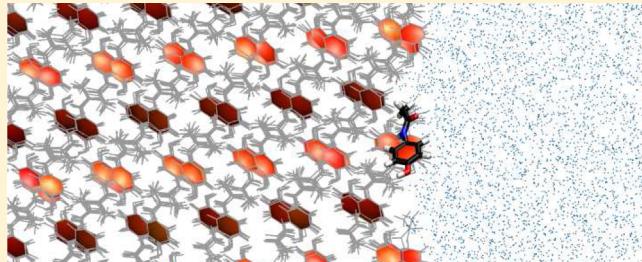


Data Filtering for Effective Analysis of Crystal–Solution Interface Molecular Dynamics Simulations

Ekaterina Elts, Maximilian M. Greiner, and Heiko Briesen*

Chair for Process Systems Engineering, Technische Universität München, 85354 Freising, Germany

ABSTRACT: Analysis of processes occurring at the solid-solution interface during crystal growth and dissolution simulations requires an effective way to detect rare, uncorrelated transitions from the liquid to the solid state or vice versa. Because of the oscillatory behavior of molecules, this is not a trivial problem. Usually, to take the thermal vibration and rotational flexibility of the molecules into account, the data (e.g., orientation, center of mass position) needed to determine the molecular state are averaged over some time interval. Then they are evaluated using some order parameters to classify the individual molecules as being either crystalline or in solution. In this case, the results can be very sensitive to the time interval, which is mostly chosen in some heuristic way. To suppress the problem of fast non-Markovian dynamics and to make the identification of the molecular state more reliable and robust, the application of a Kalman filter, optionally combined with a hysteretic approach, is proposed in this contribution. A scheme to estimate the filter parameters is introduced. To demonstrate the approach, simple and widely used order parameters based on the structural features of molecules are taken. The obtained results are clearly superior to those based on the data averaging technique and are important for the effective transition rates calculation as well as for the general analysis of the time evolution of interfaces.



1. INTRODUCTION

Dissolution and growth of crystals from solution are very important processes that are widely used in several industrial branches for the synthesis and application of pharmaceutical products, special chemicals, and so forth. The design of these processes would be greatly enhanced by an a priori prediction of crystal growth and dissolution rates. One of the greatest challenges in predicting the rates from theory alone is the atomistic nature of such processes. These involve the incorporation of individual atoms or molecules into the crystal lattice as well as separation from it, eventually making the information on such molecular level events necessary for studying crystallization and dissolution kinetics.

Given the ever increasing computational power, molecular dynamics (MD) simulations offer a promising platform to gather information on molecular behavior at the crystal–solution interface. In the past decade, a number of investigations of crystallization processes from solution using MD were reported.^{1–13} While in the majority of these works relatively small and simple particles are considered (LJ particles,¹ clathrate hydrates,^{2–4} urea molecules,^{5–8} glycine^{9–12}), the simulation of systems consisting of complex molecules with high degree of conformational flexibility still remains a challenge.¹⁴

Typically, the crystal growth rate decreases with increasing complexity of the molecule. It is more difficult for a complex molecule to arrange itself in the appropriate lattice structure.¹⁴ However, MD simulations alone cannot access time scales larger than several hundred nanoseconds with a reasonable computational effort. Recently reported MD simulations have

explored only very limited time scales from a few nanoseconds^{10,13} up to a maximum of 0.2 μ s.⁸ The computational time needed to carry out a typical 0.2 μ s MD simulation for the $5.5 \times 5.5 \times 12$ nm box filled with water and containing a $5.5 \times 5.5 \times 3$ nm urea crystal slab is about 15–20 days on 16 CentOS 5 Cray cores.⁸ The explored time scale is still far below the time scale of interest for actual process modeling. To bridge this gap and to simulate up to the microscale and beyond, where the theoretical results can be challenged experimentally and would be industrially relevant, the multiscale strategy by Piana and Gale^{5–7} can be used. According to their approach, MD simulations are carried out only to derive the rates for individual growth and dissolution events, like incorporation of a solute molecule into a kink or step site on a crystal surface. These rates are then used as inputs for kinetic Monte Carlo (kMC) simulations, which allow much larger time scales to be investigated, as only events important to changes in the macroscopic state of the system are considered.

The determination of the transition rates for all steps of growth or dissolution of the crystalline surface involves classifying the individual molecules as being either crystalline or in solution. Crystalline molecules can then be further subdivided in different substates according to their local coordination environment on the basis of the number of neighboring molecules of a given type.⁷

Classifying individual molecules needs the local crystalline order to be measured. The order parameters that can

Received: September 12, 2013

Published: March 6, 2014



discriminate between different states, for example, like the liquid/solution state and the solid, often tend to be system specific and the identification of an appropriate one is far from trivial in many cases.¹⁵ Thus, a significant number of approaches to differentiate one of the states from the other exists in the literature.^{1–5,8,12,16–24}

Steinhardt et al.¹⁶ defined local bond order parameters that are based on spherical harmonics analysis. These parameters depend only on the angles between the vectors to the neighboring particles and, therefore, are independent of a reference frame. Steinhardt order parameters, especially q_4 and q_6 , are widely used for simple systems such as monatomic fluids, highly symmetric molecules, or colloids.^{25–30} The modification of the method was proposed by Lechner and Dellago,¹⁷ as they noticed that thermal fluctuations smear out the order parameters distributions such that it may be difficult to distinguish local crystalline structures based on the Steinhardt bond order parameters. They increased the accuracy of the crystal structure determination by averaging over the bond order parameters of nearest neighbor particles.¹⁷ Similar symmetry-based order parameters have been defined and used for water and other small symmetric systems.^{2,4,18–21} However, as was mentioned by Santiso and Trout¹² and by Chen et al.,¹⁴ it is not easy in general to extend such order parameters to more complex systems, such as the ones encountered in typical industrial and technological applications.^{12,14}

Thus, many authors still follow the most simple and intuitive way and define the order parameters for crystallization by taking advantage of particular structural features of the molecule or crystal under consideration and choosing some relevant atomic coordinates or combination of coordinates.^{1,5,6,8,10} The identification of molecules as belonging either to the crystalline or to the liquid phase is then done on the basis of the known orientation of molecules^{5,6,10} or the number of nearest neighbors¹ in the crystalline structure. Because the molecules in the crystal are not completely at rest, tolerance parameters are usually used for the comparison with reference values to take thermal vibration and rotational flexibility into account.^{1,5,6} Additionally, orientations and positions are averaged over some time interval to avoid rapid back-and-forth transitions between different states caused by random fluctuations and not representing the events significant for the time evolution of interfaces.^{1,5,6} Such an approach, however, requires a substantial amount of trial and error to come up with good tolerance parameters and sampling interval. For example, Piana and Gale⁶ defined the state of urea molecule as crystal-like, if its average dipole moment orientation (during the 50 ps interval) was within 20° of the crystal *c* axis and had a standard deviation from the average orientation of less than 10°. The robustness of this definition with respect to the sampling interval (50–100 ps), the tolerance on the orientation (10–20°), and standard deviation (10–20°) was then tested. The definition was considered to be robust, as changes in the parameters within these ranges essentially yielded the same crystal-like molecules.⁶ Reilly and Briesen¹ considered the states of a Lennard-Jones solute particle to be defined by the number of nearest-neighbors that conform to the face-centered cubic (fcc) lattice. The number of neighbors of a particle was determined by averaging the position of the surrounding particles over a period of time and checking if any of the resulting averaged positions were within threshold of 0.95 Å of the correct position of a nearest

neighbor, as determined by the symmetry of the fcc lattice. Though the value of the threshold was also chosen by a trial-and-error method, the authors paid more attention to the choice of sampling interval and tried to avoid the counting of recrossing events. Namely, the time interval was chosen on the basis of the velocity autocorrelation function (VAF) of solute particles near the equilibrated interface as well as the VAF of solute particles in the bulk of the crystal slab from the same simulation. Only transitions where the particle remained in the new state for two consecutive time intervals were counted.¹

An alternative way to distinguish molecular behavior as solid-like or liquid-like is to use quantities that measure the degree of mobility of molecules in the phase, for example, the mean-squared deviation of molecular positions and orientations,³ the self-diffusion coefficient²² or the Lindemann index.^{23,24} The decision to label the molecules as solid-like or liquid-like is then made on the basis of a predetermined threshold,^{3,23,24} or two different thresholds can even be used to discriminate the stable states from transition intermediates.²² The molecular mobility, though being fundamental in determining a state label for molecules, can only be measured over a particular time window.³ As was demonstrated by Liang and Kusalik,³ long time measurements are preferred. However, in practical work, measurements from very long time segments may lose molecular behavior of interest; thus, an appropriate segment length should be chosen carefully to be short enough but at the same time yield good state discrimination.³ Moreover, as was stressed by Santiso and Trout,¹² many methods to study complex transitions require order parameters that are only functions of the coordinates of the atoms in the system at a given time, which limits the usefulness of functions that depend on time averaging.¹²

Other approaches^{8,12} depend solely on the coordinates of the atoms in the system at a given time and not on time averages. The measure of the degree of crystallinity developed by Salvalaglio et al.⁸ contains the product of two functions describing the local density and local order. The state of each molecule is defined at every time *t* in terms of such degree of crystallinity, which assumes values in the [0, 1] interval, where the lower end represents a molecule in the liquid phase and the upper end a molecule in the crystal. A method to construct more sophisticated order parameters suitable for the study of crystallization and polymorph transformations in molecular systems has been reported by Santiso and Trout.¹² The description of their set of order parameters is based on generalized pair correlation function containing all the variables relevant for the crystal structure. For any given molecular crystalline system this set accounts for distance order, bond orientational order, relative orientational order, and even internal degrees of freedom such as dihedral angles, and can be systematically defined for complex systems using information that can be obtained from simple MD simulations of the crystals.¹² Both approaches lead to a fraction of molecules characterized by a semicrystalline state, that is, by an intermediate value of degree of crystallinity.^{8,12} As observed by Salvalaglio et al.⁸ while analyzing the MD simulation of urea crystal growth, some of these “intermediate” particles were located on the crystal surface, representing the molecules interacting with the crystal lattice. Others were located either in solution, corresponding to short-lived partially ordered clusters of urea molecules, or in the crystal, representing random fluctuations of the crystalline molecules.⁸ Such diversity in behavior of “semicrystalline” particles makes these approaches

hardly applicable to classify molecules for transition rates calculation.

Thus, we can see that, despite a significant number of approaches being developed and used, problems still exist with the direct application of these for effective calculation of transition events and rates, where the main requirements are the distinct identification of molecular states and the detection of only rare, uncorrelated transition events. The strong oscillatory behavior of molecules leads to some intermediate values of order parameters, to overlaps in order parameter distribution (making the discrimination between states ambiguous), or to multiple recrossings at the boundary between two states (resulting in overcounting of transition events). The existing solution of the problem based on the averaging technique, as applied in some works,^{1,5–7} improves the situation but, at the same time, often leads to new problems like a dependence of the rate matrix on the chosen time interval.¹

The goal of this work is to present a novel approach that can easily be used in combination with different order parameters to solve the problems described above. This approach focuses on rare, uncorrelated transition events and makes the identification of the molecular states more reliable and robust. The “intermediate” states arising due to random fluctuations are avoided and the arbitrariness connected with the assignment of such to one or other state is reduced. The enhancement is achieved by employing a Kalman filter,³¹ optionally combined with a hysteretic approach,³² to eliminate oscillatory noise from the data needed for the molecular state determination. To demonstrate our approach, simple and widely used order parameters based on the structural features of molecules are taken. The resulting molecular states enable effective transition rates calculation and depend solely on the coordinates of the atoms in the system at a given time.

The rest of the paper is organized as follows. In the next section, the problem connected with strong fluctuations of molecules at the crystal surface affecting the calculation of transition events is illustrated. Then, in section 3, the way to solve this problem by applying Kalman filtering is described. In section 4, a method to estimate the filter parameters is proposed. The analysis results obtained with Kalman filtering are compared with those based on simple data averaging in section 5. Finally, in section 6, we make some concluding remarks about the presented approach. MD simulation details for the example described in the paper are given in the Appendix.

2. ANALYSIS OF MD TRAJECTORIES

The principal objective of MD simulations in the multiscale strategy by Piana and Gale⁵ is to obtain the rates for integration, disintegration, and diffusion processes on the crystal surfaces. The state transition rates in an MD simulation can be calculated as⁵

$$k_{A \rightarrow B} = \frac{1}{\Delta t} \frac{\langle n_{A \rightarrow B} \rangle}{\langle A \rangle} \quad (1)$$

where Δt is the simulation time averaged over, $\langle n_{A \rightarrow B} \rangle$ is the number of transitions from state A to state B , and $\langle A \rangle$ is the average number of molecules in state A .

The identification of molecular states A and B is usually done on the basis of order parameter(s). The process of moving from state A to state B is considered to be a rare event, which means

that on the time scale of atomic vibrations such a transition occurs only seldom. It is also assumed that the molecule stays relatively long in one state (compared to a vibrational period), and thus, there is no memory of how it got into this state. The probability to move from the present state A into the next state B is, thus, independent of whatever state preceded state A . Focusing on this Markovian state-to-state dynamics, that is, coarse-graining the time evolution to the discrete rare events, is the central idea behind a kMC simulation.³³ Thus, only such rare, uncorrelated events are of interest for MD analysis, as the observation of these allows one to report concisely the complex dynamics of crystal growth and dissolution.

The oscillatory behavior of molecules (especially of surface molecules interacting with the solution) leads to spurious recrossings at the boundary between two states, which severely complicate the reliable identification of significant transitions that are relevant over longer time scales. The analysis problem arising due to such fast non-Markovian dynamics becomes especially acute for systems consisting of complex molecules with the high degree of conformational flexibility. For such systems, more sophisticated ways to differentiate the solution and the crystal are needed,¹⁴ and advanced methods like accelerated dynamics (e.g., parallel-replica dynamics^{34,35}) or kMC often have to be applied. Thus, the accurate detection of transition events and effective calculation of transition rates play a particularly important role in this case.

To illustrate this, an explicit example from MD simulations of the most prominent paracetamol (PCM) crystal surface (100) in contact with water was chosen. The paracetamol molecule possesses a relatively high degree of conformational flexibility, which strongly complicates the assignment of states for surface molecules. This section specifically addresses trajectory analysis problems. The simulation details can be found in the Appendix.

To characterize a crystal growth or dissolution process, a scheme to divide molecules into crystal-like and solution-like must be introduced. In the present work, we define the crystal-like molecules as those having the orientations within a particular threshold from the orientations of the molecules from the reference unit cell. The reference unit cell is specified here as the initial relaxed unit cell used to construct the starting crystal structure for the MD simulation as described in the Appendix. Moreover, to properly validate the choice of the state, the number of crystal-like neighbors is calculated for each molecule. To be considered crystal-like, the molecule is required to have not only the correct orientation but also at least one crystal-like neighbor.

According to the number of neighbors, the crystal-like molecules are further subdivided into surface molecules and molecules in the bulk of the crystal. Given that a unit cell contains n molecules, for each molecule with a proper orientation $n - 1$ nearest neighbors in the same crystal plane (unit cell and the surrounding cells) and n in each of the planes directly above and below are checked. Thus, a molecule in the bulk should have from $2n$ to $3n - 1$ neighbors, and a surface molecule from 1 to $2n - 1$ neighbors. The neighbor search is controlled by the distance threshold parameter (DT), which defines the maximal allowed deviation of a molecular center of mass position from the position where a certain neighbor for the given molecule is expected. The upper limit of this parameter can be estimated as half of the minimal distance between any two molecules in the starting crystal structure, thus preventing the algorithm from considering false molecules as correct neighbors.

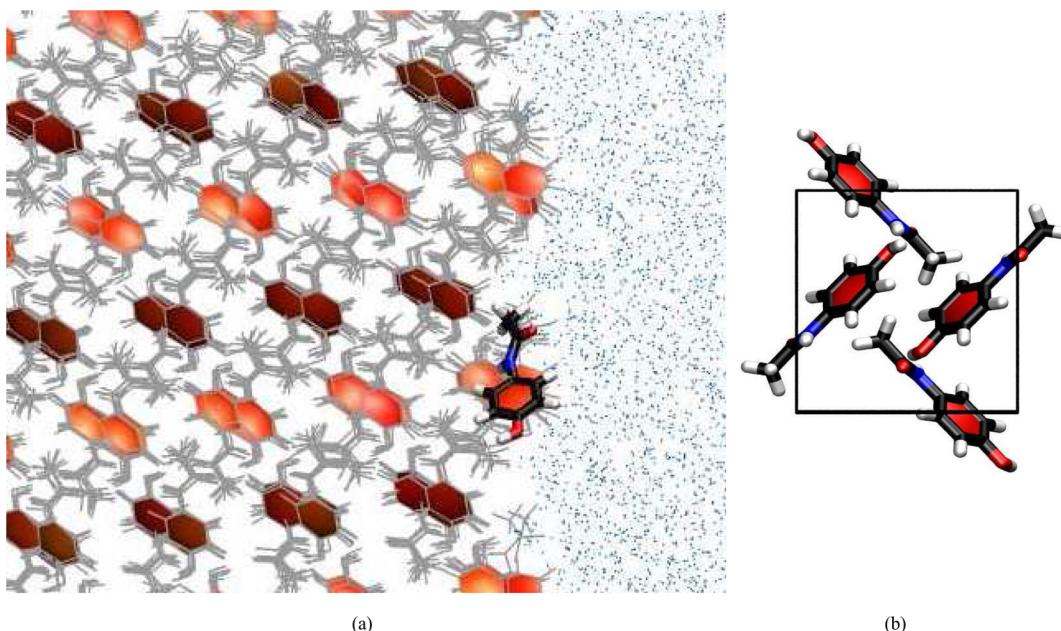


Figure 1. Orientation of a molecule of the (100) surface of a paracetamol (PCM) crystal dissolved in water (a) compared to the orientation of each of the four molecules from the reference unit cell (b) used to construct the starting crystal structure for the MD simulation.

All other molecules (with wrong orientation or without any crystal-like neighbors) are considered solution-like. The solution-like molecules that are closer to the crystal surface than the length of the reference unit cell vector c normal to this surface are identified as adsorbed.

To define molecular orientations, a coordinate system has to be fixed to the molecule. There are different possibilities to do that. For simplicity, a coordinate system fixed to the aromatic ring of a PCM molecule is used in our example. The frame is constructed by choosing the center of the aromatic ring as the origin, using the vector joining the center of the ring to the carbon atom of the ring, to which the hydroxyl group is attached, to define the x axis, and the vector normal to the aromatic ring plane to define the z axis. The y axis is then obtained as the cross product of the unit vectors in the x and z directions. The orientation of each PCM molecule from the simulation box (see Figure 1a) is compared to the orientation of each of the four molecules from the reference unit cell (see Figure 1b). Vector representation of molecular coordinate systems in global coordinate system (simulation box frame) is used. To calculate the orientation deviation, the angles between the basis vectors of a molecule under test and the reference molecule's basis vectors are evaluated. As only the relative value of the orientation deviation but not its direction is of interest, this approach is considered feasible. The orientation tolerance parameter (OT) defines the maximal allowed deviation angle for molecules to be detected as crystal-like and, thus, serves as a measure for the internal rotation flexibility of a molecule in a crystal lattice. In section 4, we will present a way to estimate the OT from a short preliminary MD simulation.

It should be noted that it is also possible to represent the molecular orientation in terms of quaternion or angle-axis coordinates. As the orientation fluctuations will lead to the fluctuations of the values corresponding to these notations, there should be no difference regarding the results and application of our approach.

Figure 2 shows the oscillatory behavior for a particular molecule of the (100) surface of a PCM crystal dissolved in

water. The deviation angles of the x , y , and z axes of the coordinate system for this molecule from the respective axes of four reference unit cell molecules are shown in Figure 2a–c. In Figure 2d the number of neighbors among crystal-like molecules for the given molecule is shown. The automatic identification of the molecular state based on the OT and DT parameters obtained as described further in section 4 is presented in Figure 2e.

From Figure 2a–c, it can be noticed that during the first 30 ns the deviations of the axes of the given molecule from the axes of all reference unit cell molecules oscillate significantly around some constant values. The deviations from the axes of the third reference unit cell molecule are less than the OT parameter. Thus, the molecule has a proper orientation corresponding to the orientation of the third molecule of the reference unit cell. From Figure 2d it can be seen that the molecule belongs to the surface, as it has three to four neighbors at that time. For the next 3 ns, the molecule is in solution, as can be noticed from the deviation plots (see Figure 2a–c). After that, it reintegrates into the crystal structure again but with the orientation corresponding to the fourth reference unit cell molecule, where it stays for the next 17 ns. Then, the molecule leaves the crystal and goes to solution again. After approximately 13 ns, it tries to integrate to the surface layer again but cannot find a correct orientation until the end of the simulation and, thus, is identified as adsorbed.

Due to strong fluctuations of the molecular position and orientation, numerous transitions between crystal-like substates with different number of neighbors (see Figure 2d) as well as from the crystal-like to the solution-like state (see Figure 2e) are registered. This causes the marker dots to sometimes form several parallel lines on these subplots. Consideration and calculation of all these transition events would lead to strong overestimation of transition rates. The number of such fast, non-Markovian transitions can be reduced by analyzing the data averaged over some time interval, as done by Piana and Gale.^{3–7} However, the resulting transition rates are very sensitive to the choice of the time interval.¹ For illustration,

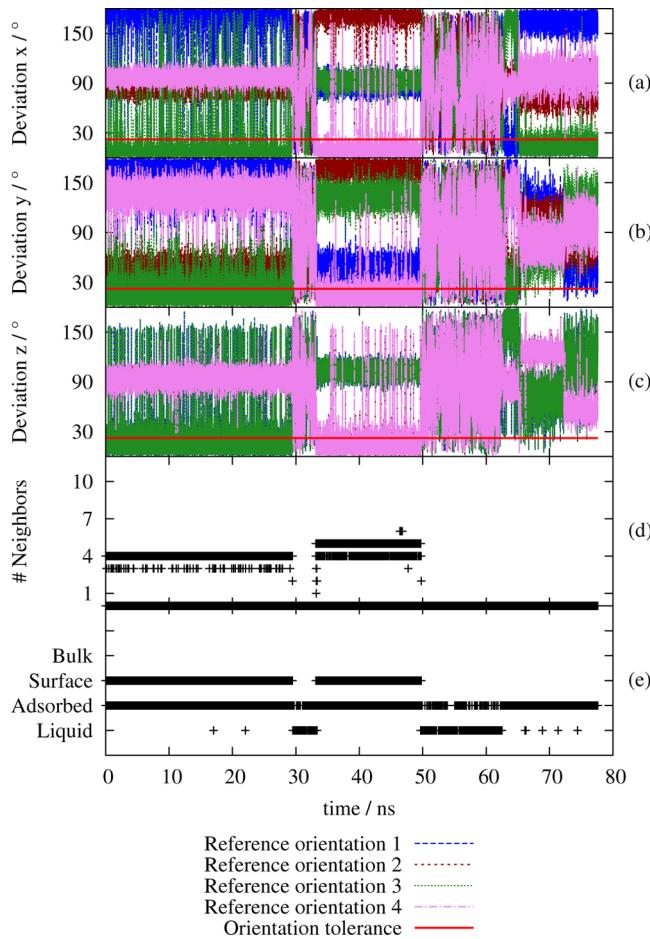


Figure 2. Unprocessed MD trajectory data saved every 2 ps and corresponding analysis results for a particular molecule of the (100) surface of a PCM crystal dissolved in water. (a)–(c) Deviation angles of the x , y , and z axes of the molecule from the respective axes of four reference unit cell molecules, correspondingly; (d) number of neighbors among the crystal-like molecules; (e) molecular state estimation. The analysis results (d) and (e) are based on the tolerance parameters $OT = 22.3^\circ$ and $DT = 1.954 \text{ \AA}$. The rapid back-and-forth transitions between different states cause the marker dots to form several parallel lines on (d) and (e).

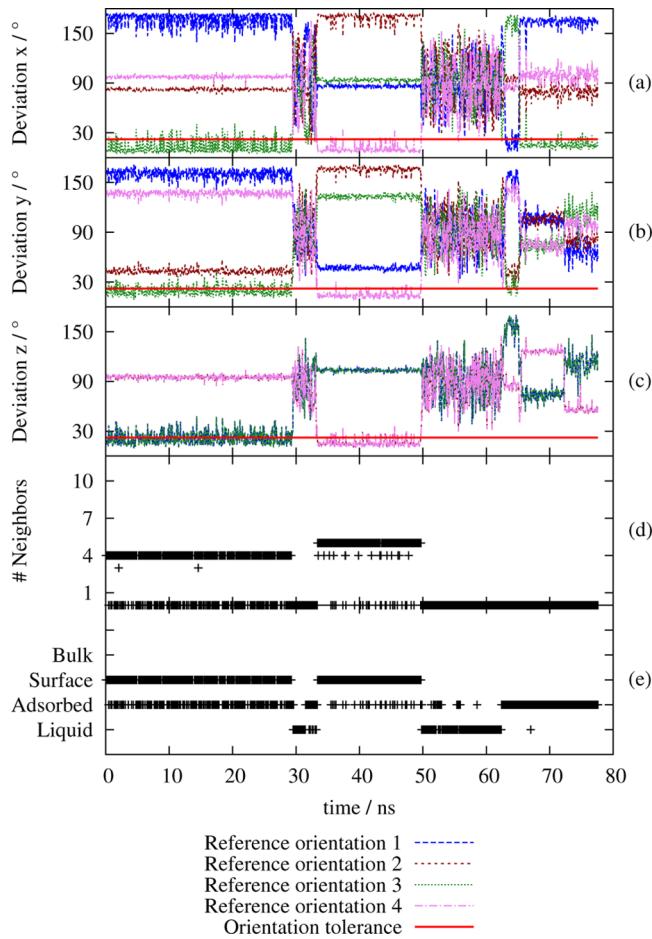


Figure 3. MD trajectory data averaged over 50 ps and corresponding analysis results for a particular molecule of the (100) surface of a PCM crystal dissolved in water. (a)–(c) Deviation angles of the x , y , and z axes of the molecule from the respective axes of four reference unit cell molecules, correspondingly; (d) number of neighbors among the crystal-like molecules; (e) molecular state estimation. The analysis results (d) and (e) are based on the tolerance parameters $OT = 22.3^\circ$ and $DT = 1.954 \text{ \AA}$. The rapid back-and-forth transitions between different states cause the marker dots to form several parallel lines on (d) and (e).

the results obtained for data averaged over 50 and 200 ps are presented in Figure 3 and Figure 4, respectively. It can be seen that the outcome is very different for these time intervals, with the best apparent results obtained for 200 ps. Although all non-Markovian transitions between the crystal-like substates with the different number of neighbors disappeared in this case, there are still many rapid back-and-forth transitions from crystal-like to solution-like states. The time interval used for averaging can be enlarged to suppress this non-Markovian dynamics. On the other hand, the selection of too large a time interval may lead to missing fast events that are important for the evolution of the surface structure and coordination states.

In the next section, it will be shown how the application of Kalman filtering to molecular trajectories allows faithful tracking only of the Markovian state-to-state dynamics without any significant loss of information.

3. KALMAN FILTERING OF ORIENTATION AND POSITION TRAJECTORIES

Kalman filtering³¹ is employed effectively in navigational and guidance systems, radar tracking, sonar ranging, and satellite orbit determination, to name just a few areas, as it not only cleans up the data measurements but also projects these onto the state estimate.³⁶ It is a recursive predictor-corrector data processing algorithm, which addresses the problem of trying to estimate the true (but unobservable) state of a system from noisy measurements, assuming that it evolves according to the transition equation³⁷

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + G\mathbf{u}_t + \mathbf{w}_t \quad (2)$$

where \mathbf{x}_t is the state vector containing the terms of interest for the system (e.g., position, velocity) at time t , \mathbf{u}_t is the vector containing any control inputs (e.g., steering angle, throttle setting), F is the state transition matrix which applies the effect of each system state parameter at time $t - 1$ on the system state at time t , G is the control input matrix which relates the optional control input \mathbf{u}_t to the state \mathbf{x}_t (e.g., applies the effect

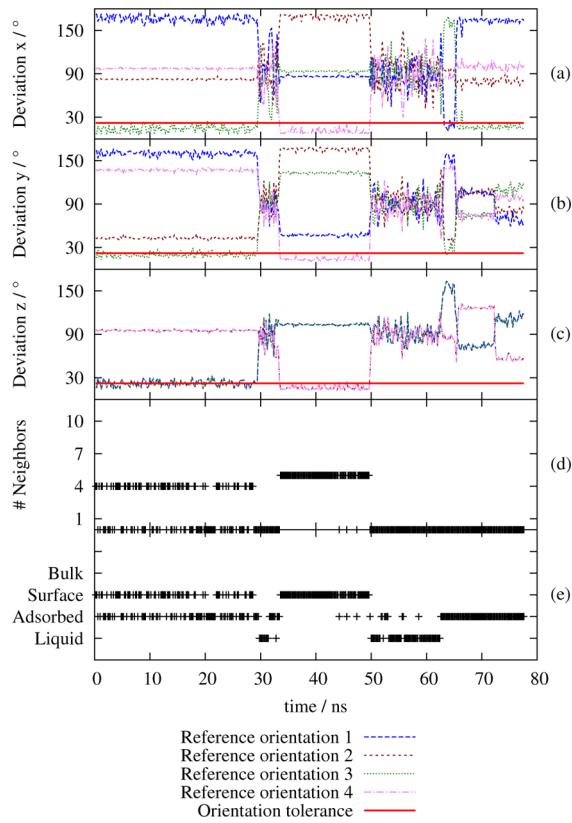


Figure 4. MD trajectory data averaged over 200 ps and corresponding analysis results for a particular molecule of the (100) surface of a PCM crystal dissolved in water. (a)–(c) Deviation angles of the x , y , and z axes of the molecule from the respective axes of four reference unit cell molecules, correspondingly; (d) number of neighbors among the crystal-like molecules; (e) molecular state estimation. The analysis results (d) and (e) are based on the tolerance parameters $OT = 22.3^\circ$ and $DT = 1.954 \text{ \AA}$.

of the throttle setting on the system velocity and position), and w_t is the vector containing the process noise terms for each parameter in the state vector.

Measurements (or observations) of the system are performed according to the model³⁷

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \quad (3)$$

where \mathbf{z}_t is the vector of measurements, \mathbf{H} is the transformation matrix that maps the state vector parameters into the measurement domain, and \mathbf{v}_t is the vector containing the measurement noise terms for each observation in the measurement vector.

The random variables w_t and v_t representing the process and measurement noise are assumed to be independent of each other, white, zero-mean, and with normal probability distributions³⁶

$$p(w_t) \sim N(0, \mathbf{Q}) \quad (4)$$

$$p(v_t) \sim N(0, \mathbf{R}) \quad (5)$$

where \mathbf{Q} and \mathbf{R} are the process noise covariance and the measurement noise covariance matrices, respectively.

The state vector \mathbf{x}_t in eq 2 has the Markov property, in that the latest value \mathbf{x}_{t-1} is all that is needed to make predictions.³⁸ Thus, in this work we propose to apply Kalman filtering to suppress the problem of fast non-Markovian dynamics. We

suggest to consider the fluctuations of orientation and position of each crystal molecule as noisy measurements of the “true” (corresponding to the Markovian molecular state) molecular orientation and position and to estimate these “true” values using the Kalman filter algorithm. On the basis of the obtained estimations, the assignment of molecular states at each time step is done.

It was already noticed by many authors that thermal motion of lattice molecules as well as fluctuations in their orientations can be approximately modeled by Gaussian functions.^{8,12,39} Hence, the Kalman filter model assumption about a zero-mean Gaussian distribution of measurement errors holds true in our case. In the case of a well-modeled, one-dimensional linear system with measurement errors drawn from a zero-mean Gaussian distribution, the Kalman filter has been shown to be the optimal estimator.^{31,37} Thus, the application of the Kalman filter should lead to the best possible estimate of the Markovian molecular state at each time step.

For each molecule we propose to filter the x , y , and z coordinates of its center of mass position as well as deviations of the x , y , z axes of the molecular frame from the respective axes of four reference unit cell molecules independently of each other. Thus, for each molecule, 15 scalar variables need to be estimated with the Kalman filter.

First, the scalar Kalman filter model must be constructed based on the eqs 2 and 3. The crystalline molecule should have a certain position and orientation; thus, we can assume that the system state \mathbf{x}_t in eq 2 does not change from step to step and $F = 1$. There is no control input, so $u_t = 0$. Fluctuations of orientation/position are considered as noisy measurements of the “true” orientation/position, so the measurement \mathbf{z}_t relates directly to the state \mathbf{x}_t and $H = 1$ in eq 3. Thus, our transition and measurement equations are simply

$$\mathbf{x}_t = \mathbf{x}_{t-1} + w_t \quad (6)$$

$$\mathbf{z}_t = \mathbf{x}_t + v_t \quad (7)$$

The obtained eqs 6 and 7 represent the well-known random walk plus noise model.^{38,40}

The Kalman filter estimates a process described by eq 6 by using a form of feedback control: the filter estimates the system state at some time and then obtains feedback in the form of noisy measurements.³⁶ Thus, the equations for the Kalman filter fall into two groups: predictor (time update) equations and corrector (measurement update) equations. Here, we just give the final equations obtained for the transition and measurement eqs 6 and 7, respectively. The detailed information how to derive these can be found in the literature.^{36,41} The predictor equations are responsible for projecting the current system state $\hat{\mathbf{x}}_{t-1}$ and the error variance P_{t-1} estimates forward in time to obtain a priori estimates $\hat{\mathbf{x}}_t^-$ and P_t^- for the next time step

$$\hat{\mathbf{x}}_t^- = \hat{\mathbf{x}}_{t-1} \quad (8)$$

$$P_t^- = P_{t-1} + Q \quad (9)$$

The corrector equations are responsible for the feedback, that is, incorporating a new measurement \mathbf{z}_t into the a priori estimate $\hat{\mathbf{x}}_t^-$ to obtain an improved a posteriori estimate $\hat{\mathbf{x}}_t$

$$K_t = P_t^- / (P_t^- + R) \quad (10)$$

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^- + K_t(\mathbf{z}_t - \hat{\mathbf{x}}_t^-) \quad (11)$$

$$P_t = (1 - K_t)P_t^- \quad (12)$$

where K_t is the so-called Kalman gain, minimizing a posteriori estimate error variance P_t .

After each predictor and corrector pair, the process is repeated with the previous a posteriori estimates used to predict the new a priori estimates.

In effect, the Kalman filter uses each new observation to update a probability distribution for the state of the system, with no need ever to refer back to any earlier observations. Indeed, to provide the best possible estimate of the “true” molecular orientation/position at each time step the information from two sources is used:³⁷ (1) prediction based on the last known “true” molecular orientation/position and (2) molecular trajectory data at a given time step. The first one is represented by Gaussian probability density function of the last estimate adjusted according to the system model. For the simple model described in this paper, this adjusted Gaussian has the same mean and the variance increased by the process noise variance Q (see eqs 8–9). The second is represented by Gaussian with the mean equal to the actual molecular orientation/position from the MD trajectory and the variance equal to the measurement noise variance R . The information from these two sources is combined by multiplying the two corresponding Gaussians together. This step can be shown to correspond to eqs 10–12.³⁷ The product of two Gaussian functions is another Gaussian, describing the obtained estimate. This is the key to the elegant recursive properties of the Kalman filter. This recursive nature is one of the very appealing features of the Kalman filter, as it makes practical implementations for real time applications much more feasible than implementations of filters, which are designed to operate on larger data sets.

It should be noted that the corrector eq 11 is very similar to an exponentially weighted moving average (EWMA) in its error-correction form³⁸

$$\hat{x}_t = \hat{x}_t^- + \lambda(z_t - \hat{x}_t^-) \quad (13)$$

where λ represents a smoothing parameter also called an EWMA weighting. Indeed, it can be shown that for the random walk plus noise model described here EWMA can be regarded as a (very simple) Kalman filter.³⁸ The Kalman gain K_t will converge to the same value as the EWMA weighting, which is a function of the signal-to-noise ratio s ³⁸

$$\lambda = 0.5(\sqrt{s^2 + 4s} - s) \quad (14)$$

where $s = Q/R$.

Still, we presented the general Kalman filter equations at the beginning of the section and used them to construct our model to provide the readers an insight that much more complex cases and models can be successfully described by the Kalman equations (e.g., considering external influence like electric fields for dipole orientations).

For the application of the Kalman filter, variance information on the process and measurement noise is needed (see eqs 4 and 5). The determination of these variance parameters Q and R from a short preliminary MD simulation will be discussed in the following section.

4. DETERMINATION OF FILTER PARAMETERS

The standard design methodology for Kalman filters requires not only a full description of the relation between the states x_t and measurements z_t but also a full description of the noise

affecting the states w_t and the measurements v_t .⁴² Both the measurement noise variance R and the process noise variance Q must be known. Poor values of Q and R lead to poor state estimation.⁴³

The process noise variance Q contributes to the overall uncertainty of the state estimate. Large Q means that the variations in the “true” state variables are assumed to be large, thus the Kalman filter tracks large changes in the data more closely than for smaller Q . The measurement noise variance R determines how much information from the measurement is used. If R is high, the Kalman filter considers the measurements as not very accurate. For smaller R , it follows the measurements more closely. Often superior filter performance can be obtained by just tuning the filter parameters Q and R . This would be similar to the heuristic search for a suitable time interval in the time-averaging approach. However, to avoid this level of arbitrariness in this work, we suggest a rational basis for choosing these parameters.

The measurement noise variance R is usually estimated prior to the operation of the filter from some off-line sample measurements.³⁶ To determine the measurement noise variance for the orientation R_o and the position R_p , as well as the tolerance parameters OT and DT, a short (80 ps) preliminary MD simulation of the (100) PCM/water interface was used (see Appendix). During such a short preliminary simulation, the crystal structure still stays stable, thus making it possible to take into account the strong fluctuations of the surface molecules. The resulting MD trajectory was taken to construct probability distributions of molecular orientations and molecular deviations from expected positions. These distributions are depicted in Figures 5 and 6, respectively.

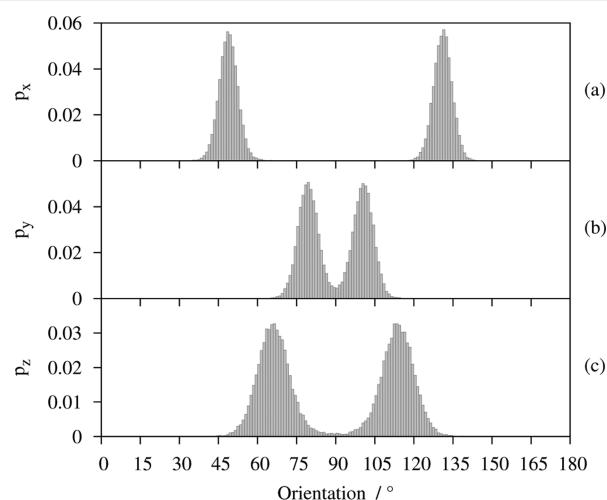


Figure 5. Probability distribution of molecular orientations calculated from a short preliminary 80 ps MD simulation of the (100) surface of PCM in water.

Figure 5a–c shows the distributions p_x , p_y , and p_z of the angles between the molecular frame axes constructed as defined in section 2 and the simulation box frame axes. It can be seen that the distributions are Gaussian-like with the mean angles corresponding to the values for the reference unit cell molecules. To estimate the OT parameter the widest p_z distribution (see Figure 5c) was considered. According to the empirical rule for a symmetric bell-shaped distribution,⁴⁴ almost all values (99.73%) lie within three standard deviations of the mean, giving a good estimation for tolerance parameter OT.

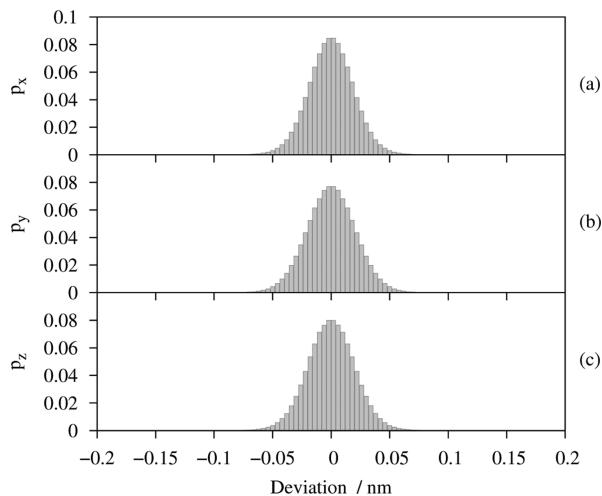


Figure 6. Probability distribution of molecular deviations from expected positions calculated from a short preliminary 80 ps MD simulation of the (100) surface of PCM in water.

The calculated value for OT was approximately 22.3° , thus leading to the standard deviation $\sigma = 7.43^\circ$. Thus, the measurement noise variance for orientation $R_o = \sigma^2$ was about 55.25.

In an analogous manner the DT parameter can be estimated. In the present work, we take an even simpler way and define the upper limit of the DT parameter as the half of the minimal distance between any two molecules in the starting crystal structure. It is the maximal value, which still prevents the neighbor-searching algorithm described in section 2 from finding false molecules at the certain neighbor positions. In our case, the calculated value of 1.954 Å was used as DT throughout the analysis. Figure 6a–c shows the probability distributions p_x , p_y , and p_z for the molecular deviations from expected positions. These deviations were calculated as the difference between the coordinates x , y , z , where the certain neighbor for the given molecule should be found and the coordinates x' , y' , z' of the actual center of mass position of this neighbor. To estimate the position measurement noise variance R_p , characterizing the thermal vibration of molecules in a crystal, the widest p_y distribution was used. 99.73% of all values were within 0.067 nm in this case, thus giving a standard deviation σ of about 0.0223 nm. It should be noted that to estimate the average squared deviation from the equilibrium position for one molecule the variance σ^2 calculated for the difference $|y - y'|$ should be divided by two, as both y and y' are independent normal random variables.⁴⁵ Thus, we obtain the position measurement noise variance $R_p = 0.5\sigma^2$ with a value of about 0.00025.

The measurement noise variances for orientation R_o and position R_p can then be used in the expectation–maximization (EM) algorithm⁴⁶ to estimate the corresponding parameters Q_o and Q_p . This algorithm is commonly used to solve the parameter estimation problems.^{47,48} It involves two steps, the expectation step and the maximization step. In the expectation step, the Kalman filter initialized with the current process noise variance estimate is applied to obtain updated state estimates. In the maximization step, an updated process noise variance estimate is obtained from the maximum likelihood calculation using the states from the expectation step

$$Q = \frac{1}{N} \sum_{t=1}^N (\hat{x}_{t+1} - \hat{x}_t)^2 \quad (15)$$

where \hat{x}_t and \hat{x}_{t+1} are system state estimates calculated by a filter in the expectation step and N is the number of simulation steps.

After calculating the Q_o and Q_p parameters for all molecules on the maximization step, minimal values are selected as the input for the next expectation step to provide the highest filtering efficiency. The expectation and maximization steps are then repeated until the convergence of the Q_o and Q_p parameter estimates is reached. This way the position process noise variance Q_p , with a value of about 1.17×10^{-6} , and the orientation process noise variance Q_o , with a value of about 0.0075, were obtained and used for filtering the positions and orientations of all molecules.

For the application of the Kalman filter (see eqs 8–12), the initial estimates for the system state \hat{x}_0 and error variance P_0 need to be chosen. The initial measurement value z_0 can be taken as initial state estimate \hat{x}_0 . The choice of the initial error variance estimate P_0 is also not critical. Almost any value except zero can be taken and the filter would eventually converge.³⁶ Choosing the zero value would cause the filter to believe that the initial state estimate \hat{x}_0 is absolutely correct. In this work, the measurement noise variance value was taken as initial error variance estimate.

5. RESULTS AND DISCUSSION

The results obtained with the Kalman filter using the estimated parameters Q and R are presented in Figure 7. It can be seen that the number of fast, non-Markovian transitions from one to another state reduced drastically compared to the results based on the initial unprocessed data (see Figure 2) as well as on the data averaged even over large time intervals (see Figure 4 for comparison). However, although rapid back-and-fourth transitions between the crystal-like substates with different number of neighbors disappeared completely, some non-Markovian transitions from crystal-like to solution-like states are still present. These occur when the deviation from the reference orientation slightly exceeds the OT value, defined as the upper limit in the 99.7% of cases during the preliminary MD simulation. To eliminate these rapid back-and-fourth transitions, the hysteretic approach described by Bernstein et al.³² or the similar “trivial recrossing suppression” scheme described by Wriggers et al.⁴⁹ additionally can be applied. The idea is to introduce a buffer region and not to count transitions until this buffer has been crossed completely. According to such an approach, the molecules are required to have the deviations from the reference orientations less than $OT \times (1 - \alpha)$ to switch from solution-like to crystal-like state, but the deviations more than $OT \times (1 + \alpha)$ to switch from crystal-like to solution-like state, where α is a parameter defining the buffer region borders. Thus, the random transition events depending solely on the particular value of the OT parameter can be successfully removed. Such an approach alone applied instead of Kalman filter would not work, as the fluctuations in the molecular orientations are too high, but in combination with the Kalman filter very good results can be achieved. In Figure 8 the results obtained using the hysteretic approach with a parameter $\alpha = 0.2$, thus defining the buffer region borders 17.84° and 26.76° , respectively, are presented. Here, the hysteretic approach was also applied to eliminate the rapid back-and-forth transitions from adsorbed to liquid state by requiring that a molecule goes

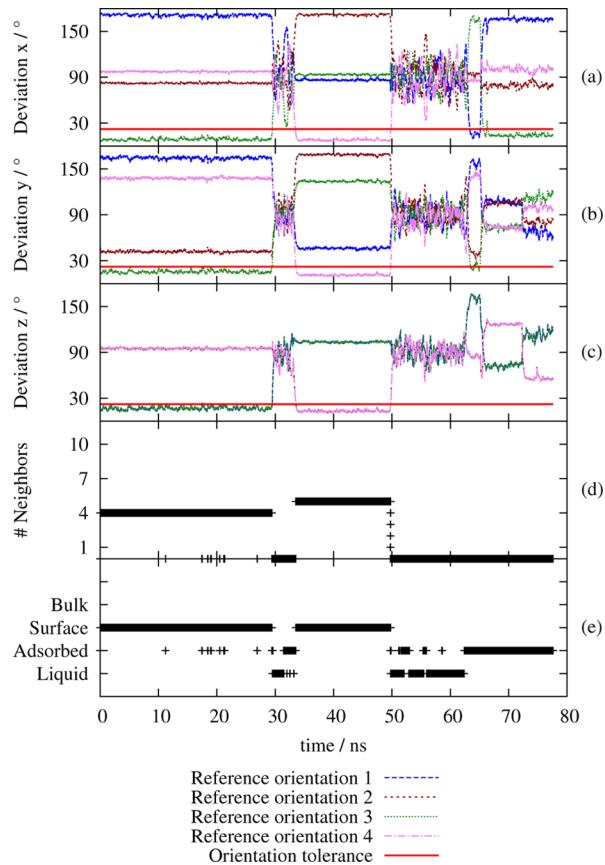


Figure 7. MD trajectory data processed with the Kalman filter and corresponding analysis results for a particular molecule of the (100) surface of a PCM crystal dissolved in water. (a)–(c) Deviation angles of the x , y , and z axes of the molecule from the respective axes of four reference unit cell molecules, correspondingly; (d) number of neighbors among the crystal-like molecules; (e) molecular state estimation. The analysis results (d) and (e) are based on the tolerance parameters $OT = 22.3^\circ$ and $DT = 1.954 \text{ \AA}$. The number of non-Markovian transitions from one to another state reduced drastically.

farther than $|c| \times (1 + \alpha)$ from the crystal surface to switch from adsorbed to liquid state and come closer than $|c| \times (1 - \alpha)$ to the surface to switch from liquid to adsorbed state, where $|c|$ is the length of the reference unit cell vector c , equal to 7.192 \AA in our case. It can be seen in Figure 8 that all non-Markovian transitions from crystal-like to solution-like states disappeared completely. The remaining transitions from adsorbed to liquid state are not relevant for crystal growth/dissolution transition rates calculation. The number of these can be hardly reduced as the molecules in the solution move chaotically.

The obtained results are useful not only for the calculation of transition rates but also for the general analysis of growth/dissolution simulations. Usually, the number of crystal, surface, and adsorbed molecules is analyzed during such simulations to draw conclusions about the evolution of growth/dissolution processes. The application of Kalman filtering significantly supports this type of investigation, making it possible to concentrate only on effective Markovian dynamics.

Figure 9 demonstrates how the number of molecules in different states changes during the simulation, analyzed by using the unprocessed data, the data averaged over 50 ps and the data processed with the Kalman filter. Expectedly, the number of liquid molecules is basically identical in all cases.

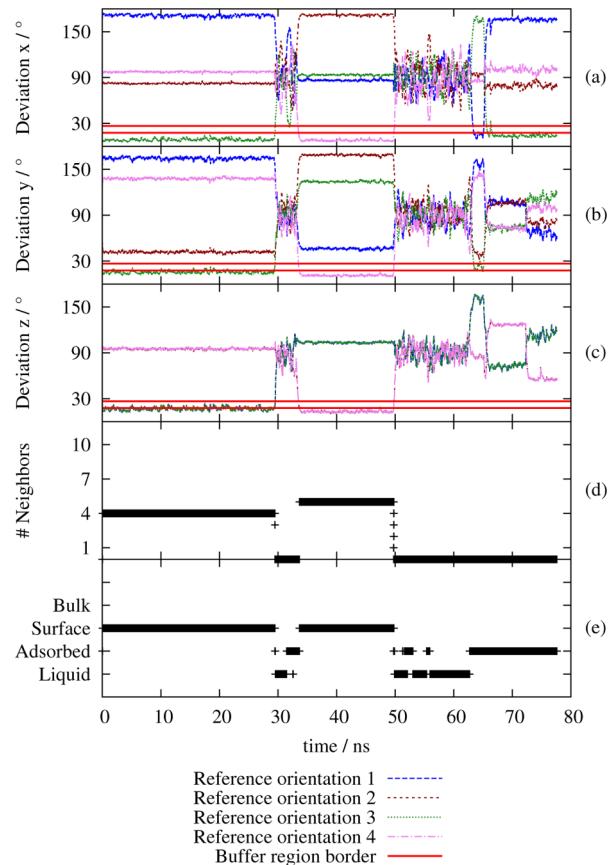


Figure 8. MD trajectory data processed with the Kalman filter combined with the hysteretic approach and corresponding analysis results for a particular molecule of the (100) surface of a PCM crystal dissolved in water. (a)–(c) Deviation angles of the x , y , and z axes of the molecule from the respective axes of four reference unit cell molecules, correspondingly; (d) number of neighbors among the crystal-like molecules; (e) molecular state estimation. The analysis results (d) and (e) are based on the buffer region borders 17.84° and 26.76° and tolerance parameter $DT = 1.954 \text{ \AA}$.

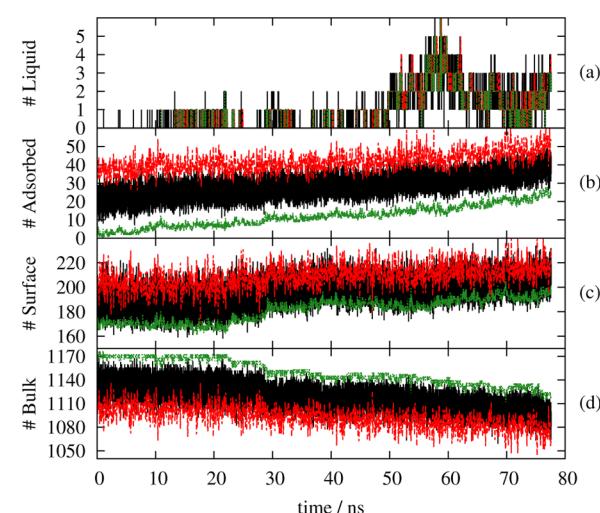


Figure 9. MD simulation of the (100) surface of a PCM crystal in contact with water. Evolution of the number of molecules in liquid (a), adsorbed (b), crystal surface (c), and crystal bulk (d) states defined using the unprocessed data (black), data averaged over 50 ps (red) and data processed with the Kalman filter (green).

The number of adsorbed molecules fluctuates strongly when analyzing the unprocessed data or averaged data. Moreover, the number of adsorbed molecules defined using data averaging is even higher than by the unprocessed data. The number of adsorbed molecules determined using the data processed with the Kalman filter is much lower and fluctuations are much weaker. The number of surface molecules is approximately the same when analyzing data with and without averaging, but the number of crystal bulk molecules is lower in case of averaging. In the case of the Kalman filter, the number of surface molecules is lower, whereas the number of molecules in the crystal bulk is higher compared to the results based on unprocessed or averaged data. Thus, the Kalman filter tends to hold the molecules in their initial states, preventing false conclusions about crystal–solution interface evolution that could be made due to strong molecular fluctuations.

To verify the robustness of the proposed approach, the analysis results are also obtained for other relevant PCM crystal surfaces, namely (001), (011), (110), in contact with water. Figure 10 presents the evolution of the number of adsorbed

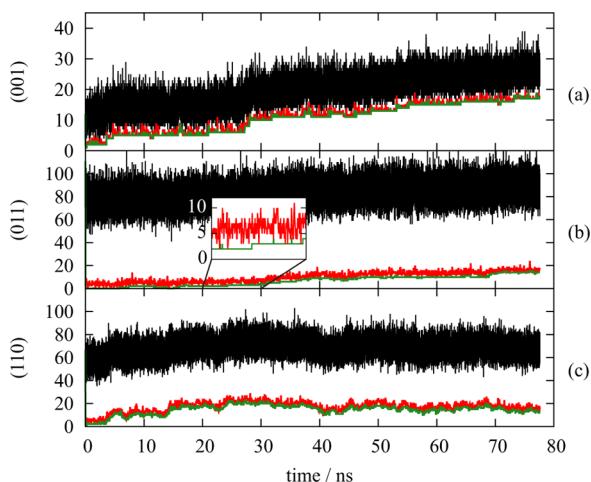


Figure 10. MD simulation of the (001), (011), and (110) surfaces of a PCM crystal in contact with water. Evolution of the number of adsorbed molecules on these surfaces defined using the unprocessed data (black), data averaged over 50 ps (red), and data processed with the Kalman filter (green).

molecules during the dissolution simulation, that is, number of molecules losing proper orientation or all crystal-like neighbors. Like in Figure 9, it is determined by using the unprocessed data, the data averaged over 50 ps and the data processed with the Kalman filter. For these surfaces, the averaging and the Kalman filtering lead to similar results; however, the number of adsorbed molecules determined using the data processed with the Kalman filter is slightly lower and fluctuations are much weaker (see Figure 10). Thus, the application of Kalman filter helps to analyze the time evolution of interfaces significantly, as only rare events important to changes in the macroscopic state of the system are considered. Further, the successful application of the proposed approach for a study of dissolution processes occurring on different surfaces of aspirin and ibuprofen crystal in contact with water is demonstrated in another paper from our group.⁵⁰

6. CONCLUSION

A new approach, based on the Kalman filtering, is proposed to improve the estimation of molecular states during crystal growth and dissolution MD simulations, making it possible to focus on rare, uncorrelated transition events, that is, effective dynamics of the Markov chain. For the application of a Kalman filter, information on the measurement noise variance R and the process noise variance Q is needed. Often, these parameters are just tuned to obtain good filter performance. To avoid this level of arbitrariness, a scheme to define all filter parameters is introduced in the present work, thus providing a way for robust and reliable molecular state definition.

The approach can be easily used in combination with any order parameters, as it is independent from the choice of these. In this work, simple and widely used order parameters (orientation and number of neighbors) based on the structural features of molecules are taken, as they are intuitive, have clear definitions and physical interpretation. A way to estimate the orientation and distance tolerance parameters is proposed. It is shown that the detection of rare, uncorrelated transition events with the help of Kalman filtering is clearly superior to the results based on the data averaging techniques currently used in the literature. The noise connected with the oscillatory behavior of molecules is eliminated almost completely and without any guesswork, thus providing benefits compared to data averaging, where the quality of the results is highly sensitive to the sampling interval selection. Moreover, the recursive nature of Kalman filter makes it in principle possible to carry out the MD analysis “on the fly” during the simulation without waiting until the whole trajectory is written. Thus, excessive MD data do not even have to be stored if one is only interested in the transition behavior. To refine the results obtained with the Kalman filtering and to remove the remaining minimal non-Markovian transitions of molecules between the states, the hysteretic approach is considered. This approach allows to avoid the strong dependence of the results on the particular orientation and distance tolerance values.

The presented results are very promising and useful not only for the effective transition rates calculation but also for the general analysis of the time evolution of interfaces. Also, methods where the event detection plays an important role, like parallel-replica dynamics,^{34,35} would benefit from the proposed approach. Moreover, the approach could also be used in many situations not related to interfaces, such as hydrogen bond fluctuations in hydration shells,⁵¹ conformational dynamics of biomolecules,^{49,52} structural rearrangements induced by external parameters, and others.

It should be noted that only the simplest Kalman filter design was used in this work, where molecular coordinates as well as the deviations from the x , y and z axis orientations of reference unit cell molecules are considered independently of each other. The process noise variance was defined solely for the molecules belonging to the crystal lattice, thus providing optimal filtering only for these. A possible next step would be to evaluate the adaptive Kalman filter with the dynamic rescaling of the process noise. This would allow using different process noise variances for different molecular states, thus leading to the most favorable filtration for each molecule.

APPENDIX: COMPUTATIONAL DETAILS

MD simulations of the (100), (001), (011), and (110) surfaces of a PCM crystal in contact with water were carried out using

the GROMACS simulation code (version 4.5.4).⁵³ The molecular interaction parameters for PCM were taken from the general Amber force field (GAFF),⁵⁴ and the TIP3P⁵⁵ model was used for water. The initial coordinates for the simulation were generated with the program GDIS⁵⁶ from a PCM unit cell determined by X-ray diffraction.⁵⁷ To relax the experimental unit cell under GAFF, first the cell vectors were optimized during a 50 ps simulation by slowly cooling the system from 100 to 0 K while applying a Berendsen barostat at zero pressure. After optimization of the cell vectors all atomic positions were relaxed once more by applying the energy minimization technique. The starting structure for the MD simulation was then built from this relaxed unit cell by repeating it periodically in space until a cell size in x , y , and z direction of $6 \times 6 \times 4$ nm was obtained. The simulation cell was then enlarged by 4 nm in the z direction and filled with water using the genbox package within GROMACS.⁵³ Solvent molecules accidentally placed into the crystal slab were manually moved into the liquid. The final system consisted of 1344 PCM molecules and 7345 water molecules. For the obtained system, energy minimization (5000 steps) and four equilibration steps of 80 ps each with temperature controlled at 310 K were performed. In the first step only water was allowed to move in an NVT ensemble as to ensure proper distribution around the crystal. In the second step, the water molecules were fixed to allow for a gentle equilibration of the PCM structure. In the third step, the whole system was released and in the last step pressure was adjusted to 1 bar in an NpT ensemble until a constant volume was reached. This last equilibration step was used to estimate the OT and DT parameters as well as the filter parameters as described in section 4. The production MD simulation was performed in an NpT ensemble. The time step for the simulation was 2.0 fs. The LINCS algorithm was used to constrain bonds.⁵⁸ Nonbonded interactions were calculated with a cutoff of 9.0 Å. The particle mesh Ewald method was used for treating the long-range electrostatic interactions. In production simulations, temperature was controlled using the Nosé–Hoover thermostat^{59,60} and the Parrinello–Rahman barostat⁶¹ was used to perform constant pressure simulations. In equilibration simulations, the Berendsen method⁶² was applied to control temperature and pressure. The simulation temperature was set to the physiological temperature of 310 K. During the production simulation, the trajectory was saved every 2 ps.

AUTHOR INFORMATION

Corresponding Author

*H. Briesen. E-mail: heiko.briesen@wzw.tum.de.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by the *Deutsche Forschungsgemeinschaft* (DFG) through grants BR 2035/4-1 and BR 2035/8-1 and has made use of the computer resources provided by the Leibniz Supercomputing Centre under the grant pr58la. The ScalaLife Competence Center (<http://www.scalalife.eu/>) is also acknowledged for the assistance with computing resources and the Gromacs code. M.G. gratefully acknowledges the support by the Faculty Graduate Center Weihenstephan of TUM Graduate School at Technische Universität München, Germany.

REFERENCES

- (1) Reilly, A.; Briesen, H. *J. Chem. Phys.* **2012**, *136*, 034704.
- (2) Hawtin, R.; Quigley, D.; Rodger, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4853–4864.
- (3) Liang, S.; Kusalik, P. G. *Chem. Phys. Lett.* **2010**, *494*, 123–133.
- (4) Jacobson, L. C.; Matsumoto, M.; Molinero, V. *J. Chem. Phys.* **2011**, *135*, 074501.
- (5) Piana, S.; Gale, J. D. *J. Am. Chem. Soc.* **2005**, *127*, 1975–1982.
- (6) Piana, S.; Gale, J. D. *J. Cryst. Growth* **2006**, *294*, 46–52.
- (7) Piana, S.; Reyhani, M.; Gale, J. D. *Nature* **2005**, *438*, 70–73.
- (8) Salvalaglio, M.; Vetter, T.; Giberti, F.; Mazzotti, M.; Parrinello, M. *J. Am. Chem. Soc.* **2012**, *134*, 17221–17233.
- (9) Banerjee, S.; Briesen, H. *J. Chem. Phys.* **2009**, *131*, 184705.
- (10) Gnanasambandam, S.; Rajagopalan, R. *CrystEngComm* **2010**, *12*, 1740–1749.
- (11) Cheong, D. W.; Boon, Y. D. *Cryst. Growth Des.* **2010**, *10*, 5146–5158.
- (12) Santiso, E. E.; Trout, B. L. *J. Chem. Phys.* **2011**, *134*, 064109.
- (13) Gao, Y.; Olsen, K. W. *Mol. Pharmaceut.* **2013**, *10*, 905–917.
- (14) Chen, J.; Sarma, B.; Evans, J. M.; Myerson, A. S. *Cryst. Growth. Des.* **2011**, *11*, 887–895.
- (15) Anwar, J.; Zahn, D. *Angew. Chem., Int. Ed.* **2011**, *50*, 1996–2013.
- (16) Steinhardt, P. J.; Nelson, D. R.; Ronchetti, M. *Phys. Rev. B* **1983**, *28*, 784–805.
- (17) Lechner, W.; Dellago, C. *J. Chem. Phys.* **2008**, *129*, 114707.
- (18) Radhakrishnan, R.; Trout, B. L. *J. Am. Chem. Soc.* **2003**, *125*, 7743–7747.
- (19) Brukhno, A.; Anwar, J.; Davidchack, R.; Handel, R. *J. Phys.: Condens. Matter* **2008**, *20*, 494243.
- (20) Leyssale, J.-M.; Delhommele, J.; Millot, C. *J. Am. Chem. Soc.* **2004**, *126*, 12286–12287.
- (21) Mettes, J. A.; Keith, J. B.; McClurg, R. B. *Acta Crystallogr.* **2004**, *60*, 621–636.
- (22) Zahn, D. *J. Phys. Chem. B* **2007**, *111*, 5249–5253.
- (23) Xu, S.; Bartell, L. *J. Phys. Chem.* **1993**, *97*, 13544–13549.
- (24) Kinney, K. E.; Xu, S.; Bartell, L. S. *J. Phys. Chem.* **1996**, *100*, 6935–6941.
- (25) Moroni, D.; ten Wolde, P. R.; Bolhuis, P. G. *Phys. Rev. Lett.* **2005**, *94*, 235703.
- (26) Coasne, B.; Jain, S. K.; Naamar, L.; Gubbins, K. E. *Phys. Rev. B* **2007**, *76*, 085416.
- (27) ten Wolde, R. P.; Ruiz-Montero, M. J.; Frenkel, D. *J. Chem. Phys.* **1996**, *104*, 9932–9947.
- (28) Auer, S.; Frenkel, D. *Nature* **2001**, *409*, 1020–1023.
- (29) Browning, A. R.; Doherty, M. F.; Fredrickson, G. H. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2008**, *77*, 041604.
- (30) Chushak, Y.; Bartell, L. S. *J. Phys. Chem. A* **2000**, *104*, 9328–9336.
- (31) Kalman, R. E. *J. Basic Eng.* **1960**, *82*, 35–45.
- (32) Bernstein, N.; Várnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csányi, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 646–656.
- (33) Reuter, K. First-Principles Kinetic Monte Carlo Simulations for Heterogeneous Catalysis: Concepts, Status, and Frontiers. In *Modeling and Simulation of Heterogeneous Catalytic Reactions: From the Molecular Process to the Technical System*; Deutschmann, O., Ed.; Wiley-VCH Verlag: Weinheim, Germany, 2011; Chapter 3, pp 71–111.
- (34) Perez, D.; Uberuaga, B. P.; Shim, Y.; Amar, J. G.; Voter, A. F. Accelerated Molecular Dynamics Methods: Introduction and Recent Developments. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Ed.; Elsevier: Amsterdam, the Netherlands, 2009; Vol. 5, Chapter 4, pp 79–98.
- (35) Buchete, N.-V.; Hummer, G. *Phys. Rev. E* **2008**, *77*, 030902.
- (36) Welch, G.; Bishop, G. *An introduction to the Kalman Filter (Technical Report TR 95-041)*; University of North Carolina: Chapel Hill, NC, 2006; http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf (accessed Sept 12, 2013).
- (37) Faragher, R. *IEEE Signal Process. Mag.* **2012**, *29*, 128–132.

- (38) Chatfield, C. *Time-series forecasting*; Chapman and Hall: Boca Raton, FL, 2000.
- (39) Murugan, N. A.; Sayeed, A. *J. Chem. Phys.* **2009**, *130*, 204514.
- (40) Durbin, J.; Koopman, S. J. *Time Series Analysis by State Space Methods*; Oxford University Press: New York, 2001.
- (41) Brown, R. G.; Hwang, P. Y. C. *Introduction to Random Signals and Applied Kalman Filtering*, 2nd ed.; John Wiley & Sons: New York, 1992.
- (42) Bos, R.; Bombois, X.; Van den Hof, P. M. J. Designing a Kalman filter when no noise covariance information is available. *Proc. IFAC World Congr., 16th* **2005**, 1–6.
- (43) Zhou, J.; Luecke, R. H. *Comput. Chem. Eng.* **1995**, *19*, 187–195.
- (44) Sternstein, M. *Statistics*; Barron's Educational Series: Hauppauge, NY, 1996.
- (45) Taboga, M. *Lectures on Probability Theory and Mathematical Statistics*, 2nd ed.; CreateSpace Independent Publishing Platform: Seattle, WA, 2012.
- (46) Einicke, G. A.; Falco, G.; Dunn, M. T.; Reid, D. C. *IEEE Signal Process. Lett.* **2012**, *19*, 275–278.
- (47) Kay, S. M. Maximum Likelihood Estimation. In *Fundamentals of Statistical Signal Processing: Estimation Theory*; Gettman, K., Ed.; Prentice Hall: Upper Saddle River, NJ, 1993; Vol. 1, Chapter 7, pp 157–214.
- (48) McLachlan, G. J.; Krishnan, T. *The EM algorithm and Extensions*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, 2008.
- (49) Wriggers, W.; Stafford, K. A.; Shan, Y.; Piana, S.; Maragakis, K.; Miller, P. J.; Gullingsrud, J.; Rendleman, C. A.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. *J. Chem. Theory Comput.* **2009**, *5*, 2595–2605.
- (50) Greiner, M.; Elts, E.; Schneider, J.; Reuter, K.; Briesen, H. J. *Cryst. Growth*, submitted.
- (51) Prada-Gracia, D.; Shevchuk, R.; Rao, F. *J. Chem. Phys.* **2013**, *139*, 084501.
- (52) Nerukh, D.; Jensen, C. H.; Glen, R. C. *J. Chem. Phys.* **2010**, *132*, 084104.
- (53) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (54) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (55) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (56) Fleming, S.; Rohl, A. Z. *Kristallogr.* **2005**, *220*, 580–584.
- (57) Bouhmaida, N.; Bonhomme, F.; Guillot, B.; Jelsch, C.; Ghermani, N. E. *Acta Crystallogr.* **2009**, *B65*, 363–374.
- (58) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J. *Comput. Chem.* **1997**, *18*, 1463–1472.
- (59) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (60) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (61) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (62) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.