reaction rates and the reactive trajectories are analyzed and compared with the standard hTST approach.

2 Methods

The general flowchart of our approach is given in Figure 1. The first step is the definition of states defined as the ensemble of structures in the vicinity of a local potential energy minimum characterizing either a reactant or a product. In practice, these configurations are sampled by running a short AIMD starting from minima identified on the Potential Energy Surface (PES). Then, using this trajectory, the function numerically defining states is obtained by SVM and well-chosen chemical descriptors. Depending on the reaction rate constant to compute, each state has to be labeled as reactant or product. A reaction coordinate (RC) is then built, for instance by using the decision functions of the classifiers previously used to define states. Once states and a RC are defined, AMS is run to obtain an estimate of the reaction rate constant of the Langevin dynamics which is assumed to model accurately the system dynamics.

As a side remark, the definition of states and of the RC can be done using any type of collective variables describing the configurations of the system. In the present approach, the chosen collective variables are the 2100-dimensional SOAP descriptor of a selected atomic environment. The role of the SVM is to build a linear combination of these collective variables.

2.1 Reaction rate constant estimation using AMS

Motivation. To compute rate constants of rare events by using the Hill relation (2), the flux of trajectories leaving the initial reactant state R (or the frequency at which trajectories leave R) must be evaluated. If the reactant state is properly defined, this quantity can be

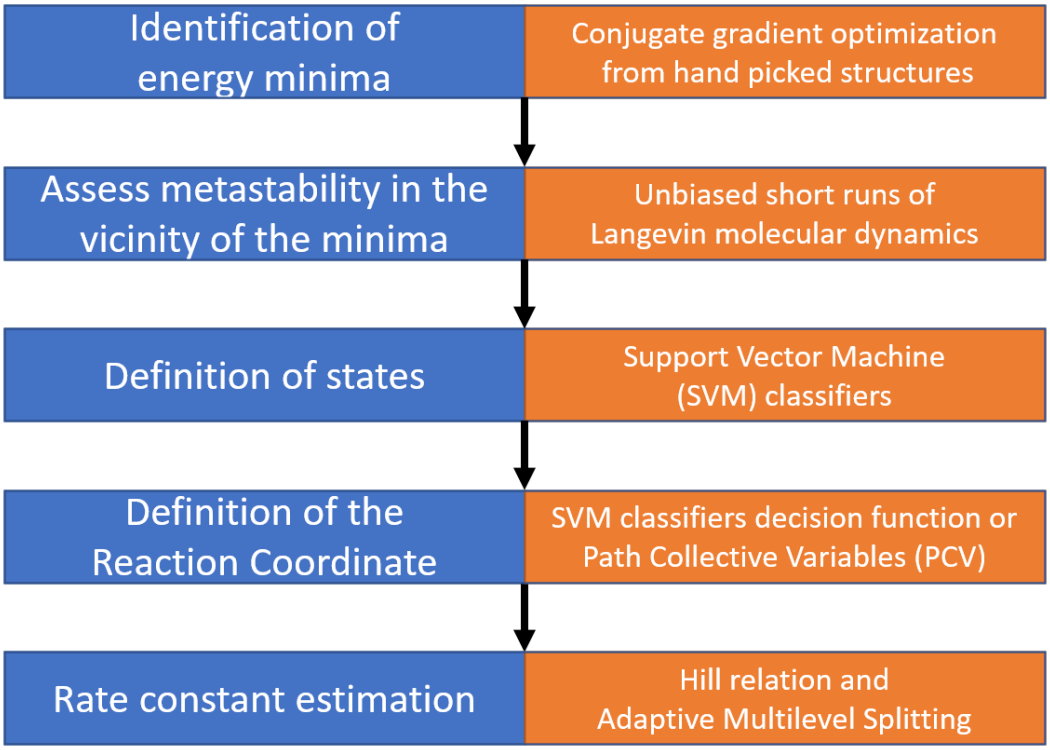


Figure 1: Global workflow to compute reaction rate constants With the Hill relation using Adaptive Multilevel Splitting and Support Vector Machine.

computed in a reasonably short time by unbiased MD. The difficulty lies in the estimation of the probability that a trajectory leaving R is reactive (i.e. goes to a product state P), since the probability $p_{R\to P}(\partial R)$ is in most cases exceedingly small. The AMS algorithm is specifically designed to evaluate low probability events.[29] The key point of AMS is to propose a method that has a good behavior in terms of variance and computational efficiency to compute the probability $p_{R\to P}(\partial R)$. This is achieved by first decomposing the rare event of interest into a succession of less unlikely events, the target probability to estimate being the product of the conditional probabilities associated with the sub-events (see SI Section 1). Moreover, the sub-events are built such that the associated conditional probabilities are all the same. This is indeed a desirable feature in order to reduce the overall variance of the estimator.[51] The mathematical analysis of the variance of the AMS estimator is provided in Ref. 29 and 54. We focus here on the presentation of the algorithm adapted to MD rare events and only mention that this algorithm is unbiased.[55] This means that, whatever the choice

of the reaction coordinate ξ and the number of replicas of the system (see below), repeating the algorithm sufficiently many times will always provide the same result in average, and this average value coincide with the target probability. On the other hand, the variance of the probability estimator depends on the quality of ξ. This opens a way to define an iterative procedure to improve the definition of reaction coordinates, using the sampled reactive trajectories to define better reaction coordinates.

Computing the flux and sampling initial conditions. A separating surface Σ_R close to R is introduced for the estimation of the flux Φ_R, to determine actual exits out of R.[24] This surface has to enclose the reactant state, so that any trajectory going from R to P has to cross Σ_R (see Fig. 2). Indeed, the location of this surface allows to select the trajectories that make actual excursions off the state R, in contrast to trajectories that would only wander out of R for a few steps and go back inside R right away. The approach to define the position of this surface is discussed precisely below. The flux Φ_R is then evaluated by starting a dynamics in the state R, counting the number of times $n_{\mathrm{loop}-R\Sigma_R R}$ it goes from R to Σ_R, crosses Σ_R and goes back to R, dividing this number by the overall time t_{tot}:

$$\Phi_R = \frac{n_{\mathrm{loop}-R\Sigma_R R}}{t_{\mathrm{tot}}} = \frac{1}{t_{\mathrm{loop}-R\Sigma_R R}}, \tag{3}$$

where $t_{\mathrm{loop}-R\Sigma_R R}$ is the average time that a trajectory takes to go out of R, cross Σ_R and go back to R. Now that the calculation of the first term Φ_R has been discussed, let us focus on the second one: $p_{R\to P}(\partial R)$. Computation of the flux Φ_R simultaneously allows to generate some positions on the surface Σ_R, which will serve to estimate the probability $p_{R\to P}(\Sigma_R)$. Indeed, the estimated quantity is $p_{R\to P}(\Sigma_R)$ instead of $p_{R\to P}(\partial R)$, this does not bias the result as far as R and Σ_R are within the same metastable basin.[24] These initial conditions must correspond to the first time a trajectory leaving R reaches the level Σ_R. As efficient calculation of the flux and the sampling of initial conditions relies on parallelization strategies, a Fleming–Viot particle process is used in our implementation of this initialization

procedure.[56] The particles undergo independent molecular dynamics which means they can be run in parallel without requiring frequent communications.

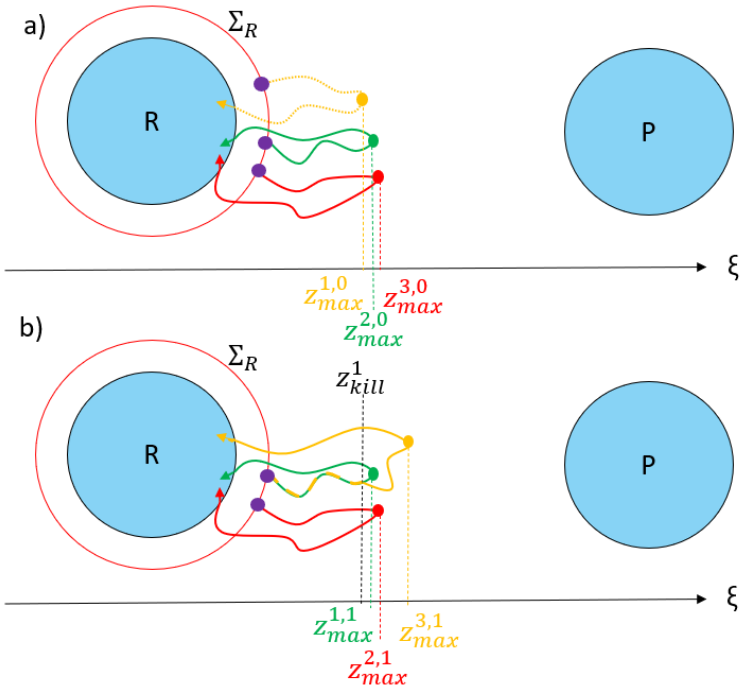


Figure 2: First iteration of the AMS algorithm with $k_{\min} = 1$ and $N_{\text{rep}} = 3$. Purple points represent the initial conditions on Σ_R. **a)** Identify the kill level $z^1_{\text{kill}} = z^{k_{\min},0}_{\max}$ and kill the replicas such that $z^{i,0}_{\max} \leq z^1_{\text{kill}}$, i.e. the orange replica. **b)** Replace the killed replicas by the trajectory of one of the remaining replicas (the green one in this example) until the level z^1_{kill} and continue the trajectory of the replica until it reaches either the state R or the state P.

AMS Requirements. To run an AMS estimation, the reactant state R and the product state P have to be defined. The surface Σ_R has to be placed such that each trajectory linking the reactant and the product state goes through Σ_R. Its distance to the boundary of R should be sufficiently small so that the sampling of initial conditions and the determination of the flux Φ_R (see the previous paragraph) is not exceedingly expensive in terms of computational cost. A number of replicas N_{rep} (or walkers) has to be defined, as well as a minimum number $k_{\min}$ of replicas to kill at each iteration of AMS. N_{rep} different initial conditions on the surface Σ_R are selected uniformly among the initial conditions sampled following the procedure described in the previous paragraph (purple points in Figure 2). Finally a

reaction coordinate ξ should be defined to index the progression along the $R \to P$ transition. It has to be consistent with the states R and P which can be generally enforced by setting $\xi(\mathbf{q}) = -\infty$ for $\mathbf{q} \in R$ and $\xi(\mathbf{q}) = +\infty$ for $\mathbf{q} \in P$.

AMS initialization. First, all the replicas are run from their initial conditions on Σ_R until either the R or P state is reached (see Fig. 2 **a)** which depicts an initialised set of three replicas). They are then iteratively updated until they all finish in the product state P. An illustration of an iteration is provided in Figure 2, the process being detailed in the next paragraph. In what follows, $\mathbf{q}_t^{i,n}$ denotes the position of the i-th replica at time t and iteration n. In particular, $\{\mathbf{q}_0^{i,n}\}_{1 \leq i \leq N_{\text{rep}}}$ are initial conditions on Σ_R. The method to estimate the probability is also summarized in the pseudo-code presented in SI Section 1.

AMS iteration. Each iteration of the main AMS loop starts by defining the largest value of the RC for each replica at the n-th iteration as $z_{\text{max}}^{i,n} = \sup_t(\xi(\mathbf{q}_t^{i,n}))$. The replicas are then reordered by increasing values $z_{\text{max}}^{i,n}$ (see Figure 2). According to the value of k_{min}, the level at which positions are killed is identified as an empirical quantile: $z_{\text{kill}}^{n+1} = z_{\text{max}}^{k_{\text{min}},n}$. This means that all the trajectories for which $z_{\text{max}}^{i,n} \leq z_{\text{kill}}^{n+1}$ are killed. The number of killed trajectories at this iteration is denoted by $\eta_{\text{killed}}^{n+1}$. Note that $\eta_{\text{killed}}^{n+1} \geq k_{\text{min}}$ by construction, but it could happen that $\eta_{\text{killed}}^{n+1} \geq k_{\text{min}}+1$ when several trajectories reach exactly the same $z_{\text{max}}^{k_{\text{min}},n}$. To keep the number of replica constant, $\eta_{\text{killed}}^{n+1}$ trajectories have to be created by randomly branching $\eta_{\text{killed}}^{n+1}$ trajectories among the remaining ones. More precisely, trajectories are duplicated until the first time they reach the level z_{kill}^{n+1} and then the dynamics is ran from these points until it reaches R or P. In fact, at each iteration, the estimated probability $1 - \frac{\eta_{\text{killed}}^{n+1}}{N_{\text{rep}}}$ is the probability for a trajectory to reach the surface $\Sigma_{z_{\text{kill}}^{n+1}}$ starting on the surface $\Sigma_{z_{\text{kill}}^{n}}$.

Any AMS iteration can be summarized by the succession of the steps illustrated in Figure 3.

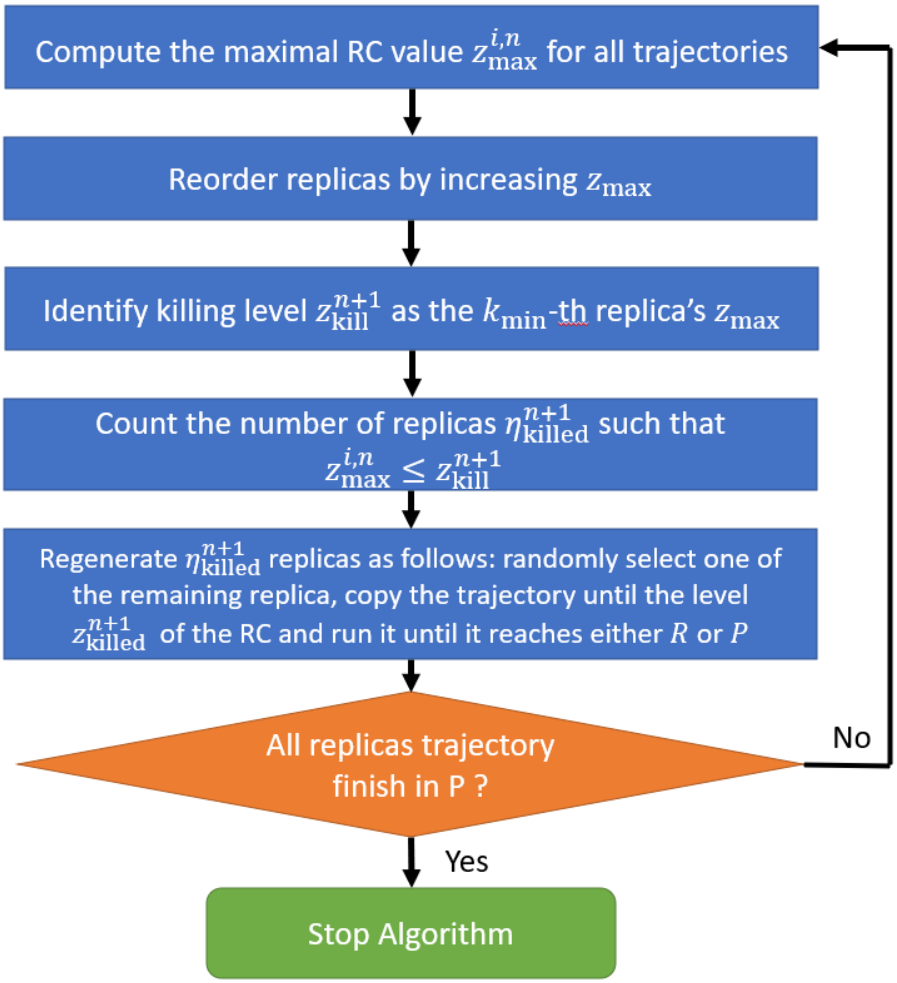


Figure 3: Flowchart of one iteration of AMS. n is an iteration index and i a replica index.

AMS termination and probability estimator. The AMS algorithm can terminate in two different manners. First, after a certain number of iterations, all the replicas reach the state P. In such a case, N_{rep} different reactive trajectories are obtained and the estimated transition probability is computed via:

$$\widehat{p}_{R\to P}(\Sigma_R) = \prod_{n=1}^{n_{\max}} \left(1 - \frac{\eta^n_{\text{killed}}}{N_{\text{rep}}}\right), \tag{4}$$

where $n_{\max}$ is the final number of iterations of the algorithm. The second option (not explicitely presented on Figure 3 since the RC is typically chosen so that this does not happen) is that at a certain iteration n, η^n_{killed} is equal to the total number of replicas of the algorithm. This can happen if at some point all the copied replicas have the same value of $z^{i,n}_{\max}$. This termination event is called "failure" as the algorithm is not able to provide reactive trajectories and the estimated probability is $\widehat{p}_{R\to P}(\Sigma_R) = 0$, consistently with expression (4). Such a situation can be encountered if the system is stuck and all the replicas are progressively replaced by the copy of a single replica. It is also possible that the replicas reach their maximum in ξ in a zone of the phase space on which the reaction coordinate ξ remains constant while the trajectories are different.

It is possible to estimate the statistical error on the estimated probability $\widehat{p}_{R\to P}(\Sigma_R)$ in (4) by repeating the estimation of the probability M_{real} times. These realisations should be independent and can take advantage of parallel architecture of current super-computers. The confidence intervals presented in the results section all correspond to a 90% confidence. More details can be found in SI Section 2.

Multiple states case. Defining the states R, P and the surface Σ_R when there are multiple metastable states requires a specific treatment to compute state to state reaction rates. Two main approaches are proposed and made precise in Section 3 of the SI. The first one samples all possible trajectories starting from a given state. The second approach more specifically focuses on the targeted transition. An illustration and comparison of the two approaches is

provided in the results section.

Implementation with a plane wave DFT code. The AMS algorithm and the sampling of initial conditions was implemented in Python scripts calling the VASP software for AIMD simulations.[57,58] All DFT simulations parameters are listed in SI Section 4.1 while AIMD parameters are presented in SI Section 4.2. Some slight modifications have been implemented in the VASP code to allow for different stopping conditions of the VASP MD runs. More details concerning the implementation can be found in the SI Section 5. The various repetitions M_{real} of the AMS estimation can be run independently in parallel. The Fleming–Viot particle scheme also allows for in dependant runs, communications are required only infrequently allowing arbitrary number of particles ran independently in parallel. The development of the scripts and the testing was mostly done on *ENER440* calculator at IFPEN. Results presented in the following section come from simulations ran on *Joliot-Curie*(Genci) and *Topaze*(CCRT).

2.2 Tools to define states and reaction coordinates

Let us conclude this section describing the methods by introducing useful tools that will be used to define the states and the reaction coordinates in the next section.

Representation of chemical structures. Reaction coordinates and states definitions must be invariant under rotation, translation and symmetries of the system as well as by permutation of identical atoms. Since description relying on Cartesian coordinates do not exhibit these properties, substantial work was conducted to find representations of atomic systems invariant by Galilean transformations and other symmetries, in particular in the field of ML empirical potentials.[59–62] We chose the smooth overlap of atomic positions (SOAP)[60] descriptor allowing to capture enough information on atomic environments to reach errors of the order of 1 meV for potential energy surface fitting.[63,64] This descriptor turned out to

be sufficient for our needs as illustrated in the result section. The detailed parameters used to compute SOAP descriptor using *dscribe* Python package[65] can be found in SI Section 4.3.

Support Vector Machine. A linear SVM model is designed to find the highest margin separation plane between two sets of labeled points. The margin denotes the minimal distance between the plane and the labeled points. The details concerning this optimization problem can be found in ML textbooks[66] or the *scikit-learn* documentation.[67] The important result for this work is that, once the optimization problem is solved, only a certain subset of the total training set is used in the definition of the plane. These are the so-called support vectors which are the closest to the separation plane. The vector normal to this plane and the scalar defining its position is thus a linear combination of the support vectors. The classifier decision function is the algebraic distance to the plane multiplied by a scaling factor chosen so that the decision function value on support vectors which are not outliers is either 1 or -1. To define multiple states using SVM, the one versus all approach was chosen, as made precise later on in the result section dedicated to the definition of states. Linear SVM models were trained using the SVC routine of *scikit-learn* package with a linear kernel.[67] The data were normalized using the standard scaler implemented in the same package. The regularization parameter was kept to the default value 1 as, after cross validation, the classification scores on the test sets were always 100%.

Path Collective Variables (PCV). The principle of PCVs is to first define a reference path for the transition as a sequence of structures $\{\mathbf{R}_i\}_{0\leq i\leq L-1}$. These structures are represented with a numerical descriptor, here the SOAP descriptor. A reaction coordinate is then constructed as:[53]

$$s(\mathbf{R}) = \frac{\sum_{i=0}^{L-1} i \ \mathrm{e}^{-\lambda d(\mathbf{R}_i,\mathbf{R})}}{\sum_{i=0}^{L-1} \mathrm{e}^{-\lambda d(\mathbf{R}_i,\mathbf{R})}}, \tag{5}$$

where d is a distance here the Euclidean norm. The parameter λ has to be of the order of variation of the inverse distances between two consecutive structures along the path. If the

structures along the path are not evenly spaced along the path according to the distance d, a sequence of values λ_i can be used instead. In the present case, we chose λ_i as:

$$\begin{cases} \lambda_i^{-1} = \frac{1}{2}\left(d(\mathbf{R}_{i-1}, \mathbf{R}_i) + d(\mathbf{R}_i, \mathbf{R}_{i+1})\right) \\ \lambda_0^{-1} = d(\mathbf{R}_0, \mathbf{R}_1);\ \lambda_{L-1}^{-1} = d(\mathbf{R}_{L-2}, \mathbf{R}_{L-1}). \end{cases} \tag{6}$$

PCVs were directly implemented in the python scripts used for the reaction coordinate evaluation during the dynamics.

3 Results and discussion

3.1 γ-Al_2O_3 models and definition of states

Model of the catalytic system. The catalytic case study chosen to benchmark the previously presented method is the transformation of a water molecule adsorbed on the (100) γ-alumina surface.[39,40,42] A representation of the γ-Al_2O_3 surface on which one water molecule is chemically adsorbed without dissociation on an aluminum Lewis site is given in Figure 4. More information about the alumina slab used is provided in SI Section 4.1.

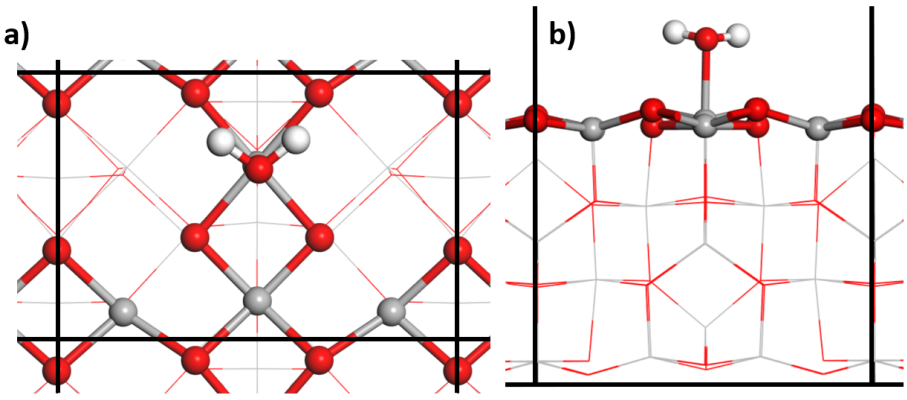


Figure 4: Representation of one water molecule adsorbed on an aluminum site of the (100)-γ-alumina surface model. Surface atoms are represented as ball and sticks while subsurface ones are represented as lines. Colors: Red: Oxygen, Grey: Aluminum, White: Hydrogen, Black: limit of the periodic cell. **a)** Top view; **b)** Side view.

The first step is to identify various potential energy minima corresponding to the metastable

states of the water molecule adsorbed on the surface either in a dissociative mode or non-dissociative modes. As described in what follows, the dissociative modes lead to the formation of two hydroxyl (OH) groups : the first one is formed upon the transfer of a H atom of the water molecule to an O site of the surface; the second one results from the native water molecule. This systematic exploration confirms previous DFT studies where the minima were identified by running multiple geometry optimizations starting from various initial conditions.[39,40]

Data set generation to learn states. Once local minima are identified, the metastability of the basins surrounding them should be assessed because these local minima should be sufficiently separated from other local minima. To quantify this, two AIMD trajectories of 1 ps each were run starting from each minimum. The first AIMD was run with a friction parameter of 5 ps^{-1} to thermalize the system faster, while the second one was run with $\gamma = 0.5$ ps^{-1}. If the system ends up in another potential energy well during this second part of the trajectory, then the initial well is not considered relevant to be qualified as a metastable state. At 300 K, multiple transitions between all basins were observed, thus all the potential wells cannot be considered metastable and relevant so as to mimic realistic chemical reactions. At 200 K, 8 genuine metastable states could be identified denoting that at this temperature, the system better mimic chemical reaction conditions. The various identified states are named A_i or D_i depending on whether the state corresponds to a non-dissociated adsorbed water molecule or to two surface hydroxyls after water dissociation, respectively (see Fig. 5). Some of these states are in fact identical as there exists a plane symmetry in this structure and thus these metastable potential energy wells should be gathered in the same state. For example, the wells D_1 and D_3 are symmetrically identical.

The numerical definition the states $A_1, A_2A_3, A_4, D_1D_3, D_2D_4$ were built using one versus all (1-vs-all) linear SVM classifiers decision function $f_{X-\text{vs}-\text{all}}$. For instance, the state A_1 is defined as $\{\mathbf{q} \mid f_{A_1-\text{vs}-\text{all}}(SOAP(\mathbf{q})) \leq -1\}$. To train these models, the data used was a

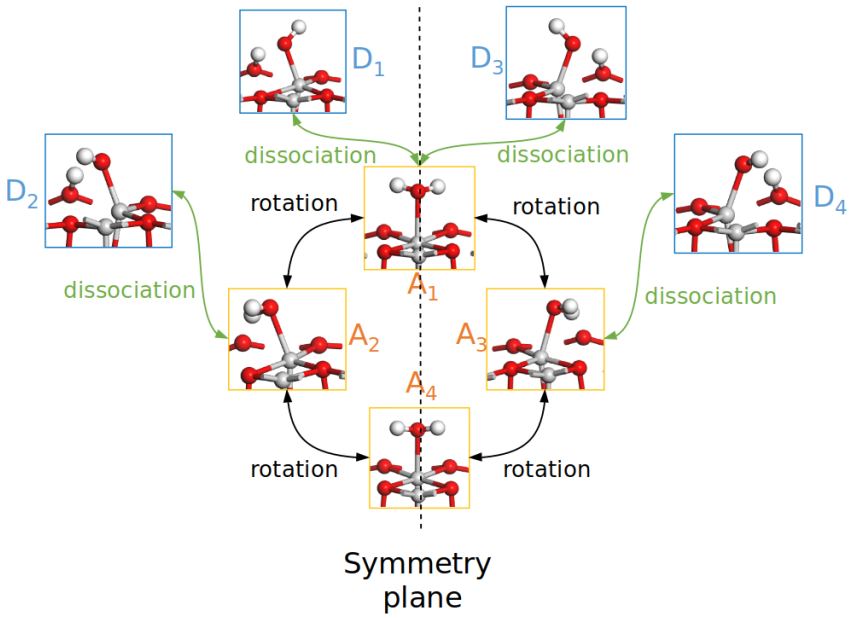


Figure 5: Representation of the main different minimum energy structures corresponding to metastable states for the water molecule adsorbed on the (100) $\gamma - Al_2O_3$ surface. Arrows represent transitions that might occur. Color legend: gray: aluminum, red; oxygen, white: hydrogen

1 ps MD trajectory at 50 K starting from each local minimum. The point of running a MD at a lower temperature was to obtain points close to the minimum of the potential energy well. The dynamics starting was run with a friction parameter of 5 ps^{-1} during 1 ps for equilibration, then run during 1 ps with $\gamma = 0.5$ ps^{-1}.

The production runs of these trajectories were used to train the SVM classifiers. Only one SOAP descriptor centered on the oxygen atom of the adsorbed water molecule was used as features in the training set, thus $SOAP$ is the function that maps the positions $\mathbf{q}$ to the SOAP descriptor of this oxygen atom environment $SOAP(\mathbf{q})$. With the parameters mentioned in the SI Section 4.4, this leads to an array of size 2100 to describe each structure. Before training the model, the variation of each dimension of the SOAP descriptors were scaled to have zero mean and unit variance. The test score of the SVM model was 100% in every case, which indicates that the set of structures represented with SOAP descriptors are linearly separable. On the other hand, trying to separate the SOAP descriptor of the trajectories starting from two symmetric minima such as D_1 and D_3 systematically led to

smaller test scores. This indicates that the well surrounding these two minima are indeed similar in the sense of the SOAP descriptor.

As a side remark, in a situation where symmetries of states are unknown, using this kind of approach can help to identify some similarities. In Figure 6, an histogram of the decision function of a A_1-vs-D_1 SOAP-SVM classifier is plotted. The various colors represent the different labelled states. It is clear that this CV allows to differentiate the A and D states. Moreover, according to this criterion, the A_2 and A_3 as well as well as the D_1/D_3 and D_2/D_4 groups of points bear some similarities for reason of symmetry.

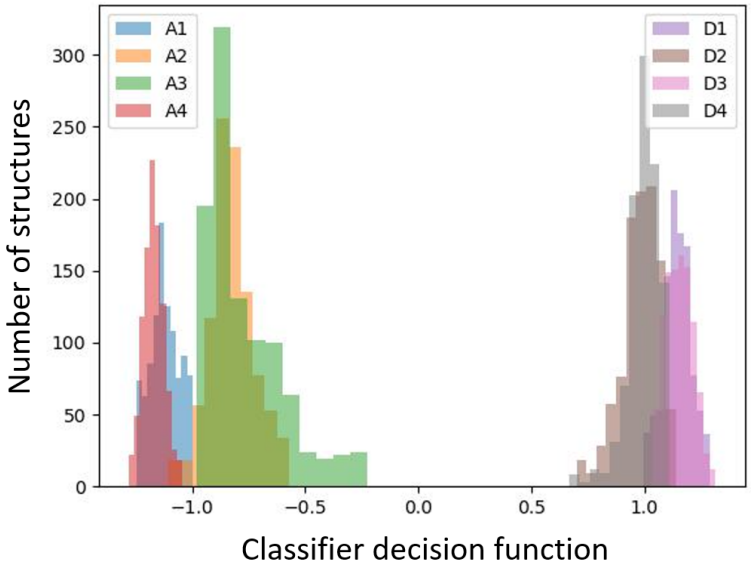


Figure 6: Histogram of A_1-vs-D_1 SOAP-SVM CV on the whole labelled dataset.

Definitions of the boundary surface Σ_R. For every metastable state X, a boundary Σ_X is defined from the 1 ps trajectory generated to assess the metastability of the basin X. This surface Σ_X is taken as a levelset of the decision function $f_{X-\text{vs}-\text{all}}$ of the SVM classifier trained to distinguish this state from the others. This level is chosen so that during the 1 ps MD the trajectory goes 10 times above this level. In practice, for all the states $A_1, A_2A_3, A_4, D_1D_3, D_2D_4$, the surface Σ_X was chosen as $\{\mathbf{q} \,|f_{A_1-\text{vs}-\text{all}}(SOAP(\mathbf{q})) = -0.95\}$. This choice implies that Σ_R for AMS is directly related to the choice of R since the two sets are determined from the level sets of the same function.

Definition of reaction coordinates (RCs). The first RCs used to perform AMS simulations are the various 1-vs-all SVM decision functions. These RCs are therefore named "1-vs-all SOAP-SVM RC" in the following sections. Some more specific RCs are built using the same approach targeting a specific transition from a state to another. In this case, the decision function is obtained by separating only the two targeted states. The corresponding RCs are termed "1-vs-1 SOAP-SVM RC". Finally, a Path Collective Variables (PCV), termed "SOAP-PCV" is also used as reaction coordinate to index the progression of AMS replicas. The SOAP-PCV RCs differ depending on the reference path. We consider here paths built by an interpolation of the z matrix representations of the minima of two metastable basins.[68] The associated RCs are termed "interpolated SOAP-PCV".

3.2 Analysis of AMS rate constants

In this section, we analyze first the sensitivity of the reaction rates to two key parameters, the number of replicas (N_{rep}) and the number of repetitions of the probability estimation (M_{real}). These parameters also govern the computational cost and how this cost can be distributed on multiple CPUs, taking advantage of the parallel architecture of current supercomputers. Then, the other impacting choices on the precision of the reaction rate constant are the RC and the states, also investigated in what follows. The reaction rate constants obtained for each observed transition are finally compared to values computed from hTST.

Parallel calculations against precision. The effect of the number of replicas (N_{rep}) and the number of AMS repetitions (M_{real}) is evaluated for a fixed number of initial conditions $N_{\text{rep}}M_{\text{real}}$, which roughly corresponds to a fixed computational cost. Indeed, assuming that every branching during one AMS realization has the same cost in average and that η^n_{killed} is constant and equals $k_{\min}$ at all steps of the AMS realization, the cost of one AMS realisation is given by the product of the number of AMS iterations ($n_{\max}$) and the number of killed

replicas ($k_{\min}$). Under these assumptions, the AMS estimator (4) writes:

$$\widehat{p} = \left(1 - \frac{k_{\min}}{N_{\text{rep}}}\right)^{n_{\max}} . \tag{7}$$

Assuming that $\frac{k_{\min}}{N_{\text{rep}}}$ is small, the computational cost of a single AMS simulation is:

$$k_{\min} n_{\max} \approx -N_{\text{rep}} \ln(\widehat{p}) . \tag{8}$$

Taking into account number of repetitions of the algorithm M_{real}, the final cost of a reaction rate constant estimation is $-M_{\text{real}} N_{\text{rep}} \ln(\widehat{p})$. Considering the current implementation of AMS, M_{real} realisation of AMS can be run in parallel. The objective is to find the minimal value of N_{rep} to better distribute the computational cost on multiple parallel realisations. With too few replicas, the intrinsic variance of the AMS estimator can be so large that the confidence interval of the estimated probability contains 0, leading to not interpretable results. Table 1 reports the evolution of water rotation rate constants $k_{A_1 \to A_2 A_3}$ calculated with AMS for various values of N_{rep} and M_{real} by using the "A_1-vs-all-SOAP-SVM" reaction coordinate and states defined as $R = A_1$ and $P = A_2A_3 \cup A_4 \cup D_1D_3 \cup D_2D_4$.

Table 1: Estimation of probability, rate and the corresponding accuracy at 90% confidence for water rotation. The number of initial conditions $M_{\text{real}} N_{\text{rep}}$ was varying M_{real} and N_{rep}. $R = A_1$, $P = A_2A_3 \cup A_4 \cup D_1D_3 \cup D_2D_4$, $\xi = A_1$-vs-all SOAP SVM RC.

M_{real}	N_{rep}	$t_{\text{loop}-R\Sigma_{A1}R}$ (fs)	$p_{A_1 \to A_2A_3}(\Sigma_{A_1})$	$k_{A_1 \to A_2A_3}$ (s^{-1})
5	400	108 ± 5	$(3.73 \pm 3.03)10^{-3}$	$(3.67 \pm 2.99)10^{10}$
10	200	110 ± 5	$(3.38 \pm 1.56)10^{-3}$	$(3.08 \pm 1.43)10^{10}$
20	100	101 ± 5	$(3.47 \pm 1.96)10^{-3}$	$(3.21 \pm 1.82)10^{10}$

By definition t_{loop} is not impacted by N_{rep} or M_{real}. The target value of probability and rate are little impacted in the present case, which is not the case for the variance. The choice of $N_{\text{rep}} = 200$ and $M_{\text{real}} = 10$ is sufficient to obtain a A_1 to A_2A_3 water rotation rate of 3.1 10^{10} s^{-1} with the 90% confidence interval of [1.65 10^{10}s^{-1}, 4.51 10^{10}s^{-1}]. Similar precision can be obtained with $N_{\text{rep}} = 100$ and $M_{\text{real}} = 20$. Therefore, it is important to perform

the AMS simulations a certain number of times (M_{real}) in order have a proper variance estimation. Hence, for a similar computational cost in CPU time, satisfying accuracy can be obtained using $M_{\text{real}} \geq 10$.

Impact of the definition of reaction coordinates and states. The definitions of the states R and P determine the type of trajectories that can be sampled by the algorithm. The choice of the reaction coordinate impacts the quality of this sampling. For instance, exploring all types of trajectories from A_1 to any other states, requires to sample initial conditions on Σ_{A_1}, set $R = A_1$ and $P = A_2A_3 \cup A_4 \cup D_1D_3 \cup D_2D_4$. Using the A_1-vs-All SOAP-SVM RC to sample trajectories ending in P leads to the results presented in Table 2. This approach

Table 2: Transition rates leaving A_1 estimated using A_1-vs-all SOAP-SVM RC, $N_{\text{rep}} = 200$, $M_{\text{real}} = 10$, $R = A_1$ and $P = A_2A_3 \cup A_4 \cup D_1D_3 \cup D_2D_4$. As the results come from the same AMS $t_{\text{loop}-R\Sigma_R R}$ is constant and equal to 110 ± 5 fs.

Transition	$p_{\text{Transition}}(\Sigma_{A_1})$	$k_{\text{Transition}}$ (s^{-1})
$A_1 \to A_2A_3$	$(3.38 \pm 1.56)\ 10^{-3}$	$(3.17 \pm 1.43)\ 10^{10}$
$A_1 \to D_1D_3$	$(1.79 \pm 1.86)\ 10^{-3}$	$(1.63 \pm 1.70)\ 10^{10}$
$A_1 \to A_4$	$(3.66 \pm 6.02)\ 10^{-7}$	$(3.44 \pm 5.50)\ 10^{6}$

allows to sample the transition $A_1 \to A_2A_3$ with a reasonable accuracy according to the estimate of the rate constant's variance. However, the less probable transitions ($A_1 \to D_1D_3$ and $A_1 \to A_4$) are under-sampled and the rate estimations are not precise enough as the 90% confidence interval contains 0. Moreover, the direct transition from A_1 to D_2D_4 being so rare that it has not even been sampled. To more accurately quantify the transition $A_1 \to D_1D_3$, more specific RCs must be used. The results obtained with two other RCs are compared in Table 3. Changing the reaction coordinate A_1-vs-all SOAP-SVM into A_1-vs-D_1 SOAP-SVM for AMS does not significantly improve the rate constant precision as the estimated variance is still so large that 0 is contained in the confidence interval. This is due to the fact that in view of the definition of R and P, AMS still samples trajectories that are of no interest such as the rotation $A_1 \to A_2A_3$. To observe only $A_1 \to D_1D_3$ reactive trajectories, one possibility would be to set $R = A_1$ and $P = D_1D_3$. However, as the AMS iteration

stops once the trajectories finish either in R or P, a trajectory including the $A_1 \rightarrow A_2A_3$ rotation would consume too much computational time before going to R or P as the state A_2A_3 is metastable. Hence R and P must be defined differently. Considering transition starting from A_1, with the choice $R = A_1 \cup A_2A_3 \cup A_4 \cup D_2D_4$, $P = D_1D_3$, and initial conditions sampled on Σ_{A_1}, AMS is compelled to sample $A_1 \rightarrow D_1D_3$ trajectories. The difference here with the previous case is that, if a rotation $A_1 \rightarrow A_2A_3$ is observed in the course of the algorithm, then it will be stopped once it enters the A_2A_3 state and be considered as a non reactive trajectory. Such trajectories will ultimately be discarded and replaced by trajectories having higher values $z_{\max}$ of the chosen reaction coordinate defined in a way so as to enhance the sampling of trajectories between the desired metastable states. Both the quality of the reaction coordinate and the choice of the R and P states are important to obtain precise results for the $A_1 \rightarrow D_1D_3$ transition (see Table 3). In our case study, the necessity to change the definition of R and P might be due to the difference of the transition probability between the rotation $A_1 \rightarrow A_2A_3$ and the water dissociation $A_1 \rightarrow D_1D_3$. Indeed, the half size of confidence intervals is larger than the target rate in the case where any type of rotations can be sampled, while constraining the AMS to sample only $A_1 \rightarrow D_1D_3$ trajectories leads to smaller confidence intervals. In the present case the interpolated SOAP-PCV RC is not significantly better than the A_1-vs-D_1-SOAP-SVM RC in terms of variance as the 90% confidence error represents 97% of the target values while for the A_1-vs-D_1-SOAP-SVM RC it is 89%.

Table 3: Variation of the RC and reactant / product states R and P to sample the $A_1 \rightarrow D_1D_3$ transition with $N_{\text{rep}} = 200$, $M_{\text{real}} = 10$ and initial conditions sampled on Σ_{A_1}

RC	$t_{\text{loop}-R\Sigma_{A_1}R}$ (fs)	$p_{A_1 \rightarrow D_1D_3}(\Sigma_{A_1})$	$k_{A_1 \rightarrow D_1D_3}$ (s^{-1})
$R = A_1$; $P = A_2A_3 \cup A_4 \cup D_1D_3 \cup D_2D_4$			
A_1-vs-all-SOAP-SVM	110 ± 5	$(1.79 \pm 1.86)\ 10^{-3}$	$(1.63 \pm 1.70)\ 10^{10}$
A_1-vs-D_1-SOAP-SVM	105 ± 3	$(1.81 \pm 1.98)\ 10^{-5}$	$(1.72 \pm 1.88)\ 10^{8}$
interpolated SOAP-PCV	104 ± 4	$(1.95 \pm 2.26)\ 10^{-4}$	$(1.87 \pm 2.17)\ 10^{9}$
$R = A_1 \cup A_2A_3 \cup A_4 \cup D_2D_4$; $P = D_1D_3$			
A_1-vs-D_1-SOAP-SVM	105 ± 2	$(3.31 \pm 2.97)\ 10^{-4}$	$(3.15 \pm 2.83)\ 10^{9}$
interpolated SOAP-PCV	108 ± 2	$(1.78 \pm 1.73)\ 10^{-4}$	$(1.64 \pm 1.59)\ 10^{9}$

Comparison of the rate constants calculated with AMS and with hTST. Various rate constants involved in the reaction networks of Figure 5 were computed using AMS. Various reaction coordinates and various definitions of the states R and P were used to obtain the results presented in Table 4. For the sake of clarity, the choice of R, P, RC and AMS parameters for each transition are listed in SI Table 1. These rates obtained by AMS are directly compared to the reaction rate constants computed from the static hTST approach. Activation free energies calculated with hTST are reported in SI Table 3. and they qualitatively compare with previously published DFT data.[44]

Table 4: Transition rate constants for all the transitions observed in this study with 90% confidence interval for AMS results.

Transition	$k_{\text{Transition}-\text{AMS}}$ (s^{-1})	$k_{\text{Transition}-\text{hTST}}$ (s^{-1})
Water rotations		
$A_1 \rightarrow A_2A_3$	$(3.08 \pm 1.43)\ 10^{10}$	$7.55\ 10^{10}$
$A_2A_3 \rightarrow A_1$	$(1.49 \pm 0.46)\ 10^{11}$	$2.06\ 10^{12}$
$A_2A_3 \rightarrow A_4$	$(4.33 \pm 2.20)\ 10^{10}$	$3.64\ 10^{10}$
$A_4 \rightarrow A_2A_3$	$(2.35 \pm 0.87)\ 10^{11}$	$5.66\ 10^{11}$
$A_1 \rightarrow A_4$	$(3.34 \pm 6.56)\ 10^{6}$	$2.04\ 10^{8}$
$A_4 \rightarrow A_1$	$(1.34 \pm 0.68)\ 10^{10}$	$8.65\ 10^{10}$
Hydroxyl rotation		
$D_1D_3 \rightarrow D_2D_4$	$\varnothing$	$2.38\ 10^{9}$
$D_2D_4 \rightarrow D_1D_3$	$(2.86 \pm 4.71)\ 10^{8}$	$4.15\ 10^{9}$
Formation and dissociation of water		
$A_1 \rightarrow D_1D_3$	$(1.64 \pm 1.59)\ 10^{9}$	$3.37\ 10^{11}$
$D_1D_3 \rightarrow A_1$	$(2.32 \pm 1.59)\ 10^{10}$	$1.13\ 10^{12}$
$A_2A_3 \rightarrow D_2D_4$	$(7.86 \pm 7.53)\ 10^{9}$	$5.45\ 10^{13}$
$D_2D_4 \rightarrow A_2A_3$	$(1.28 \pm 0.54)\ 10^{11}$	$1.17\ 10^{13}$
$A_2A_3 \rightarrow D_1D_3$	$\varnothing$	$\varnothing$
$D_1D_3 \rightarrow A_2A_3$	$(2.33 \pm 3.14)\ 10^{8}$	$\varnothing$

Reaction rate constants obtained by harmonic approximation are consistently higher than those obtained via the Hill relation and AMS for the Langevin dynamics, with one single exception for the $A_2A_3 \rightarrow A_4$ rotation. Assuming that the friction parameter is set so that Langevin dynamics reproduces accurately the system's dynamics, the AMS rate constants should be more precise than the TST ones due to the intrinsic overestimation of rates of TST as mentioned in the introduction. The harmonic approximation of the

potential energy surface for fast approximations of free energies can lead to large errors. In particular, entropic effects are usually mistreated by hTST approaches as it was underlined by previous theoretical studies based on transition path sampling and blue moon ensemble simulations[49] or other approaches.[12] In the present case, this might be the reason for the important overestimation of the rates of formation and dissociation events. Especially in the case of the $A_2A_3 \rightarrow D_2D_4$ transition, the approximation of the TS free energy is so bad that the activation free energy is negative (as reported in SI Table 3) which leads to the large overestimation of the rates.

Under the assumption of a correctly parameterized dynamics, the values presented in Table 4 allow to realize that most water rotations are at least one order of magnitude faster than dissociation events. Only the direct $A_1 \rightarrow A_4$ rotation seems to occur less frequently. The formation of water happens on the same timescale as the fast water rotations depending on the hydroxyls conformation. This ordering has to be compared to the one from hTST rate constants. The quickest changes are the water formation and dissociation events. The slowest formation event occur as frequently as the fastest water rotation.

Using the presented approach to compute various reaction rate constants, especially those of forward and backward reactions, one can deduce also reaction free energies:

$$K_{R\rightarrow P} = \frac{k_{R\rightarrow P}}{k_{P\rightarrow R}}, \tag{9}$$

and

$$\Delta G_{R\rightarrow P}(T) = -\mathcal{N}_{\mathrm{A}} k_{\mathrm{B}} T \ln\left(K_{R\rightarrow P}\right), \tag{10}$$

where $\mathcal{N}_{\mathrm{A}}$ is the Avogadro number and $K_{R\rightarrow P}$ is the reaction equilibrium constant.

The values of the reaction free energies of Table 5 allow to identify that according to the harmonic approximation, the most stable state should be the D_1D_3, while the most stable one identified with the AMS method is A_1 for T=200 K. Previous ab initio thermodynamic studies within harmonic approximations also identified that the dissociative state is more

Table 5: Reaction heats at 200 K computed from Table 4 and hTST

	AMS Value (kJ.mol^{-1})	hTST Value (kJ.mol^{-1})
Water rotations		
$\Delta G_{A_1 \to A_2A_3}$	2.62 ± 2.66	5.50
$\Delta G_{A_2A_3 \to A_4}$	2.81 ± 2.83	4.56
$\Delta G_{A_1 \to A_4}$	13.8 ± 4.43	10.1
Hydroxyl rotations		
$\Delta G_{D_1D_3 \to D_2D_4}$	$\varnothing$	0.93
Water dissociations		
$\Delta G_{A_1 \to D_1D_3}$	4.41 ± 3.88	-2.56
$\Delta G_{A_2A_3 \to D_2D_4}$	4.64 ± 3.54	2.01

favored.[39,40] Here also, one may suspect that entropic contributions may be at the origin of the change in the stability order. In particular, within the harmonic approximation, it is assumed that the adsorbed water molecule in A_1 state and in D_1D_3 has similar rotational and transnational degree of freedoms. We cannot exclude that this assumption leads to errors as AIMD simulation reveals numerous rotational movements of the adsorbed water. This effect influences the entropy change and stabilizes the non dissociated A_1 state with respect to the dissociated one D_1D_3. This thermodynamic analysis may also be consistent with the previous kinetic observation. Indeed, the thermodynamic stabilization of the non dissociated reactant states with AMS induces that water dissociation rate constants are significantly smaller with AMS than with hTST.

3.3 Analysis of AMS reactive trajectories.

In addition to computing reaction rates, we show in this section how the AMS method allows to sample reactive trajectories. The overall AMS trajectories lengths are in the order of 200 ps. Qualitatively speaking, some chemically relevant trends can be identified. We identify there are two pathways for the rotation $A_4 \to A_1$. The first, and the less likely one, is similar to the path identified by the NEB static approach. The second one seems to be more similar to a $A_4 \to A_2A_3 \to A_1$ rotation, where the trajectory does not actually enter the A_2A_3 state but approaches it for a few femtoseconds before continuing toward the

A_1 state. The same type of paths are observed in the few trajectories where a transition $D_1D_3 \rightarrow A_2A_3$ occurs. However, such a systematic analysis of each reactive trajectory might become rapidly tedious and not safe enough to capture the overall chemical trends, since more than 2000 $A_1 \rightarrow D_1D_3$ trajectories are sampled by the AMS algorithm. An automated method is therefore necessary to analyze all of them and some dimensionality reduction is useful to this end.

Clustering reactive trajectories. In the case of the $A_4 \rightarrow A_1$ rotation, two paths exist which can be identified by a visual inspection of many reactive trajectories. A more systematic way to proceed would be to rely on clustering methods, which are specially designed to identify groups within a dataset. Among the various possible approaches, we used here an approach based on the K-means algorithm as implemented in *scikit-learn*.[67] To make numerical representation of each trajectory independant on its length, each trajectory was represented as the intersections of the trajectory and five isolevels of the A_4-vs-all SOAP-SVM RC. The details of the procedure to perform this clustering are presented in SI Section 7. It is important to mention that K-means method requires to know *a priori* the number of clusters to find thus various values should be tested. The two types of paths can be identified by visual inspection of the trajectory closest to each cluster's centroid even though all the trajectories are not perfectly assigned by this approach. Of course resorting to other clustering methods could be more efficient but such a systematic study is beyond the scope of the present work. The "top" path (see Figure 7 and trajectories supplied in electronic supplementary materials) qualitatively looks similar to the path found by the NEB. The fact that this path is less sampled than the "side" path indicates that this transition is rarer.

Stochastic Transition State estimation. One possibility is to consider only one structure per trajectory instead of the whole trajectory. The most important structure $\mathbf{q}$ along a trajectory can be defined as the one such that the committor probability is $p_{R\rightarrow P}(\mathbf{q}) = 0.5$, (where $p_{R\rightarrow P}(\mathbf{q})$ is the probability that a molecular dynamics trajectory starting from $\mathbf{q}$

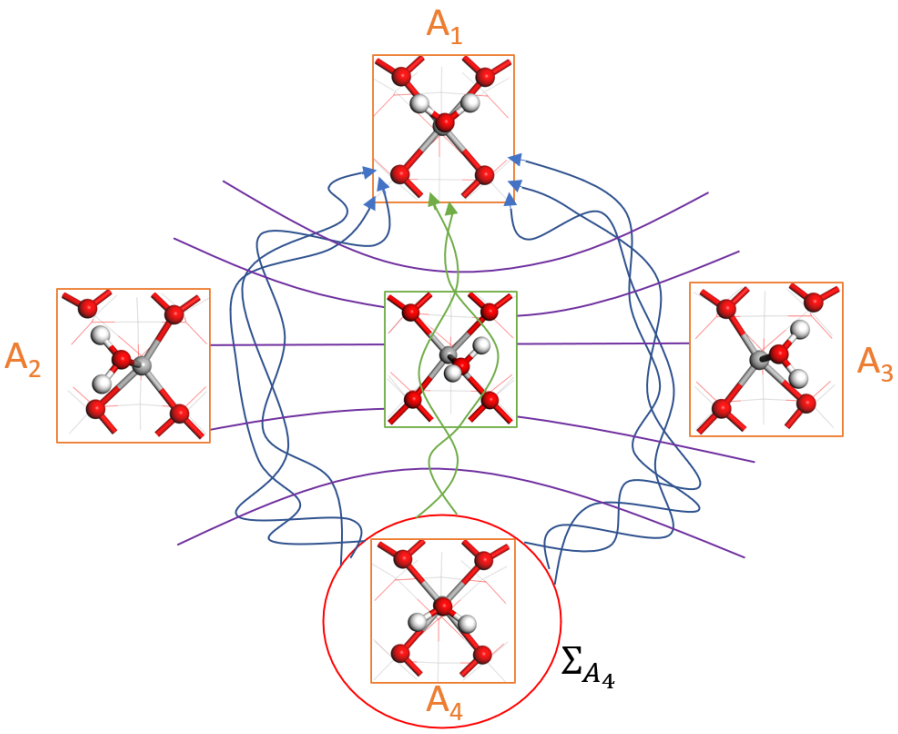


Figure 7: Schematic representation of the two types of paths for the $A_4 \rightarrow A_1$ rotation. The first path (blue) is named "side" while the second one (green) is named "top". The purple line represent the RC isolevels used to represent the trajectories.

reaches first the P state rather than R. According to the IUPAC goldbook,[8] in the part of the TS definition referring to a surface, all the structures satisfying the $p_{R\rightarrow P}(\mathbf{q}) = 0.5$ conditions are part of the transition state. This definition of transition state as a "set of states (each characterized by its own geometry and energy)" is indeed not consistent with the following part of the definition "The transition state is characterized by one and only one imaginary frequency" which presents it as a first order saddle point on the potential energy surface. The various structures $\mathbf{q}$ such that $p_{R\rightarrow P}(\mathbf{q}) = 0.5$ are not necessarily identical to the saddle points identified via the NEB method and harmonic frequencies calculations, although some resemblance is expected. We propose to investigate this point in what follows for one water dissociation on the alumina surface.

As mentioned in the methods section, the estimated probability for a trajectory to reach the surface $\Sigma_{z^{n+1}_{\text{kill}}}$ starting on the surface $\Sigma_{z^{n}_{\text{kill}}}$ is $1 - \frac{\eta^{n+1}_{\text{killed}}}{N_{\text{rep}}}$. By identifying the level $n_{0.5}$ such

that

$$\prod_{n=n_{0.5}}^{n_{\max}} \left(1 - \frac{\eta^n_{\text{killed}}}{N_{\text{rep}}}\right) = 0.5, \tag{11}$$

one can define the iso-level $\Sigma_{0.5}$ of the reaction coordinate. The configurations corresponding to reactive trajectories crossing this surface are such that $\widehat{p}_{R\to P}(\mathbf{q}) = 0.5$. There should be at least one structure corresponding to this condition per reactive trajectory. Considering only the first structure crossing the iso-level $\Sigma_{0.5}$, the mean structure is computed, in the sense of the SOAP descriptor. This analysis was applied for the various realizations of AMS that were run.

Stochastic Transition State of water dissociation. For the dissociation event $A_1 \to D_1D_3$, the interpolated SOAP-PCV reaction coordinate with the reactant and product states defined as $R = A_1 \cup A_2A_3 \cup A_4 \cup D_2D_4$ and $P = D_1D_3$ conducts to a mean structure of configuration such that $p_{R\to P}(\mathbf{q}) = 0.5$ qualitatively similar to the saddle point of the PES determined with the NEB method, as represented in Figure 8 (the corresponding trajectory is provided in electronic supplementary information). From a quantitative viewpoint, some slight structural differences can be noted regarding the O-H distances involving the transferred H atom. For the saddle point, the broken O-H bond is 0.14 Å shorter than for AMS, whereas the newly formed O-H bond is 0.14 Å larger. This difference might come from the fact that the momenta can bear a certain importance in the committor. Indeed the committor values estimated bears dynamical information while the saddle point is defined only with the positions.

The quality of this analysis depends on the quality of the sampling of the reaction path. Indeed considering the reactive trajectories sampled from the AMS done with: $R = A_1$ and $P = \cup\, A_2A_3 \cup A_4 \cup D_1D_3 \cup D_2D_4$ the definition of the stochastic TS is of poor quality. This comes from the fact that this AMS mostly samples $A_1 \to A_2A_3$ trajectories (and only rarely $A_1 \to D_1D_3$ trajectories). The best approximation of a stochastic TS is on the most sampled path (Region 1 in Fig. 9). This is in line with the obtained results for the

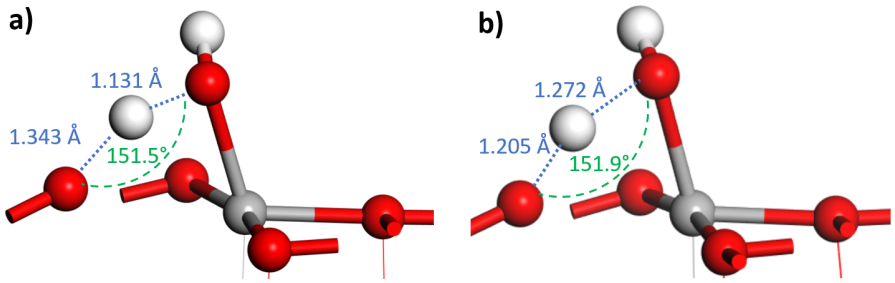


Figure 8: Ball and sticks representation of **a)** saddle point on the PES and **b)** mean structures such that $p_{R\to P}(\mathbf{q}) = 0.5$ on AMS using interpolated SOAP PCV RC. Color legend, red: oxygen, gray: aluminum, white: hydrogen.

confidence interval of the reaction rate constants (see Table 3). In Figure 9, the green curve represents the $\Sigma_{0.5}$ iso-level of the reaction coordinate, while the red one is the $\Sigma_{0.5}$ iso-level of the committor function. As these levels do not match perfectly on the whole space, the best approximation of the stochastic TS is in the region of space where most of the reactive trajectories concentrate (Region 1). This issue in the analysis of the reactive trajectories of less probable transitions is recurrent when multiple paths are sampled. An alternative approach to automatically identify whether multiple paths leading to a single product are present within a set of sampled trajectories would be desirable.

4 Conclusion

We proposed and implemented a theoretical approach based on the Hill relation to compute the exact reaction rate constants using rare event sampling and support vector machines. It is illustrated on various chemical events occurring at an oxide material surface. A key algorithm to this end is the Adaptive Multilevel Splitting, which estimates reaction probabilities and samples reactive trajectories by using ab initio molecular dynamics. For that purpose, SVM was used to define the chemically relevant states and reaction coordinates to index the transition from reactant to products. It allows to compute the exact reaction rates for the dynamics at hand and makes possible a detailed analysis of reaction mechanisms via the inspection of reactive trajectories. The implementation done so as to communicate with a

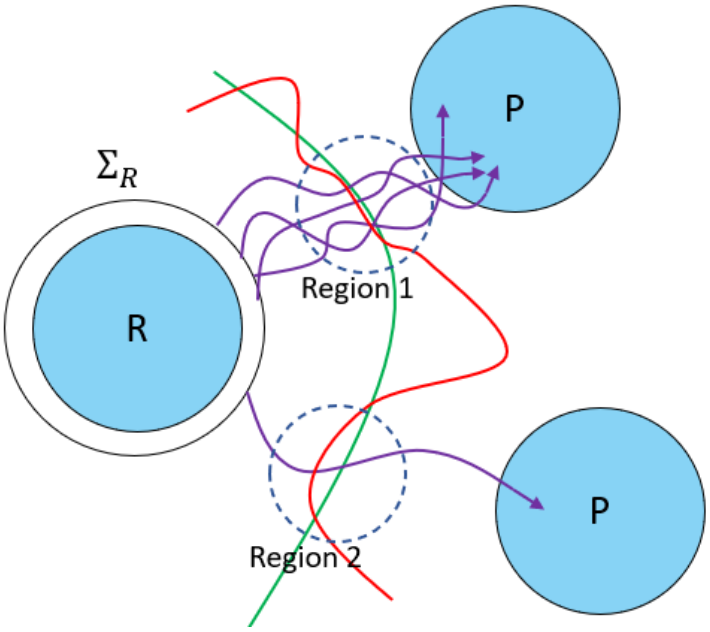


Figure 9: Schematic representation of poor match of the $\Sigma_{0.5}$ iso-level of the committor function (red) and the reaction coordinate (green). The green iso-level is placed after an AMS sampling of some reactive trajectories from R to P where a majority of the trajectories has gone via Region 1.

plane-wave DFT software allowed to illustrate the approach by studying the reactivity of a water molecule adsorbed on the γ-alumina (100) surface. The computed reaction rate constants were discussed and compared to those of a static hTST approach.

The methods precision is impacted by the choice of reaction coordinate, the choice of reactants and products in a multiple state situation, the number of repetitions of the probability estimation and the number of replicas intrinsic to AMS. The hTST approach does not make assumptions on the system's dynamics, but relies on strong assumptions concerning the shape of the potential energy surface, implying uncontrolled approximation of entropy. The proposed methodology allows to alleviate these limitations at the expense of an increased computational cost. Assuming that the Langevin dynamics accurately models the system's dynamics (which involves in particular having a relevant value of the friction coefficient), the presented approach should be more precise than TST approaches. In the case considered here, hTST reaction rate constants are always higher than the ones estimated via AMS and the Hill relation. The relative stability of states is also different. In particular, we show

that hTST underestimates the thermodynamic stability of adsorbed water molecules, and simultaneously overestimates rate constants of water dissociation and formation. On top of that, the analysis of reactive trajectories allows to identify possible paths that are not clearly identified via the NEB approach.

This method used in combination with *ab-initio* molecular dynamics can be computationally expensive. This issue might be alleviated by the use of Machine Learning Force Fields (MLFF), which can approach the accuracy of DFT force calculations at a much smaller computational cost. It also provides the opportunity to accurately describe nuclear quantum effects using path integral molecular dynamics such as in Ref. 69. Some active learning schemes to train MLFFs have been proposed recently and they could articulate well with the present method.[70,71] In particular, in contrast to standard MD, the presented approach favor a sampling of transition regions which are crucial to the description of chemical event. The study of a specific system could be done by first using jointly AMS and active learning to generate an accurate MLFF. Then, it could be used to evaluate accurately reaction rate constants and sample reactive trajectories.

Acknowledgement

This project was realized in the framework of the joint laboratory IFPEN–Inria Convergence HPC/AI/HPDA for the energetic transition. Calculations were performed using the following HPC resources: Jean Zay and Occigen from GENCI-CINES, Joliot-Curie (Irene) from TGCC/CEA (Grant A0120806134), ENER440 from IFP Energies nouvelles and Topaze from CCRT-CEA. The work of TL and GS was funded in part by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (project EMC2, grant agreement No 810367).

References

(1) Broadbelt, L. J.; Snurr, R. Q. Applications of molecular modeling in heterogeneous catalysis research. *Appl. Catal. A: Gen.* **2000**, *200*, 23–46.

(2) Chizallet, C.; Raybaud, P. Density functional theory simulations of complex catalytic materials in reactive environments: beyond the ideal surface at low coverage. *Catal. Sci. Technol.* **2014**, *4*, 2797–2813.

(3) Chen, B. W. J.; Xu, L.; Mavrikakis, M. Computational methods in heterogeneous catalysis. *Chem. Rev.* **2020**, *121*, 1007–1048.

(4) Piccini, G.; Lee, M.-S.; Yuk, S. F.; Zhang, D.; Collinge, G.; Kollias, L.; Nguyen, M.-T.; Glezakou, V.-A.; Rousseau, R. Ab initio molecular dynamics with enhanced sampling in heterogeneous catalysis. *Catal. Sci. Technol.* **2022**, *12*, 12–37.

(5) Feynman, R. P. Forces in molecules. *Phys. Rev.* **1939**, *56*, 340–343.

(6) Eyring, H. The activated complex in chemical reactions. *J. Chem. Phys.* **1935**, *3*, 107–115.

(7) Bennett, C. H. *Algorithms for Chemical Computations*; American Chemical Society: Washington, DC, USA, 1977; Chapter 4, pp 63–67.

(8) *The IUPAC Compendium of Chemical Terminology*; International Union of Pure and Applied Chemistry (IUPAC), 2014; p 10.1351/goldbook.t06468.

(9) Hänggi, P.; Talkner, P.; Borkovec, M. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **1990**, *62*, 251–341.

(10) Evans, M. G.; Polanyi, M. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.* **1935**, *31*, 875.

(11) Wigner, E. The transition state method. *Trans. Faraday Soc.* **1938**, *34*, 29.

(12) Collinge, G.; Yuk, S. F.; Nguyen, M.-T.; Lee, M.-S.; Glezakou, V.-A.; Rousseau, R. Effect of collective dynamics and anharmonicity on entropy in heterogenous catalysis: Building the case for advanced molecular simulations. *ACS Catal.* **2020**, *10*, 9236–9260.

(13) Chandler, D. Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **1978**, *68*, 2959.

(14) Chipot, C., Pohorille, A., Eds. *Free energy calculations*; Springer Berlin Heidelberg, 2007.

(15) Rousset, M.; Stoltz, G.; Lelièvre, T. *Free Energy Computations*; Imperial College Press, 2010.

(16) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced sampling in molecular dynamics. *J. Chem. Phys.* **2019**, *151*, 070902.

(17) Horiuti, J. On the statistical mechanical treatment of the absolute rate of chemical reaction. *Bull. Chem. Soc. Jpn.* **1938**, *13*, 210–216.

(18) Keck, J. Statistical investigation of dissociation cross-sections for diatoms. *Faraday Discuss.* **1962**, *33*, 173.

(19) Vanden-Eijnden, E.; Tal, F. A. Transition state theory: Variational formulation, dynamical corrections, and error estimates. *J. Chem. Phys.* **2005**, *123*, 184103.

(20) Miller, W. H.; Schwartz, S. D.; Tromp, J. W. Quantum mechanical rate constants for bimolecular reactions. *J. Chem. Phys.* **1983**, *79*, 4889–4898.

(21) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Advances in Chemical Physics*; John Wiley & Sons, Inc., 2003; pp 1–78.

(22) Mandelli, D.; Hirshberg, B.; Parrinello, M. Metadynamics of paths. *Phys. Rev. Lett.* **2020**, *125*, 026001.

(23) Hill, T. *Free Energy Transduction in Biology: The Steady-State Kinetic and Thermodynamic Formalism*; Elsevier Science and Technology Books, 2012.

(24) Baudel, M.; Guyader, A.; Lelièvre, T. On the Hill relation and the mean reaction time for metastable processes. *Stoch Process Their Appl* **2023**, *155*, 393–436.

(25) Lelièvre, T.; Ramil, M.; Reygner, J. Estimation of statistics of transitions and Hill relation for Langevin dynamics. *arXiv:2206.13264* **2022**, to appear in Annales de l'Institut Henri Poincaré.

(26) van Erp, T. S.; Moroni, D.; Bolhuis, P. G. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* **2003**, *118*, 7762–7774.

(27) Allen, R. J.; Warren, P. B.; ten Wolde, P. R. Sampling rare switching events in biochemical networks. *Phys. Rev. Lett.* **2005**, *94*, 018104.

(28) Huber, G.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **1996**, *70*, 97–110.

(29) Cérou, F.; Guyader, A. Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.* **2007**, *25*, 417–443.

(30) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised learning methods for molecular simulation data. *Chem. Rev.* **2021**, *121*, 9722–9758.

(31) Chen, M. Collective variable-based enhanced sampling and machine learning. *Eur. Phys. J. B* **2021**, *94*, 211.

(32) Gkeka, P.; Stoltz, G.; Farimani, A. B.; Belkacemi, Z.; Ceriotti, M.; Chodera, J. D.; Dinner, A. R.; Ferguson, A. L.; Maillet, J.-B.; Minoux, H.; Peter, C.; Pietrucci, F.;

Silveira, A.; Tkatchenko, A.; Trstanova, Z.; Wiewiora, R.; Lelièvre, T. Machine learning force fields and coarse-grained variables in molecular dynamics: Application to materials and biological systems. *J. Chem. Theory Comput.* **2020**, *16*, 4757–4775.

(33) Ferguson, A. L. Machine learning and data science in soft materials engineering. *J. Condens. Matter Phys.* **2017**, *30*, 043002.

(34) Sultan, M. M.; Pande, V. S. Automated design of collective variables using supervised machine learning. *J. Chem. Phys.* **2018**, *149*, 094106.

(35) Pozun, Z. D.; Hansen, K.; Sheppard, D.; Rupp, M.; Müller, K.-R.; Henkelman, G. Optimizing transition states via kernel-based machine learning. *J. Chem. Phys.* **2012**, *136*, 174101.

(36) Christiansen, M. A.; Mpourmpakis, G.; Vlachos, D. G. Density functional theory - Computed mechanisms of ethylene and diethyl ether formation from ethanol on γ-Al_2O_3(100). *ACS Catal.* **2013**, *3, 9*, 1965–1975.

(37) Larmier, K.; Nicolle, A.; Chizallet, C.; Cadran, N.; Maury, S.; Lamic-Humblot, A.-F.; Marceau, E.; Lauron-Pernot, H. Influence of coadsorbed water and alcohol molecules on isopropyl alcohol dehydration on γ-alumina: Multiscale modeling of experimental kinetic profiles. *ACS Catal.* **2016**, *6*, 1905–1920.

(38) Hass, K. C.; Schneider, W. F.; Curioni, A.; Andreoni, W. The chemistry of water on alumina surfaces: Reaction dynamics from first principles. *Science* **1998**, *282*, 265–268.

(39) Digne, M.; Sautet, P.; Raybaud, P.; Euzen, P.; Toulhoat, H. Hydroxyl groups on γ-alumina surfaces: A DFT study. *J. Catal.* **2002**, *211*, 1–5.

(40) Digne, M.; Sautet, P.; Raybaud, P.; Euzen, P. Use of DFT to achieve a rational understanding of acido-basic properties of γ-alumina surfaces. *J. Catal.* **2004**, *226*, 54–68.

(41) Wischert, R.; Laurent, P.; Copéret, C.; Delbecq, F.; Sautet, P. γ-Alumina: The essential and unexpected role of water for the structure, stability, and reactivity of "defect" sites. *J. Am. Chem. Soc.* **2012**, *134*, 14430–14449.

(42) Pigeon, T.; Chizallet, C.; Raybaud, P. Revisiting γ-alumina surface models through the topotactic transformation of boehmite surfaces. *J. Catal.* **2022**, *405*, 140–151.

(43) Lu, Y.-H.; Wu, S.-Y.; Chen, H.-T. H_2O Adsorption/Dissociation and H_2 generation by the reaction of H_2O with Al_2O_3 materials: A first-principles investigation. *J. Phys. Chem. C* **2016**, *120*, 21561–21570.

(44) Pan, Y.; Liu, C.-J.; Ge, Q. Adsorption and protonation of CO_2 on partially hydroxylated γ-Al_2O_3 surfaces: A density functional theory study. *Langmuir* **2008**, *24*, 12410–12419.

(45) Ngouana-Wakou, B. F.; Cornette, P.; Valero, M. C.; Costa, D.; Raybaud, P. An atomistic description of the γ-alumina/water interface revealed by ab initio molecular dynamics. *J. Phys. Chem. C* **2017**, *121*, 10351–10363.

(46) Réocreux, R.; Jiang, T.; Iannuzzi, M.; Michel, C.; Sautet, P. Structuration and dynamics of interfacial liquid water at hydrated γ-alumina determined by ab initio molecular simulations: Implications for nanoparticle stability. *ACS Appl. Nano Mater.* **2017**, *1*, 191–199.

(47) Lo, C. S.; Radhakrishnan, R.; Trout, B. L. Application of transition path sampling methods in catalysis: A new mechanism for CC bond formation in the methanol coupling reaction in Chabazite. *Catal. Today* **2005**, *105*, 93–105.

(48) Bucko, T.; Benco, L.; Dubay, O.; Dellago, C.; Hafner, J. Mechanism of alkane dehydrogenation catalyzed by acidic zeolites: Ab initio transition path sampling. *J. Chem. Phys.* **2009**, *131*, 214508.

(49) Rey, J.; Bignaud, C.; Raybaud, P.; Bucko, T.; Chizallet, C. Dynamic features of transition states for *beta*-scission reactions of alkenes over acid zeolites revealed by AIMD simulations. *Angew. Chem., Int. Ed. Engl.* **2020**, *59*, 18938–18942.

(50) Roet, S.; Daub, C. D.; Riccardi, E. Chemistrees: Data-driven identification of reaction pathways via machine learning. *J. Chem. Theory Comput.* **2021**, *17*, 6193–6202.

(51) Lopes, L. J. S.; Lelièvre, T. Analysis of the adaptive multilevel splitting method on the isomerization of alanine dipeptide. *J. Comput. Chem.* **2019**, *40*, 1198–1208.

(52) Teo, I.; Mayne, C. G.; Schulten, K.; Lelièvre, T. Adaptive multilevel mplitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time. *J. Chem. Theory Comput.* **2016**, *12*, 2983–2989.

(53) Branduardi, D.; Gervasio, F. L.; Parrinello, M. From A to B in free energy space. *J. Chem. Phys.* **2007**, *126*, 054103.

(54) Cérou, F.; Delyon, B.; Guyader, A.; Rousset, M. On the Asymptotic Normality of Adaptive Multilevel Splitting. *SIAM-ASA J. Uncertain. Quantif.* **2019**, *7*, 1–30.

(55) Bréhier, C.-E.; Gazeau, M.; Goudenège, L.; Lelièvre, T.; Rousset, M. Unbiasedness of some generalized adaptive multilevel splitting algorithms. *J. Appl. Probab.* **2016**, *26*, 3559 – 3601.

(56) Binder, A.; Lelièvre, T.; Simpson, G. A generalized parallel replica dynamics. *J. Comput. Phys.* **2015**, *284*, 595–616.

(57) Kresse, G.; Hafner, J. Ab-initio molecular dynamics for liquid metals. *Phys. Rev. B* **1993**, *47*, 558–561.

(58) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(59) Behler, J.; Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401.

(60) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.

(61) Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. *Phys. Rev. B* **2019**, *99*, 014104.

(62) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A critical review of machine learning of energy materials. *Adv. Energy Mater.* **2020**, *10*, 1903242.

(63) Bartók-Pártay, A. *The Gaussian Approximation Potential*; Springer Berlin Heidelberg, 2010.

(64) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine learning a general-purpose interatomic potential for silicon. *Phys. Rev. X* **2018**, *8*, 041048.

(65) Himanen, L.; Jäger, M. O.; Morooka, E. V.; Canova, F. F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **2020**, *247*, 106949.

(66) Murphy, K. P. *Probabilistic Machine Learning: An introduction*; MIT Press, 2022.

(67) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Édouard Duchesnay, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(68) Fleurat-Lessard, P. `http://pfleurat.free.fr/ReactionPath.php`.

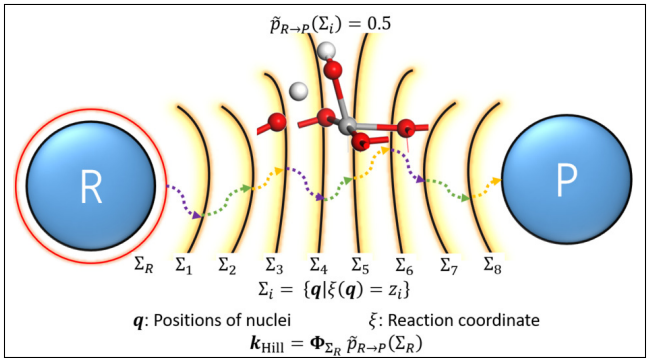
(69) Bocus, M.; Goeminne, R.; Lamaire, A.; Cools-Ceuppens, M.; Verstraelen, T.; Speybroeck, V. V. Nuclear quantum effects on zeolite proton hopping kinetics explored

with machine learning potentials and path integral molecular dynamics. *Nat. Commun.* **2023**, *14*.

(70) Vandermause, J.; Torrisi, S. B.; Batzner, S.; Xie, Y.; Sun, L.; Kolpak, A. M.; Kozinsky, B. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *Npj Comput. Mater.* **2020**, *6*.

(71) Jinnouchi, R.; Miwa, K.; Karsai, F.; Kresse, G.; Asahi, R. On-the-fly active learning of interatomic potentials for large-scale atomistic simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 6946–6955.

TOC Graphic


$\tilde{p}_{R\to P}(\Sigma_i) = 0.5$
R
P
Σ_R Σ_1 Σ_2 Σ_3 Σ_4 Σ_5 Σ_6 Σ_7 Σ_8
$\Sigma_i = \{\boldsymbol{q}|\xi(\boldsymbol{q}) = z_i\}$
$\boldsymbol{q}$: Positions of nuclei
ξ: Reaction coordinate
$\boldsymbol{k}_{\mathrm{Hill}} = \boldsymbol{\Phi}_{\Sigma_R}\, \tilde{p}_{R\to P}(\Sigma_R)$