



HAL
open science

Timescale prediction of complex multi-barrier pathways using flux sampling molecular dynamics and 1D kinetic integration: Application to cellulose dehydration

Pierre-Louis Valdenaire, R.J.M. Pellenq, Franz Joseph Ulm, Adri C.T. van Duin, Jean-Marc Leyssale

► To cite this version:

Pierre-Louis Valdenaire, R.J.M. Pellenq, Franz Joseph Ulm, Adri C.T. van Duin, Jean-Marc Leyssale. Timescale prediction of complex multi-barrier pathways using flux sampling molecular dynamics and 1D kinetic integration: Application to cellulose dehydration. *Journal of Chemical Physics*, 2020, 152 (2), pp.024123. 10.1063/1.5126391 . hal-03018519

HAL Id: hal-03018519

<https://hal.science/hal-03018519>

Submitted on 27 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Timescale prediction of complex multi-barrier pathways using flux sampling molecular dynamics and 1D kinetic integration: application to cellulose dehydration

Pierre-Louis Valdenaire,¹ Roland J. M. Pellenq,^{1,2} Franz J. Ulm,^{1,2} Adri C. T. van Duin,³ and Jean-Marc Leyssale^{1,4}

¹*CNRS/MIT/Aix Marseille Univ. joint lab "MultiScale Materials Science for Energy and Environment", Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA^{a)}*

²*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA*

³*Department of Mechanical and Nuclear Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*

⁴*Univ. Bordeaux, CNRS, Bordeaux INP, ISM, UMR 5255, F-33400, Talence, France^{b)}*

ABSTRACT

Reactive molecular dynamics simulations, especially those employing acceleration techniques, can provide useful insights on the mechanism underlying the transformation of buried organic matter, yet, so far, it remains extremely difficult to predict the timescales associated to these processes at moderate temperatures (*i.e.* when such timescales are considerably larger than those accessible to MD). We propose here an accelerated method based on flux sampling and kinetic integration along a 1D order parameter that can considerably extend the accessible timescales. We demonstrate the utility of this technique in an application to the dehydration of crystalline cellulose at temperatures ranging from 1900 K to 1500 K. The full decomposition is obtained at all temperatures apart from $T=1500$ K, showing the same distribution of the main volatiles (H_2O , CO and CO_2) as recently obtained using replica exchange molecular dynamics. The kinetics of the process is well fitted with an Arrhenius law with $E_a = 93$ kcal/mol and $k_0 = 9 \times 10^{19} \text{ s}^{-1}$, which are somehow larger than experimental reports. Unexpectedly, the process seems to considerably slow down at lower temperatures, severely departing from the Arrhenius regime, probably because of an inadequate choice of the order parameter. Nevertheless, we show that the proposed method allows considerable time sampling at low temperature compared to conventional MD.

I. INTRODUCTION

Reactive molecular dynamics (RMD) simulations based on empirical potentials like ReaxFF^{1,2} allow investigating the dynamics of complex reactive systems at a

fraction of the cost of, more accurate, DFT calculations³. Since the development of ReaxFF, an important field of application has been the transformation of organic products undergoing either lab pyrolysis - for the production of bio-sourced materials or fuels⁴⁻⁸ - or geological burial^{4,9-17}. In these studies, the pyrolysis of some form of organic matter (OM) - including coals^{9,11,12,15,17}, immature kerogen^{14,16}, wood derivatives like lignin and cellulose⁵⁻⁸, or other fossil OM^{10,13} - is simulated at either constant or increasing temperature and the distributions of created byproducts (gas or liquid) are analyzed and discussed as a function of the type of organic matter^{9,13,16}, the temperature^{4,8-10,12-14,17} or heating rate⁶, or the time^{5,11,13,17}. Some authors report detailed chemical mechanisms (*i.e.*, sequences of individual chemical reactions) of the early pyrolysis stages^{4-7,9,10,12,14}. Few studies also report on rate law predictions, either for the entire pyrolysis process or for some important steps, based on timescale estimates at several temperatures and the assumption of first order kinetics (Arrhenius law)^{4,13}.

Because reactive MD simulations are typically limited to timescales of the order of a nanosecond, simulations are usually performed at much higher temperatures, generally around or above 2000 K, than those prevailing in geochemistry or in pyrolysis experiments. Geochemical OM conversion generally takes place over million years at about 400 K while pyrolysis experiments are run for minutes up to 700 K. It is however relatively unclear how this huge gap in temperature between experiments and simulations may affect the results, especially the distribution of products in such different thermodynamic conditions.

Recently, an accelerated scheme, the replica exchange molecular dynamics (REMD) method¹⁸, was used by Atmani *et al.*, together with ReaxFF, to investigate the degradation of cellulose under geological conditions¹⁹. REMD also uses reactive events obtained at high temperature, however, those are connected to the low temperature replica, the one of interest, on a statistical me-

^{a)}Electronic mail: plvalden@mit.edu

^{b)}Electronic mail: jean-marc.leyssale@u-bordeaux.fr

chanics basis. This is achieved by exchanging the configurations of adjacent replicas (in terms of temperature) using a Metropolis scheme ensuring that each replica progressively converges towards its own equilibrium state. REMD thus allows for a much faster convergence of the low temperature replicas than what would be obtained in conventional MD. Using this technique Atmani *et al.* showed that the degradation of cellulose proceeds in two steps. First, cellulose filaments almost entirely desintegrate into very small molecular entities (counting less than twelve C atoms) while producing a massive amount of water and to a lesser extent CO and CO₂. Second, a macromolecular carbonaceous phase (kerogen/coal) grows and progressively matures - *i.e.* evolves from H/O rich aliphatic to C rich aromatic - while producing additional H/O rich molecules¹⁹.

The work of Atmani *et al.*, is the only simulation work on the topic which operates under actual geological conditions (*i.e.* low temperature). It is also the only one, to our knowledge, to report the full conversion of OM into fluid and char. The composition of the fluid phase is consistent with pyrolysis experiments²⁰. The char evolution, both in terms of composition (H:C *vs* O:C ratios) and structure is also consistent with knowledge from natural samples and pyrolysis experiments²¹⁻²³. However, these simulations do not allow predicting any time-based information (timescale or rate of the different steps), nor producing a real dynamic pathway, *i.e.* a typical sequence of reactive events at the considered temperature.

In what follows we propose a computational framework, based on flux sampling MD trajectories and a 1D kinetic integration, to investigate the dynamics of macroscopic transformations in condensed phases involving multiple reversible barrier crossings. This method is then applied to investigate the fragmentation, or dehydration, of cellulose at moderate temperatures, *i.e.* at temperatures at which the process cannot be simulated using conventional MD simulations.

II. PATH SAMPLING METHODS

Rare event sampling methods are used to simulate physical or chemical transitions between well-defined states, separated by high free energy barriers compared to the thermal energy of the systems. In such situations, the transitions occur too slowly to be observed on the timescales of regular MD simulations. Several methods have been proposed in the literature to address this difficulty, from the pioneering work of Olender and Elber²⁴ to the maxflux method²⁵, the activation-relaxation technique (ART)²⁶, the hyperdynamics²⁷, parallel replicas dynamics (PRD)²⁸ and temperature accelerated dynamics²⁹ methods of Voter and co-workers, metadynamics³⁰, milestoning³¹ and transition path sampling (TPS)³² based methods including transition interface sampling (TIS)³³ and forward flux sampling (FFS)^{34,35}. While some of these techniques require mod-

ifications of the dynamical equations^{24,25}, and hence of the used MD package, or of the potential energy surface^{27,30}, some, on the contrary rely on the use of short, unperturbed MD trajectories^{28,31,33,34} and are thus somehow easier to implement. Amongst them, the most straightforward is certainly the parallel replica dynamics (PRD) method²⁸. In the PRD method, M MD simulations (called replicas) starting at the same initial macrostate yet with slightly different microstates - for instance different distributions of velocities - are run in parallel until a reaction is detected in one of them. Noting τ the time at which the transition is obtained in the successful replica, $t = M\tau$ is a properly drawn escape time for the considered transition time. Repeating the calculation with different initial conditions allows to determine the average escape time $\langle t \rangle$ and reaction rate $k = \langle t \rangle^{-1}$. Also, PRD is an efficient and robust method to investigate a sequence of barrier crossings for which an estimate of the global timescale is simply the sum of the individual escape times. However, the acceleration is limited, *i.e.* the total physical time simulated is the sum of time simulated on the different replicas. The gain in timescale is thus simply the number of parallel replicas used, for instance two orders of magnitude for 100 replicas.

Another class of accelerated method, derived from Transition Path Sampling (TPS)^{32,36}, consists in sampling preferentially the region of phase space separating reactants and products rather than wasting computer resources in sampling these long-lived states. Of particular interest is the Forward Flux Sampling (FFS) method^{34,37,38}.

As other TPS-based methods, FFS requires that the initial state A and final state B of the system are unambiguously defined. In addition, FFS requires some order parameter λ defined as a series of states λ_i evolving monotonically from λ_0 ($\equiv A$) to λ_n ($\equiv B$). The rate constant of the entire transition ($\lambda_0 \rightarrow \lambda_n$) can thus be expressed as

$$k^{\lambda_0 \rightarrow \lambda_n} = \Phi_{0,1} \cdot P_{\lambda_n|\lambda_1} \quad (1)$$

where $\Phi_{0,1}$ is the flux from λ_0 to λ_1 , meaning the number of trajectories leaving A and crossing the first interface (λ_1) per time and volume unit, and $P_{\lambda_n|\lambda_1}$ is the probability that a trajectory that left A , reaches B without coming back to A . This quantity can be decomposed in a product of probabilities to cross each interface from the previous one, without passing through A :

$$P_{\lambda_n|\lambda_1} = \prod_{i=1}^{n-1} P_{\lambda_{i+1}|\lambda_i}. \quad (2)$$

By splitting a large (diffusive) barrier into simple steps, it is possible to simulate extremely rare processes with only few hundreds or thousands MD trajectories of few ps each. FFS has been a very popular technique in the past years to investigate phase transitions such as the capillary evaporation of water³⁹ or crystal nucleation in supercooled liquids, for which the rate is of the order of 1

$s^{-1}m^{-3}$ ^{40,41}. This represents a major timescale bridging considering that such rates are obtained from a total of about 1 μs of MD simulation of a volume of about 10 nm^3 . However, by construction, FFS is only adapted to the case of a unique free energy barrier. At each step i , it is required that the different replica reach quickly (within a few ps) either the next interface ($i+1$) or return to the initial state (A). Complex reactive systems, presenting rough (free) energy landscapes for which the system may end up trapped for long times at i or at any intermediate state j ($0 < j < i$) cannot efficiently be treated with FFS.

Finally, the so-called **milestoning methods**, by Elber and coworker^{31,42–44}, are also particularly appropriate to investigate rates associated to rare events. Despite some variability in its implementations, milestoning, as FFS or TIS, is founded on the spatial separation of the transition in a series of discrete states, called milestones, along a reaction coordinate or order parameter λ separating reactants and products. The system is then simulated at each milestone λ_i to investigate its distribution (*i.e.* the distribution of states over all the hypersurface of internal coordinates normal to λ). Starting from these states, simulations are launched to determine the distributions of local first passage times (LFPTD) $K_{ij}(t)$ for the transition from λ_i to λ_j ($j = \pm 1$) where t is time. Assuming that the system is in the initial state at $t = 0$ and that the final state is absorbing (configurations reaching the final state will stick there), expressions for the probability to make a transition to milestone λ_i at t , $Q_i(t)$ and the probability to find the system at milestone λ_i at time t , $P_i(t)$ were derived by Faradjian and Elber³¹. Shalloway and Faradjian⁴² then showed that the mean first passage time at the absorbing final state λ_f can be obtained as

$$\langle t_1(f) \rangle = \int_t \frac{\partial P_f}{\partial t} dt \quad (3)$$

The transition rate is then simply given by $\langle t_1(f) \rangle^{-1}$. It has to be noted that a major difference between milestoning and transition path sampling methods (TIS or FFS) is that milestones are sampled independently from each other. While it allows for parallelizing over both time and space, it implies a memory loss at the milestones^{31,45} whereas trajectories in path sampling methods are continuous.

III. METHOD

Following the idea of the FFS and milestoning methods, we propose a new computational scheme that is also based on describing the macroscopic transformation along a 1D reaction coordinate, or order parameter, λ . This transformation is also splitted into a finite number of pieces, or **milestones**, not necessarily known *a priori*, which, in a reactive setup, consists in a series of long-lived minima λ_i along the reaction path (see Figure 1). The

approach presented here consists first into determining the forward and backward fluxes at each λ_i using parallel reactive molecular dynamics simulations. The resulting fluxes are then integrated in a second step to determine the first passage times at each λ_i , and hence, the transition rate. Two approaches, a stochastic one and an analytic one, are proposed for the integration stage.

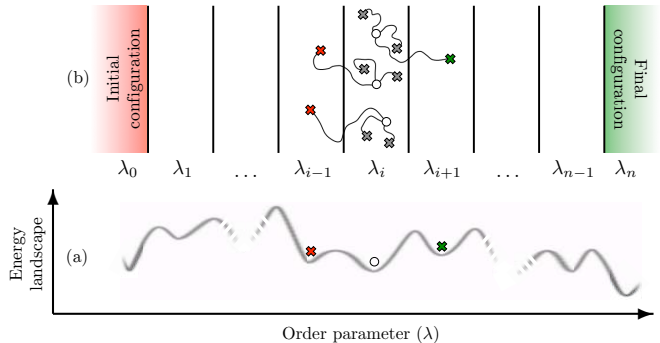


Figure 1. Schematic description of (a) an energy landscape divided in (b) a series of long-lived states λ_i separated by high energy barriers. λ_0 and λ_n represent the initial and final states, in red and green, respectively. At state λ_i , a series of MD simulations (or replicas) are run for a given time τ using different initial configurations and velocity distributions, hence producing different trajectories. Probabilities of reaching the next state (λ_{i+1}), returning to the former state (λ_{i-1}) or remaining at the current one in time τ can then be estimated. The trajectories represented in the figure have either different initial configurations (circles), obtained from different trajectories at the former step, or different initial velocity distributions, and hence trajectories, from the same initial configuration.

A. Flux sampling

As shown schematically in Figure 1, flux sampling consists in determining the respective probabilities that a system located in the long-lived state λ_i either reaches state λ_{i+1} ($P_{\lambda_i \rightarrow \lambda_{i+1}}(\tau)$), returns to state λ_{i-1} ($P_{\lambda_i \rightarrow \lambda_{i-1}}(\tau)$) or remains in state λ_i ($P_{\lambda_i \rightarrow \lambda_i}(\tau)$) after a MD trajectory of duration τ . As indicated in Figure 1, these probabilities can be obtained by running a series of n_i simulations of duration τ starting from a set of representative configurations of step λ_i . A set of trajectories is obtained either by selecting different initial configurations obtained from successive trajectories at the former state, or by assigning different velocity distributions to a given initial configuration (NB: we define here a configuration as the set of atomic coordinates).

Assuming that successive steps are independent from each other, that is that the system stays sufficiently long in λ_i to rethermalize (lose the memory of the formerly visited state), the probability to have a transition from λ_i to λ_{i+1} in a time shorter than τ is linked to its rate

constant $k^{\lambda_i \rightarrow \lambda_{i+1}}$ according to

$$P_{\lambda_i \rightarrow \lambda_{i+1}}(\tau) = 1 - e^{-k^{\lambda_i \rightarrow \lambda_{i+1}}\tau} \simeq \frac{N_i^{\lambda_{i+1}}}{n_i} \quad (4)$$

where n_i is the total number of replicas and $N_i^{\lambda_{i+1}}$ is the number of replicas that reached λ_{i+1} . A similar expression for the probability to have a backward transition

from λ_i to λ_{i-1} is given by:

$$P_{\lambda_i \rightarrow \lambda_{i-1}}(\tau) = 1 - e^{-k^{\lambda_i \rightarrow \lambda_{i-1}}\tau} \simeq \frac{N_i^{\lambda_{i-1}}}{n_i} \quad (5)$$

where $k^{\lambda_i \rightarrow \lambda_{i-1}}$ is its rate constant and $N_i^{\lambda_{i-1}}$ is the number of replicas that returned to λ_{i-1} . Finally, the probability of having no transition is:

$$P_{\lambda_i \rightarrow \lambda_i}(\tau) = 1 - P_{\lambda_i \rightarrow \lambda_{i+1}}(\tau) - P_{\lambda_i \rightarrow \lambda_{i-1}}(\tau) = -1 + e^{-k^{\lambda_i \rightarrow \lambda_{i+1}}\tau} + e^{-k^{\lambda_i \rightarrow \lambda_{i-1}}\tau} \simeq \frac{N_i^{\lambda_i}}{n_i} \quad (6)$$

where $N_i^{\lambda_i} = n_i - N_i^{\lambda_{i+1}} - N_i^{\lambda_{i-1}}$ is the number of trajectories that remained at λ_i .

To summarize, forward and backward fluxes for all λ_i are computed according to the following algorithm:

-
- 1 - Run n_i MD replicas from state λ_i with different initial configurations and/or different Maxwell-Boltzmann velocity distributions
 - 2 - Stop all replicas after time τ
 - 3 - Compute $P_{\lambda_i \rightarrow \lambda_{i+1}}(\tau)$; $P_{\lambda_i \rightarrow \lambda_{i-1}}(\tau)$; $P_{\lambda_i \rightarrow \lambda_i}(\tau)$
 - 4 - Distribute the successful $N_i^{\lambda_{i+1}}$ final configurations of step λ_i to serve as n_{i+1} initial configurations at step λ_{i+1}
-

We describe in what follows how the set of probabilities described above can be integrated to determine the average first passage time $\langle t_1(\lambda_i) \rangle$ at each step λ_i .

B. Stochastic integration

A very straightforward integration approach is based on constant time step kinetic Monte Carlo (KMC) tra-

jectories. Starting from λ_0 at time $t = 0$, a stochastic dynamical trajectory with time step Δt can be produced by iterating the following algorithm:

-
- 1 - Draw a random number α in $[0 : 1]$
 - 2 - Update the time according to $t = t + \Delta t$
 - 3 - Update the current value of the order parameter λ according to:

$$\begin{aligned} \lambda(t) = \lambda_{i-1} & \quad \text{if} & \quad \alpha \leq P_{\lambda_i \rightarrow \lambda_{i-1}}(\Delta t) \\ \lambda(t) = \lambda_i & \quad \text{if} & \quad P_{\lambda_i \rightarrow \lambda_{i-1}}(\Delta t) < \alpha \leq P_{\lambda_i \rightarrow \lambda_{i-1}}(\Delta t) + P_{\lambda_i \rightarrow \lambda_i}(\Delta t) \\ \lambda(t) = \lambda_{i+1} & \quad \text{if} & \quad \alpha > P_{\lambda_i \rightarrow \lambda_{i-1}}(\Delta t) + P_{\lambda_i \rightarrow \lambda_i}(\Delta t) \end{aligned}$$
 - 4 - if first passage at λ_i then $t_1(\lambda_i) = t$
 - 5 - if $\lambda(t) = \lambda_n$ then stop, else go to 1
-

Within this scheme, it is assumed that Δt is small enough to reproduce correctly the residence time in each state (in other words that $P_{\lambda_i \rightarrow \lambda_i}(\Delta t) \gg P_{\lambda_i \rightarrow \lambda_{i-1}}(\Delta t) + P_{\lambda_i \rightarrow \lambda_{i+1}}(\Delta t)$). In general, choosing the length of the MD trajectories (τ) as time step for the integration Δt is a reasonable option. Statistical averages of first passage times $\langle t_1(\lambda_i) \rangle$ can be obtained by averaging over independent KMC trajectories (that is, using different series of random numbers).

C. Analytic integration

The stochastic integration, despite its simplicity, may become numerically demanding for some energy landscape showing a stiffly increasing profile (*i.e.*, for which $P_{\lambda_i \rightarrow \lambda_{i-1}}(\tau) \gg P_{\lambda_i \rightarrow \lambda_{i+1}}(\tau)$). To overcome this difficulty, we propose an analytical expression of $\langle t_1(\lambda_i) \rangle$ based on the equilibrium distribution of replicas on the coarse-grained energy landscape. A practical way is to

divide the phase space (along the reaction coordinate) in subsets sub_i of increasing dimension from the initial state only sub_0 , to the full phase space sub_n , as schematized in Fig. 2. The first passage time $\langle t_1(\lambda_i) \rangle$ is thus the sum of the escape times from the successive subsets leading to state i and hence, of the inverse of the rate constants for leaving the successive subsets:

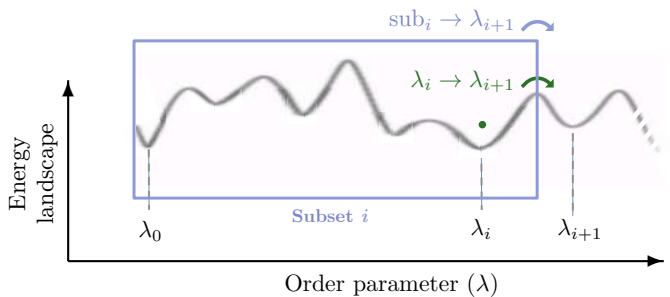


Figure 2. Schematics illustrating the difference between the escaping from state λ_i or subset i (sub_i) to state λ_{i+1} , associated to rates $k^{\lambda_i \rightarrow \lambda_{i+1}}$ and $k^{\text{sub}_i \rightarrow \lambda_{i+1}}$, respectively.

At equilibrium, the escape rate from subset j simply reads:

$$k^{\text{sub}_j \rightarrow \lambda_{j+1}} = P_{\lambda_j | \text{sub}_j} \cdot k^{\lambda_j \rightarrow \lambda_{j+1}} \quad (8)$$

where $P_{\lambda_j | \text{sub}_j}$ is the equilibrium probability of having the system in state j within subset j . From Eq. 7 and Eq. 8 we get the final expression for $\langle t_1(\lambda_i) \rangle$:

$$\langle t_1(\lambda_i) \rangle = \sum_{j=0}^{i-1} \frac{1}{k^{\text{sub}_j \rightarrow \lambda_{j+1}}} \quad (7)$$

$$\langle t_1(\lambda_i) \rangle = \sum_{j=0}^{i-1} \frac{1}{P_{\lambda_j | \text{sub}_j} \cdot k^{\lambda_j \rightarrow \lambda_{j+1}}} \quad (9)$$

Recalling that

$$\frac{\partial P_{\lambda_l | \text{sub}_j}}{\partial t} = P_{\lambda_{l-1} | \text{sub}_j} \cdot k^{\lambda_{l-1} \rightarrow \lambda_l} + P_{\lambda_{l+1} | \text{sub}_j} \cdot k^{\lambda_{l+1} \rightarrow \lambda_l} - P_{\lambda_l | \text{sub}_j} \cdot (k^{\lambda_l \rightarrow \lambda_{l-1}} + k^{\lambda_l \rightarrow \lambda_{l+1}}) \quad (10)$$

and assuming equilibrium $\frac{\partial P_{\lambda_l | \text{sub}_j}}{\partial t} = 0$ and normalization $\sum_{l=0}^j P_{\lambda_l | \text{sub}_j} = 1$ allows determining $P_{\lambda_j | \text{sub}_j}$ and hence $\langle t_1(\lambda_i) \rangle$.

D. Comparison with FFS and milestoneing

The proposed method constructs a tree of configurations connecting the initial state to the final state with new branches initiated at all considered values of the order parameter (milestone). In a similar fashion to FFS it allows producing true dynamical trajectories. However, the sampling quality is uneven with respect to λ as, while backward moves are accounted for to compute fluxes, their final configurations are not considered in describing the system at λ_{i-1} (*i.e.* no simulations are launched from these configurations). A very practical feature of the method is that, unlike FFS or milestoneing, for which all flux sampling simulations need to either terminate at the next interface or return back to the initial state (FFS), or at the next or previous milestones (milestoneing), flux sampling in our method is of limited duration and only a small part of the simulations have to change their values of the order parameter, which is

very suited for reactive systems like the one considered in this work where simulations can be trapped within a milestone. This somehow reminds of the PRD method where only one successful replica is required. As in FFS, configurations at λ_i are only sampled when visited from a forward trajectory initiated at λ_{-1} . Also, again like in the FFS method, fluxes at λ_i can only be sampled after the calculations λ_{i-1} , which does not allow for parallelization of the simulation along λ , a major advantage of milestoneing. Regarding the integration, the analytical approach presented in section III C is similar in essence to the *QP* approach presented by Faradjian and Elber³¹.

E. Computational details

The proposed method is applied to a small crystal of cellulose containing four periodic cellulose fibers ($\text{C}_{12}\text{H}_{20}\text{O}_{10}$) arranged according to crystal structure I- β . A major difficulty of the proposed methodology is the

definition of the order parameter (or reaction coordinate) λ , which has to be an easy to compute one-dimensional parameter, while being able to describe the complex evolution of OM under pyrolysis. In this work we decided to focus on the initial stage of the pyrolysis process of cellulose which, according to REMD simulations, consists in the fragmentation of the fibers and the massive production of small molecules (mostly water)¹⁹. Therefore, we adopted the number of created (or released) molecules as the order parameter for the present work, namely the difference between the total number of molecules and the initial number (four). This means that $\lambda_i \rightarrow \lambda_{i+1}$ (respectively $\lambda_i \rightarrow \lambda_{i-1}$) represents an increase (resp. decrease) by one of the number of released molecules.

The flux sampling simulations use short MD simulations performed in the NPT ensemble to determine the sets of reaction rates required in the kinetic integration stage. As in former work by Atmani *et al.*¹⁹, for comparison purpose, a combination of parameters from the 2013 C-C ReaxFF⁴⁶ with C-H, C-O, H-H, H-O and O-O parameters from Refs 47,48 is used to describe interatomic interactions.

Even though flux sampling significantly accelerates the dynamics, we cannot expect here to access geological values for time and temperatures. As we show in the results section, direct MD sampling of cellulose dehydration can be performed at temperatures as low as about 1700 K. Therefore, we apply the flux sampling at temperatures ranging from 1900 K to 1500 K, by steps of 100 K. A consequence of these values for the temperature is that under geological pressures, even at the crystalline state, the system would certainly evaporate. We thus perform the simulations at a much higher value for the pressure, 2.5 GPa. This value was chosen as it gives for a temperature of 1900 K the same equilibrium volume for cellulose as the one obtained under geological conditions (423 K, 25 MPa). In the MD simulations, the Nosé-Hoover thermostat and Nosé-Hoover-Andersen barostat⁴⁹ are used to fix temperature and pressure, respectively. Time constants of 0.05 and 0.5 ps are used for temperature and pressure relaxations, respectively. Equations of motion are integrated with a velocity-verlet integration scheme, using a timestep of 0.1 fs.

Initial atomic velocities are drawn according to Maxwell-Boltzmann distributions at the considered temperatures. Care is taken to make sure that velocities of different replicas are drawn with different random number seeds, to make sure of having different trajectories, even for replicas sharing the same initial atomic positions. The duration of the MD runs was set to 10 ps, which on one side, is long enough to have negligible perturbation from the random distribution of velocities at the beginning of each iteration and, on the other side, is short enough to ensure a low probability of reaction and thus ensure that the probability of having successive reactive events in one MD run is very low. **Even though the flux sampling runs are relatively short (10 ps) with respect to the barostat frequency (0.5 ps), we verified that the**

computed virial pressures were within statistical errors ($\leq 2\%$) of the applied pressure.

At the end of a MD run, the geometry of the final configuration is minimized with a conjugate gradient method to ease the clustering of atoms into molecules. The latter is achieved using a standard cluster analysis scheme based on distance cutoffs of 1.95 (C-C), 1.5 (C-H), 2.0 (C-O), 1.05 (H-H), 1.4 (H-O) and 1.9 (O-O) Å. It allows determining the total number of molecules present in the system, and hence, the corresponding value of the order parameter $\lambda = 4 - n_{mol}$, where n_{mol} is the total number of identified molecular clusters. The number of parallel MD replicas per iteration is adjusted (by series of 100), in order to obtain at least five forward and backward events. Within this constraint we had to perform from 100 to 1400 replicas, depending on temperature and iteration, to meet the desired statistics. **The obtained number of successive attempts at λ_i , and thus of independent realizations at λ_{i+1} varies between 5 and 36, depending on λ_i and T. These configurations are then distributed evenly during the flux sampling simulations at λ_{i+1} . On average, the number of independent realizations increases from 5.4 (1500 K) to 21.3 (1900 K). Following the description in section III A, the fluxes at λ_i , $P_{\lambda_i \rightarrow \lambda_{i+1}}(\tau)$, $P_{\lambda_i \rightarrow \lambda_{i-1}}(\tau)$ and $P_{\lambda_i \rightarrow \lambda_i}(\tau)$ are obtained by dividing the numbers of forward, backward and unreacted trajectories, respectively, by the total number of attempts (see Eq. 4,5 and 6).**

Finally, the first chemical reaction from perfect periodic fibers appeared to be very rare at the considered temperatures, and, furthermore, does not lead to the creation of any extra molecule. The starting point in this work will thus be a cellulose crystal with an already broken C-O bond (i. e., the weakest in the considered system). The Arrhenius parameters for this initial reaction have been estimated by fitting from the results of direct MD at higher temperatures (1800, 2000 and 2100 K): $E_a \approx 55$ kcal/mol and $k_0 \approx 10^{14}$ s⁻¹. This reaction is excluded in the following results.

The KMC integration is performed with a constant timestep of 10 ps (*i.e.*, equal to the length of MD simulations in the flux sampling stage). Average properties are obtained by averaging over 4×10^4 independent KMC trajectories. All the atomistic simulations, including MD and energy minimizations were performed with LAMMPS⁵⁰. One CPU unit per replica was used throughout this work and a hundred replicas were run in parallel during the flux sampling simulations.

IV. RESULTS

Fig. 3 shows the forward and backward rate constants obtained by applying the flux sampling method to the cellulose crystal at five temperatures ranging from 1500 K to 1900 K. As can be seen, all reported rates are in the 10^9 to 3×10^{10} s⁻¹ range which corresponds to reaction times varying from 30 ps to a ns. Note that by

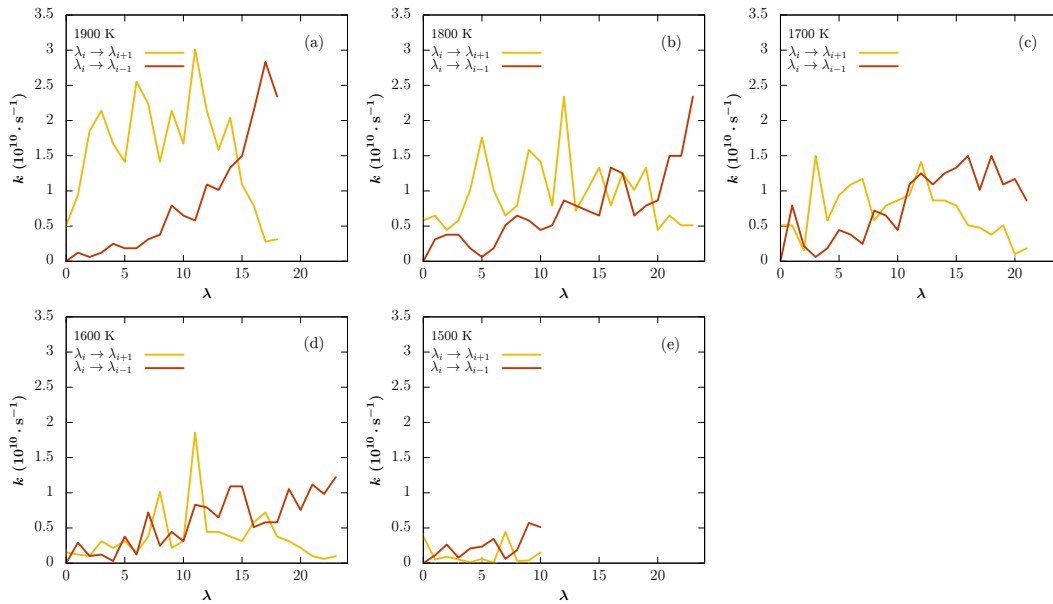


Figure 3. Evolutions of the forward (yellow) and backward (red) rate constants with the number of released molecules (order parameter λ) at (a) 1900 K, (b) 1800 K, (c) 1700 K, (d) 1600 K and (e) 1500 K.

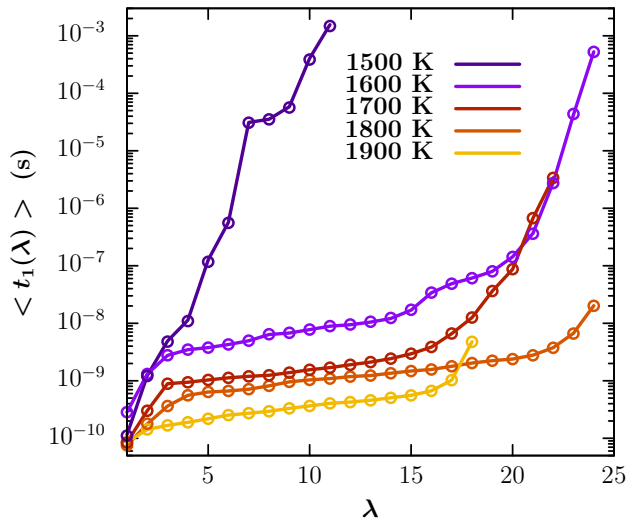


Figure 4. First passage time $\langle t_1 \rangle$ as a function of the number of released molecules (λ) obtained for temperatures ranging from 1500 K to 1900 K. Symbols: stochastic integration; lines: analytic integration (Eq. 9).

definition, the backward rate at $\lambda = 0$ is null, and thus not considered here. Overall, the reaction rates decrease with decreasing temperature. Despite some necessarily large error bars, unquantified here due to insufficient amount of data, the forward rates are generally larger than the backward rates for most reported values of the order parameter, thus indicating a spontaneous process (*i.e.* that the system's free energy decreases with increasing λ). Conversely, at the final values of λ we system-

atically observe that the backward rate is the largest, indicating an increase in free energy with increasing λ .

Fig. 4 shows the evolutions of the first passage time $\langle t_1 \rangle$ with the number of released molecules as computed using the stochastic and analytic integration methods for all the investigated temperatures. First, we note that the two methods lead to absolutely identical results, confirming that the equilibrium hypothesis, on which the analytic method is based, actually holds. Second, we observe that albeit some initialization at low λ values, $\langle t_1 \rangle$ evolves as a power law of λ for λ in the range 3 - 16 (note that data at 1500 K are incomplete due to high computational cost at this temperature). Above 16 produced molecules, $\langle t_1 \rangle$ quickly increases with λ , indicating a significant slowing down of the production of small molecules.

Fig. 5 shows the initial cellulose fiber and a configuration at $\lambda = 20$ of the flux sampling method performed at 1600 K. One can note a significant volume expansion (of 35 to 45 %), due to the production of fluid. Interestingly, and as obtained at much lower temperature using the REMD technique (see Atmani et al.¹⁹), the cellulose crystal has been entirely fragmented into relatively small molecular species, among which are found a high proportion of water and to a lesser extent CO and CO₂.

Fig. 6 splits the data of Fig. 4 into five panels (one per temperature) and add the main contributions to λ in terms of the average numbers of H₂O, CO and CO₂ molecules. It shows that for all temperatures, and as expected from Ref. 19, the process starts with the production of water in the early stages. The water production then quickly accelerates when the system starts releasing CO and CO₂. Apart from the incomplete data

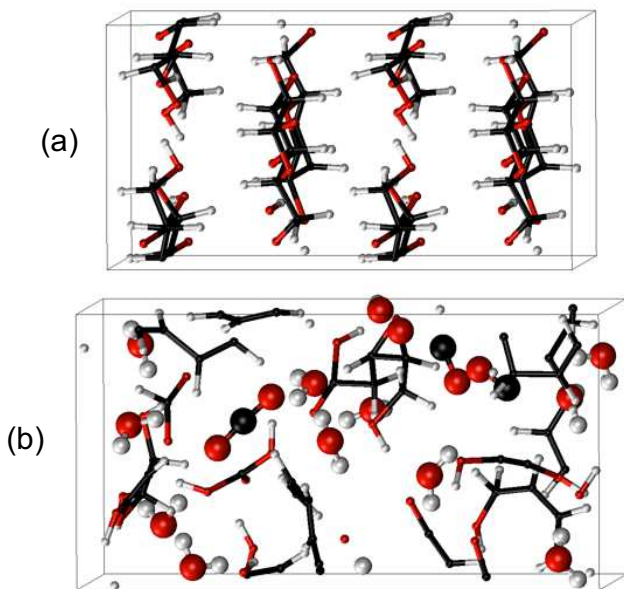


Figure 5. Snapshots of (a) the initial cellulose fiber with main axis normal to the screen (top) and (b) a configuration at $\lambda = 20$ at 1600 K (bottom). Common small molecules (H_2O , CO and CO_2) are highlighted with larger spheres, boxes represent the periodic boundary conditions and the color code is carbon (black), oxygen (red) and hydrogen (white).

at 1500 K, one observes that the production of small molecules plateaus, or at least its rate significantly decreases, above a total number of around 16. At this stage, in addition to the extremely small molecules (with zero or one C atom), some short oxygen-rich fragments and few hydrocarbons (C_6H_6 , C_5H_7 , C_5H_6) are still present in the system. Interestingly, the composition of the system near the plateau marking the end of the dehydration process, is very similar to the one obtained at 423 K using REMD¹⁹, which at equivalent system size would read: 13 H_2O , 2.4 CO and 1.6 CO_2 . Identifying the average first passage time required to produce 13 water molecules can thus provide a definition for the characteristic time of dehydration.

The evolution of the dehydration time with temperature is shown in Fig. 7, together with the corresponding Arrhenius fit. As can be seen the dehydration time increases by more than two orders of magnitude, from $\sim 3 \times 10^{-10}$ s to $\sim 7 \times 10^{-8}$ s, when T decreases from 1900 to 1600 K. These data can be fitted with an Arrhenius rate law with activation energy of 93 kcal/mol and pre-exponential factor of $\approx 9 \times 10^{19}$ s^{-1} . We note that both activation energy and pre-exponential factor are significantly larger than published data for cellulose pyrolysis, with typical average values of about 55 kcal/mol and 10^{14} s^{-1} , for E_a and k_0 , respectively (see the detailed review of experimental data in Ref. 51). Moreover, while the fitted rate law predicts a dehydration time slightly lower than 10^{-6} s at $T = 1500$ K, the simulation data in Fig. 4 show that dehydration remains incomplete after 10^{-3}

s at this temperature, clearly showing that our method fails at lower temperatures.

To better quantify the issue with the proposed method, we compare in Fig. 8 the evolution of λ with $\langle t_1(\lambda) \rangle$ obtained with the proposed method and brute force MD. As can be seen, first passage times at given λ are generally lower with the brute force MD and the difference between the two methods increases when T decreases. Note that we could not produce the brute force MD data at T lower than 1700 K, due to high computational cost.

Explaining why the method fails at simulating dehydration at low temperatures is not straightforward, however, there are good reasons to consider that it all stems from the definition of the 1D order parameter. Indeed, the choice made for the definition of λ , *i.e.* a global increase in the number of molecules, may be overconstraining. This implies that the dehydration process goes backward when the total number of molecules decreases. Inspection of some $\lambda_i \rightarrow \lambda_{i-1}$ events in the flux sampling simulations at 1500 K shows that this is not necessarily the case. Indeed, many events actually correspond to reactive events within the fluid, for instance the combination of a formaldehyde molecule with a water molecule to form methanediol. Such a reaction corresponds to the maturation of the created fluid which accompanies dehydration. It is wrongly accounted as a backward transition because of the definition of the order parameter. While these states are properly accounted for in the direct MD simulations, they are discarded from the reactive paths in the accelerated method as configuration obtained from backward moves are rejected (they are not accounted for in the description of λ_{i-1}). Indeed, as other accelerated methods (FFS and TIS for instance), the present method produces a tree of reaction paths, meaning that many possible λ_{i+1} states are considered when progressing forward but that backward moves are assumed to return to the limited set of already visited configurations at λ_{i-1} . Therefore, the predicted rates are obtained from a limited subset of the possible reactive pathways, for instance those that do not consider maturation within the fluid, and hence are necessarily underestimated. This is a consequence of the fact that the adopted, system-specific, order parameter does not satisfy the assumption of being monotonously increasing from the initial to the final state. From the results in Fig. 8, showing that the results mostly depart from brute force MD at large λ , we can guess that fluid maturation mostly become important close to the end of the dehydration process.

In order to check for possible size effects we repeated the calculations at 1700 K with systems of 336 and 504 atoms obtained by replicating the cellulose crystal along the smallest cell dimension (the vertical direction in Fig. 5). Fig. 9 shows the dehydration ratio - the average number of produced water molecules $n_{\text{H}_2\text{O}}$ normalized by the expected amounts of water $n_{\text{H}_2\text{O},\text{REMD}}$ at the end of the dehydration process (from REMD results), respectively 13, 26 and 39 molecules for system sizes of 168, 336 and 504 atoms, respectively - as a function of first passage

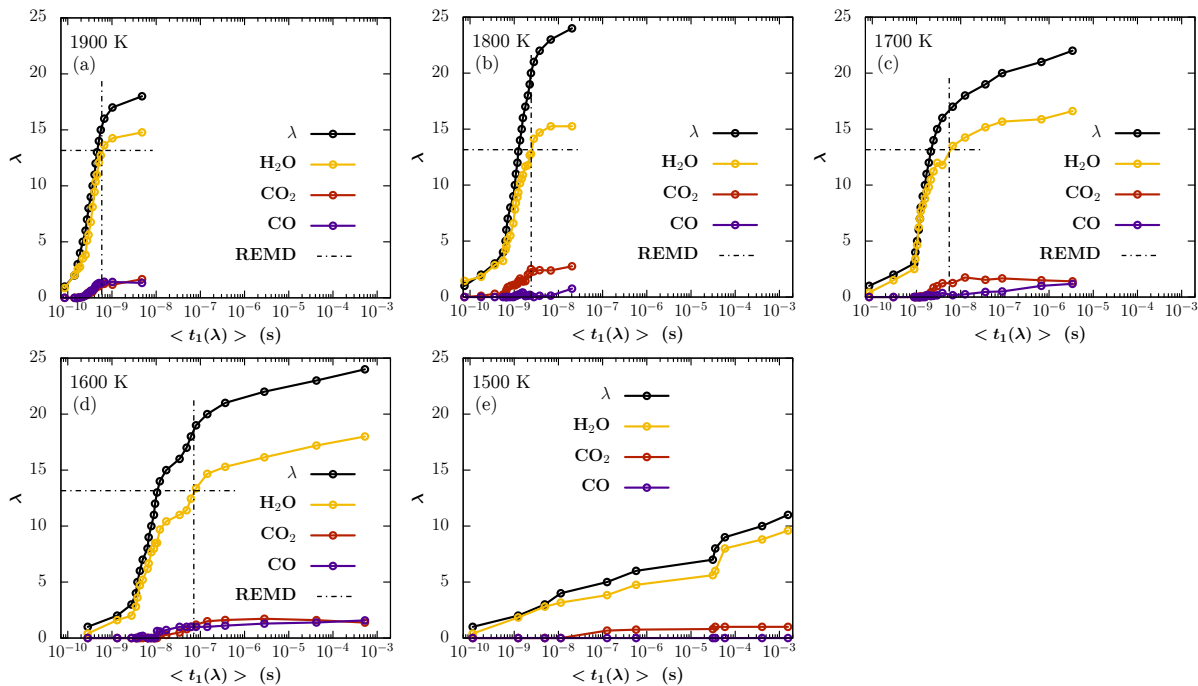


Figure 6. Evolution of λ with $\langle t_1(\lambda) \rangle$ with added average numbers of H₂O, CO and CO₂ molecules formed during the process at (a) 1900 K, (b) 1800 K, (c) 1700 K, (d) 1600 K and (e) 1500 K. Dashed lines represent the corresponding number of water molecules, 13 at equivalent system size, obtained at the end of the dehydration stage in a REMD simulation at 423 K¹⁹.

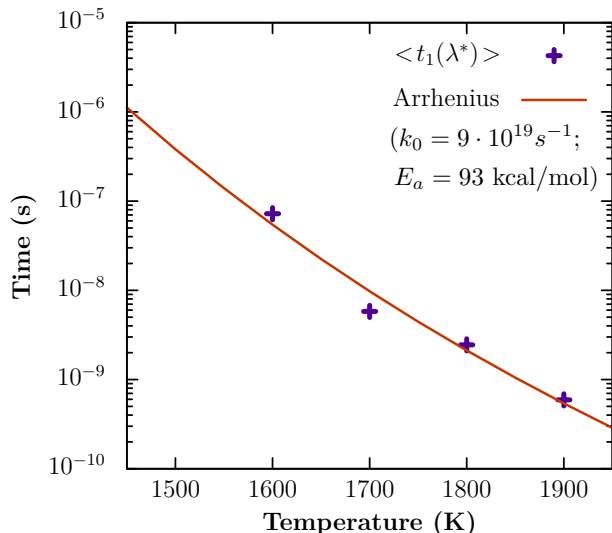


Figure 7. Evolution of the dehydration time with temperature. Crosses: simulation data; line: Arrhenius fit.

time. As can be seen, the obtained results are very similar for the three system sizes, showing no systematic deviation with size even if not totally superimposed due to statistical errors. The obtained dehydration times, obtained by interpolating $\langle t_1 \rangle$ at $\frac{n_{\text{H}_2\text{O}}}{n_{\text{H}_2\text{O}, \text{REMD}}} = 1$ are within 3-6 ns. We note however that the computational cost increases linearly with the number of atoms as the

number of molecules produced, and hence of milestones, is extensive.

Finally, Fig. 10 shows the physical times that we were able to simulate at each temperature (*i.e.*, the first passage time at the largest λ_i displayed in Fig. 4) compared to the effective simulated MD times over all replicas and per replica. The considerable acceleration obtained with the proposed integrated flux sampling method is obvious, especially at reduced temperature. Note that the purple and yellow curves would correspond to the times accessible to PRD and serial MD simulations, respectively, within the same wall clock time.

V. CONCLUSION

We have proposed a novel method based on flux sampling and 1D kinetic integration aimed at enhancing the reach of current reactive MD simulations by orders of magnitude. As a case study, the method has been applied to the decomposition (dehydration) of cellulose upon pyrolysis in confined conditions, *i. e.* at constant temperature, pressure and mass. The number of independent molecular clusters was chosen as the order parameter describing the progress of the process. Using this method we were able to produce a tree of molecular trajectories connecting the initial cellulose crystal to a myriad of fully decomposed configurations at various moderate temperatures, *i. e.* temperatures at which the process could not, or at least hardly, be simulated with standard MD simu-

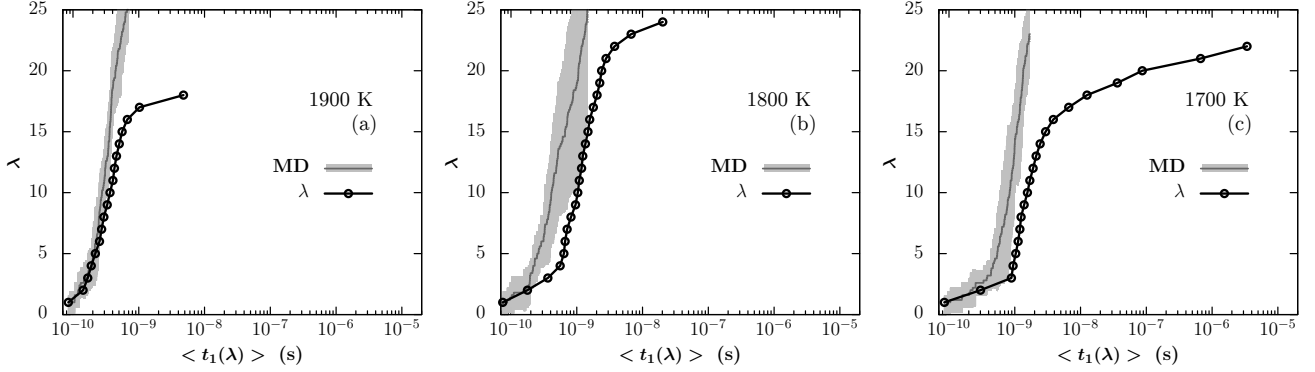


Figure 8. Comparison of the evolution of λ with $\langle t_1(\lambda) \rangle$ as obtained with the proposed method and brute force MD simulations at (a) 1900 K, (b) 1800 K and (c) 1700 K. In grey are shown the standard deviation amongst five independent MD runs at each temperature.

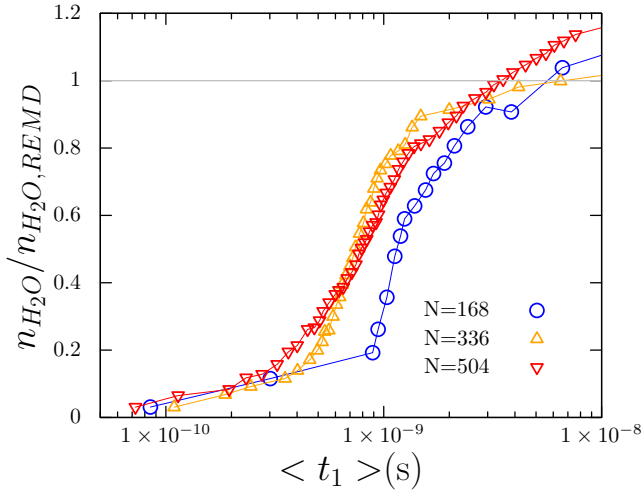


Figure 9. Dehydration ratio $n_{H_2O}/n_{H_2O,REMD}$ as a function of first passage time obtained at 1700 K for systems of 168, 336 and 504 atoms. n_{H_2O} and $n_{H_2O,REMD}$ are the number of water molecules in the simulations and at the end of the dehydration process (from REMD), respectively.

lations, demonstrating the acceleration capability of the method. Interestingly, the proportions of some important volatile molecules produced during the process, namely H_2O , CO and CO_2 were found in excellent agreement with results from replica exchange MD simulations¹⁹. We have also been able to estimate the characteristic times required for dehydration down to 1600 K, for which a value of about 100 ns was obtained, however, at lower temperatures, the simulation results depart significantly from the fitted Arrhenius law at higher temperatures and the dehydration timescale could not be estimated. Comparison with brute force MD simulations at higher temperatures also show an overestimation of the predicted times. While the accuracy and statistical significance of the presented results could be improved with enhanced sampling, at the price of a much higher computational

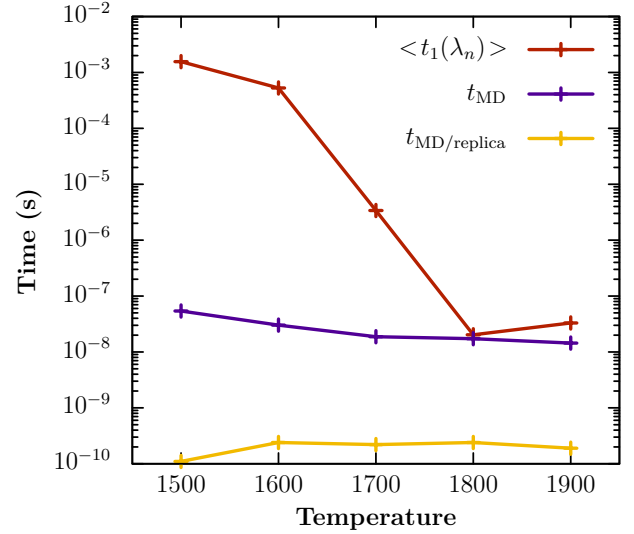


Figure 10. Comparison, at every temperature, of (red) the computed physical time to reach the final state in Fig. 4, (purple) the total simulated MD time (summed over replicas) and (yellow) the simulated MD time per replica.

cost, the main reason for the discrepancy between the accelerated and direct MD results on the dehydration times is certainly due to the choice of the order parameter λ . We have shown that the order parameter used in the simulations is not appropriate as, for instance, it does not allow to capture fluid maturation effects, which appears to accelerate dehydration. Choosing an appropriate order parameter for such complex systems, which are both reactive and highly disordered (low symmetry) is extremely difficult. Some techniques, like the maxflux algorithm²⁵, exist to define on the fly the best order parameter, yet they have never been applied, to our knowledge, to such complex systems. Although further work on the definition of a more suitable order parameter is definitely required to solve the considered problem (cel-

lulose dehydration), the methodology presented in this work can easily be transferred to other systems and especially those for which it is easier to identify a transition coordinate. This method can certainly compete with existing ones like FFS or milestoneing, especially for systems showing rough energy landscapes for which the proposed flux sampling strategy may prove to be more efficient.

ACKNOWLEDGEMENTS

This work was supported by the Game-Changer Shell project enabled through MIT Energy Initiative and by CNRS and Aix-Marseille University foundation (AMIDEX). It was granted access to the HPC resources of [TGCC/CINES/IDRIS] under the allocation 2016-A0040910065 made by GENCI (Grand Equipement National de Calcul Intensif). Some of the MD simulations presented here were also performed using the facilities of the Mésocentre de Calcul Intensif en Aquitaine (MCIA).

REFERENCES

- ¹A. van Duin, S. Dasgupta, F. Lorant, and W. Goddard, "ReaxFF: A reactive force field for hydrocarbons," *J. Phys. Chem. A* **105**, 9396–9409 (2001).
- ²T. P. Senftle, S. Hong, M. M. Islam, S. B. Kylasa, Y. Zheng, Y. K. Shin, C. Junkermeier, R. Engel-Herbert, M. J. Janik, H. M. Aktulga, T. Verstraelen, A. Grama, and A. C. T. van Duin, "The reaxff reactive force-field: development, applications and future directions," *npj Comput. Mater.* **2**, 15011 (2016).
- ³H. M. Aktulga, S. A. Pandit, A. C. T. VanDuin, and A. Y. Grama, "Reactive molecular dynamics: numerical methods and algorithmic techniques," *J. Sci. Comput.* **34**, C1–C23 (2012).
- ⁴Y. Zhang, X. Wang, Q. Li, R. Yang, and C. Li, "A reaxff molecular dynamics study of the pyrolysis mechanism of oleic-type triglycerides," *Energy Fuels* **29**, 5056–5068 (2015).
- ⁵M. Zheng, Z. Wang, X. Li, X. Qiao, W. Song, and L. Guo, "Initial reaction mechanisms of cellulose pyrolysis revealed by reaxff molecular dynamics," *Fuel* **177**, 130–141 (2016).
- ⁶T. Zhang, X. Li, X. Qiao, M. Zheng, L. Guo, W. Song, and W. Lin, "Initial mechanisms for an overall behavior of lignin pyrolysis through large-scale reaxff molecular dynamics simulations," *Energy Fuels* **30**, 3140–3150 (2016).
- ⁷T. Zhang, X. Li, and L. Guo, "Initial reactivity of linkages and monomer rings in lignin pyrolysis revealed by reaxff molecular dynamics," *Langmuir* **33**, 11646–11657 (2017).
- ⁸S. C. Rismiller, M. M. Groves, M. Meng, Y. Dong, and J. Lin, "Water assisted liquefaction of lignocellulose biomass by reaxff based molecular dynamic simulations," *Fuel* **215**, 835 – 843 (2018).
- ⁹E. Salmon, A. C. van Duin, F. Lorant, P.-M. Marquaire, and W. A. Goddard III, "Early maturation processes in coal. part 2: Reactive dynamics simulations using the reaxff reactive force field on morwell brown coal structures," *Org. Geochem.* **40**, 1195 – 1209 (2009).
- ¹⁰E. Salmon, A. C. van Duin, F. Lorant, P.-M. Marquaire, and W. A. Goddard, "Thermal decomposition process in algaenan of botryococcus braunii race I. part 2: Molecular dynamics simulations using the reaxff reactive force field," *Org. Geochem.* **40**, 416 – 427 (2009).
- ¹¹F. Castro-Marciano, M. F. Russo, A. C. van Duin, and J. P. Mathews, "Pyrolysis of a large-scale molecular model for illinois no. 6 coal using the reaxff reactive force field," *J. Anal. Appl. Pyrolysis* **109**, 79–89 (2014).
- ¹²J.-H. Zhan, R. Wu, X. Liu, S. Gao, and G. Xu, "Preliminary understanding of initial reaction process for subbituminous coal pyrolysis with molecular dynamics simulation," *Fuel* **134**, 283–292 (2014).
- ¹³C. Zou, S. Raman, and A. C. T. van Duin, "Large-scale reactive molecular dynamics simulation and kinetic modeling of high-temperature pyrolysis of the gloeocapsomorpha prisca microfossils," *J. Phys. Chem. B* **118**, 6302–6315 (2014).
- ¹⁴X. Liu, J.-H. Zhan, D. Lai, X. Liu, Z. Zhang, and G. Xu, "Initial pyrolysis mechanism of oil shale kerogen with reactive molecular dynamics simulations," *Energy Fuels* **29**, 2987–2997 (2015).
- ¹⁵M. Zheng, X. Li, F. Nie, and L. Guo, "Investigation of overall pyrolysis stages for liulin bituminous coal by large-scale reaxff molecular dynamics," *Energy Fuels* **31**, 3675–3683 (2017).
- ¹⁶G. Pawar, P. Meakin, and H. Huang, "Reactive molecular dynamics simulation of kerogen thermal maturation and cross-linking pathways," *Energy Fuels* **31**, 11601–11614 (2017).
- ¹⁷M. Gao, X. Li, and L. Guo, "Pyrolysis simulations of fugo coal by large-scale reaxff molecular dynamics," *Fuel Process. Technol.* **178**, 197–205 (2018).
- ¹⁸Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.* **314**, 141–151 (1999).
- ¹⁹L. Atmani, C. Bichara, R. J.-M. Pellenq, H. Van Damme, A. C. T. van Duin, Z. Raza, L. A. Truffandier, A. Obliger, P. Kralert, F. J. Ulm, and J.-M. Leyssale, "From cellulose to kerogen: molecular simulation of a geological process," *Chem. Sci.* **8**, 8325–8335 (2017).
- ²⁰M. R. Hajaligol, J. B. Howard, J. P. Longwell, and W. A. Peters, "Product compositions and kinetics for rapid pyrolysis of cellulose," *Ind. Eng. Chem. Process Des. Dev.* **21**, 457–465 (1982).
- ²¹M. Monthieux, P. Landais, and J.-C. Monin, "Comparison between natural and artificial maturation series of humic coals from the mahakam delta, indonesia," *Org. Geochem.* **8**, 275–292 (1985).
- ²²F. Béhar and M. Vandenbroucke, "Chemical modelling of kerogens," *Org. Geochem.* **11**, 15–24 (1987).
- ²³M. Vandenbroucke and C. Largeau, "Kerogen origin, evolution and structure," *Org. Geochem.* **38**, 719 – 833 (2007).
- ²⁴R. Olender and R. Elber, "Calculation of classical trajectories with a very large time step: Formalism and numerical examples," *J. Chem. Phys.* **105**, 9299–9315 (1996).
- ²⁵S. Huo and J. E. Straub, "The maxflux algorithm for calculating variationally optimized reaction paths for conformational transitions in many body systems at finite temperature," *J. Chem. Phys.* **107**, 5000–5006 (1997).
- ²⁶G. T. Barkema and N. Mousseau, "Event-based relaxation of continuous disordered systems," *Phys. Rev. Lett.* **77**, 4358–4361 (1996).
- ²⁷A. F. Voter, "A method for accelerating the molecular dynamics simulation of infrequent events," *J. Chem. Phys.* **106**, 4665–4677 (1997).
- ²⁸A. F. Voter, "Parallel replica method for dynamics of infrequent events," *Phys. Rev. B* **57**, R13985–R13988 (1998).
- ²⁹M. R. Sorensen and A. F. Voter, "Temperature-accelerated dynamics for simulation of infrequent events," *J. Chem. Phys.* **112**, 9599–9606 (2000).
- ³⁰A. Laio and M. Parrinello, "Escaping free-energy minima," *Proc. Nat. Ac. Sci. USA* **99**, 12562–12566 (2002).
- ³¹A. K. Faradjian and R. Elber, "Computing time scales from reaction coordinates by milestoneing," *J. Chem. Phys.* **120**, 10880–10889 (2004).
- ³²C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, "Transition path sampling and the calculation of rate constants," *J. Chem. Phys.* **108**, 1964–1977 (1998).

- ³³T. S. van Erp, D. Moroni, and P. G. Bolhuis, "A novel path sampling method for the calculation of rate constants," *J. Chem. Phys.* **118**, 7762–7774 (2003).
- ³⁴R. J. Allen, P. B. Warren, and P. R. ten Wolde, "Sampling rare switching events in biochemical networks," *Phys. Rev. Lett.* **94**, 018104 (2005).
- ³⁵R. J. Allen, C. Valeriani, and P. R. ten Wolde, "Forward flux sampling for rare event simulations," *J. Phys.: Condens. Matter.* **21**, 463102 (2009).
- ³⁶C. Dellago, P. Bolhuis, and P. Geissler, "Transition path sampling methods," in *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, edited by M. Ferrario, G. Ciccotti, and K. Binder (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006) pp. 349–391.
- ³⁷R. J. Allen, D. Frenkel, and P. R. ten Wolde, "Simulating rare events in equilibrium or nonequilibrium stochastic systems," *J. Chem. Phys.* **124**, 024102 (2006).
- ³⁸R. J. Allen, D. Frenkel, and P. R. ten Wolde, "Forward flux sampling-type schemes for simulating rare events: Efficiency analysis," *J. Chem. Phys.* **124**, 194111 (2006).
- ³⁹S. Sharma and P. G. Debenedetti, "Evaporation rate of water in hydrophobic confinement," *Proc. Nat. Ac. Sci. USA* **109**, 4365–4370 (2012).
- ⁴⁰T. Li, D. Donadio, and G. Galli, "Nucleation of tetrahedral solids: A molecular dynamics study of supercooled liquid silicon," *J. Chem. Phys.* **131**, 224519 (2009).
- ⁴¹A. Haji-Akbari and P. G. Debenedetti, "Direct calculation of ice homogeneous nucleation rate for a molecular model of water," *Proc. Nat. Ac. Sci. USA* **112**, 10582–10588 (2015).
- ⁴²D. Shalloway and A. K. Faradjian, "Efficient computation of the first passage time distribution of the generalized master equation by steady-state relaxation," *J. Chem. Phys.* **124**, 054112 (2006).
- ⁴³A. M. A. West, R. Elber, and D. Shalloway, "Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide," *J. Chem. Phys.* **126**, 145104 (2007).
- ⁴⁴E. Vanden-Eijnden, M. Venturoli, G. Ciccotti, and R. Elber, "On the assumptions underlying milestoning," *J. Chem. Phys.* **129**, 174102 (2008).
- ⁴⁵K. Kuczera, G. S. Jas, and R. Elber, "Kinetics of helix unfolding: Molecular dynamics simulations with milestoning," *J. Phys. Chem. A* **113**, 7461–7473 (2009).
- ⁴⁶S. G. Srinivasan, P. Ganesh, and A. C. T. van Duin, "Development of a reaxff potential for carbon condensed phases and its application to the thermal fragmentation of a large fullerene," *J. Phys. Chem. A* **119**, 571–580 (2015).
- ⁴⁷Y. K. Shin, L. Gai, S. Raman, and A. C. T. van Duin, "Development of a reaxff reactive force field for the Pt-Ni alloy catalyst," *J. Phys. Chem. A* **120**, 8044–8055 (2016).
- ⁴⁸F. Tavazza, T. P. Senftle, C. Zou, C. A. Becker, and A. C. T. van Duin, "Molecular dynamics investigation of the effects of tip-substrate interactions during nanoindentation," *J. Phys. Chem. C* **119**, 13580–13589 (2015).
- ⁴⁹W. Shinoda, M. Shiga, and M. Mikami, "Rapid estimation of elastic constants by molecular dynamics simulation under constant stress," *Phys. Rev. B* **69**, 134103 (2004).
- ⁵⁰S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J. Comput. Phys.* **117**, 1–19 (1995).
- ⁵¹Y.-C. Lin, J. Cho, A. Tompsett, P. R. Westmoreland, and G. W. Huber, "Kinetics and mechanism of cellulose pyrolysis," *J. Phys. Chem. C* **113**, 20097–20107 (2009).