Development of collective variables for the molecular description of crystallization processes of naphthalene

ETH zürich

Department of Information Technology and Electrical Engineering

Semester Project

Development of collective variables for the molecular description of crystallization processes of naphthalene

Thilo Weber
thiweber@student.ethz.ch

Supervisor
Zoran Bjelobrk
zoranb@ipe.mavt.ethz.ch

Professor
Prof. Dr. Vanessa Wood
wood@iis.ee.ethz.ch

January 3, 2017

Acknowledgements

I would like to thank Prof. Vanessa Wood for enabling me to do this thesis as a part of my studies. I want to thank my supervisor Zoran Bjelobrk. He introduced me to the vast and absorbing field of computational physics and was always very motivated to help me throughout my experiments. I am grateful to Prof. Marco Mazzotti for his support and for allowing me to write this thesis in his group. Last but not least I would like to thank Konstantin Wernli who helped me with the mathematical considerations in Appendix B.

Abstract

The recent success in simulating crystal growth and nucleation of urea in [11] and [12] by means of enhanced sampling aided molecular dynamics (MD) simulations motivates to apply the same approach to other organic compounds. The large free energy barriers of crystallization processes make it impossible to study them with conventional MD simulations. A crucial part in the enhanced sampling methods plays the choice of collective variables (CVs). In this thesis the way of adapting the CVs used for the urea simulations for naphthalene is described. A first success in computationally growing and dissolving naphthalene crystal layers is documented and a route for further investigations is presented. Further, we develop other CVs for a more general case of applications.

Statement of Originality

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor. For a detailed version of the declaration of originality, please refer to Appendix F.

Thilo Weber,
Zurich, January 3, 2017

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Outline 1

2 Preliminaries **3**

2.1 Molecular Dynamics 3

2.2 Collective Variables 3

2.3 Prediction of Crystal Shape 3

2.4 Enhanced Sampling 4

3 Naphthalene Polymorph I Analysis **6**

3.1 Naphthalene Molecule and Notations 6

3.2 Naphthalene Polymorph I 6

3.3 Analysis Methods 7

3.3.1 Radial Distribution Function 7

3.3.2 "Running" Coordination Number 8

3.3.3 Angular Distribution 8

4 Collective Variables **10**

4.1 Gsmac 10

4.2 GsmacSmooth 12

4.3 Radsmac 12

4.4 Maxsmac 12

4.5 Maxradsmac 13

5 Experimental Setups **14**

5.1 Analysis Setup 14

5.2 Crystal Growth Setup 15

6 Results **17**
6.1 Analysis Simulations 17
6.1.1 Radial Distribution Function Results 17
6.1.2 "Running" Coordination Number Results 17
6.1.3 Angular Distribution Results 18
6.2 CV Parameters 21
6.2.1 Gsmac 21
6.2.2 Radsmac 21
6.2.3 Maxsmac and Maxradsmac 21
6.3 CV Distinctness Performance 21
6.4 CV Crystal Growth Performance 22

7 Discussion **25**
7.1 Analysis Simulations 25
7.2 CV Performance 25
7.2.1 Distinctness 25
7.2.2 Crystal Growth 25
7.2.3 Computational Cost 26

8 Conclusion and Outlook **27**

A Rotation of Unit Cell **28**

B Analytical Expression for Angular Distribution **30**

C CV Derivatives **32**
C.1 Inner Derivatives 32
C.2 Gsmac Derivatives 33

D CV Evaluations for Further Values **35**

E Task Description **38**

F Declaration of Originality **43**

Acronyms **45**

Bibliography **45**

List of Figures

2.1 Schematic plot of the FES $F(S)$ as a function of a CV S of a crystal growth simulation. S_0, S_1 and S_2 are the locations of three local minimums. 4

3.1 Naphthalene molecule with enumeration of the atoms. 7
3.2 Naphthalene Polymorph I. 8
3.3 Vectors $v_{i,89}$ and $v_{i,1257}$ and angles θ_{89} and θ_{1257} of the naphthalene polymorph I. 9

4.1 SWFs $\rho_i = \rho(n_i)$ (left) and $f_{ij} = f(r_{ij})$ (right). 11

5.1 Simulation boxes for crystal analysis. 15
5.2 Simulation box for crystal growth. The naphthalene crystal in the middle of the box is placed in a solution of naphthalene solved in ethanol (ethanol molecules are omitted). 16

6.1 RDF of the naphthalene polymorph I and metastable liquid naphthalene at three different times of the NPT simulation. 18
6.2 RCN of the naphthalene polymorph I and metastable liquid naphthalene at three different times of the NPT simulation. 19
6.3 Angular distributions of the naphthalene polymorph I and metastable liquid naphthalene at three different times of the NPT simulation. . . 20
6.4 Steps for finding parameters $\bar{\theta}_k$ and σ_k of the function ϕ (as defined in 4.4) from angular distribution. 22
6.5 Value distributions of the different local CVs S_i. 23
6.6 Crystal growth and dissolution simulations. The ethanol molecules are omitted. 24

A.1 Rotation of unit cell: The grey area is the (1-10) face of the old unit cell and will become the (001) face in the new unit cell. 29

B.1 Sampled angular distribution and fitted function. 31
B.2 Spherical coordinates. 31

D.1 Value distributions of the different local CVs S_i for $r_{cut} = 5.5\,\text{nm}$ and $n_{cut} = 3$. 36
D.2 Value distributions of the different local CVs S_i for $r_{cut} = 8\,\text{nm}$ and $n_{cut} = 9$. 37

Chapter 1

Introduction

1.1 Motivation

Today, controlling the shape of an organic crystal is a hard task because we still have very little understanding of the underlying growth mechanisms. The ability of controlling crystal shapes would be a benefit for numerous applications in different fields, e.g. in pharmaceutical industries or in nanotechnology. One way of getting a deeper insight into crystal growth phenomena is with the help of MD. There has been a first success of this approach in the case of urea [11], [12], where the crystal shapes under different growth conditions could be accurately predicted. This success motivates to achieve similar results for naphthalene, a molecule with a monoclinic unit cell structure rather than a primitive one as in the case of urea. In [7] different crystal shapes of naphthalene crystals grown under different conditions are determined experimentally. The first goal is now to be able to reproduce these experimental results using MD. In case of success, new conditions can be simulated in order to find favorable crystal shapes and consecutively the results can be verified experimentally.

A key role in the MD simulation of crystal growths play CVs, which are used for understanding the current state of the simulated system, as well as for accelerating the whole simulation process through enhanced sampling [14]. The goal of this thesis is to develop a CV which can be used to accurately simulate the crystal growth mechanisms of naphthalene.

1.2 Outline

In Chapter 2, Preliminaries, the role of CVs in MD and enhanced sampling is elaborated briefly. In Chapter 3 the polymorph I of naphthalene [10] and different structure analysis methods are presented. In Chapter 4, Collective Variables, different CVs are

introduced. In Chapter 5, Experimental Setup, the concrete simulation setups used for the crystal analysis as well as for the evaluation of the CVs by growing a crystal is described. In Chapter 6 the results of the analysis and crystal growth simulations are presented. In Chapter 7 these results are discussed. Finally, in Chapter 8 we give a conclusion of this thesis and an outlook to future work.

Chapter 2

Preliminaries

2.1 Molecular Dynamics

MD is a computer simulation method for studying the movements of atoms and molecules. The trajectories of the atoms are determined by numerically solving Newton's equations of motion, where the forces between the atoms are computed using molecular mechanics force fields.

2.2 Collective Variables

In MD one is normally interested in a certain property of the simulated system, e.g. in our case if a given molecule is in liquid or solid phase. The purpose of CVs is to describe such a property in only a low number of dimensions. Mathematically, a CV S is a function from the space of all atomic coordinates $\mathcal{R}$ to the real numbers:

$$S : \mathcal{R} \to \mathbb{R} : \mathbf{r} \mapsto S(\mathbf{r}),$$

where $\mathbf{r} \in \mathcal{R}$ is the collection of all atom coordinates in system. An appropriate CV should be able to distinguish between all relevant stable and metastable states of the system.

2.3 Prediction of Crystal Shape

The shape of a crystal under certain growth conditions can be predicted from the growth velocities of the different crystal faces, whereas these velocities can be determined from the free energy barrier between two crystal layers in the direction of the given face [12]. The goal is therefore to find a CV which can distinguish between the

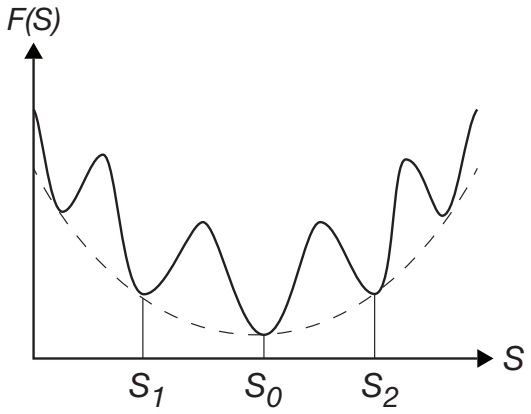


Figure 2.1: Schematic plot of the FES $F(S)$ as a function of a CV S of a crystal growth simulation. S_0, S_1 and S_2 are the locations of three local minimums.

number of crystalline layers in direction of a given face and compute the free energy as a function of the CV, the so called free energy surface (FES) $F(S)$.

Figure 2.1 shows a schematic plot of the FES $F(S)$ as a function of a CV of a crystal growth simulation. At point S_0 the crystal-solution system is in equilibrium state, therefore the FES is at its minimum. At point S_1 the crystal has one molecule layer less which can be seen from a smaller CV value. Here, the solution around the crystal is supersaturated and the FES is therefore higher than in the equilibrium state. At point S_2 has an additional molecule layer. Here, the solution is undersaturated. The different crystal layer states are separated by activation energy barriers.

2.4 Enhanced Sampling

The FES can be computed by sampling the equilibrium probability distribution $p(S)$ of the CV values:

$$F(S) = -\frac{1}{k_B T} \log p(S) = -\frac{1}{k_B T} \log \int d\mathbf{r} \delta(S - S(\mathbf{r})) e^{-\frac{U(\mathbf{r})}{k_B T}}.$$

To this end we have to simulate the system as long as any possible state has occurred enough times to be able to make a good estimate of the distribution $p(S)$.

The emergence and disappearance of crystal layers under certain conditions is a rare event in the order of milliseconds. This is much too long for conventional MD simulations where systems can be simulated in the order of pico- to nanoseconds within a reasonable time. This is where enhanced sampling methods, e.g. variational enhanced sampling [14], comes into play. The idea is to add a variational bias potential

$V(S)$ as a function of the CV S during the simulation which enhances the system to visit any possible state in the CV space.

A pseudo-code algorithm of MD with enhanced sampling is shown in Algorithm 1. First the coordinates $\mathbf{r}$ and the velocities $\mathbf{v}$ of all atoms and the bias potential V are initialized. In every MD step t the current CV S_t is computed, the force $\mathbf{F}_i$ acting on every atom is calculated from the force field potential $U(\mathbf{r})$ and an additional bias term $V(S)$ and the atoms positions and velocities are propagated Δt seconds. Every *nth* MD step the bias potential V is updated in order to pull the system in a new region of the CV space. In order to compute the atomic forces the derivatives of the CVs are needed, therefore the CVs have to be differentiable. Normally, these derivatives are calculated analytically.

Algorithm 1: Enhanced sampling

set initial $\mathbf{r}, \mathbf{v}$
set $V = 0$
every MD step:
    compute CV values:
        $S_t = S(\mathbf{r})$
    every n MD steps:
        update bias potential:
            $V = V(S_t)$
    compute atomic forces:
        $\mathbf{F}_i = -\frac{\partial U(\mathbf{r})}{\partial \mathbf{r}_i} - \frac{\partial V(S)}{\partial S}|_{S=S_t} \frac{\partial S(\mathbf{r})}{\partial \mathbf{r}_i}$
    propagate $\mathbf{r}, \mathbf{v}$ by Δt

Chapter 3

Naphthalene Polymorph I Analysis

In this chapter the Polymorph I of naphthalene [10], the experimental subject of this thesis, is presented. Further, three methods for analyzing the simulation systems are introduced. The purpose of the analysis methods is firstly to assess if the chosen force field is capable of reconstructing the X-ray diffraction (XRD) measured crystal structure in the MD simulations. Secondly, they will be used for the design and parameter estimation of the CVs. The concrete simulation setups for the analysis are presented in Chapter 5 and the results are presented in Chapter 6.

3.1 Naphthalene Molecule and Notations

Naphthalene is an organic compound with formula $C_{10}H_8$. Figure 3.1 shows a naphthalene molecule where the atoms are enumerated. Throughout this thesis the coordinates of a specific atom of a specific naphthalene molecule will be referred to as $r_{\mathrm{NA}i,k}$, where $i = 1, \ldots, N$ and $k = 1, \ldots, 18$. The geometrical center between to atoms is written as $\langle\{r_{\mathrm{NA}i,k}, r_{\mathrm{NA}i,l}\}\rangle$. With the vector between two atoms defined as $v_{i,kl} = r_{\mathrm{NA}i,l} - r_{\mathrm{NA}i,k}$ we denote the angle between the molecules $r_{\mathrm{NA}i}$ and $r_{\mathrm{NA}j}$ as $\measuredangle(v_{i,kl}, v_{j,kl})$, where

$$\measuredangle(v_i, v_j) = \cos^{-1} \frac{v_i \cdot v_j}{||v_i|| \cdot ||v_j||}.$$

3.2 Naphthalene Polymorph I

In Figure 3.2 the Polymorph I of naphthalene and the unit cell with the (001) face at the top are shown. There are two Molecules in one unit cell.

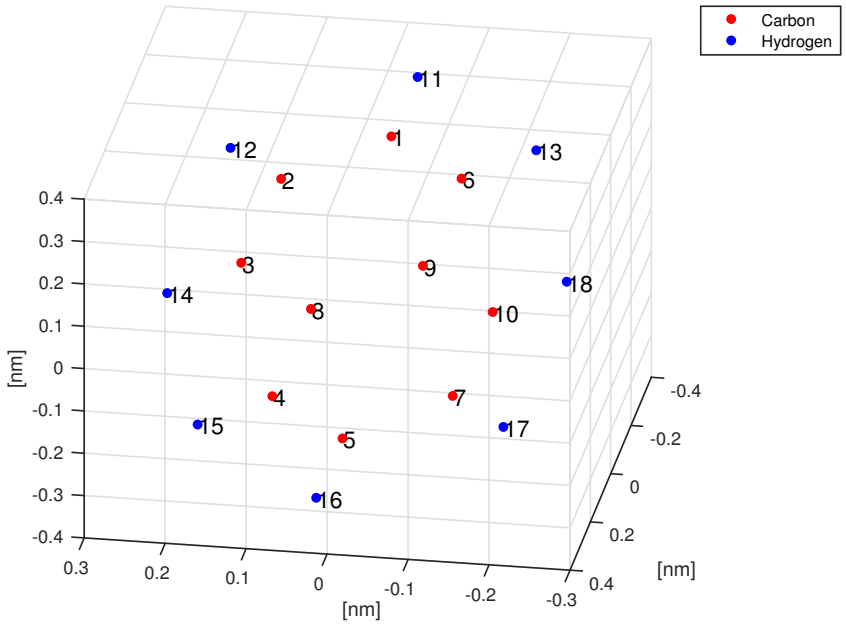


Figure 3.1: Naphthalene molecule with enumeration of the atoms.

3.3 Analysis Methods

In the context of crystalline systems three analysis functions are of special interest which all have the form of an ensemble average of the correlation between a molecule in the crystal and its neighbors. The Radial Distribution Function (RDF) [13] and the "Running" Coordination Number (RCN) [13] relate distances between molecules and the angular distribution compares the angles between two molecules.

3.3.1 Radial Distribution Function

The RDF $g(r)$ is a measure of the probability of finding two molecules at distance r apart under the conditions of the canonical ensemble. It is defined as:

$$g(r) = \frac{1}{\rho}\frac{(N-1)\langle\delta(r-r')\rangle}{4\pi r^2} = \frac{\langle\rho_{shell}(r)\rangle}{\rho} \approx \frac{1}{\rho}\frac{\langle N(r\pm\frac{\delta r}{2})\rangle}{V(r\pm\frac{\delta r}{2})},$$

where ρ is the density, N is the number of molecules and $\rho_{shell}(r)$ is the density on a shell at distance r. In practice the RDF is obtained by a histogram of counting the number of molecules $N(r\pm\frac{\delta r}{2})$ within distance $\pm\frac{\delta r}{2}$ divided by the volume $V(r\pm\frac{\delta r}{2})$ of a shell of thickness $\pm\frac{\delta r}{2}$.

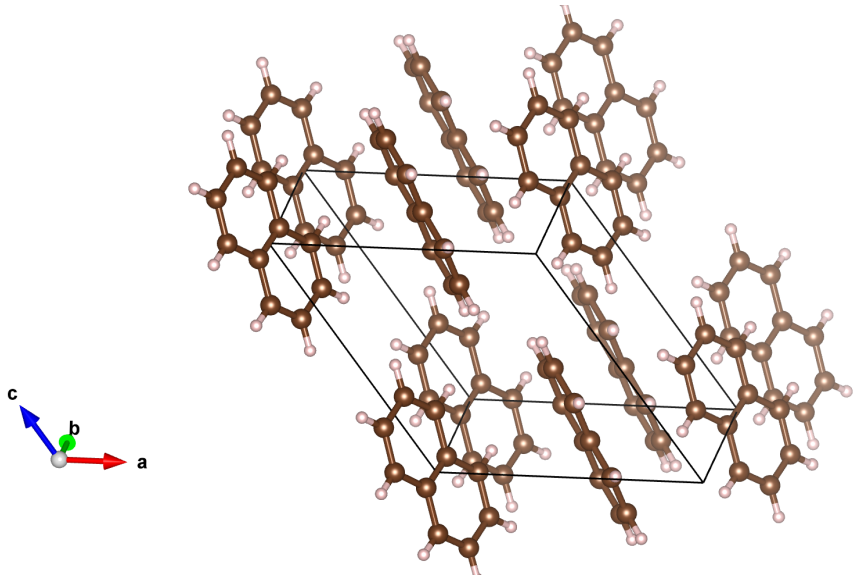


Figure 3.2: Naphthalene Polymorph I.

3.3.2 "Running" Coordination Number

The RCN $N(r)$ is defined as the average number of particles in a sphere of radius r around a molecule:

$$N(r) = 4\pi\rho \int_0^r r'^2 g(r) dr' = 4\pi \int_0^r r'^2 \rho_{shell}(r') dr' \approx \langle N(r' < r) \rangle.$$

The RCN is obtained simple by accumulating the number of molecules $N(r' < r)$ within a distance r.

3.3.3 Angular Distribution

In order to measure an angle between two molecules a characteristic vector in the molecule is defined. In the case of naphthalene the vector between atom $r_{\mathrm{NA}i,8}$ and atom $r_{\mathrm{NA}i,9}$, defined as

$$v_{i,89} = r_{\mathrm{NA}i,9} - r_{\mathrm{NA}i,8}$$

and the vector from the center of the atoms $r_{\mathrm{NA}i,1}$ and $r_{\mathrm{NA}i,2}$ to the center of atoms $r_{\mathrm{NA}i,5}$ and $r_{\mathrm{NA}i,7}$, defined as

$$v_{i,1257} = \langle \{ r_{\mathrm{NA}i,5}, r_{\mathrm{NA}i,7} \} \rangle - \langle \{ r_{\mathrm{NA}i,1}, r_{\mathrm{NA}i,2} \} \rangle$$

are of special interest as they point along symmetric axes of the molecule. In Figure 3.3 these two vectors and the corresponding angles between two molecules

$$\theta_{89} = \measuredangle(v_{i,89}, v_{j,89})$$

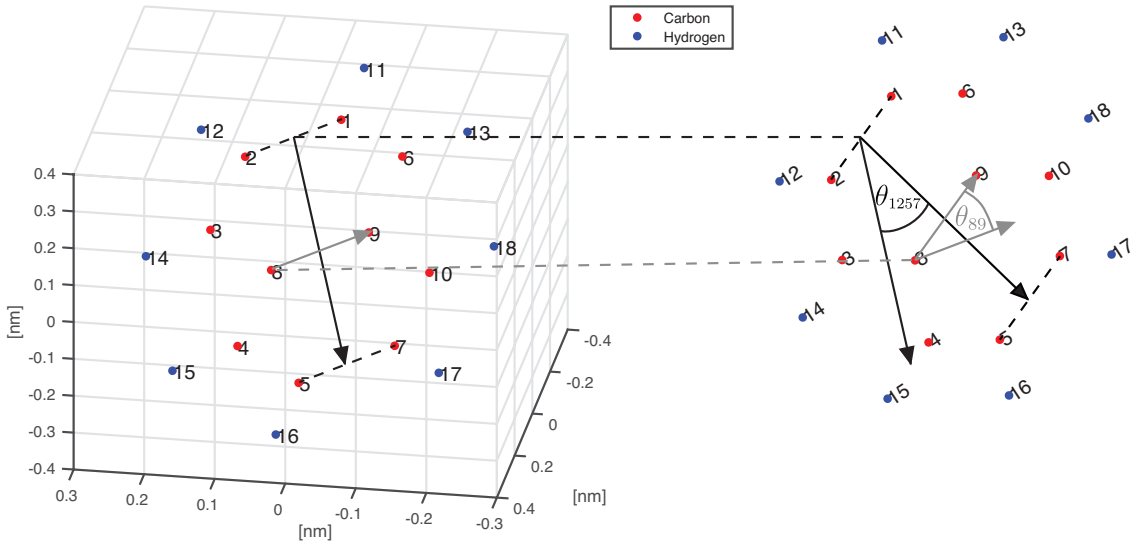


Figure 3.3: Vectors $v_{i,89}$ and $v_{i,1257}$ and angles θ_{89} and θ_{1257} of the naphthalene polymorph I.

and

$$\theta_{1257} = \measuredangle(v_{i,1257}, v_{j,1257})$$

are marked.

Now we are interested in the probability distribution of these angles in the system:

$$p(\theta) \approx \frac{\langle N(\theta \pm \frac{\delta\theta}{2}\rangle}{\delta\theta \cdot N},$$

where $\langle N(\theta \pm \frac{\delta\theta}{2}\rangle$ is the ensemble average number of angles θ in the interval $[\theta - \frac{\delta\theta}{2}, \theta + \frac{\delta\theta}{2}]$ of length $\delta\theta$ and N is the total number of angles in the system. This distribution can be sampled with a normalized histogram.

Chapter 4

Collective Variables

There are several approaches to the development CVs [14]. One important property is the ability to distinguish clearly between the different states of interest. The success of the so called Gsmac CV [6] in [11] and [12] suggest to adapt the same method to the case of naphthalene at first. Further, other CVs building up on the ideas of Gsmac are developed. In this chapter the CVs are stated formally. Parameter values and performance results are presented in Chapter 6.

4.1 Gsmac

The idea of this CV is to determine the degree of crystallinity for each molecule of interest in the system locally as $S_i \in [0, 1]$, where a value of $S_i = 1$ means that the molecule i is part of some crystalline structure and a value of $S_i = 0$ means that the molecule i is not part of any crystalline structure. This local CV for one molecule is defined as

$$S_i = \frac{\rho_i}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} f_{ij} \underbrace{\sum_{k=1}^{K} e^{-\frac{(\theta_{ij} - \bar{\theta}_k)^2}{2\sigma_k^2}}}_{\phi_{ij} = \phi(\theta_{ij})} . \tag{4.1}$$

Here, θ_{ij} is the angle between molecule i and molecule j as described in 3.3.3. ϕ_{ij} is a function that compares the angles θ_{ij} to K reference angles $\bar{\theta}_k$, the angle values of the expected crystal structure, and expresses the similarity as a value between 0 and 1. The values of ϕ_{ij} are summed over all molecules j in the system weighted with a switching function (SWF) $f_{ij} = f(r_{ij})$, a function of the distance r_{ij} between molecule i and molecule j that decreases smoothly from 1 to 0 at a certain cutoff distance r_{cut}. In practice introducing the SWF f_{ij} enables a more efficient implementation by using a nearest neighbor implementation, e.g. with a Verlet list [15], where we don't have

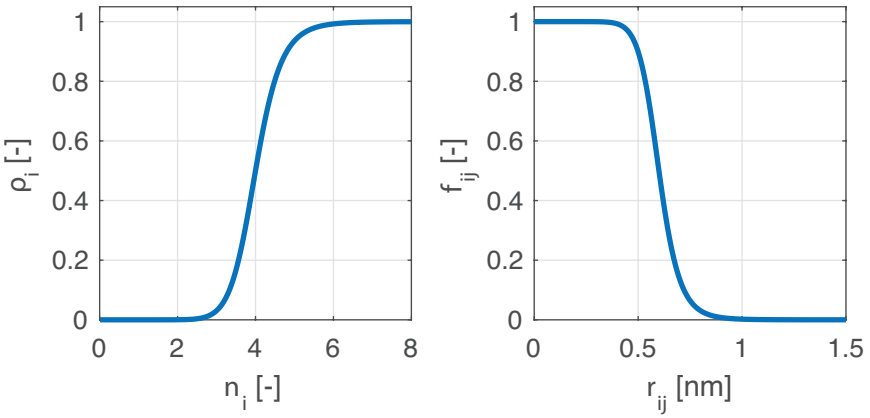


Figure 4.1: SWFs $\rho_i = \rho(n_i)$ (left) and $f_{ij} = f(r_{ij})$ (right).

to sum over all molecules j in the system but only over the nearest neighbors. The SWF $f_{ij} = f(r_{ij})$ could e.g. be of the form

$$f(r_{ij}) = \frac{1 - \left(\frac{r_{ij}}{r_{\text{cut}}}\right)^a}{1 - \left(\frac{r_{ij}}{r_{\text{cut}}}\right)^b}, \tag{4.2}$$

as plotted in Figure 4.1.

Finally, the whole sum is normalized by the coordination number n_i, which is defined as

$$n_i = \sum_{j=1}^{N} f_{ij}$$

and multiplied by another SWF ρ_i on the coordination number n_i, e.g.

$$\rho_i = \frac{1 - \left(\frac{n_i}{n_{\text{cut}}}\right)^{-a}}{1 - \left(\frac{n_i}{n_{\text{cut}}}\right)^{-b}}. \tag{4.3}$$

An example of ρ_{ij} is as well plotted in Figure 4.1. The SWF ρ_{ij} checks if there are enough nearest neighbors around a molecule for it to be part of a crystal structure.

The CVs S_i of can be combined to one CV S_{Gsmac} describing the whole system by summing over i:

$$S_{Gsmac} = \sum_{i=1}^{N} S_i. \tag{4.4}$$

S_{Gsmac} takes values in the interval $[0, N]$ and can be seen as an approximation of the amount of molecules in the system in a crystal phase. For the calculation of the bias potential, as described in Algorithm 1, we need the derivatives $\frac{dS(\mathbf{R})}{d\mathbf{R}}$ which is still dependent on the local structure of S and not only on its summed up value. The derivatives are elaborated in Appendix C.

4.2 GsmacSmooth

The first idea for an improvement of this variable is to replace the sum in Equation 4.4 with a weighted average over all the nearest neighbors of molecule i. The resulting CV

$$S_{GsmacSmooth} = \sum_{i=1}^{N} \frac{1}{n_i + 1}\big(S_i + \sum_{\substack{j=1 \\ j\neq i}}^{N} f_{ij} S_i\big) = \frac{1}{n_i + 1}\big(S_{Gsmac,i} + \sum_{\substack{j=1 \\ j\neq i}}^{N} f_{ij} S_{Gsmac,i}\big)$$

is a smoothed version of S_{Gsmac} which takes the local correlation of the molecules in a crystalline structure more into account.

4.3 Radsmac

The next idea is to compare the distances between two molecules with expected reference values instead of the angles between them. This CV bases on the spacial structure of a crystal rather than on the orientation. The local CV for each molecule can be formulated as

$$S_i = \frac{\rho_i}{n_i} \sum_{\substack{j=1 \\ j\neq i}}^{N} f_{ij} \sum_{k=1}^{K} e^{-\frac{(r_{ij} - \bar{r}_k)^2}{2\sigma_k^2}}.$$

Like for Gsmac we can combine the CVs S_i either by a sum or by a sum of smoothed averages.

4.4 Maxsmac

The idea of this section is to formulate a CV which doesn't compare characteristic variables of the system to reference values but rather to each other. Such a CV could be used for more general problems without introducing any hyper parameters $\bar{\theta}_k$ or $\bar{r}_k$. Here, the more general assumption about a crystal structure is, that for every molecule pair $\{\mathrm{NA}_i, \mathrm{NA}_j\}$ there is at least one molecular pair $\{\mathrm{NA}_i, \mathrm{NA}_k\}$ with a similar angle:

$$\theta_{ij} = \measuredangle(v_{i,lm}, v_{j,lm}) \approx \theta_{ik} = \measuredangle(v_{i,lm}, v_{k,lm}),$$

where

$$v_{i,lm} = r_{\mathrm{NA}i,m} - r_{\mathrm{NA}i,l}.$$

This assumption can be employed to a CV by the use of the max function:

$$S_i = \frac{\rho_i}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} \max_k \big(\sum_{k=1}^{N} e^{-\frac{(\theta_{ij}-\theta_{ik})^2}{2\sigma_k^2}} f_{ij} f_{ik} \big).$$

With a smooth approximation of the max function,

$$\max(x_1, \ldots, x_n) \approx \frac{1}{p} \log \Big(\sum_{i=1}^{N} \exp\big(p x_i\big) \Big) \quad \text{for} \quad p \gg 1,$$

we can define the a CV as

$$S_i = \frac{\rho_i}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{p} \log \Big(\sum_{k=1}^{N} \exp\big(p \exp(-\frac{(\theta_{ij}-\theta_{ik})^2}{2\sigma_k^2}) f_{ij} f_{ik}\big) \Big).$$

4.5 Maxradsmac

By applying the same step as from Gsmac to Radsmac we can define another CV as

$$S_i = \frac{\rho_i}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{p} \log \Big(\sum_{k=1}^{N} \exp\big(p \exp(-\frac{(r_{ij}-r_{ik})^2}{2\sigma_k^2}) f_{ij} f_{ik}\big) \Big).$$

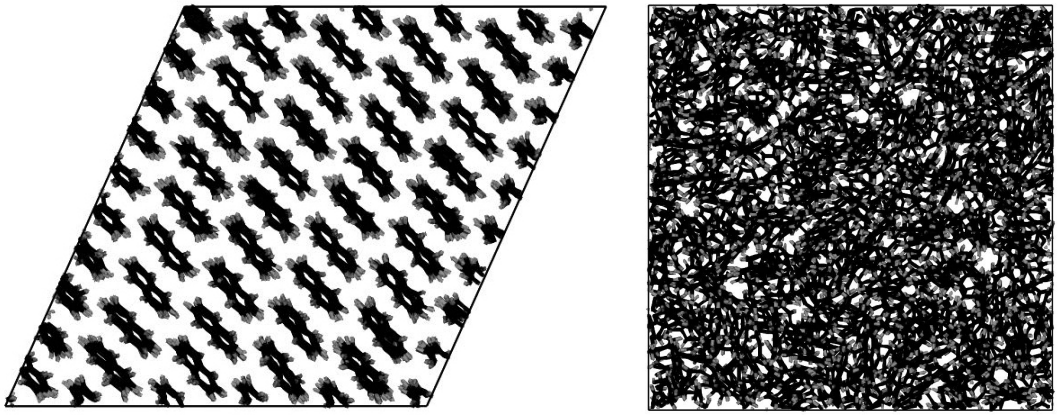
Chapter 5

Experimental Setups

In this chapter the different simulation setups used for the computational experiments are presented. Two analysis setups (Section 5.1) are used to compare the structure differences of a naphthalene in crystal phase and in a metastable liquid phase with the methods defined in Section 3.3 and to evaluate the distinguishing performance of the CVs. The crystal growth setup (Section 5.2) is used to grow a crystal with an enhanced sampling simulation as described in Section 2.4 in order to evaluate the enhanced sampling performance of the CVs. All simulations are performed with GROMACS 5.1.1 [2] under periodic boundary conditions (PBCs) and PLUMED 2.4 [3] is used for the enhanced sampling. As a force field general AMBER force field (GAFF) described in [5] is used for all simulations. In GROMACS a non-bonded cutoff of 1 nm is chosen. Long range electrostatic interactions are calculated with the particle mesh Ewald approach. The LINCS algorithm for bond lengths constraints enables to use time steps of 2 fs. First, an energy minimization (with the conjugate gradient algorithm and a maximum force tolerance of 50 kJ/mol/nm) as well as 100 ns pressure and temperature relaxations to 1 bar and 300 K (using the velocity rescale thermostat and the semi-isotropic Berendsen barostat [1]) are performed for all simulation boxes. Then the systems are simulated with the velocity rescale thermostat [4] and the Parrinello-Rahman barostat [8].

5.1 Analysis Setup

For the crystal analysis a simulation box of size 4.05 nm $\times$ 2.98 nm $\times$ 3.57 nm of crystalline naphthalene, containing 250 molecules, is constructed by concatenations of the unit cell described in Section 3.2 and in [10]. For the liquid analysis a similar simulation box containing metastable liquid naphthalene is constructed by placing molecules at random positions and with random orientations in the box. The metastable liquid



(a) Crystal box. (b) Metastable liquid box.

Figure 5.1: Simulation boxes for crystal analysis.

is taken as a worst case of a solution of naphthalene in terms of distinctness to a crystal.

The systems are simulated for 10 ns under NPT conditions at 300 K and 1 bar. The two boxes are shown in Figure 5.1.

5.2 Crystal Growth Setup

For the crystal growth and dissolution simulations two boxes are constructed with help of the genbox utility of GROMACS with one naphthalene crystal in the middle of the box immersed in a solution of naphthalene and ethanol (see Figure 5.2). In the first box the (001) face of the crystal (which is supposed to grow more slowly than the (1-10) face) and in the second box the (1-10) face (which is supposed to grow faster than the (001) face) are exposed to the solution. The initial (001) crystal is composed of $5 \times 7 \times 4$ unit cells. The (1-10)-faced-up crystal is constructed of $6 \times 4 \times 9$ properly rotated unit cells. The details of the unit cell rotation are described in Appendix A. The (001) system contains totally 580 naphthalene molecules (280 in the crystal and 300 solved in ethanol) and 1'332 ethanol molecules. This makes an initial supersaturation of 4.710. The (1-10) system contains 652 naphthalene molecules (432 in the crystal and 220 in the solution) and 1'733 ethanol molecules which gives an initial supersaturation of 1.278. The simulations are performed for 90 ns and are biased with variationally enhanced sampling (VES) [14], [9] (using Legendre polynomials of order 20, a iteration step size of 0.01 and each iteration is 500 MD steps long).

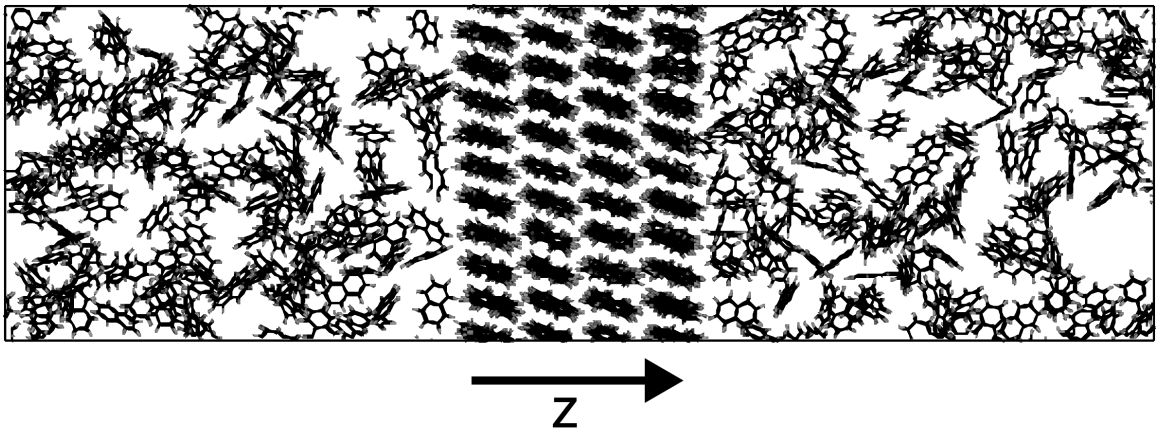


Figure 5.2: Simulation box for crystal growth. The naphthalene crystal in the middle of the box is placed in a solution of naphthalene solved in ethanol (ethanol molecules are omitted).

Chapter 6

Results

In Section 6.1 the results of the analysis simulations described in Section 5.1 are presented. From these results the parameters of the CVs described in Chapter 4 are derived in Section 6.2. In Section 6.3 the performance of the CVs in distinguishing between crystal and liquid naphthalene is evaluated. Finally, in Section 6.4, the results of the crystal growth simulations as described in Section 5.2 are presented.

6.1 Analysis Simulations

6.1.1 Radial Distribution Function Results

In Figure 6.1 the RDFs of the simulations of crystal naphthalene and metastable liquid naphthalene are plotted. For comparison we also have the RDF of the XRD measured crystal used as initialization of for the crystal simulation in the plot. There is no pairwise distance in any system lower than 0.3 nm, which visualizes the fact that two molecules can not be arbitrarily close to each other. The crystal systems have as expected a clearer spacial structure than the metastable liquid system. The peaks of the simulated crystal coincide well with the peaks of the measured crystal.

6.1.2 "Running" Coordination Number Results

Figure 6.2 shows the RCN of the three systems. From this cumulative version of the RDF it is clearly visible how many neighbors a molecule in the crystal structure has at a certain distance. E.g., there are four nearest neighbors at 0.5 nm and another two neighbors at about 0.6 nm.

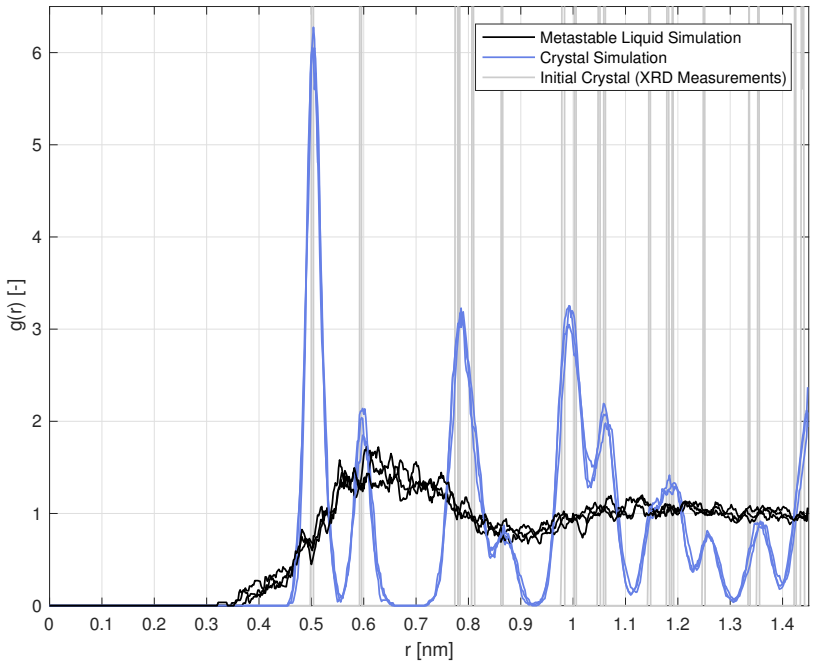


Figure 6.1: RDF of the naphthalene polymorph I and metastable liquid naphthalene at three different times of the NPT simulation.

6.1.3 Angular Distribution Results

Figure 6.3 shows the angular distributions of the three systems. The angular distribution can be written by function of the form (see Appendix B)

$$f_\Theta(\theta) = f_{\Theta'}(\theta) sin(\theta), \quad \theta \in [0, \pi].$$

In the case of a metastable liquid $f_{\Theta'}(\theta')$ is a uniform distribution on the interval $[0, \pi]$. In case of a crystal $f_{\Theta'}(\theta')$ can be approximated by a function of the form

$$f_{\Theta'}(\theta') = \frac{1}{K} \sum_{k=1}^{K} c_k \cdot e^{\frac{-(\theta' - \bar{\theta}'_k)^2}{2\sigma'^2_k}}, \quad \theta' \in [0, \pi],$$

where $\bar{\theta}'_k,\ k = 1, \ldots, K$ are the expected angles taken from the XRD measurements (angular distribution of the measured crystal in Figure 6.3), c_k are constants which normalize each term in the sum such that they integrate to one and σ_k determine the with of the angle distribution depending on the temperature.

With these considerations we can conclude that in the metastable liquid the angles are uniform distributed. In the crystal we have to peaks in the distributions (the angles are mirrored at $\pi/2$, therefore we have no angles above $\pi/2$), one at $\theta = \bar{\theta}'_1 = 0$

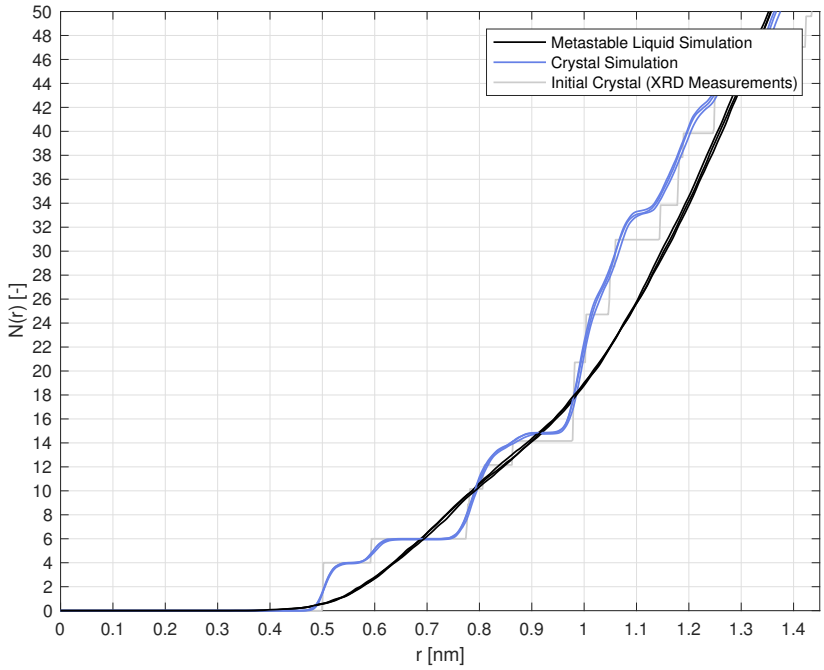
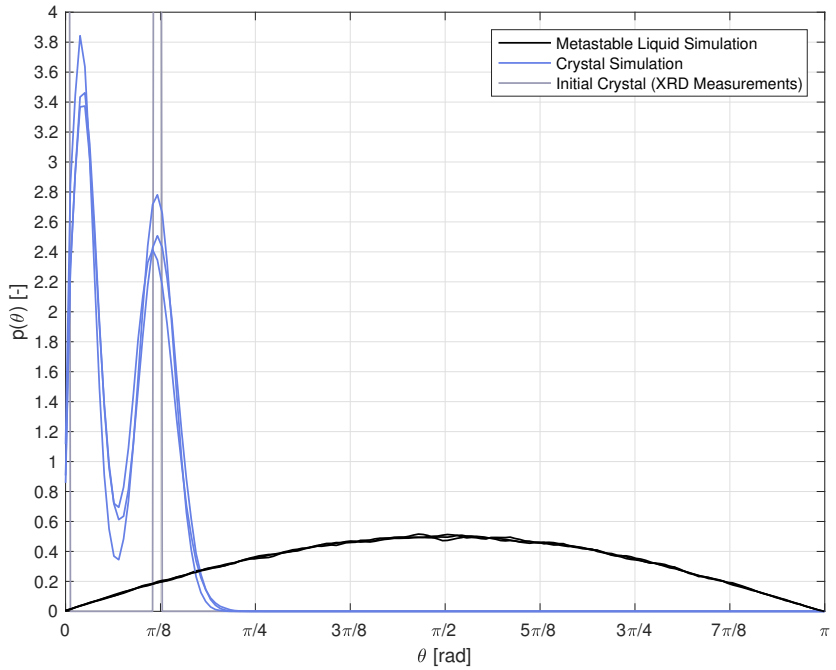
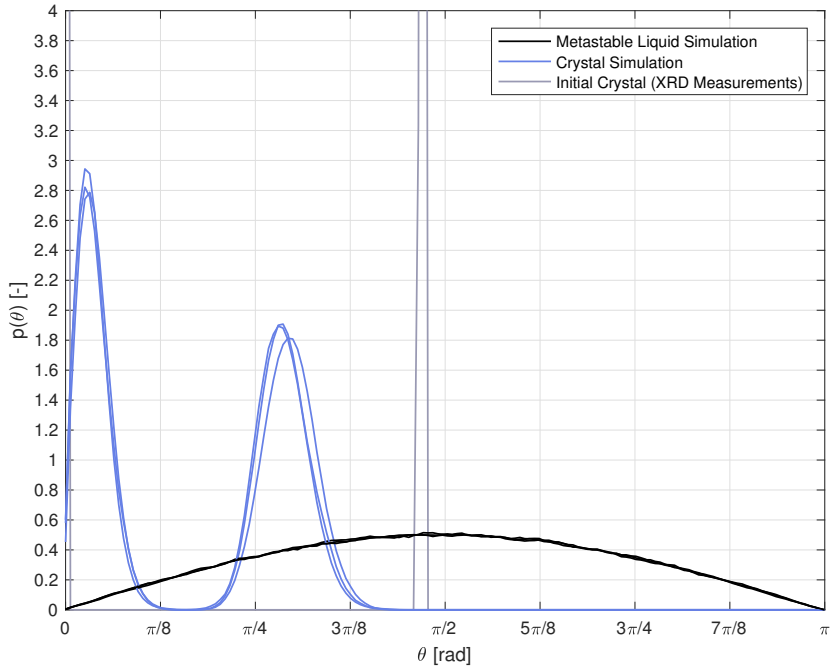


Figure 6.2: RCN of the naphthalene polymorph I and metastable liquid naphthalene at three different times of the NPT simulation.

of all molecules which are orientated in the same direction and another one $\theta = \bar{\theta}'_2$ of all molecules which are orientated in a second direction. From comparing simulated crystal distribution with the measured distribution we see that the θ_{1257} angle is well reconstructed by the simulation but the second θ_{89} angle is shifted to the left during the simulation. This indicates that energy minimum of the GAFF of the naphthalene crystal remarkably differs from reality and that there has to be done further investigations in finding an appropriate force field. For this thesis, for practical reasons, we will stay with this force field while being aware, that the used GAFF may not reproduce the growth rates of crystal faces measured in reality.



(a) Distribution of θ_{1257}.



(b) Distribution of θ_{89}.

Figure 6.3: Angular distributions of the naphthalene polymorph I and metastable liquid naphthalene at three different times of the NPT simulation.

6.2 CV Parameters

The goal of this section is to find the parameters defined in Chapter 4 with help of the analysis results from the previous sections of this chapter.

6.2.1 Gsmac

The steps to find the parameters $\{\bar{\theta}_k\}$ and $\{\sigma_k\}$ of the function ϕ (as defined in 4.1) are plotted in Figure 6.4. We start with the angular distribution of the simulated crystal from the previous section and fit a function of the form

$$\left(c_1 \cdot e^{-\frac{(\theta-\bar{\theta}_1)^2}{2\sigma_1^2}} + c_2 \cdot e^{-\frac{(\theta-\bar{\theta}_2)^2}{2\sigma_2^2}}\right) \cdot \sin(\theta)$$

to it. From this fitting we directly get the parameters $\{\bar{\theta}_k\} = \{0, 0.3674\}$ and $\{\sigma_k\} = \{0.0730, 0.0778\}$. The resulting function $\phi(\theta)$ is plotted in bold blue. The CV is not very sensitive on the values of σ_k. Therefore can also be slightly adjusted.

The parameters r_{cut} and n_{cut} can be determined with help of the RCN in Figure 6.2. The computational cost of the CV is highly sensitive on the value of r_{cut} (and the related size of the Verlet list), therefore we should choose it as small as possible. We find that a value of $r_{cut} = 0.65$ nm is a good trade-off between CV performance and computational cost. From the RCN in Figure 6.2 we directly see that there are 6 neighbors within a distance of 0.65 nm. To be on the save side we therefore can choose $n_{cut} = 5$.

6.2.2 Radsmac

Similar to the previous section the parameter values for Radsmac from the RDF plot (Figure 6.1). Here, we take the values $\{\bar{r}_k\} = \{0.503, 0.596\}$ and $\{\sigma_k\} = \{0.02, 0.02\}$

6.2.3 Maxsmac and Maxradsmac

For the σ_k, r_{cut} and n_{cut} values of Maxsmac and Maxradsmac we can take the same values as for Gsmac and Radsmac respectively. For the p we take a value of 300.

6.3 CV Distinctness Performance

In Figure 6.5 the value distributions of the local CVs S_i with the parameter values from the previous section are shown. (In Appendix D the distributions with other parameter values are shown.) In Gsmac the distributions of the metastable liquid

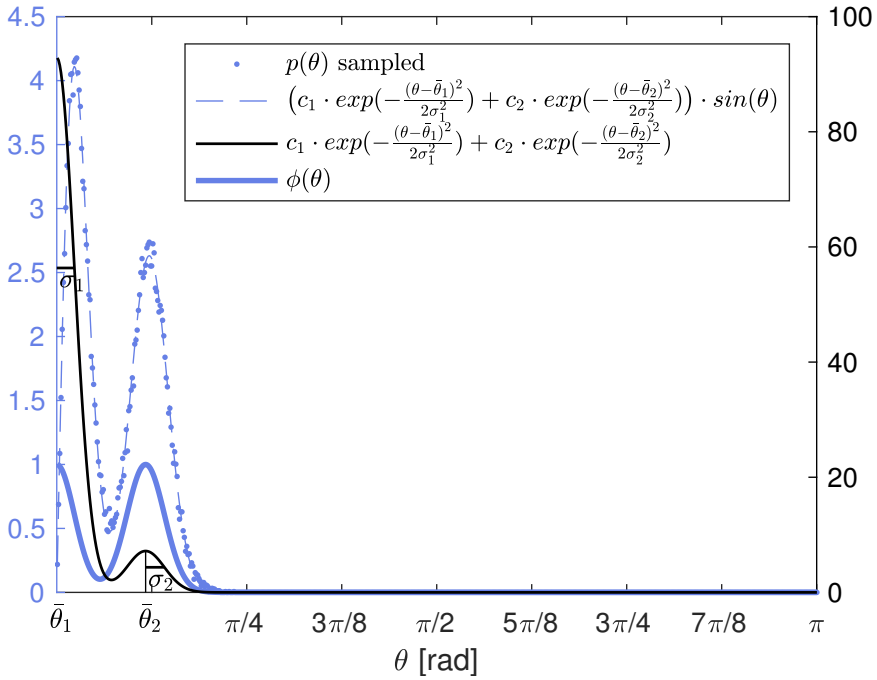
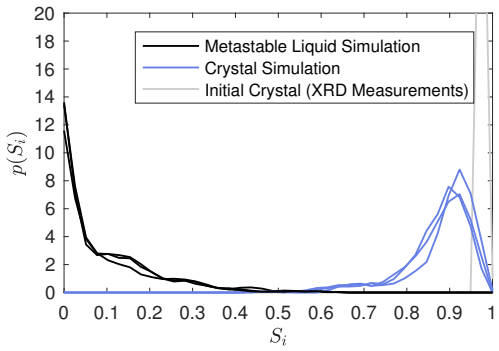


Figure 6.4: Steps for finding parameters $\bar{\theta}_k$ and σ_k of the function ϕ (as defined in 4.4) from angular distribution.

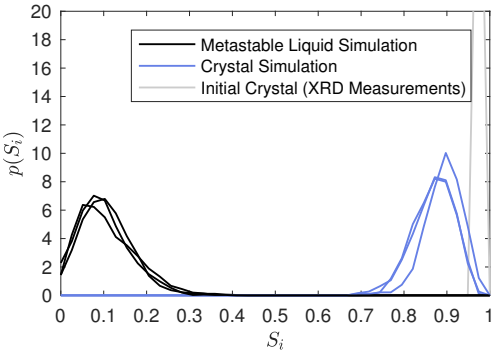
and the crystal systems have still a small overlapping region. The smoothing in GsmacSmooth eliminates this overlapping remarkably. All other CVs are also capable of distinguishing between the crystal and the lipuid system. If we compare the distance CVs (Radsmac and Maxradsmac) with their angle counterpart (GsmacSmooth and Maxsmac), the distributions of the liquid and the crystal are a little bit less separated.

6.4 CV Crystal Growth Performance

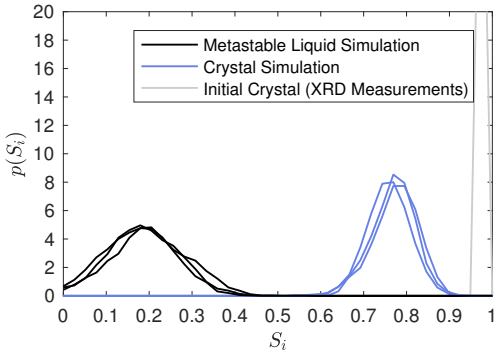
In Figure 6.6 the initial and final configurations of the crystal growth and dissolution simulations is shown. It is clearly visible from comparing the first two figures how one layer has grown completely and one layer is about to be built in the (001) face simulation box. Also the naphthalene concentration in the solution is remarkably reduced. From the two lower figures one can see how in the (1-10) face simulation one crystal layer has been dissolved.



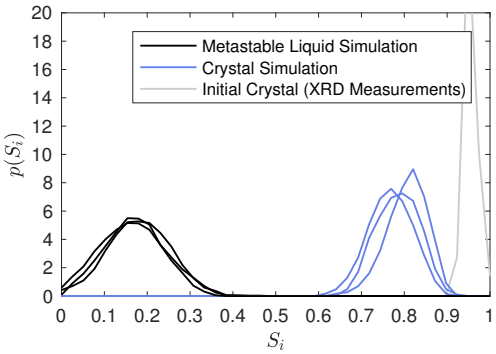
(a) Gsmac.



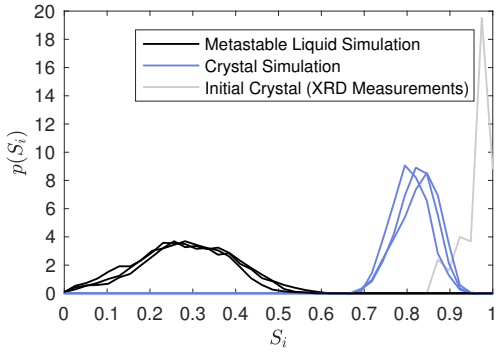
(b) GsmacSmooth.



(c) Radsmac (smoothed).

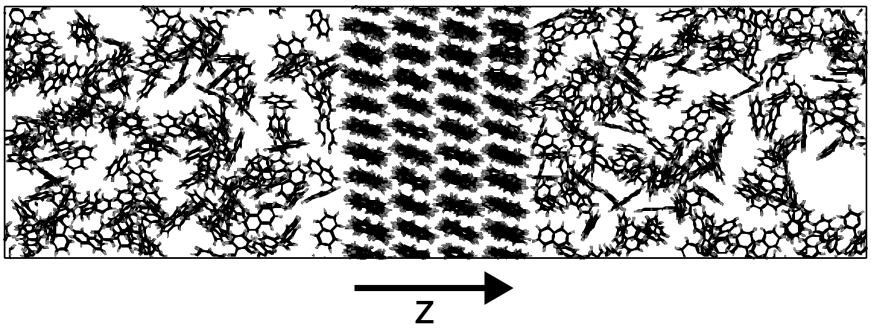


(d) Maxsmac (smoothed).

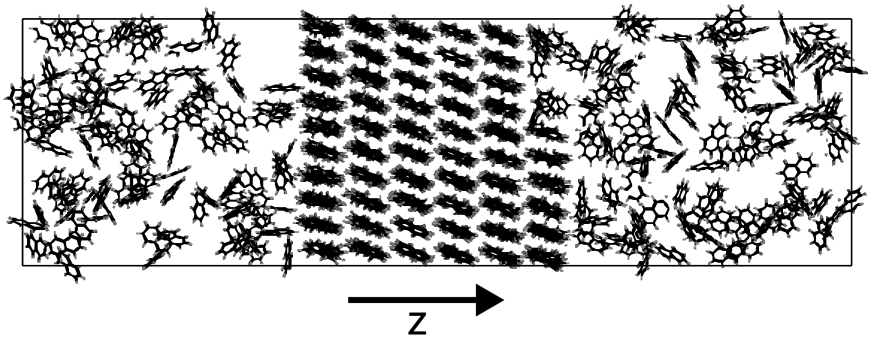


(e) Maxradsmac (smoothed).

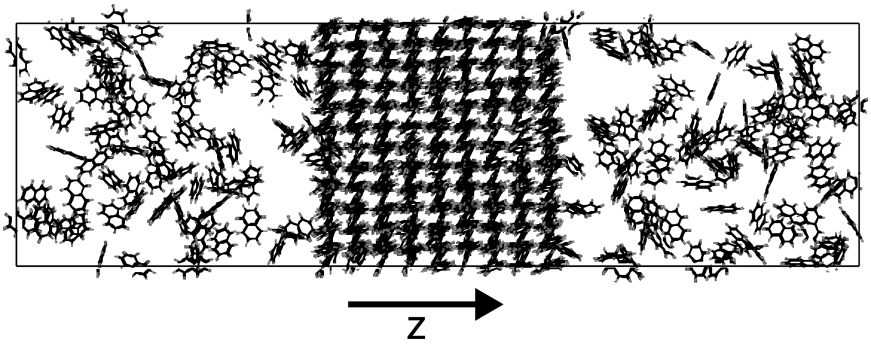
Figure 6.5: Value distributions of the different local CVs S_i.



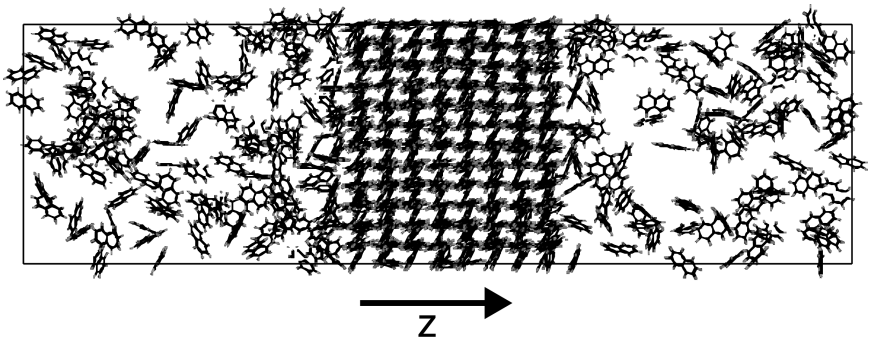
(a) (001) face initial configuration.



(b) (001) face final configuration.



(c) (1-10) face initial configuration.



(d) (1-10) face final configuration.

Figure 6.6: Crystal growth and dissolution simulations. The ethanol molecules are omitted.

Chapter 7

Discussion

7.1 Analysis Simulations

The distribution of the θ_{89} angle (Figure 6.3b) manifests some significant differences between the simulations of the naphthalene polymorh I to the XRD measurements (see Section 6.1.3) caused by the choice of the force field. Nevertheless, the good correlations between the simulations and the measurements in the RDF analysis (Figure 6.3a) and the distribution of the θ_{1257} angle (Figure 6.1) suggest that the GAFF can be used for some first crystal growth simulation experiments, being aware of the expected discrepancies to real measurements.

In Section 6.2 it is described how the parameters for the different CVs can be determined from the performed analysis methods.

7.2 CV Performance

7.2.1 Distinctness

From Figure 6.5 it is clear that all CV are capable of distinguishing between a crystal and a metastable liquid system. The difference to a solution will be even more explicit as the impact of the SWF ρ_i further comes into play when the density of naphthalene molecules decreases, as it is the case in a solution. The smoothing of the local CVs S_i brings a notable improvement in distinctness of the two systems.

7.2.2 Crystal Growth

Figure 6.6 shows that we have been able to grow and dissolve a naphthalene crystal for two different crystal faces. This process is only possible due to the enhanced

sampling. It shows that Gsmac adapted for naphthalene is a proper CV for the task. For the sampling of the FES longer simulation runs have to be performed.

7.2.3 Computational Cost

The computational cost is a critical aspect of a CV. The cost of Gsmac and Radsmac (whithout smoothing) is in the order of

$$\mathcal{O}(N^2)$$

and of Maxsmac and Maxradsmac in the order of

$$\mathcal{O}(N^3),$$

where N is the number of molecules in the system. However by the use of a Verlet list all CVs can be computed in order of

$$\mathcal{O}(N),$$

where Gsmac and Radsmac have a linear overhead for the number of neighbors in the Verlet list and the other methods a quadratic one. If the smoothed versions are used to combine the local CVs S_i instead of summation Gsmac and Radsmac have a quadratic overhead for the number of neighbors and Maxsmac and Maxradsmac have a qubic one. However all CVs seem to be feasible as they show already a good performance for a small number of considered neighbors. The number of neighbor molecules in the Verlet list is very sensitive to the chosen cutoff distance used to build the list. Therefore it could be beneficial to investigate in optimizing the cutoff of the Verlet list compared to the cutoff r_{cut} of the SWF f_{ij} as a trade-off between computational cost and accuracy.

Chapter 8

Conclusion and Outlook

We have managed to grow and dissolve a naphthalene crystal in a naphthalene-ethanol solution. This is definitely a first success confirming the power of the VES apporach. Further investigations have to be done. Different CVs have been developed, so far they still have to be implemented in PLUMED and their performance in an enhanced sampling simulation has to be tested. In terms of computational performance one could try to optimize the parameters of the SWF f_{ij} and the Verlet list. Another critical point is to find a proper force field for the simulations. When these issues are solved one can start with the simulation experiments for learning more about the crystallization process of naphthalene.

Appendix A

Rotation of Unit Cell

In this appendix the rotation of the unit cell of the naphthalene polymorph I such that the (1-10) face becomes the (001) face is described. All quantities from the old and the new unit cells are marked in Figure A.1. We have:

$$a' = c \qquad b' = y' \qquad c' = b \qquad \alpha' = \phi$$

$$\gamma' = \sin^{-1}\frac{x'}{c} \qquad c_z = c\sin\beta \qquad c_x = c\cos\beta \qquad \phi = \tan^{-1}\frac{a}{b}$$

$$\psi = \cos^{-1}\frac{(\begin{pmatrix}-a\\-b\\0\end{pmatrix}\times\begin{pmatrix}c_x\\0\\c_z\end{pmatrix})\cdot\begin{pmatrix}0\\0\\1\end{pmatrix}}{||\begin{pmatrix}-a\\-b\\0\end{pmatrix}\times\begin{pmatrix}c_x\\0\\c_z\end{pmatrix}||\cdot||\begin{pmatrix}0\\0\\1\end{pmatrix}||} = \cos^{-1}\frac{ac_z}{\sqrt{b^2c_z^2+a^2c_z^2+b^2c_x^2}}$$

$$x' = \frac{c_z}{\sin\psi} \qquad y' = \sqrt{a^2+b^2} \qquad z' = b\sin\phi\sin\psi$$

We have to rotate the unit cell first around the z-axis by ϕ and then around the y-axis by $-\psi$, which means we have to multiply every point by two rotation matrices:

$$r_{new} = R_Y R_Z r_{old},$$

where

$$R_Z = \begin{pmatrix}\cos\phi & -\sin\phi & 0\\ \sin\phi & \cos\phi & 0\\ 0 & 0 & 1\end{pmatrix}$$

and

$$R_Z = \begin{pmatrix}\cos -\phi & 0 & \sin -\phi\\ 0 & 1 & 0\\ -\sin -\phi & 0 & \cos -\phi\end{pmatrix}.$$

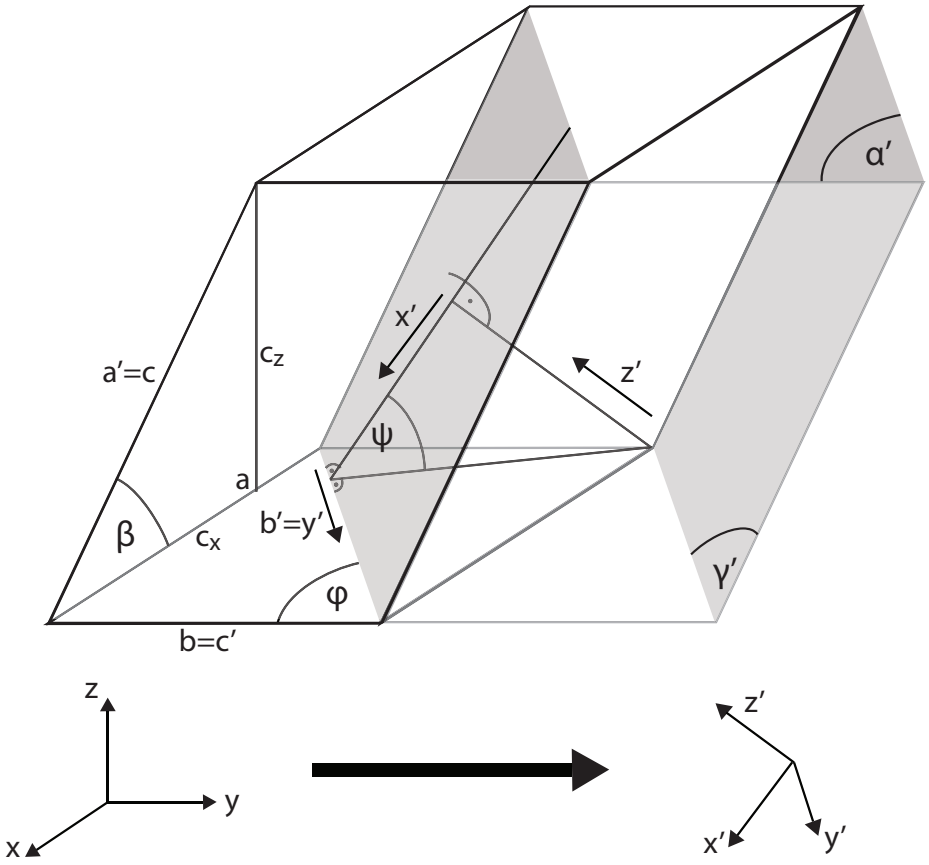


Figure A.1: Rotation of unit cell: The grey area is the (1-10) face of the old unit cell and will become the (001) face in the new unit cell.

Appendix B

Analytical Expression for Angular Distribution

In this appendix it is augmented that an angular distribution can be written as a function of the form

$$f_\Theta(\theta) = f_{\Theta'}(\theta) sin(\theta),$$

where $f_{\Theta'}(\theta')$ is an arbitrary probability distribution on the interval $[0, \pi]$, as shown in Figure B.1.

If we want to find the angular distribution between two random oriented vectors, we always can translate and turn the two vectors, such that both vectors start from the origin and the first vector is oriented along the z-axis, without affecting the distribution. We can also normalize both vectors. Now the angle between the vector coincides with the polar angle θ of the spherical coordinates (Figure B.2) and the random orientation of the second vector is a point density on the unit sphere. If we assume the two independent random angles Θ' and Φ', where Θ' is distributed in the yz-plane according to $f_{\Theta'}(\theta')$ and Φ' is distributed in the xy-plane according to $f_{\Phi'}(\phi')$, we find the probability of a point on the unit sphere contained in a differential area according to

$$f(\theta', \phi') sin(\theta') d\theta' d\phi' = f_{\Theta'}(\theta') f_{\Phi'}(\phi') sin(\theta') d\theta' d\phi',$$

where $sin(\theta') d\theta' d\phi'$ is the surface area element of the spherical coordinates. Therefore we can get the probability distribution of θ from

$$f_\Theta(\theta) d\theta = \int_\phi f_{\Theta'}(\theta) f_{\Phi'}(\phi') sin(\theta) d\theta' d\phi' = f_{\Theta'}(\theta) sin(\theta) d\theta.$$

In particular, if we assume that the angle distribution of two molecule vectors in the plane spanned by this two vectors can be approximated by a cutout of a Gaussian

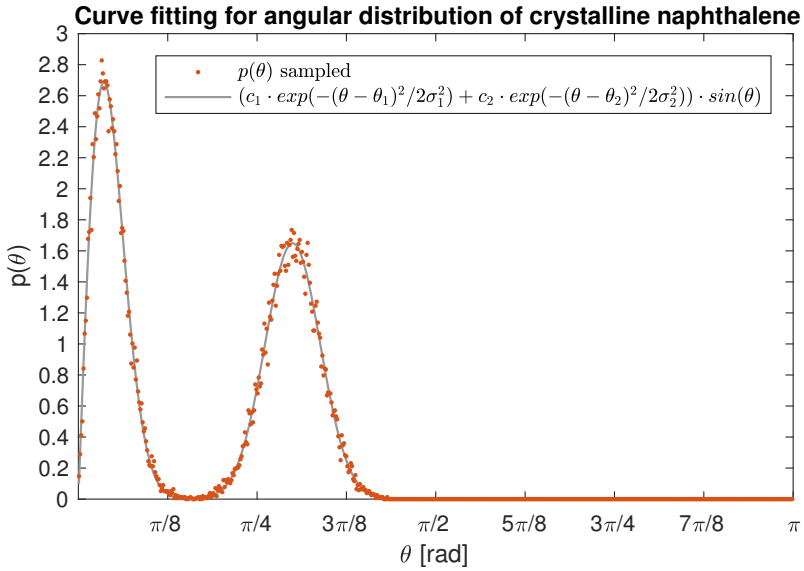


Figure B.1: Sampled angular distribution and fitted function.

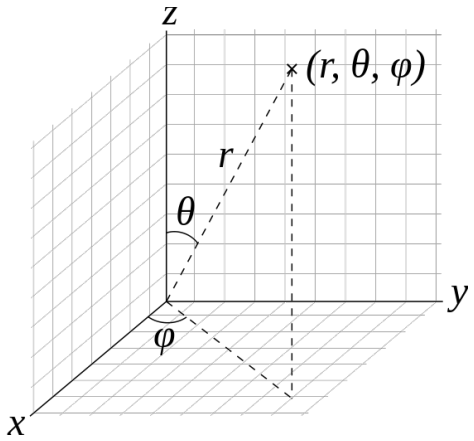


Figure B.2: Spherical coordinates.

distribution, $c \cdot exp(-(\theta' - \theta'_0)^2/2\sigma'^2)$, $\theta' \in [0, \pi]$, the angular distribution of the two vectors follows to be

$$f_\Theta(\theta) = c \cdot e^{\frac{-(\theta-\theta'_0)^2}{2\sigma'^2}} \cdot sin(\theta).$$

Appendix C

CV Derivatives

For the derivatives of the CVs only the terms containing of the center atoms $r_{i,center}$ and the start atoms $r_{i,start}$ and end atoms $r_{i,end}$ of the vector defined for the intermolecular angle are different from zero.

C.1 Inner Derivatives

Derivative of ϕ_{ij}:

$$\begin{aligned}\frac{\partial}{\partial r^{(d)}_{i,start}}\phi(\theta_{ij}) &= \sum_{k=1}^{K} -\frac{\theta_{ij}-\bar{\theta}_k}{\sigma_k^2}\cdot\phi_k(\theta_{ij})\cdot\frac{\partial}{\partial r^{(d)}_{i,start}}\theta_{ij} \\ &= \sum_{k=1}^{K} -\frac{\theta_{ij}-\bar{\theta}_k}{\sigma_k^2}\cdot\phi_k(\theta_{ij})\cdot\frac{-1}{\sqrt{1-\arg(\theta_{ij})}}\cdot\frac{\partial}{\partial r^{(d)}_{i,start}}\arg(\theta_{ij}) \\ &= \frac{-1}{\sqrt{1-\arg(\theta_{ij})}}\Big(-\frac{v_j^{(d)}}{||v_i||\cdot||v_j||}+\frac{v_i^{(d)}(v_i\cdot v_j)}{||v_i||^3\cdot||v_j||}\Big)\sum_{k=1}^{K}-\frac{\theta_{ij}-\bar{\theta}_k}{\sigma_k^2}\phi_k(\theta_{ij}),\end{aligned}$$

where

$$\arg(\theta_{ij}) = \frac{v_i\cdot v_j}{||v_i||\cdot||v_j||}$$

and

$$v_i = \begin{pmatrix} v_i^{(1)} \\ v_i^{(2)} \\ v_i^{(3)}\end{pmatrix} = \begin{pmatrix} r^{(1)}_{i,end}-r^{(1)}_{i,start} \\ r^{(2)}_{i,end}-r^{(2)}_{i,start} \\ r^{(3)}_{i,end}-r^{(3)}_{i,start}\end{pmatrix}_{PBC}$$

is the PBC-distance vector from atom $r_{i,start}$ to $r_{i,start}$.

Further

$$\frac{\partial}{\partial r^{(d)}_{i,end}}\phi(\theta_{ij}) = -\frac{\partial}{\partial r^{(d)}_{i,start}}\phi(\theta_{ij}).$$

Derivative of f_{ij}:

$$\frac{\partial}{\partial r^{(d)}_{i,center}} f(||r_{ij}||) = -\frac{r^{(d)}_{i,j}}{||r_{i,j}||} \cdot f'(||r_{ij}||),$$

where

$$v_i = \begin{pmatrix} r^{(1)}_{j,center} - r^{(1)}_{i,center} \\ r^{(2)}_{j,center} - r^{(2)}_{i,center} \\ r^{(3)}_{j,center} - r^{(3)}_{i,center} \end{pmatrix}_{PBC}$$

is the PBC-distance vector from atom $r_{i,center}$ to $r_{j,center}$.

C.2 Gsmac Derivatives

$$\frac{\partial}{\partial r^{(d)}_{i,a}} S = \frac{\partial}{\partial r^{(d)}_{i,a}} S_i + \sum_{\substack{l=1 \\ l \neq i}}^{N} \frac{\partial}{\partial r^{(d)}_{i,a}} S_l,$$

where

$$\begin{aligned}
\frac{\partial}{\partial r^{(d)}_{i,center}} S_i &= \frac{\rho_i}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} \Big(\frac{\partial}{\partial r^{(d)}_{i,center}} f_{ij} \Big) \phi_{ij} \\
&\quad - \frac{\rho_i}{n_i^2} \sum_{\substack{m=1 \\ m \neq i}}^{N} \Big(\frac{\partial}{\partial r^{(d)}_{i,center}} f_{im} \Big) \sum_{\substack{j=1 \\ j \neq i}}^{N} f_{ij} \phi_{ij} \\
&\quad + \frac{1}{n_i} \rho'(n_i) \sum_{\substack{m=1 \\ m \neq i}}^{N} \Big(\frac{\partial}{\partial r^{(d)}_{i,center}} f_{im} \Big) \sum_{\substack{j=1 \\ j \neq i}}^{N} f_{ij} \phi_{ij} \\
&= \frac{1}{n_i} \sum_{\substack{m=1 \\ m \neq i}}^{N} \Big(\frac{\partial}{\partial r^{(d)}_{i,center}} f_{im} \Big) \Bigg(\rho_i \Big(\phi_{ij} - \frac{1}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} f_{ij} \phi_{ij} \Big) + \rho'(n_i) \sum_{\substack{j=1 \\ j \neq i}}^{N} f_{ij} \phi_{ij} \Bigg),
\end{aligned}$$

$$\frac{\partial}{\partial r^{(d)}_{i,center}} S_l = -\frac{\partial}{\partial r^{(d)}_{i,center}} S_i, \qquad \text{for } l \neq i,$$

$$\frac{\partial}{\partial r^{(d)}_{i,start}} S_i = \frac{\rho_i}{n_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} f_{ij} \Big(\frac{\partial}{\partial r^{(d)}_{i,start}} \phi(\theta_{ij}) \Big),$$

$$\frac{\partial}{\partial r^{(d)}_{i,start}} S_l = \frac{\rho_l}{n_l} f_{li} \Big(\frac{\partial}{\partial r^{(d)}_{i,start}} \phi(\theta_{li}) \Big),$$

$$\frac{\partial}{\partial r_{i,end}^{(d)}} S_i = -\frac{\partial}{\partial r_{i,start}^{(d)}} S_i$$

and

$$\frac{\partial}{\partial r_{i,end}^{(d)}} S_l = -\frac{\partial}{\partial r_{i,start}^{(d)}} S_l.$$

Appendix D

CV Evaluations for Further Values

In this appendix the CV evaluations from different parameters than used in Section 6.3 are presented. All CV are evaluated for $r_{cut} = 5.5$ nm and $n_{cut} = 3$ in Figure D.1 and for $r_{cut} = 8$ nm and $n_{cut} = 9$ in Figure D.2. For Gsmac the parameters

$$\{\bar{\theta}_k\} = \{0, 0.3776\}$$

and

$$\{\sigma_k\} = \{0.1, 0.1\}.$$

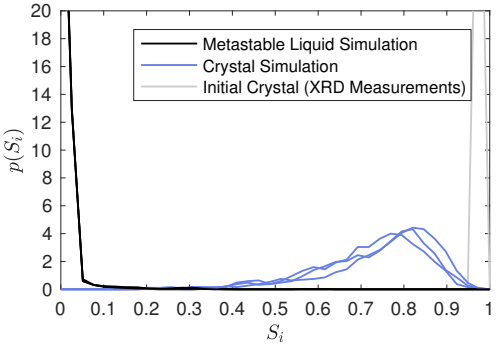
are chosen. For Radsmac the parameters:

$$\{\bar{r}_k\} = \{0.503, 0.596, 0.778, 0.782\}$$

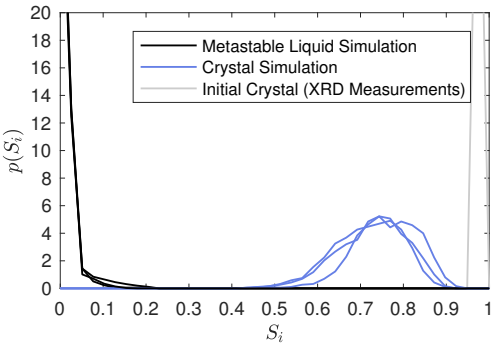
and

$$\{\sigma_k\} = \{0.02, 0.02, 0.004, 0.004\}.$$

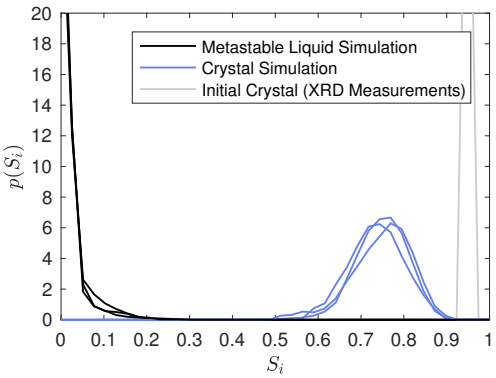
For Maxsmac the parameter values are $\sigma = 0.1$ and $p = 300$ and for Maxradsmac $\sigma = 0.02$ and $p = 300$.



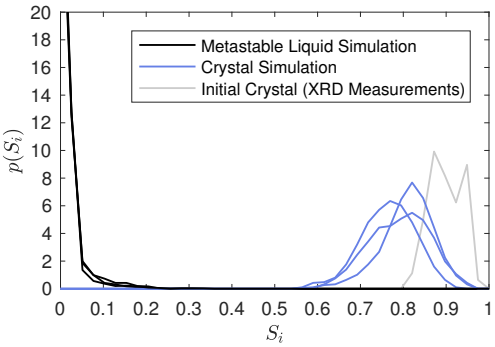
(a) Gsmac.



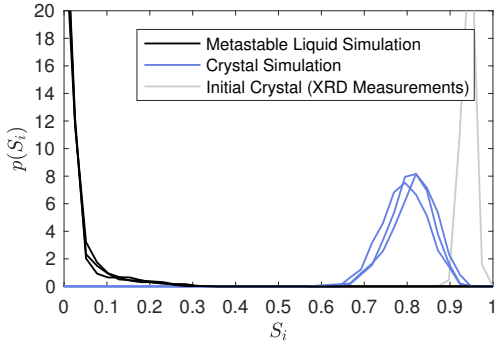
(b) GsmacSmooth.



(c) Radsmac (smoothed).

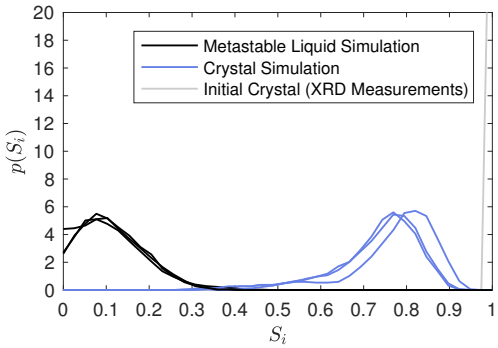


(d) Maxsmac (smoothed).

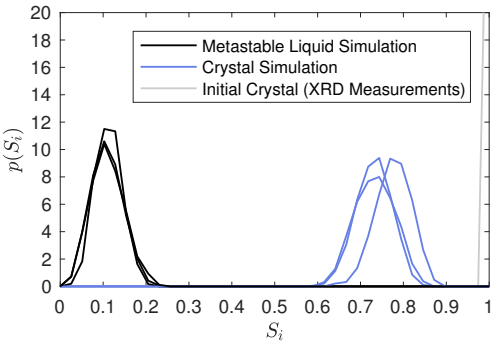


(e) Maxradsmac (smoothed).

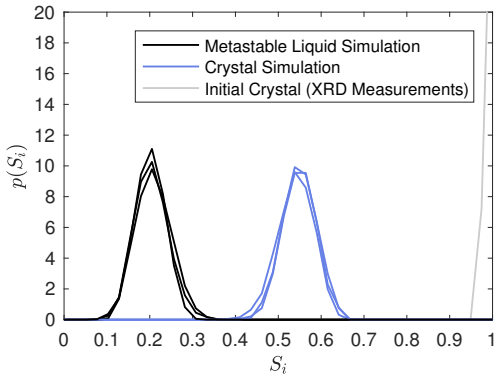
Figure D.1: Value distributions of the different local CVs S_i for $r_{cut} = 5.5\,\text{nm}$ and $n_{cut} = 3$.



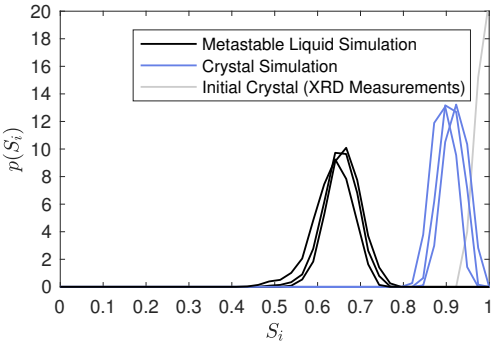
(a) Gsmac.



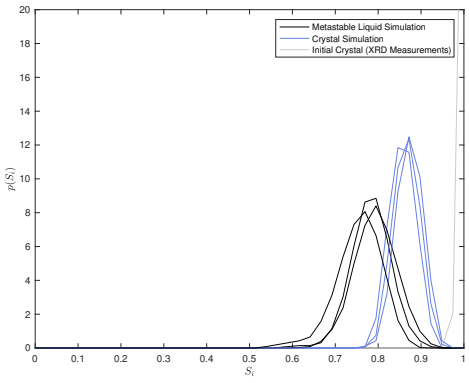
(b) GsmacSmooth.



(c) Radsmac (smoothed).



(d) Maxsmac (smoothed).



(e) Maxradsmac (smoothed).

Figure D.2: Value distributions of the different local CVs S_i for $r_{cut} = 8$ nm and $n_{cut} = 9$.

Appendix E

Task Description

Outline semester thesis: Development of collective variables for the molecular description of crystallization processes for paracetamol

September 22, 2016

1 Literature research

i) Crystal morphology of paracetamol including experimentally observed polymorphs.

ii) Force fields of paracetamol as well as ethanol, in particular of the type AMBER.

iii) Summary on the solubility of paracetamol in water and ethanol, and the melting point of each of the paracetamol polymorphs.

2 Development of the collective variable (CV) with the Gsmac algorithm

The CV should be able to quantify the degree of crystallinity of each individual paracetamol molecule. The concepts of the Gsmac algorithm should be adapted and tested as a quantitative measure of crystallinity [1].

The adaption of the Gsmac algorithm to paracetamol comprises following points:

Density:

i) Choice of appropriate center of mass for the paracetamol molecule.

ii) Calculation of the distance for all paracetamol molecules within the crystal lattice of the corresponding polymorph.

iii) Choice of the appropriate cut-off distance for the coordination number of each molecule within the lattice.

Order:

i) Choice of appropriate vectors and corresponding angles of the vectors between pairs of molecules within the lattice of each polymorph.

ii) Choice of a reasonable width of the vector angles defined in the Gaussian function within the Gsmac algorithm.

If Gsmac does not provide a satisfactory measure for the crystallinity of the paracetamol molecules, Gsmac should be modified with new concepts like a different weighting method for each angular contribution or introducing a lower bound in addition to the upper bound for the nearest neighbor weighting. Also an approach containing the concept of Landau's theory of phase transition [2] can provide a good phase discriminating CV for paracetamol.

3 Nucleation simulations of paracetamol

Crystallization processes can be investigated by simulation in an extensive manner once a suitable CV for the phase discrimination is found. Under the condition a good CV for the crystallinity of paracetamol was provided, nucleation simulations of paracetamol from the melt and in ethanol will be performed. Variationally enhanced sampling will be used to accelerate the sampling of the CV space of interest. Nucleation simulations provide quantitative results on the critical nuclei size, the height of the activation barrier for an nucleation event, and the pathway of the nucleation mechanism.

References

[1] F. Giberti et al. *Chem. Eng. Sci.* **121** (2015) 51.

[2] L. Landau, *Phys. Z. Sowjet.* **11** (1937) 26.

Schedule semester thesis: Development of collective variables for the molecular description of crystallization processes

October 26, 2016

1 Gsmac simulations

18.09.2016 - 31.11.2016

1.1 Literature research

30.10.2016

a) Search for force fields of naphthalene

b) Find method for the extraction of the transition pressure from MD simulations

1.2 Simulation of the naphthalene polymorph I

20.11.2016

Simulate the stability of the naphthalene polymorph I

a) Measure the distances and angles of the polymorph I

b) Write code which measures the average distances and angles of the naphthalene polymorph I simulation

c) Compare the average distances and angles between the simulation and XRD measurement

1.3 Transition pressure simulations

20.11.2016

Calculate the transition pressure between the two naphthalene polymorphs

1.4 Gsmac tests

30.11.2016

a) Find the right vector definition within the naphthalene molecule

b) Find the right Gsmac parameters for the naphthalene polymorph I

c) Compare the Gsmac values for liquid and solid naphthalene

d) Improvement test for the Gsmac algorithm:
   - ◇ Change the angle function (e.g. step functions instead of Gaussians)
   - ◇ Include lower and upper boundaries for the measurement of the number of nearest neighbors
   - ◇ Further improvement tests

2 Development of novel CV concepts for crystallization processes

01.12.2016 - 16.01.2017

Write general C++ code for the distinction between different polymorphs and liquid using Landau's principle of symmetry breaking [1],[2] or other principles like time averaging or space averaging recursive techniques

References

[1] L. Landau, *Phys. Z. Sowjet.* **11** (1937) 26

[2] F. Giberti et al. *Chem. Eng. Sci.* **121** (2015) 51

Appendix F

Declaration of Originality

ETH

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Developement of collective variables for the molecular description of crystallisation processes of naphthalene

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Weber	Thilo

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zürich, 3.1.2017	Thilo Weber

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

Acronyms

CV collective variable.

FES free energy surface.

GAFF general AMBER force field.

MD molecular dynamics.

PBC periodic boundary condition.

RCN "Running" Coordination Number.

RDF Radial Distribution Function.

SWF switching function.

XRD X-ray diffraction.

Bibliography

[1] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684, 1984.

[2] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.*, 91:43–56, 1995.

[3] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R.A. Broglia, and M. Parrinello. Plumed: a portable plugin for free energy calculations with molecular dynamics. *Comp. Phys. Comm.*, 180(1961), 2009.

[4] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007.

[5] Wendy D. Cornell, Piotr Cieplak, Christopher I. Bayly, Ian R. Gould, Kenneth M. Merz, David M. Ferguson, David C. Spellmeyer, Thomas Fox, James W. Caldwell, and Peter A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 113:5179–5197, 1995.

[6] Federico Giberti, Matteo Salvalaglio, Marco Mazzotti, and Michele Parinello. Insight into the nucleation of urea crystals from the melt. *Chemical Engineering Science*, 121:51–59, 2015.

[7] Michael A. Lovette and Michael F. Doherty. Predictive modeling of supersaturation-dependent crystal shapes. *J. Am. Chem. Soc.*, 12:656–669, 2012.

[8] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.*, 52:7182–7190, 1981.

[9] Pablo M. Piaggi, Omar Valsson, and Michele Parinello. A variational approach to nucleation simulation. *The Royal Society of Chemistry*, 12:38–50, 2016.

[10] V.I. Ponomarev, O.S. Filipenko, L.O. Atovmyan, V.I. Ponomarev, O.S. Filipenko, and L.O. Atovmyan. *Kristallografiya*, 21:392, 1976.

[11] Matteo Salvalaglio, Thomas Vetter, Federico Giberti, Marco Mazzotti, and Michele Parinello. Uncovering molecular details of urea crystal growth in the presence of additives. *J. Am. Chem. Soc.*, 134:17221–17233, 2012.

[12] Matteo Salvalaglio, Thomas Vetter, Marco Mazzotti, and Michele Parinello. Controlling and predicting crystal shapes: The case of urea. *Angew. Chem. Int. Ed.*, 52:13369–13372, 2013.

[13] Mark Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, 2010.

[14] Omar Valsson, Pratyush Tiwary, and Michele Parinello. Controlling and predicting crystal shapes: The case of urea. *Annu. Rev. Phys. Chem.*, 67:159–184, 2016.

[15] L. Verlet. Computer 'experiments' on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, 1967.