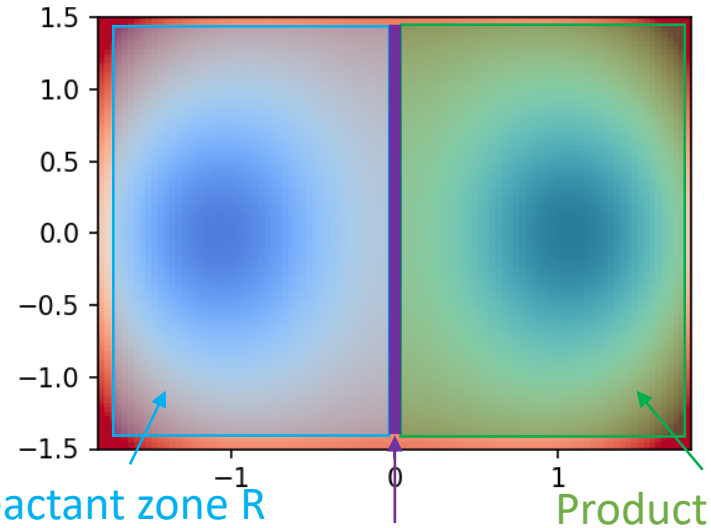
COMPUTING SURFACE REACTION RATE USING MACHINE LEARNING INTER-ATOMIC POTENTIAL

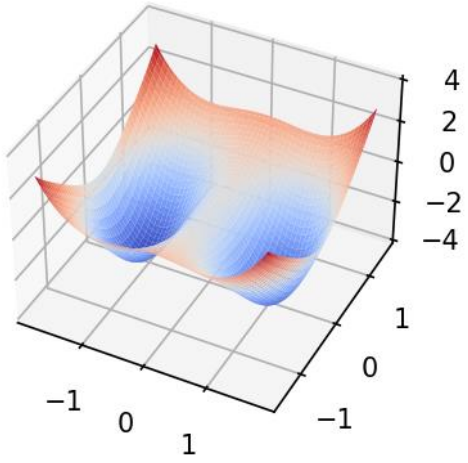
THOMAS PIGEON

2 THE HILL RELATION

Transition State Theory



2-dimensional potential

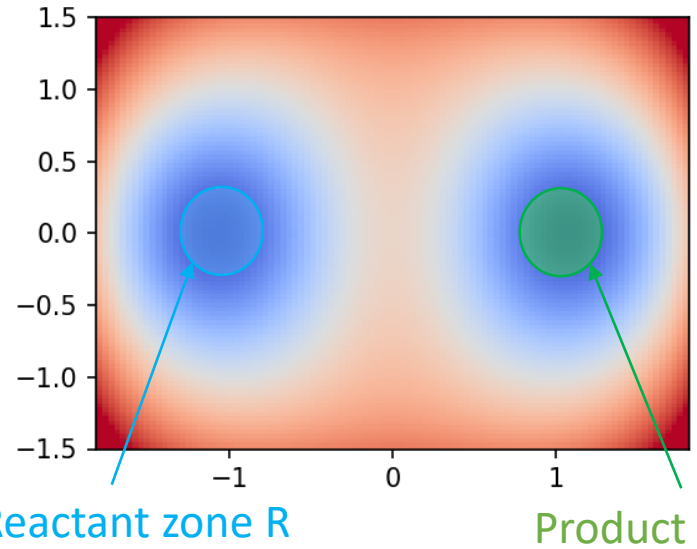


Rate = probability of being in TS with respect to R
× frequency of decomposition to P

$$\boldsymbol{k}^{TST} = \mathbb{P}(TS \mid R)\ \phi_{TS \to P}$$

Sensitive to the TS definition
TST overestimates rates (κ)[1]

Hill relation[2]



Rate = probability of reaching P before R starting from ∂R × exit frequency from R basin

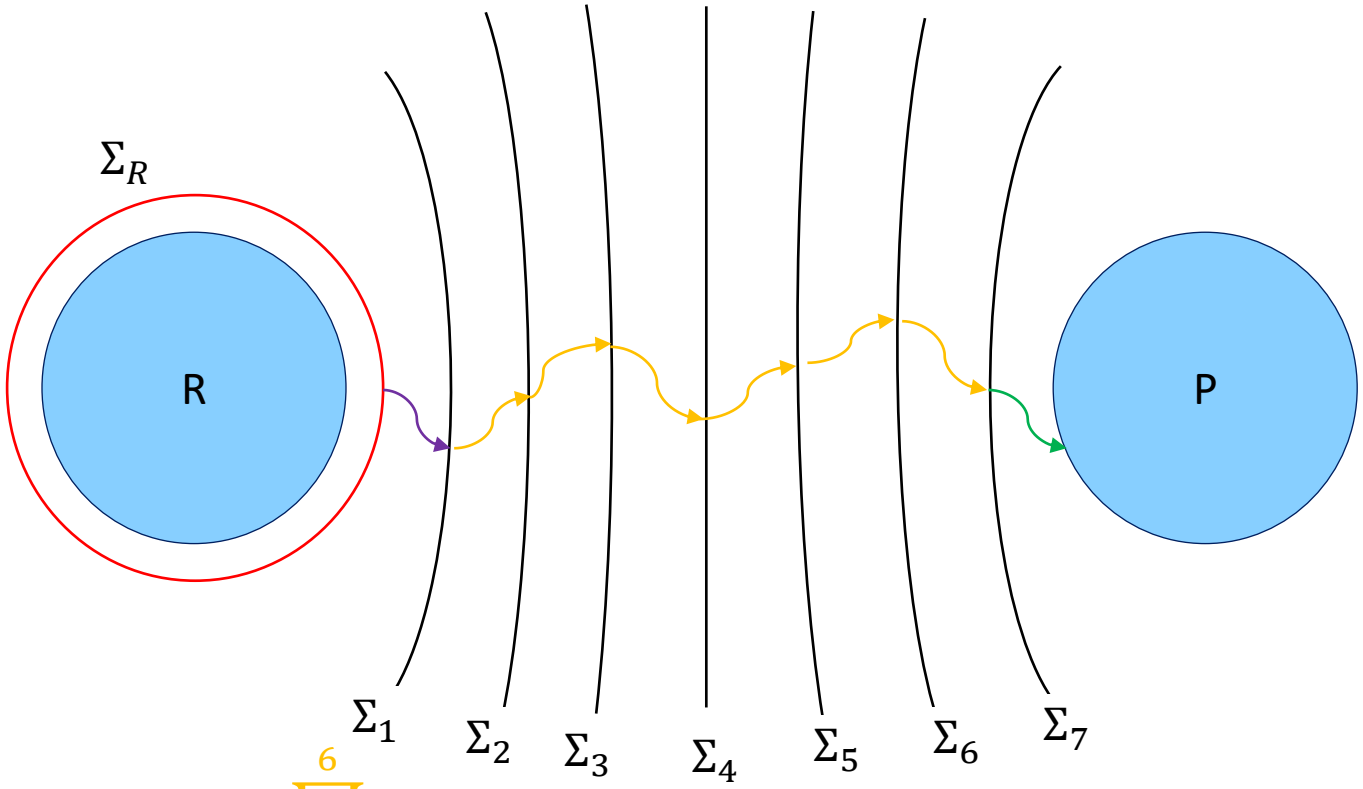
$$\boldsymbol{k}^{Hill} = \mathbb{P}^{\partial R}(\tau_P < \tau_R)\ \phi_R$$

Not extremely sensitive to the definition of R and P

[1] Hänggi, P. Talkner, P. Borkovec, M. (1990) *Reviews of Modern Physics* , 62, 2, 251
[2] Hill, T. (2012) Free energy transduction in biology: The steady-state kinetic and thermodynamic formalism. *Elsevier Science and Technology Books*

I. ADAPTIVE MULTI-LEVEL SPLITTING

What is a Multilevel Splitting estimator?



$$\mathbb{P}^{\Sigma_R}(\tau_{\Sigma_1} < \tau_R) \prod_{i=1}^{6} \mathbb{P}^{\Sigma_i}(\tau_{\Sigma_{i+1}} < \tau_R | \tau_{\Sigma_i} < \tau_R)\, \mathbb{P}^{\Sigma_7}(\tau_P < \tau_R | \tau_{\Sigma_7} < \tau_R)$$

$$= \mathbb{P}^{\Sigma_R}(\tau_P < \tau_R)$$

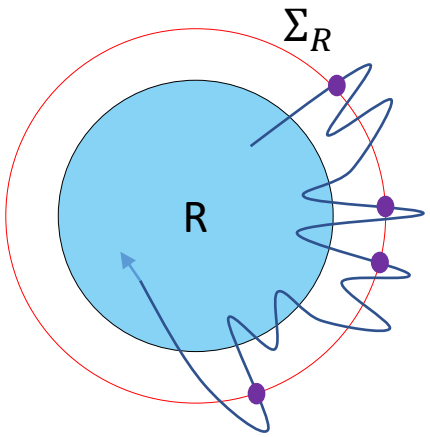
How to place Σ_i and compute $\mathbb{P}^{\Sigma_i}(\tau_{\Sigma_{i+1}} < \tau_R | \tau_{\Sigma_i} < \tau_R)$?

I. ADAPTIVE MULTI-LEVEL SPLITTING

AMS can be split in 3 steps:

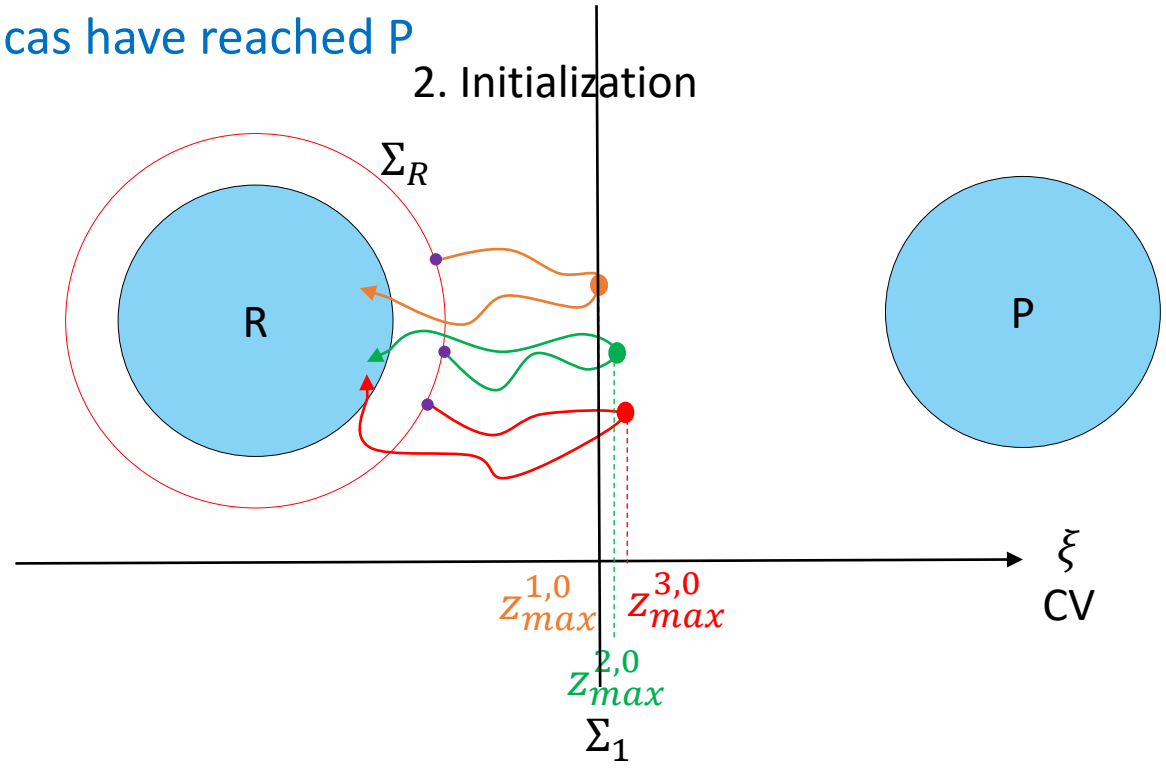
1. Generate initial conditions on Σ and estimate $t_{R-\Sigma_R-R} = \frac{1}{\phi_R}$
2. Initialize N_{rep} replicas by running an unbiased dynamics until it reaches R or P. Classify all the replicas by increasing z_{max}
3. Apply the AMS loop until all replicas have reached P

1. Initial conditions and flux



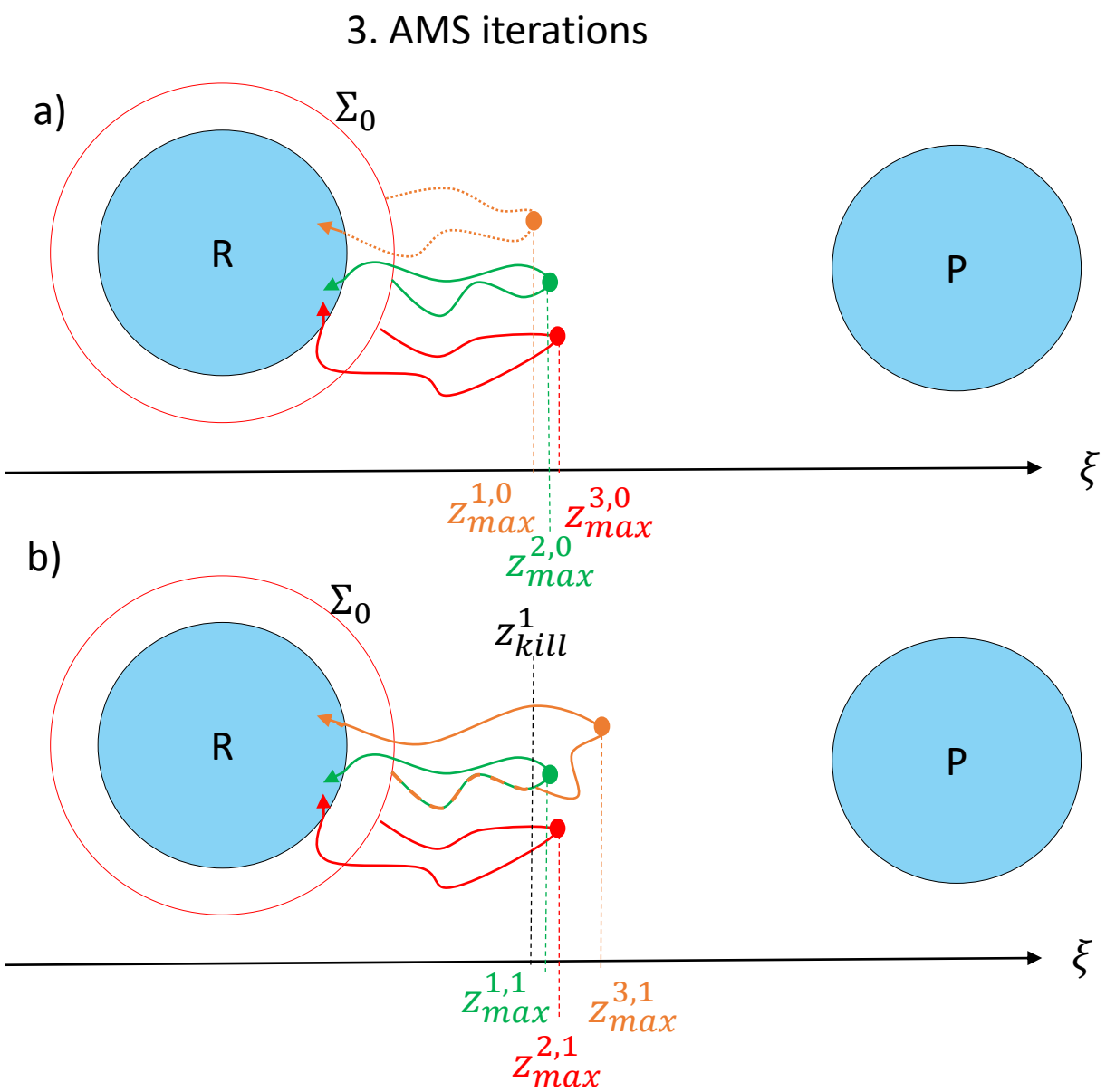
MD for Initial conditions.

2. Initialization



[1] F. Cérou, A. Guyader, (2007) *Stochastic Analysis and Applications* 25, 417
[2] L. J. S. Lopes, T. Lelièvre, (2019) *J. Comput. Chem* 40, 1198

I. ADAPTIVE MULTI-LEVEL SPLITTING



3. AMS iterations: $i \geq 0$

a) Save the smallest ($z_{max}^{1,i}$) as z_{kill}^{i+1} and delete all the trajectories that did not "go above" z_{kill}^{i+1}

b) Randomly select one trajectory within the remaining ones. Copy it until it reaches z_{kill}^{i+1} and continue it until it reaches R or P.

c) Classify all the replicas by increasing z_{max}

$$\tilde{p} = \prod_{i=1}^{i_{max}-1} \mathbb{P}^{\Sigma_i}(\tau_{\Sigma_{i+1}} < \tau_R | \tau_{\Sigma_i} < \tau_R) = \left(1 - \frac{1}{N}\right)^{i_{max}}$$

Unbiased estimator: $\mathbb{E}[\tilde{p}] = p_{R-P}(\Sigma_R)$

Variance depends on CV: $\mathrm{Var}[\tilde{p}] = f(\xi)$

→ M_{real} estimations to compute $\mathbb{E}[\tilde{p}]$ and estimate $\mathrm{Var}[\tilde{p}]$

→ N_{rep} impacts $\mathrm{Var}[\tilde{p}]$ as well

II. VASP MACHINE LEARNING FORCE FIELD

VASP MLFF:[1,2]

2 and 3 body (SOAP-like) atom centered descriptors

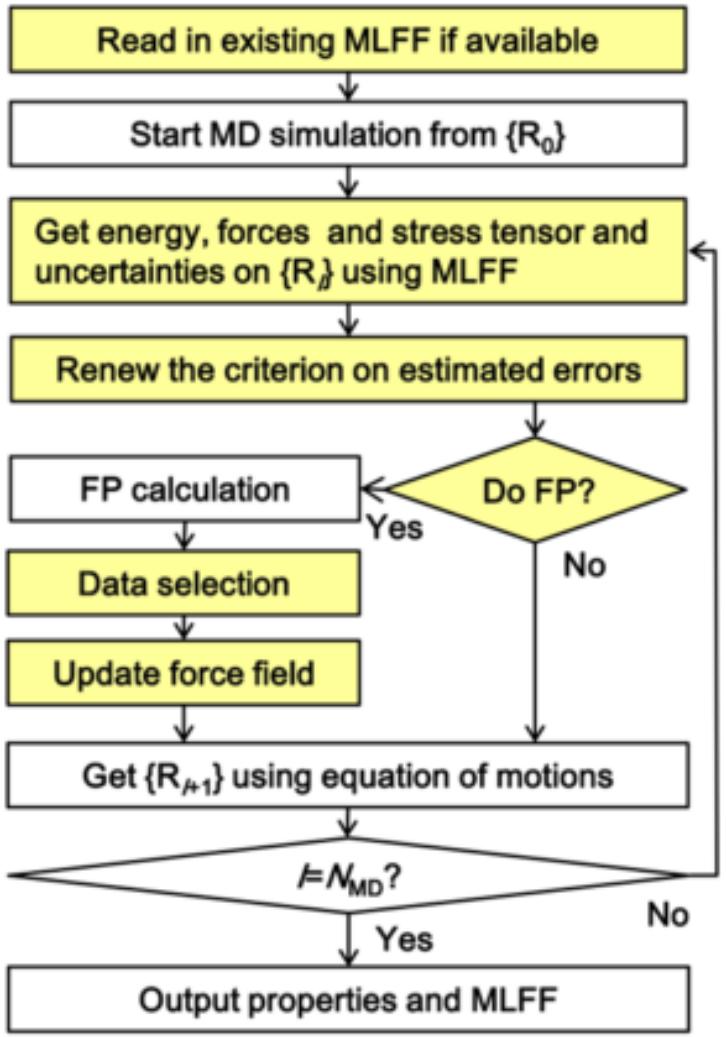
$$\rho_i^{(2)}(r) = \frac{1}{\sqrt{4\pi}} \sum_{n=1}^{N_{\mathrm{R}}^0} c_{n00}^i \chi_{n0}(r)$$

$$\rho_i^{(3)}(r,s,\theta) = \sum_{l=1}^{L_{\max}} \sum_{n=1}^{N_{\mathrm{R}}^l} \sum_{\nu=1}^{N_{\mathrm{R}}^l} \sqrt{\frac{2l+1}{2}} p_{n\nu l}^i \chi_{nl}(r)\, \chi_{\nu l}(s)\, P_l(\cos\theta)$$

$\chi_{\nu l}$: spherical Bessel functions
P_l : Legendre polynomials

Bayesian linear regression

VASP active learning scheme:[1,2]



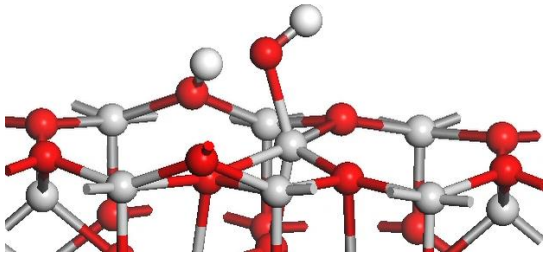
1 Jinnouchi, R.; Karsai, F.; Kresse, G. (2019) *Phys. Rev. B* 100, 014105
2 Jinnouchi, R.; Miwa, K.; Karsai, F.; Kresse, G.; Asahi, R. (2020) *J Phys. Chem. Lett.* 11, 6946

II. VASP MACHINE LEARNING FORCE FIELD

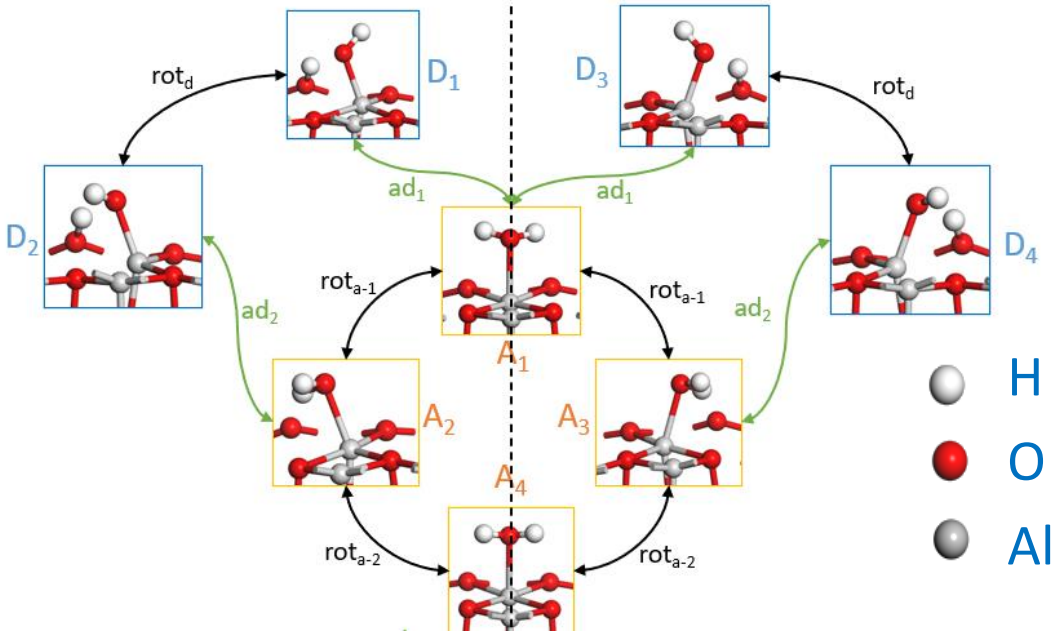
Method to train a MLFF:

1. Identify the various metastable states (intermediates)
   → dissociated (D_i) or associated (A_i)
2. Run active learning dynamics in these states to sample Potential Energy Surface (PES) around the minima

D_1 dissociated structure



dt = 1 fs
Total time = 50 ps

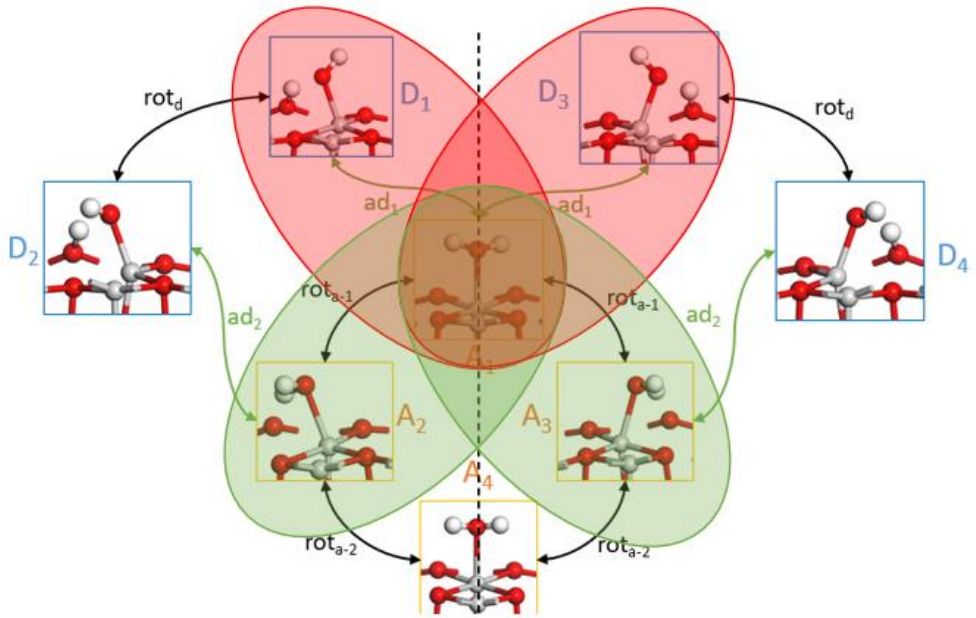


Identified structures and **intuitively** plausible transitions

3. Concatenate the dataset (containing E, Forces, and positions)
4. AMS with Active learning
5. Re-fit the force field
   → Rate can be estimated with AMS with MLFF without active learning

II. VASP MACHINE LEARNING FORCE FIELD

$A_1 \rightarrow D_1 D_3$ Dissociation



with N_{rep} = 200 and M_{real} = 10

DFT

$k_{A_1 \rightarrow D_1 D_3} = \quad (1.64 \pm 1.59)\ 10^9\ s^{-1}$

MLFF

$k_{A_1 \rightarrow D_1 D_3} = \quad (2.76 \pm 3.81) 10^9\ s^{-1}$

with N_{rep} = 800 and M_{real} = 10

$k_{A_1 \rightarrow D_1 D_3} = \quad (2.43 \pm 1.15)\ 10^9\ s^{-1}$

Test MAE forces =76 meV/Å
Test MAE energy = 20 meV
1000 configurations randomly drawn
from sampled reactive trajectories

CONCLUSIONS

AMS on H_2O / (100) γ-Al_2O_3

- AMS was implemented with AIMD, using basic ML tools allows to identify the necessary function to estimate rates and sample paths
- MLFF-MD and DFT-MD AMS rates are consistent
- MLFF used in prediction mode drastically reduces the computational cost
- Current implementation of AMS with VASP limits the application of active learning

  →Restart does have an important cost for the active learning