MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Finding the stability region of CO_2-hydrates using computer simulations"

verfasst von / submitted by

Jakob Daniel Michl

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna, 2018

Acknowledgments

First I would like to thank my supervisor, Prof. Christoph Dellago, who made it possible for me to work in the research group of computational physics and who created a work environment in which not only scientific discovery but also the joy of research is central.

A special thanks goes to Dr. Marcello Sega. He always had an open door whenever I ran into trouble and he spent hours explaining the principles of molecular dynamics simulations to me.

I would also like to thank my fellow master students, Matthias Kiss, Igor Kraut and Nils Clees who have always taken time to discuss problems.

Furthermore, I want to thank my friends, my girlfriend and my family. Their support means more to me than words can express.

Abstract

Although clathrate hydrates occur in large amounts in the deep see and in permafrost regions, the phase diagram of a CO_2-H_2O binary system has not been completely discovered yet. Different experiments showed the existence of a CO_2-hydrate phase different from all known phases. Recently Amos *et al.* provided the structure of this phase and called it χ-phase[1]. In this study, we used molecular dynamics simulations to find the stability region of the χ-phase. Using rigid nonpolarizable models (TIP4P/Ice for water and TraPPE for carbon dioxide), first the melting line of both, the *sI*-phase and the χ-phase were determined. Afterwards, using a method based on thermodynamic integration, the solid-solid coexistence line between the χ-phase and the *sI*-phase and the solid-solid coexistence line between the χ-phase and ice VI + CO_2 I have been computed. Knowing the different coexistence lines, it was possible to determine the stability region of the χ-phase in the phase diagram.

Zusammenfassung

Obwohl Gashydrate in großen Mengen in der Tiefsee und im Permafrost vorkommen und als mögliche Energiequelle der Zukunft gelten, ist das Phasendiagramm des binären CO_2-H_2O-Systems noch immer nicht zur Gänze bekannt. Verschiedene Experimente zeigen die Existenz einer Phase, die von allen bekannten Phasen verschieden ist. Vor kurzem wurde eine Struktur der Phase von Amos *et al.* publiziert und diese wurde aufgrund einer auftretenden chiralen Struktur „χ-Phase" genannt [1]. In dieser Arbeit verwenden wir Molekulardynamik-Simulationen, um den Stabilitätsbereich der χ-Phase zu finden. Dazu wurden starre, nicht polarisierbare Modelle für Wasser (TIP4P/Ice) und für CO_2 (TraPPE) verwendet. Zuerst wurde die Schmelzlinie der χ-Phase und der *sI*-Phase berechnet. Danach wurde eine Methode, die auf thermodynamischer Integration beruht, verwendet, und die Koexistenzlinie zwischen der χ-Phase und der *sI*-Phase, um die Koexistenzlinie zwischen der χ-Phase und Eis VI + CO_2 I zu bestimmen. Mit diesen Koexistenzlinien war es möglich, den Stabilitätsbereich im Phasendiagramm einzugrenzen.

Contents

List of Figures

List of Abbreviations

χ-hydrate	χ-phase of CO_2-hydrate
sI-hydrate	*sI*-phase of CO_2-hydrate
EOM	Equation of Motion
EPM2	Elementary Physical Model 2
FI	Filled ice
GROMACS	GROningen MAchine for Chemical Simulations
HDA	High density amorphous
LINCS	LINear Constraint Solver
L-J	Lennard-Jones
LDA	Low density amorphous
lr	long-range
MD	Molecular Dynamics
NH	Nosé-Hoover
PBC	Periodic Boundary Conditions
PME	Particle-mesh Ewald
PR	Parrinello-Rahman
sr	short-range
TIP4P	Transferable Intermolecular Potential with 4 Points
TraPPE	Transferable Potentials for Phase Equilibria
ZD	Zhang Duan potential

1. INTRODUCTION

Clathrate hydrates are crystalline solids based on water. Physically they resemble to ice in which gas molecules are trapped inside cages of hydrogen bonded water molecules. The water network which operates as host would in absence of gas molecules, called guests, be unstable and collapse into ice or water. However, clathrate hydrates do not form a compound in a chemical definition, because the formation process is a first order phase transition and not a chemical reaction.

Research on clathrate hydrates has been very intense in the last years, because of the huge amount of methane (CH_4) that has been discovered in hydrate deposits in permafrost regions and in the sea floor of continental margins [2–5]. Kvenvolden [2] and Makogon [6] assume that that the gas reserves in known hydrate reservoirs are twice as large as those contained in fossil fuel reserves. Therefore clathrate hydrates could be a key future energetic source, but many technical challenges concerning exploitation have to be solved first [7–9]. A second remarkable scientific and practical interest in clathrate hydrates is the possibility to replace CH_4 with CO_2 in natural clathrate hydrates [10–12]. Since CO_2 is responsible for about 64% of the enhanced greenhouse effect [13], a major environmental challenge is to reduce the amount of CO_2 released into the atmosphere. Recent experimental studies propose that the replacement of CH_4 with CO_2 is possible, which could make it possible to store CO_2 permanently [14–17].

In this context the research in CO_2-hydrates became more important in the last years. A powerful tool to study the microscopic structure of CO_2 clathrate hydrates and the correlation with macroscopic properties are molecular-scale computer simulations. Phenomena which occur at an atomistic scale, for example the hydrate nucleation, metastability or the growth mechanism of hydrates, can often not be observed from experiments. Therefore these phenomena have been studied intensively [18] using computer simulations and lead to a fundamental physicochemical understanding of many processes. In the center of this simulations have been the *sI*-phase of clathrate hydrates, which can be a host for both CO_2 and CH_4 molecules.

Very recent experiments suggest that there exists a different type of CO_2 clathrate hydrate occurring at higher pressures [19, 20]. Massani *et al.* showed that this new type of clathrate hydrate can be produced by a first order phase transition starting from *sI*-phase of a CO_2-hydrate [19]. The first structural description of this new type of clathrate hydrate was made by Amos *et al.* in 2017 using neutron and X-ray diffraction [1]. The water network is unrelated to any experimentally known ice, silica, or zeolite network. Instead it is related to two Zintl compounds. The guest molecules are not trapped into cages, but rather move freely inside of large

open spiral channels. In these channels, the CO_2 molecules are forming chiral spirals, therefore the phase is also called chiral-phase (χ-phase). Since the guest molecules can move freely in the channels the guest to host ratio is very variable, an energetically optimal ratio is not proposed so far.

Until now, no Molecular Dynamics simulation has been performed on this new type of clathrate hydrate. In this work the melting curve of the *sI*-phase and the χ-phase of CO_2-hydrates have been calculated for a fixed guest to host molecule ratio performing two-phase coexistence simulations using GROMACS, a modern Molecular Dynamics simulation software package[21]. Subsequently, the solid-solid coexistence line of the *sI*-phase and the χ-phase and the solid-solid coexistence line of the χ-phase and a system, in which ice VI and solid CO_2 occur separated, were computed. The position of the coexistence line were limited to certain pressure areas which allowed the comparison to experimental results.

Summarizing, the main aim of this thesis is the computation of the stability region of the *sI*-phase and the χ-phase of CO_2-clathrate hydrates and of ice VI and CO_2 separated. This allows a classification of the phase diagram of binary CO_2-H_2O-systems and lead to a deeper understanding of the χ-phase of CO_2-clathrate hydrates.

2. STRUCTURES

2.1 Clathrate Hydrates

Gas hydrates have been discovered by Sir Humphry Davy in 1810 when he noticed that a solution of water and chlorine gas can form a solid above the melting temperature of ice [22]. The first practical relevance of gas hydrates was found by Hammerschmidt in 1934. He discovered that gas pipelines in Candada have been plugged by hydrates [23]. This is a problem which still plagues oil and gas industry since gas hydrates cause blockages in pipelines [24].

Clathrate hydrates are (nonstoichiometric) crystalline solids formed by water and gas molecules [25]. There is a range of possible guest molecules such as methane, carbon dioxide, ethane, propane, isobutane or nitrogen [26]. The crystalline structure is formed by a water host lattice encaging guest molecules. This leads to a very stable structure which could be stable at temperatures where water normally is expected to be liquid [27].

Many different gas hydrates are known, the three most common structures are *sI*, *sII* and *sH* [25, 27]. *sI* already exists at low pressures as many experiments show [20, 28, 29]. Hirai et al. used Raman spectroscopy and XRD to show the transformation of *sI* to ice VI and *sII* at ~ 0.5 GPa [30]. On further compressions the structure transforms to ice VII and ice VIII and solid CO_2 [30]. At even higher pressures a transformation to the *sH* structure occurs. The for this work relevant *sI* structure is described in more detail in section 2.1.1.

Even though many different gas hydrates are known, the variety seems small compared to the riches of other network-forming materials as for example zeolites. The reason for this is that four-membered rings more likely formed in Si−O-based tetrahedral networks than in water networks [1]. The tendency to form four-membered rings increases in water networks at high pressures [31]. Therefore it seems reasonable that more hydrates remain to be discovered at high pressures.

One new structure proposed by Amos et. al. is based on neither a known ice structure nor a hydrate, but instead related to a Zintl phase [1]. This structure has a chiral network topology and hence differs qualitatively from clathrate hydrates. A detailed description of the chiral hydrates follows in section 2.1.2.

2.1.1 *sI* Phase of CO_2-Hyrdate

At most natural pressures, temperatures and gas compositions the *sI*-structure is the most stable phase of binary CO_2-H_2O-systems [32]. The discovery of the

structure of the *sI*-phase is a process which took more than two decades of X-ray diffraction experiments [25]. In the late 1940s and early 1950s, von Stackelberg *et al.* summarized different diffraction experiments [33, 34] and together with other works from von Stackelberg and Müller [35], Claussen [36, 37], and Pauling and Marsh [38] the two clathrate phases *sI* and *sII* were determined.

The structure is a Weaire-Phelan structure which includes that it is cubic and has space group $Pm\bar{3}n$. The lattice parameter has a length of 12.07 Å[39] measured at $0°C$ with CO_2 guest molecules. The unit cell contains six tetradecathedron cages, $5^{12}6^2$, and two pentagonal dodecathedron cages 5^{12} [40]. These eight cavities are build by 46 H_2O molecules. One unit cell can contain up to 8 gas molecules, experimental results suggest that between 80% and 100% of the cages are occupied [39, 41, 42].

(a)   (b)
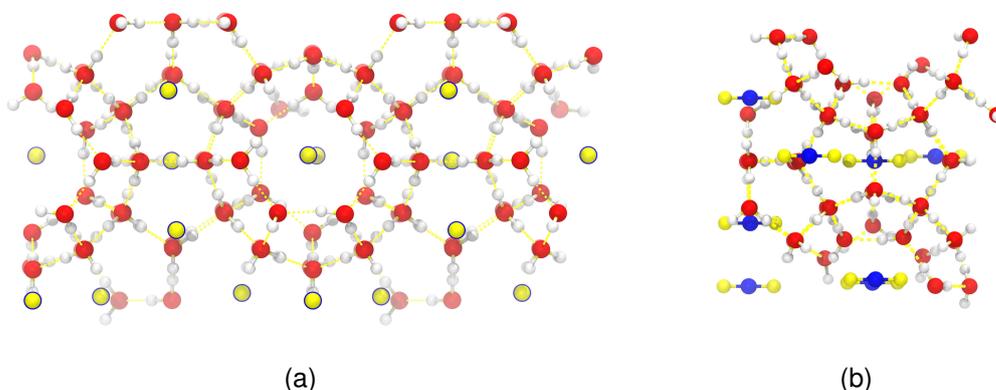
Figure 1: Projections of *sI*-Clathrate created with VMD [43]: (a) x-z view, (b) y-z view. This configuration is composed by two unit cells, each containing 46 H_2O (H-white and O-red) molecules and 8 CO_2 (C-blue and O-yellow) molecules.

2.1.2   Chiral Phase of CO_2-Hydrate

The existence of a phase which is not the *sI*-phase in a binary CO_2-H_2O-system around a pressure of approximately 1 GPa was known from several experiments. For example, Massani *et al.*[19] and Bollengier *et al.* [20] demonstrated its existence and called the structure filled ice, but they could not provide a structural description. The first description of the structure was proposed by Amos *et al.* in 2017 [1], using neutron and X-ray diffraction and calculations based on density functional theory (DFT). They recognized the water network, which plays the role of the host in the clathrate-hydrate structure, and described it as ice XVII (also denoted as C_0). Ice XVII was known before from Raman spectroscopy and it was known that it can trap H_2 molecules [44]. However, that it can also form a stable phase with trapped CO_2 molecules was unknown until 2017.

The structure has the feature, that the CO_2 molecules are not trapped in cages, but they are trapped in channels. In these channels the CO_2 molecules are ordered along

(a)   (b)
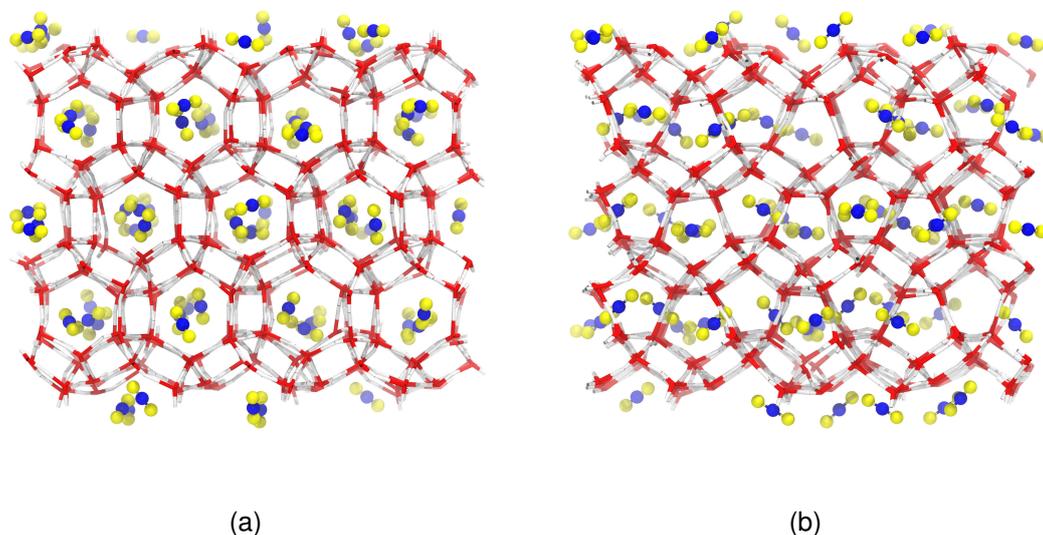
Figure 2: Projections of χ-Clathrate created with VMD: (a) x-z view, (b) y-z view. This configuration is composed by $4 \times 4 \times 4$ unit cells, each containing 6 H_2O (H-white and O-red) molecules and alternating 1 CO_2 (C-blue and O-yellow) or 2 CO_2 molecules.

a spiral which lead to a chiral structure, this is why the new phase is denoted as χ-phase. The empty ice XVII structure is a crystalline structure which has hexagonal symmetry and spacegroup $P6_122$.

This topology is not related to any other hydrate phase, ice phase, silica phase or zeolite network, but found in the Zintl phases $NaGaSn_5$ and Na_2ZnSn_5 [45][46]. The unit cell is hexagonal/trigonal with $a = 6.2753(5)$ Å and $c = 6.2988(6)$ Å. Oxygen of CO_2 is on $6b(0.08618, 0.1724, 0.25)$, $12c(0.189(3), 0.061(4), 0.044(3))$, carbon is on $12c(0.233(3), 0.017(3), 0.044(3))$.

The arrangement of the χ-phase including the channel structure allows in principle any guest molecule to host molecule ratio. Amos *et al.* showed that the maximum guest to host ratio is 1:3.5. A higher ratio would lead to unstable hydrates because the repulsion between the CO_2 molecules is too strong [1].

2.2 Ice

The phase diagram of water is very complex. There are numerous triple points leading to many different ice forms. Using cluster analysis ice can be divided into high-pressure ices (ice VII, ice VIII, ice X), low-pressure ices (ice Ih, ice Ic, ice XI) and others, which are in a modest pressure range between 200 MPa and 2000 MPa [47]. Ices from the last cluster are the most important ones for this work, because they occur in the same temperature and pressure range as *sI*-hydrate and χ-hydrate do.

Ice VI plays an outstanding role in examining hydrates, because there are suspicions that a phase transition between hydrates and ice VI is possible [19].
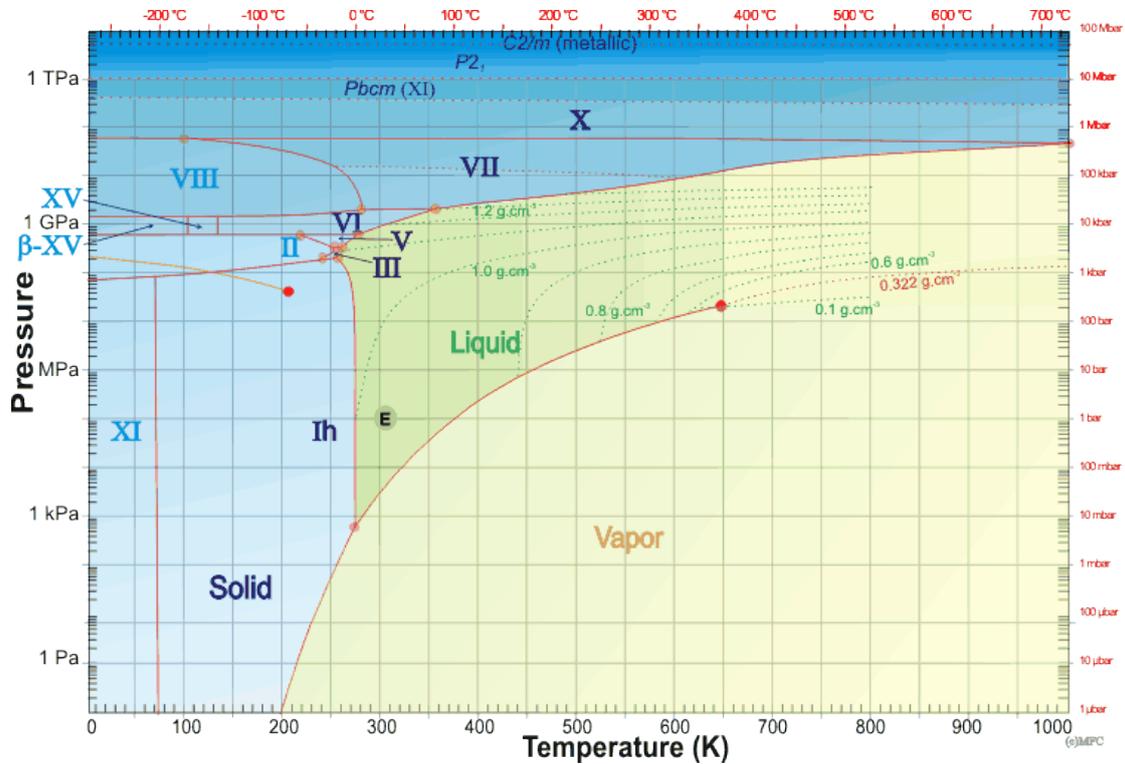
Figure 3: Pressure-temperature phase diagram of water. Note that in this phase diagram only the stable phases are shown. Ice XVII is not shown in this diagram, because without guest molecules it is not stable. The orange dots show triple points and the red dots indicate critical points. The orange lines refer to low density amorphous (LDA) and high density amorphous (HDA) ice. Figure from [48].

Figure 3 shows the phase diagram of water [48]. Only the stable phases are shown, therefore ice XVII, which forms the host network of the χ-phase of CP$_2$-hydrates as discussed in section 2.1.2 but is unstable without a guest molecule, is not shown in this diagram.

The central part of the phase diagram (in a region around 1 GPa) is very complex. Since this is the region which is relevant for the average surface conditions on Earth, Mars and Venus there is a special interest in understanding this part of the phase diagram. A phase diagram of only this region can be found in section 4.2.1 where it is compared to different water models.

2.2.1   Ice VI

The unit cell of ice VI is tetragonal with dimensions $a = 6.27$ Å and $c = 5.79$ Å and contains ten water molecules. It is in space group $P4_2/nmc$ and has a density of $1.31\,\mathrm{g\,cm^{-3}}$ [49]. There are 45 symmetrically distinct configurations leading to 576

possible arrangements of the water molecules in the unit cell under consideration of the ice-rules [50]

Ice VI has at least four triple-points (three of them are experimentally well established), potentially there could be two more triple points if there exists a low temperature proton-ordered ice [50]. The triple point with ice II and ice VI is estimated at a temperature of $-55\,^{\circ}$C and a pressure of 620 MPa. The triple point with ice V and liquid water is at a temperature of $-0.16\,^{\circ}$C and a pressure of 632.4 MPa. The triple point of ice VI, ice VII and ice VIII is at a temperature of 5 $^{\circ}$C and a pressure of 2.1 GPa and the triple point of ice VI, ice VII and liquid water is at a temperature of 118 $^{\circ}$C and a pressure of 2.216 GPa.
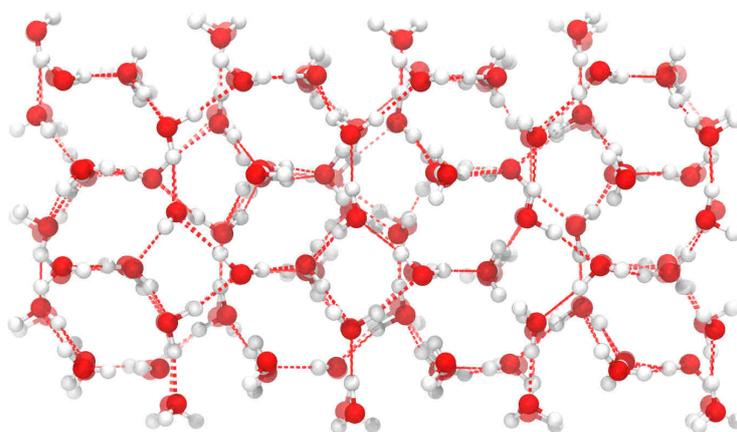
Figure 4: Projection Ice VI created with VMD: This configuration is composed of 4×2 unit cells, each containing 10 H_2O (H-white and O-red) molecules.

2.3 Carbon Dioxide

The first systematic investigation of carbon dioxide (CO_2) was made by Joseph Black in 1757, he called the gas "fixed air". In 1782 Antoine Lavoisier identified "fixed air" as a combination of carbon and oxygen. At the end of the 18th century the solubility of CO_2 in water was studied. This was the starting point for the industrial use of CO_2 for mineral water and soft drinks. In the 1830th Charles Thilorier was the first one able to produce solid CO_2 which occurred during the evaporation of liquid CO_2. Using compressors for higher pressure the production of solid CO_2 was improved and the first dry ice industry was built up in 1925 [51]. In the 20th century new crystalline structures for CO_2 were discovered, at the moment there are seven different structures known. Below a pressure of 10 GPa the only stable crystalline structure is dry ice, which is scientifically known as CO_2 phase I [52]. This structure

can also exist at modest pressure and temperature regions in which also clathrates and ice VI can exist.

2.3.1 CO_2 I

The first structural description of CO_2 phase I was made by Bart Olinger [38] in 1952. He used a high pressure X-ray diffraction technique which allowed him to measure CO_2 at 296 K from 1 GPa to 10 GPa. He compared the results with different CO_2 I models and was so able to find the space group of CO_2 phase I. The structure is a face centered cubic structure and is part of symmentry group $Pa3$. There are 4 CO_2 molecules per unit cell with 20 external degrees of freedom (12 pseudo-translational vibrations and 8 pseudo-rotational librations) [53, 54].
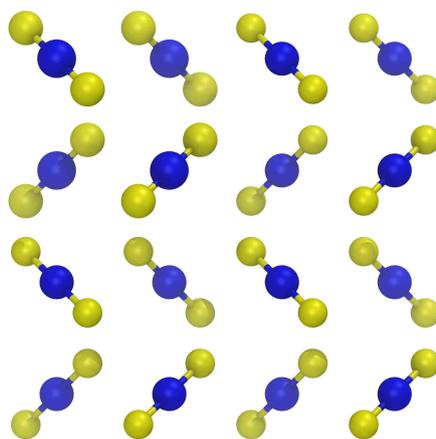
Figure 5: Projection CO_2 I: This configuration is composed of 2×2 unit cells, each containing 4 CO_2 (C-blue and O-yellow) molecules.

Despite a fundamental understanding of the phase diagram of CO_2 being important for several reasons, many aspects of solid CO_2 are unknown so far [52]. CO_2 I is one of nature's most common molecular crystals, it occurs in the Earth's atmosphere and is ubiquitous on other planets, and therefore very relevant in the field of astrochemistry. Nevertheless there are many gaps in the phase diagram of CO_2 and therefore there are intense experimental efforts today to explore the high-pressure temperature range of the phase-diagram [55].

For CO_2 I the transition mechanism which allows a transformation to CO_2 phase III, phase VI, phase IV and maybe phase II is still not completely understood yet [56]. Li *et al.* assume that the most accurate phase diagram of CO_2 is a composite from experimental data of Litasov *et al.* [57], Giordano *et al.* [58, 59] and Iota *et al.* [60, 61]. Even though a intensive experimental effort is made, the prediction of the phase

diagram is hard due to large hysteresis effects and the metastability of phase III and more research in the field is necessary.

3. BACKGROUND

In this chapter the theoretical background for understanding Molecular Dynamics simulations is provided. In section 3.1 necessary physical principles are discussed. The main emphasis in this section is on statistical mechanics, a detailed introduction to statistical mechanics is provided in *Elementary Principles of Statistical Mechanics* by J. W. Gibbs [62]. Thermodynamic principles are not discussed in this chapter, the book *Thermodynamics: An Engineering Approach* by Y. A. Cengel and M. A. Boles [63] serves as a good introduction.

In section 3.2 the principles of Molecular Dynamics simulations are discussed. *Understanding Molecular Simulation: From Algorithms to Applications* by D. Frenkel and B. Smit [64] and *The Art of Molecular Dynamics* by D. C. Rapaport [65] provide detailed introductions to Molecular Dynamics simulations and serve as reference for section 3.2.

3.1 Statistical Mechanics

3.1.1 Fundamental Principles

Statistical Mechanics is used to study systems with a large degree of freedom. In particular a concept is given which describes the relation between macroscopic observations (macro states) and microscopic configurations of a certain system (micro states). Since time and length scales of macroscopic phenomena are much larger than molecular dimensions not individual micro states but an ensemble of all micro states is considered.

A micro state determines the state of a classical N-particle system under the Born-Oppenheimer approximation with two $3N$ dimensional vectors

$$
\begin{aligned}
\mathbf{r}^N &= (\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, ..., \mathbf{r}_N) \\
\mathbf{p}^N &= (\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, ..., \mathbf{p}_N),
\end{aligned}
\tag{1}
$$

where N is the total number of particles in the system, $\mathbf{r}_i$ is position of the i-th particle and $\mathbf{p}_i$ is the momentum of the i-th particle. $\mathbf{r}^N$ and $\mathbf{p}^N$ span a $6N$ dimensional phase space, each micro state is represented by a single point in the phase space.

The total energy of a system is given by the Hamiltonian function

$$H(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N).$$

(2)

Using Hamiltonian equations of motion for Cartesian coordinates

$$\dot{\mathbf{r}} = \frac{\partial H}{\partial \mathbf{p}} \qquad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{r}},$$

(3)

the Newtonian equations of motion are following directly:

$$\dot{\mathbf{r}}_i = \frac{\mathbf{p}_i}{m_i},$$
$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i(\mathbf{r}^N) = -\nabla U(\mathbf{r}^N).$$

(4)

This is a set of $6N$ coupled ordinary first-order differential equations which can be solved in principle with initial conditions $\{\mathbf{r}^N(0), \mathbf{p}^N(0)\}$. The solution yields to a trajectory $\{\mathbf{r}^N(t), \mathbf{p}^N(t)\}$ in the phase space, which completely determines the system for all times t.

Since many Hamiltonian equations can not be solved analytically it is often necessary to solve the equations approximately with numerical methods.

3.1.2 Time Average, Ensemble Average and Ergodicity

In Molecular Dynamics simulations it is often important to compute averages of certain quantities, because macroscopic phenomena are related to the averages on a microscopic scale.

The time average of an observable $A(\mathbf{r}^N, \mathbf{p}^N)$ of a trajectory is given by

$$\overline{A}_t = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau dt\, A(\mathbf{r}^N(t), \mathbf{p}^N(t)).$$

(5)

A second way of calculating averages is the ensemble average. An ensemble is the collection of all possible micro states which can occur under given conditions. The ensemble average of an observable $A(\mathbf{r}^N, \mathbf{p}^N)$ is given by:

$$\langle A \rangle = \int d\mathbf{r}^N d\mathbf{p}^N f(\mathbf{r}^N, \mathbf{p}^N) A(\mathbf{r}^N, \mathbf{p}^N),$$

(6)

where $f(\mathbf{r}^N, \mathbf{p}^N)$ is the probability density function of micro state $\{\mathbf{r}^N, \mathbf{p}^N\}$.
In Statistical Mechanics it is often necessary to accept the Ergodic Hypothesis. The Ergodic Hypothesis claims that for large time scales the volume of a phase space

region is proportional to the time a system spends in micro states with the same energy. An alternative formulation is that a system is called ergodic, when the time average is equivalent to the ensemble average on a large time scale:

$$\overline{A}_t = \langle A \rangle. \tag{7}$$

Proving ergodicity is often impossible though it is often reasonable and necessary to assume that a system is ergodic.

3.1.3 Correlation Functions

Since MD simulations are discrete and finite all computed quantities are affected by statistical errors. Trajectories are calculated by solving the equations of motion for a finite number of steps M and for a finite step size Δt. Therefore (5) can be approximated by calculating

$$\overline{A} = \frac{1}{M} \sum_{i=1}^{M} A(\mathbf{r}_i^N(t), \mathbf{p}_i^N(t)), \tag{8}$$

where $\overline{A}$ denotes the sample average. It is easy to show that $\overline{A}$ is an unbiased estimator of the expectation of A by showing $\langle \overline{A} \rangle = \langle A \rangle$, where A_i is the abbreviation for $A(\mathbf{r}_i^N(t), \mathbf{p}_i^N(t))$.

The variance measures how far a set of numbers is spread out from their average. In statistics variance plays a central role because it contains a lot of information about the quality of a test set. Mathematically variance is defined as the expectation of the squared deviation of a random variable from its mean:

$$Var(\overline{A}) = \langle \overline{A}^2 \rangle - \langle \overline{A} \rangle^2 \tag{9}$$

Here it was assumed that the configurations in the sample are statistically independent. In MD simulations, however, configurations are correlated to each other which needs to be taken into account for the calculation of statistical errors. Since one assumes that the trajectory is stationary, the expression $\langle A_i A_j \rangle - \langle A_i \rangle \langle A_j \rangle$ should only depend on the separation between time step i and time step j the variance can be rewritten as

$$Var(\overline{A}) = \frac{1}{M^2} \sum_{i=1}^{M} \sum_{k=1-i}^{M-i} [\langle A_i A_{i+k} \rangle - \langle A_i \rangle \langle A_{i+k} \rangle]. \tag{10}$$

By summing from $k = -\infty$ to $k = \infty$, the following definition of normalized time autocorrelation function is reasonable:

$$C_\text{A}(k) = \frac{\langle \text{A}(\mathbf{r}_i^N)\text{A}(\mathbf{r}_{i+k}^N)\rangle - \langle \text{A}(\mathbf{r}_i^N)\rangle\langle \text{A}(\mathbf{r}_{i+k}^N)\rangle}{\langle \text{A}^2\rangle - \langle \text{A}\rangle^2}. \tag{11}$$

The correlation length τ_A is defined by

$$\tau_\text{A} = \sum_0^\infty C_\text{A}(k). \tag{12}$$

The variance of the sample average can be finally written as

$$Var(\overline{\text{A}}) = \frac{2\tau_\text{A}}{M} Var(\text{A}). \tag{13}$$

From (13) it can be concluded that a large sample size improves statistical accuracy. Correlations effectively reduce the sample size by a factor $\frac{1}{2\tau_\text{A}}$.

3.1.4 Thermodynamic Ensembles

Each Molecular Dynamics simulation generates a configurations which correspond to a certain thermodynamic ensemble. The most conventional MD algorithm as it is discussed in 3.2 studies the evolution of N particles in a Volume V. The energy is preserved due to Hamiltonian mechanics. Since N, V and E are constrained the microcanonical ensemble is also called (NVE)-ensemble. If the Ergodicity Hypothesis is used, then the time averages in such MD simulations are equivalent to the ensemble averages of a microcanonical ensemble.
For experimentalists on the other hand it is often easier to control parameters as temperature or pressure. Therefore other ensembles are designed in which other macroscopic parameters than N, V and E are constrained. One ensemble is assigned to each combination of constraints.
In the thermodynamic limit $N \to \infty$ and $V \to \infty$ such that $\frac{N}{V} = \rho = $ const fluctuations cease and different ensembles become indistinguishable.
Each ensemble can be represented by a partition function. The partition function is a dimensionless function of state variables of a system. Partition functions are often useful, because most thermodynamic variables can be expressed in therms of it.

Microcanonical Ensemble

The microcanonical ensemble is also denoted as (NVE)-ensemble and corresponds to an isolated system (a system which can not exchange energy or particles with the environment) with a specified total energy. Per definition all micro states with

(a) Micro-canonical   (b) Canonical   (c) Isothermal-isobaric
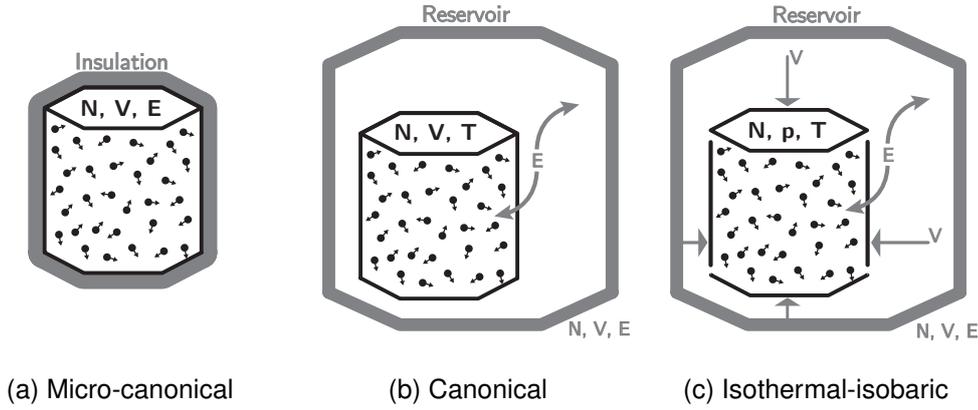
Figure 6: Overview of the relevant statistical ensembles of this work. Figure from Master thesis of Elija Feigl [66]

N-particles, a volume V and an energy $\left[E - \frac{\Delta}{2}, E + \frac{\Delta}{2}\right]$ have the same probability to occur. This probability is described by the probability density function f_{NVE}:

$$f_{NVE} = \frac{1}{\int d\mathbf{r}^N d\mathbf{p}^N \delta\left(H(\mathbf{r}^N, \mathbf{p}^N) - E\right)} \delta\left(H(\mathbf{r}^N, \mathbf{p}^N) - E\right). \tag{14}$$

The partition function Ω_{NVE} of the microcanonical ensemble describes the total number of micro states with energy E. It is expressed as

$$\Omega_{NVE} = \frac{1}{N!(2\pi\hbar)^{3N}} \int d\mathbf{r}^N d\mathbf{p}^N \delta\left(H(\mathbf{r}^N, \mathbf{p}^N) - E\right). \tag{15}$$

Here the factor $\frac{1}{N!}$ is needed due to the assumption that the N particles are indistinguishable.

Canonical Ensemble

A systems with a fixed volume V and a fixed number of particles N, but which,in contrast to the microcanonical, can exchange thermal energy with a heat bath can be described with a canonical ensemble. Due to the contact of the system with the heat bath the temperature T is a constraint. Therefore the canonical ensemble is also denoted as (NVT)-ensemble.

The probability to find a system in a certain state is described by the probability density function f_{NVT}:

$$f_{NVT} = \frac{1}{\int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta H(\mathbf{r}^N, \mathbf{p}^N)}} e^{-\beta H(\mathbf{r}^N, \mathbf{p}^N)}. \tag{16}$$

$e^{-\beta H(\mathbf{r}^N, \mathbf{p}^N)}$ is called Boltzmann-factor, $\beta = \frac{1}{k_B T}$ is called reciprocal temperature and k_B is the Boltzmann constant.

The assigned partition function Q_{NVT} is expressed as:

$$Q_{NVT} = \frac{1}{N!(2\pi\hbar)^{3N}} \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta H(\mathbf{r}^N, \mathbf{p}^N)}. \tag{17}$$

Isothermal-Isobaric Ensemble

A system which is in contact with a heat-bath and which can expand or shrink, but which can not exchange particles with its environment can be described with an isothermal-isobaric ensemble. Since volume and energy can change temperature and pressure fluctuate around constant equilibrium values. Therefore the isothermal-isobaric ensemble is also denoted as (NpT)-ensemble.

The isothermal-isobaric ensemble is important in chemistry because chemical reactions normally take place under constant pressure and temperature.

The probability to find a system in a certain state is given by the density function f_{NpT}:

$$f_{NpT} = \frac{1}{\int_0^\infty dV \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta[H(\mathbf{r}^N, \mathbf{p}^N) + pV]}} e^{-\beta[H(\mathbf{r}^N, \mathbf{p}^N) + pV]}. \tag{18}$$

The assigned partition function Δ_{NpT} is given by:

$$\Delta_{NpT} = \frac{\beta p}{N!(2\pi\hbar)^{3N}} \int dV \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta[H(\mathbf{r}^N, \mathbf{p}^N) + pV]} \tag{19}$$

3.2   Molecular Dynamics Simulations

The central purpose of a MD simulation is solving the equations of motion to study the time evolution of a classical many-body system. There are many different ways of implementing MD simulations, however the basic structure has always a similar form as in algorithm 1.

Each MD simulation starts with the initialization according to the provided input data. Central for the global algorithm is the input of the potential energy V as a function of the position of the particles in the system. Other quantities which my be initialized are the box size, masses, the number of particles, periodic boundary conditions, etc. Further information for the initialization is provided in section 3.2.2.

After the initialization, in a for-loop over M time steps of size Δt the three subroutines `force`, `integrate` and `averages` carry out the main part of the MD simulation. The subroutine `force` computes the forces on all particles. This includes

---

Algorithm 1 Global MD algorithm

```
1: call init()                                         ▷ initialization
2: t= 0
3: for M steps do
4:     call force()                                    ▷ compute forces
5:     call integrate()                                ▷ integrate EOM
6:     t += Δt                                         ▷ increase time step
7:     call averages()                                 ▷ compute averages
8: end for
```
---

the calculation of the forces of all non-bonded particles $\mathbf{F}_i = \sum_j \mathbf{F}_{ij}$ plus the forces occurring due to bonded interactions plus restraining and external forces. Further information for the computation of the forces can be found in section 3.2.3.

After the computation of the forces the subroutine `integrate` integrates the EOM:

$$\dot{\mathbf{r}} = \frac{\partial H}{\partial \mathbf{p}} \qquad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{r}}. \tag{20}$$

There are many different possibilities how to integrate the EOM. The relevant aspects of the integration step are further discussed in section 3.2.4. The two subroutines `force` and `integrate` compute the time evolution of the initialized system and represent the core of the algorithm.

Finally the subroutine `averages` computes all kind of averages which are necessary for analyzing properties of the system each time step.

3.2.1 Instability, Shadow Hamiltonian

Before discussing each step in algorithm 1 in detail it is necessary to prove if the whole procedure can reproduce the dynamics of a simulated system. For the prove one can start with the Hamiltonian in Cartesian coordinates:

$$H(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N). \tag{21}$$

Using the Hamiltonian EOM (3) the following equations of motion have to be solved

$$\begin{aligned} \dot{\mathbf{r}}_{\mathbf{i}} &= \frac{\mathbf{p}_{\mathbf{i}}}{m_i}, \\ \dot{\mathbf{p}}_i &= \mathbf{F}_i(\mathbf{r}^N) = -\nabla_{\mathbf{r}_i} U(\mathbf{r}^N), \end{aligned} \tag{22}$$

where $\mathbf{F}_i$ is the force exerted on particle i. For a N-body-system there are $6N$ coupled first order differential equations.

Starting from some initial condition $x_0^N = \{\mathbf{r}_0^N, \mathbf{p}_0^N\}$ the equations of motion 22 are

solved for small time steps Δt and produce a discrete trajectory $x_t^N(x_0^N)$ in phase space. For a small time step the hope is that the discrete trajectory is close to the real trajectory of the particle.

A displacement $\delta x_0^N = \{\delta \mathbf{r}_0^N, \delta \mathbf{p}_0^N\}$ in the initial condition leads to a new trajectory with starting condition $x_0^{N\prime} = x_0^N + \delta x_0^N$. It turns out that the initially small displacement grows exponentially

$$\left| \delta x_t^N \right| \sim \left| \delta x_0^N \right| e^{\lambda t}, \tag{23}$$

where λ is the Lyapunov exponent. In a system of N particles there are $6N$ such Lyapunov exponents. Consequently small displacements which are unavoidable due to the numerical integration of the equations of motion grow exponentially and lead to completely different trajectories. The growth of the displacement is also called Lyapunov instability and due to the instability accurate long-time predictions in many real world applications are impossible.

To understand why MD simulations are nevertheless meaningful, despite the displacement of the analytical trajectories grows exponentially, it is necessary to define time reversibility of an algorithm.

As discussed in section 3.1.1 Newtonian and Hamiltonian mechanics are time reversible and conserving phase space volume and energy. Hence integration algorithms should also obey these fundamental symmetries. Considering a trajectory $x^N(t)$ in phase space as solution of the equations of motion, if the very same trajectory but time reversed is also a solution of the same equations of motions, then the dynamics are said to be time reversible.

Integration methods which are conserving the energy exactly do not exist, however, there are integration methods which are time reversible and which are conserving phase space volume. Algorithms which are conserving phase space volume are also called symplectic algorithms. One example for a symplectic algorithm is the velocity Verlet algorithm which is discussed in section 3.2.4.

For a integration method it is important to be symplectic not only because one would like to obey the fundamental properties of the Hamiltonian dynamics but also because symplectic algorithms have a shadow Hamiltonian H_S.

By backward error analysis it is possible to show that starting from a specific initial state and using a symplectic algorithm of order n with finite time step Δt for the Hamiltonian $H(\mathbf{r}^N, \mathbf{p}^N)$ one gets a discrete trajectory that lies on the exact continuous trajectory that has the same initial state and which is generated by a shadow Hamiltonian $H_S = H + \Delta H$, where $\Delta H = \mathcal{O}(\Delta t^n)$. This holds if the potential energy function $U(\mathbf{r}^N)$ is analytic.

During a MD simulation with a symplectic algorithm one does not obtain a trajectory that conserves H but one that conserves H_S, which differs from the original Hamiltonian by a term of oder $\mathscr{O}(\Delta t^N)$. Therefore in contrast to non symplectic algorithms, symplectic algorithms show no long-time drift of the energy but only fluctuations of order $\mathscr{O}(\Delta t^N)$.

Whenever one deals with quantities that do not depend on slightly different Hamiltonians, as it is often the case in MD calculations, the existence of the shadow Hamiltonian guarantees correct predictions.

3.2.2 Initialization

The first step of each simulation is the initialization. During the initialization the initial positions and velocities are assigned to all particles in the system. The positions of the particles must be chosen in a way to be compatible with the structure which should be simulated. If velocities can not be provided it is possible to generate them. There are different possibilities how to generate velocities, in GROMACS the Maxwell-Boltzmann distribution for the velocities v_i, $i = 1,\dots,3N$ at temperature T is chosen:

$$p(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} e^{-\frac{m_i v_i^2}{2k_B T}}, \tag{24}$$

where k_B is Boltzmann's constant. This is accomplished by the generation of 12 random numbers R_k in the range $0 \leq R_k < 1$, subtracting 6 of the sum and multiplying the result with the standard deviation of the velocity distribution $\sqrt{\frac{k_B T}{m_i}}$. The in this way generated energy does not correspond perfectly to the temperature T, therefore it is necessary to remove the center-of-mass motion and scale the velocities in a way that the total energy corresponds to the temperature T.

3.2.3 Calculating Forces

The computation of the forces acting on each particle in the system is typically the computationally most expensive part of MD simulations. Considering force models which are pairwise additive (which are used in this work, see section 4.1) each particle can interact with $N-1$ neighbors leading to a total of $\frac{N(N-1)}{2}$ interactions for a system of N particles. This implies that the time which is necessary to evaluate the forces scales as $\mathscr{O}(N^2)$. Fortunately there are techniques to make the computation more efficient. It is possible to distinguish between short-range and long-range interactions in a way that for long-range forces the time scales as $\mathscr{O}(N)$, rather than $\mathscr{O}(N^2)$. This is discussed in more detail in section 3.2.6 and section 3.2.7. For finding all particles in the short-range area it is necessary to use neighbor searching

algorithms. There exist different efficient techniques to generate lists in which the neighbors of each particle can be stored. The generation of this lists is in principle $\mathcal{O}(N^2)$ which is quite expensive, but the lists have the advantage that it is not necessary to recompute them each time step. Furthermore it is possible to reduce the costs to $\mathcal{O}(N)$ if a cutoff-scheme is used as discussed in section 3.2.6.

3.2.4 Integrating the Equations of Motion

There are numerous methods for solving the set of differential equations, for example Runge-Kutta methods, Predictor-corrector method or Linear multistep methods.
All of them have different advantages and disadvantages, generally they should be fast, require less memory and should permit the use of a large time step.

Velocity Verlet Algorithm

The velocity verlet algorithm is a commonly used integration method which is also used in the context of this work. It can be derived from the Taylor expansion of $\mathbf{r}(t + \Delta t)$ and the time reversed Taylor expansion of $\mathbf{r}(t)$:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m}\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{25}$$

$$\mathbf{r}(t) = \mathbf{r}(t + \Delta t) - \mathbf{v}(t + \Delta t)\Delta t + \frac{\mathbf{F}(t + \Delta t)}{2m}\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{26}$$

By inserting (25) in (26) one gets the following equation

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t) + \mathbf{F}(t + \Delta t)}{2m}\Delta t + \mathcal{O}(\Delta t^2) \tag{27}$$

which, together with (25, forms the velocity Verlet algorithm. The velocity Verlet algorithm is time reversible, as can be shown by applying the algorithm using inverted momenta which is necessary to obey Hamilton mechanics.

For proving phase space volume conservation it is necessary to use the Liouville's theorem (see chapter 4, *Frenkel & Smit* [64]).

3.2.5 Periodic Boundary Conditions

Usually MD simulations should predict properties of macroscopic systems ($\sim 10^{23}$ particles), however present-day computer simulations can handle simulations in a range of a few hundred to a few million particles. For such small systems the ratio between surface and internal particles is much higher than the ratio in macroscopic

systems, therefore it is not safe to assume that the choice of the boundary conditions is negligible.

In order to simulate bulk phases it is essential to assume that the simulated N-particle system is surrounded by infinite bulk. Therefore periodic boundary conditions (PBC) are applied. Using PBC, the volume containing the N particles of the simulation, called simulation box, is treated as primitive cell. By duplicating this primitive cell an infinite periodic lattice of identical cells is created (see figure 7). A certain particle can now not only interact with the particle in the simulation box but also with all other particles in this infinite arrangement. Assuming interactions to be pairwise additive, the total potential energy of the particles in the simulation box is given by

$$U_{tot} = \frac{1}{2} \sum_{i,j,n}' u\left(\left|\mathbf{r}_{ij} + \mathbf{n}L\right|\right), \tag{28}$$

where L is the diameter of the simulation box (which for convenience is assumed to be cubic) and $\mathbf{n} = (n_x, n_y, n_z)$ with $n_x, n_y, n_z \in \mathbb{N}$ addressing a certain periodic image. The prime indicates that the term $i = j$ is excluded when $\mathbf{n} = \vec{0}$.
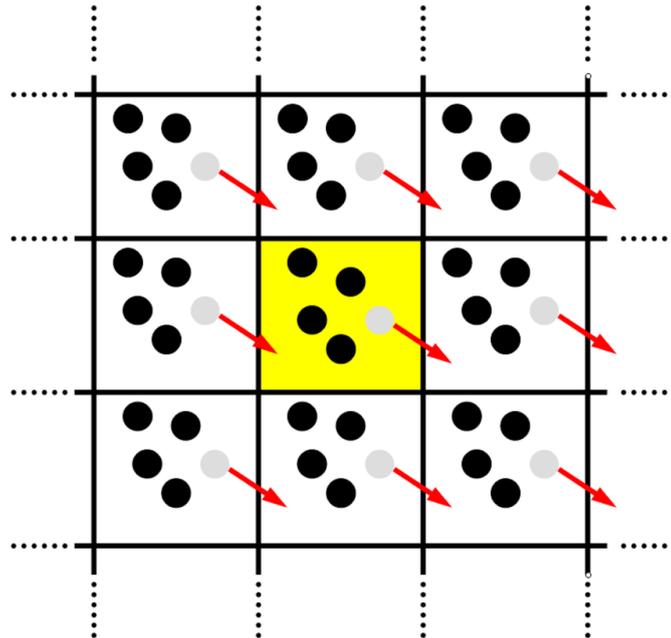
Figure 7: Two dimensional visualization of periodic boundary conditions. A particle (gray) leaving the primitive cell (yellow) along the right arrow and enters the primitive cell on the opposing side of the primitive cell. Figure from *M. O. Steinhauser* [67].

Although using PBC is a very effective method for simulating certain systems, it is necessary to be aware of correlations produced by PBC which are not occurring in macroscopic systems.

3.2.6   Truncation of Interactions

Since often only short-range interactions are important, infinite sums as in 28 become finite by truncating interactions beyond a certain cutoff distance r_c. Interactions between a particle and its neighboring particles are only computed, when their distance is smaller than r_c. The resulting error can be made arbitrary small by choosing r_c sufficiently large. Using a cutoff radius (28) can be rewritten as

$$U_{tot} = \sum_{i<j} u_c(r_{ij}) + U_L, \tag{29}$$

where u_c stands for the truncated potential energy function and U_L for the remaining long-range interaction correction.

There are different methods to truncate the potential. The simplest truncation is the one explained above by just calculating interactions within a certain cutoff distance r_c. A more common choice in MD simulations is to truncate and shift the potential, such that the potential vanishes at the cutoff radius:

$$u_{tr-sh}(r) = \begin{cases} u(r) - u(r_c), & \text{for } r \leq r_c \\ 0, & \text{for } r > r_c \end{cases}. \tag{30}$$

The additional shift has the advantage that there are no discontinuities in the potential and hence the intermolecular forces are always finite.

A third way of truncation is using minimal image convention. In this case the truncation is not spherical, instead the interaction of a particle i with particle j is always with the particle j in the nearest periodic image.

3.2.7   Long-Range Correction

Using a truncation as described in section 3.2.6 there are different correction techniques for including long-range interactions. Corrections are necessary, because otherwise serious differences occur. The simplest correction is a tail correction. The pair potential $u(r)$ gets truncated and a remaining contribution can be estimated as:

$$U_{tail} = \frac{N\rho}{2} \int_{r_c}^{\infty} \mathrm{d}r\, u(r) 4\pi r^2. \tag{31}$$

This expression can be solved in certain cases analytically. If this is not possible advanced methods are necessary, in this work Particle-mesh Ewald summation (PME) was used.

Particle-Mesh Ewald Summation

Particle-mesh Ewald (PME) is a method developed by *Tom Darden, Darrin York* and *Lee Pedersen* in 1994 for evaluation electrostatic energies and forces of large periodic systems [68]. It is based on Ewald summation, which is named after *Paul Peter Ewald* who has used this method in 1921 to calculate optical and electrostatic potentials of grid potentials [69]. PME has the advantage that it scales with $N \log N$ and is therefore superior to classical Ewald summation for almost all systems.

The main ideal of Ewald summation is to rewrite the used potential U as a sum of two potentials, one for short-range interactions U_{sr} and one for long-range interactions U_{lr}:

$$U_{tot} = U_{sr} + U_{lr}. \tag{32}$$

This has a huge advantage since the short-range term converges quickly in real space, the long-range term on the other side converges quickly in Fourier (reciprocal) space. The short-range term can be typically computed easily since a suitable cutoff makes the system small. For the long-range part it is necessary to find a convenient mathematical form to handle the computational costs. The chosen form is most typically a Gaussian charge distribution around all occurring charges. The position of the charges can be described with a Dirac delta distribution. In principle the long-range potential is the sum of all energies in a unit cell and the energies with all other particles in the system:

$$E_{lr} = \int \int d\mathbf{r} d\mathbf{r}' \rho_{tot}(\mathbf{r}) \rho_{uc}(\mathbf{r}') \varphi(\mathbf{r} - \mathbf{r}'), \tag{33}$$

where $\rho_{uc}(\mathbf{r}')$ is the charge density of the unit cell, $\rho_{tot}(\mathbf{r})$ is the charge density of the total system and $\varphi(\mathbf{r} - \mathbf{r}')$ is the chosen form of the potential. $\rho_{uc}(\mathbf{r}')$ and $\rho_{tot}(\mathbf{r})$ can be described by using Dirac delta distribution:

$$\rho_{uc}(\mathbf{r}) = \sum_k q_k \delta(\mathbf{r} - \mathbf{r}_k), \tag{34}$$

where k represents charged particles and q_k is the charge of particle k and

$$\rho_{tot}(\mathbf{r}) = \sum_{n_1, n_2, n_3} \sum_k q_k \delta(\mathbf{r} - \mathbf{r}_k - n_1 \mathbf{a}_1 - n_2 \mathbf{a}_2 - n_3 \mathbf{a}_3) \tag{35}$$

where n_1, n_2, n_3 are integers and $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are the lattice vectors of the unit cell. $\rho_{tot}(\mathbf{r})$ can be represented by the convolution of the lattice function $L(\mathbf{r})$ and the $\rho_{uc}(\mathbf{r})$ which is in Fourier space simply the product:

$$\tilde{\rho}_{tot}(\mathbf{k}) = \tilde{L}(\mathbf{k}) \cdot \tilde{\rho}_{uc}(\mathbf{k}), \tag{36}$$

where

$$L(\mathbf{r}) = \sum_{n_1, n_2, n_3} \delta(\mathbf{r} - n_1 \mathbf{a}_1 - n_2 \mathbf{a}_2 - n_3 \mathbf{a}_3) \tag{37}$$

has in Fourier space the form

$$\tilde{L}(\mathbf{k}) = \frac{(2\pi)^3}{\Omega} \sum_{m_1, m_2, m_3} \delta(\mathbf{r} - m_1 \mathbf{b}_1 - m_2 \mathbf{b}_2 - m_3 \mathbf{b}_3), \tag{38}$$

where $\mathbf{b}_1, \mathbf{b}_2$ and $\mathbf{b}_3$ are the reciprocal lattice vectors and Ω is the volume of the unit cell. The single-particle potential $v(\mathbf{r})$

$$v(\mathbf{r}) = \int d\mathbf{r}' \rho_{uc}(\mathbf{r}') \varphi(\mathbf{r} - \mathbf{r}') \tag{39}$$

can be presented in Fourier space as

$$\tilde{V}(\mathbf{k}) = \tilde{\rho}_{uc}(\mathbf{k}) \cdot \tilde{\Phi}(\mathbf{k}). \tag{40}$$

This allows to write the long-range energy as

$$E_{lr} = \int d\mathbf{r} \rho_{tot}(\mathbf{r}) v(\mathbf{r}). \tag{41}$$

Using Parseval's theorem the long-range energy can be written as sum

$$E_{lr} = \frac{1}{\Omega} \sum_{m_1, m_2, m_3} \left| \tilde{\rho}_{uc}(\mathbf{k}) \right|^2 \tilde{\Phi}(\mathbf{k}). \tag{42}$$

3.2.8 Constraints

Considering the simulation of molecules, a problem which can occur is that intermolecular motions such as bond vibrations are on an other time scale than the interaction of molecules. For simulating bond vibrations a very small time step would be necessary which has the disadvantage that the dynamics of the whole system are evolving very slow. There are different approaches for handling this problem. One approach is to assume that molecules are rigid. This approximation is reasonable, because due to the typically small amplitude of bond vibrations the shape of the molecule is influenced only little. For water additionally the stretching mode is almost always in the quantum mechanical ground state which justifies the approximation of a rigid molecule. The most common ways to guarantee rigid molecules are using quaternions or constraint algorithms.

Using a constraint algorithm one imagines that each constraint is maintained by a constraint force. It is then possible to solve the EOM considering the constraint forces. Considering n constraints, each constraint is described by a function

$$\sigma_\alpha(\mathbf{r}^N) = 0 \tag{43}$$

where $\alpha = 1, \ldots, n$. The constraint algorithm is based on the method of Lagrange multipliers, the modified Lagrangian has the following form:

$$\mathcal{L}' = \mathcal{L} - \sum_{\alpha=1}^{n} \lambda_\alpha \sigma_\alpha(\mathbf{r}^n) \tag{44}$$

where λ_α are Lagrange multipliers. Using the Lagrangian EOM

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \mathcal{L}'}{\partial \dot{\mathbf{r}}} - \frac{\partial \mathcal{L}'}{\partial \mathbf{r}} = 0 \tag{45}$$

one gets

$$m_i \ddot{\mathbf{r}}_i = -\nabla_i U - \sum_{\alpha=1}^{n} \lambda_\alpha \nabla_i \sigma_\alpha = \mathbf{F}_i + \sum_{\alpha=1}^{n} \mathbf{G}_i^\alpha. \tag{46}$$

where $\mathbf{G}_i^\alpha$ denote the constraint forces. These are the adopted EOM which have to be solved for the dynamics of the system. Still it is necessary to find the Lagrange multipliers which are undefined so far. It is possible to show, that the Lagrange multipliers are given by

$$\Lambda = \mathcal{M}^{-1}(\mathcal{F} + \mathcal{T}) \tag{47}$$

where

$$\mathcal{F}_\alpha = \sum_i \frac{1}{m_i} \mathbf{F}_i \cdot \nabla_i \sigma_\alpha \tag{48}$$

$$\mathcal{M}_{\alpha\beta} = \sum_i \frac{1}{m_i} \nabla_i \sigma_\alpha \cdot \nabla_i \sigma_\beta \tag{49}$$

$$\mathcal{T}_\alpha = \mathrm{Tr}\Big(\sum_{ij} \nabla_i \nabla_j \sigma_\alpha \dot{\mathbf{r}}_i \dot{\mathbf{r}}_j\Big). \tag{50}$$

The solution of the adapted EOM in formula 46 together with the Lagrange multiplier obtained by formula 47 give the solution of the exact EOM. However, it is not guaranteed that these solutions hold for the approximated MD simulations. Therefore some different methods have been developed which maintain the constraints exactly in the approximation of MD simulations. The most famous algorithm is the SHAKE algorithm [70]. In this work the chosen method was using

the constraint algorithm LINCS [71] developed by *Hess, Bekker, Berendsen* and *Fraaije* in 1997.

LINCS

Using the notation used by *Hess et al.* [71] the Newtonian EOM can be written as

$$\frac{\mathrm{d}^2 \mathbf{r}}{\mathrm{d} t^2} = \mathbf{M}^{-1} \mathbf{F} \tag{51}$$

where $\mathbf{F}$ is a $3N$ force vector and $\mathbf{M}$ is a $3N \times 3N$ diagonal matrix which contains the masses of the particles. There are K independent constraint equations

$$g(\mathbf{r})_i = \left| \mathbf{r}^{i_1} - \mathbf{r}^{i_2} \right| - d_i = 0 \tag{52}$$

where $i = 1, \ldots, K$ and d_i is the bond length between particle i_1 and particle i_2. Using again the method of Lagrange multiplier the adopted EOM has the form

$$-\mathbf{M} \frac{\mathrm{d}^2 \mathbf{r}}{\mathrm{d} t^2} + \mathbf{B}^\mathsf{T} \lambda + \mathbf{F} = 0 \tag{53}$$

where $\mathbf{B}_{hi} = \frac{\partial g_h}{\partial \mathbf{r}_i}$. Replacing λ the EOM to be solved has the form

$$\frac{\mathrm{d}^2 \mathbf{r}}{\mathrm{d} t^2} = (\mathbf{I} - \mathbf{TB}) \mathbf{M}^{-1} \mathbf{F} - \mathbf{T} \frac{\mathrm{d} \mathbf{B}}{\mathrm{d} t} \frac{\mathrm{d} \mathbf{r}}{\mathrm{d} t} \tag{54}$$

where $\mathbf{I}$ is the unity and $\mathbf{T} = \mathbf{M}^{-1} \mathbf{B}^\mathsf{T} (\mathbf{B} \mathbf{M}^{-1} \mathbf{B}^\mathsf{T})^{-1}$ is the matrix which transforms motions in the constrained coordinates into motions in Cartesian coordinates. Using a Leap-Frog integration scheme to solve this equation the new positions can be computed by

$$\mathbf{r}_{n+1} = \mathbf{r}_{n+1}^{unc} - \mathbf{M}^{-1} \mathbf{B}_n (\mathbf{B}_n \mathbf{M}^{-1} \mathbf{B}_n^\mathsf{T})^{-1} (\mathbf{B}_n \mathbf{r}_{n+1}^{unc} - \mathbf{d}) \tag{55}$$

here $\mathbf{d}$ is the vector containing all constrained bond lengths. Inverting $(\mathbf{B}_n \mathbf{M}^{-1} \mathbf{B}_n^\mathsf{T})$ needs half of the CPU time and has to be done every time step. Therefore the following transformation is made

$$(\mathbf{B}_n \mathbf{M}^{-1} \mathbf{B}_n^\mathsf{T})^{-1} = \mathbf{S}(\mathbf{I} - \mathbf{A}_n)^{-1} \mathbf{S} \tag{56}$$

where

$$\mathbf{S} = \mathrm{Diag}\left(\sqrt{\frac{1}{m_{1_1}} - \frac{1}{m_{1_2}}}, \cdots, \sqrt{\frac{1}{m_{K_1}} - \frac{1}{m_{K_2}}} \right). \tag{57}$$

In this form the following expansion can be used

$$(\mathbf{I} - \mathbf{A}_n)^{-1} = \mathbf{I} + \mathbf{A}_n + \mathbf{A}_n^2 + \dots \tag{58}$$

which reduces the computational costs extremely. In GROMACS this expansion was improved by using the approximation

$$(\mathbf{I} - \mathbf{A}_n)^{-1} = \mathbf{I} + \mathbf{A}_n + \dots + \mathbf{A}_n^{N_i} + (\mathbf{A}_n^* + \dots \mathbf{A}_n^{*N_i})\mathbf{A}_n^{N_i} \tag{59}$$

where A_n^* all elements except those that couple constraints within rigid triangles are zero and $N - i$ is the order or the expansion. A detailed explanation of this approximation can be found in the P-LINCS paper by *Hess* [72].

3.2.9 Temperature and Pressure Coupling

Applying the procedure discussed so far in section 3.2 will result in a simulation of the microcanonical ensemble, since the number of particles N can not change, the volume V of the simulation box is fixed and the energy E is constrained by the EOM.

For performing simulations in the canonical ensemble *(NVT)* or in the isothermal-isobaric ensemble *(NVT)* it is necessary to control temperature and pressure explicitly. For temperature control it is necessary to use a thermostat, for pressure control a barostat is needed.

There are two different approaches for sampling a different ensemble. Either the time evolution from the Newtonian EOM gets combined with appropriate stochastic steps or the Hamiltonian gets modified resulting in modified equations of motion. The second approach was introduced by *Andersen* [73] in 1979 and is the most common used in MD simulations. There exist many different thermostats and barostats, for this work Nosé-Hoover thermostat and Parrinello-Rahman barostat were used.

Nosé-Hoover Thermostat

The Nosé-Hoover thermostat is an algorithm to keep a constant temperature in a MD simulation. Originally it was developed by *Nosé* in 1984 [74], one year later it was improved bei *Hoover* [75]. A heat bath consisting of only one imaginary particle achieves constant temperature conditions during the simulation. The imaginary particle introduces an additional degree of freedom which checks if the temperature is too low or too high compared to the target temperature. This degree of freedom is denoted as s, the corresponding momentum is denoted as p_s. The adapted

Hamiltonian which includes the additional degree of freedom has the following form:

$$H_{NH}(\mathbf{r}^N, \mathbf{p}^N, p_s, s) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i s^2} + U(\mathbf{r}^N) + \frac{p_s^2}{2Q} + g k_B T \ln(s). \tag{60}$$

The parameter Q is called thermostat mass and has the units of energy·time^2. It determines how quickly the thermostat relaxes. T is the desired temperature and g has to be chosen in a way, that the microcanonical distribution in the $6N + 2$ phase space yields a canonical distribution in the $6N$ phase space spanned by N real particles.

Using virtual coordinates and defining a friction coefficient ζ

$$
\begin{aligned}
\mathbf{r}'_i &= \mathbf{r}_i \\
\mathbf{p}'_i &= \frac{\mathbf{p}_i}{s} \\
t' &= \int^t \frac{\mathrm{d}t}{s} \\
\zeta &:= \frac{p_s}{Q}
\end{aligned}
\tag{61}
$$

the EOM to the Hamiltonian H_{NH} have the following form:

$$
\begin{aligned}
\frac{\mathrm{d}\mathbf{r}'_i}{\mathrm{d}t'} &= \frac{\mathbf{p}'_i}{m_i} \\
\frac{\mathrm{d}\mathbf{p}'_i}{\mathrm{d}t'} &= \mathbf{F}_i - \zeta \mathbf{p}'_i \\
\frac{\mathrm{d}\zeta}{\mathrm{d}t'} &= \frac{1}{Q}\left[\sum_{i=1}^{N} \frac{\mathbf{p}_i'^2}{m_i} - 3N k_B T\right] \\
\frac{\mathrm{d}\ln(s)}{\mathrm{d}t'} &= \zeta.
\end{aligned}
\tag{62}
$$

Introducing ζ in the Hamiltonian H_{NH} and finding $g = 3N$ the Hamiltonian has the form:

$$H_{NH}(\mathbf{r}^N, \mathbf{p}^N, \zeta, s) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i s^2} + U(\mathbf{r}^N) + \frac{\zeta^2 Q}{2} + 3N k_B T \ln(s). \tag{63}$$

For this Hamiltonian it can be easily shown that the energy is conserved. and that the resulting EOM in equation 62 are unique.

Using Nosé-Hoover EOM it is possible to understand the way the thermostat works. The third equation in (62) determines the sign of the friction coefficient by comparing (twice) the kinetic energy $\sum_{i=1}^{N} \frac{\mathbf{p}_i'^2}{m_i}$ to (twice) the average kinetic energy $3N k_B T$. If the kinetic energy is larger than the average kinetic energy for temperature

T, than the rate of the change of the friction coefficient ζ is positive. Equation two in (62) couples the friction coefficient to the momentum. If the friction coefficient is positive, then the momenta get decelerated and the kinetic energy of the system decreases.

Equation 62, line 3, furthermore shows that the thermostat does not react instantaneously, but the time derivative reacts. This mechanism is called integral feedback. The reaction time of the thermostat is controlled by the parameter Q. For large values of Q the thermostat reacts quickly resulting in a larger disturbance in the dynamics of the system. The goal is finding a Q as large as possible affecting the dynamics as little as possible. As a measure for the disturbance of the dynamics one can use the velocity autocorrelation function. The thermostat will change the velocity autocorrelation function but it should not be changed qualitatively.

Parrinello-Rahman Barostat

For a isothermal-isobaric ensemble it is not only necessary to control temperature but also to control pressure. The Parrinello-Rahman barostat is named after *M. Parrinello* and *A. Rahman* who have designed the barostat in 1981 [76]. The Parrinello-Rahman barostat is commonly used in combination with Nosé-Hoover-thermostat. For simplification here only the Parrinello-Rahman modification of the NH-EOM are shown.

Vectors of the simulation box are represented in a matrix **b** which obeys the following EOM:

$$\frac{\mathrm{d}^2\mathbf{b}}{\mathrm{d}t^2} = V\mathbf{W}^{-1}\mathbf{b}^{\mathsf{T}-1}(\mathbf{P} - \mathbf{P}_{ref}). \tag{64}$$

V denotes the volume of the simulation box, **W** is a matrix parameter which controls the strength of the coupling. The matrix **P** denotes the current pressure and the matrix $\mathbf{P}_{ref}$ denotes the reference pressure. The Newtonian EOM is then modified in the following way:

$$\frac{\mathrm{d}^2\mathbf{r}_i}{\mathrm{d}t^2} = \frac{\mathbf{F}_i}{m_i} - \mathbf{M}\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} \tag{65}$$

$$\mathbf{M} = \mathbf{b}^{-1}\left[\mathbf{b}\frac{\mathrm{d}\mathbf{b}^{\mathsf{T}}}{\mathrm{d}t} + \frac{\mathrm{d}\mathbf{b}}{\mathrm{d}t}\mathbf{b}^{\mathsf{T}}\right]\mathbf{b}^{\mathsf{T}-1} \tag{66}$$

If the strength of the coupling $\mathbf{W}^{-1}$ is zero, then **M** is zero and the adopted EOM 65 is unchanged compared to the Newtonian EOM. The strength of the coupling depends

on the isothermal compressibility matrix β, the pressure time constant τ_p and the largest box matrix element L:

$$\mathbf{W}^{-1} = \frac{4\pi^2\beta}{3\tau_p^2 L}. \tag{67}$$

Often it is recommendable to use the PR-barostat only for equilibrated systems. If the barostat is used in a system with a pressure far away from the equilibration pressure the system starts to oscillate and could even crash.

3.3   Software

3.3.1   GROMACS

Gromacs, *GROningen MAchine for Chemical Simulations* [21] [77] is a MD software package mainly designed for biophysical applications. The development started in 1991 at the University of Groningen, Netherlands, in the Biophysical Chemistry department. GROMACS is free, open-source and released under the GNU Lesser General Public License (LGPL).

GROMACS has implemented all standard algorithms needed for MD simulations (e.g. Nosé-Hoover thermostat, Parrinello-Rahman barostat, PME, PBC, velocity-Verlet-algorithm, etc.). It can run in parallel, which was crucial for using the Vienna Scientific Cluster (VSC), a supercomputer cluster operated by *University of Vienna, TU Wien* and *University of Natural Resources and Life Sciences, Vienna (BOKU)*. Furthermore many different analysis tools are available. A documentation of all provided algorithms is available at `http://manual.gromacs.org/documentation/`.

For this thesis GROMACS version 2016.4, released on September 15, 2017, was used.

4. MOLECULAR MODELS AND SIMULATION DETAILS

For MD simulations of CO_2-hydrates it is necessary to model water molecules and to model CO_2 molecules. The modelling consists of two parts: On the one hand it is necessary to define how the molecules interact, for that it is necessary to define different potentials to simulate occurring force fields. This is presented in section 4.1. On the other hand it is necessary to consider static properties of molecules. Such static properties the atoms which form a molecule, the masses, virtual particles, etc. are defined in the topology of a molecule. The used water model is discussed in section 4.2, the model for the CO_2 molecules is discussed in section 4.3.

4.1 Interaction Functions and Force Fields

As discussed in section 3.2 it is necessary to calculate the forces on each particle in a MD simulation. Therefore the choice of the potentials in a system has a large impact on the dynamics of the system. As mentioned before the true interaction is given by solutions of quantum mechanical differential equations as the Schrödinger equation or the Dirac equation for all electrons and nuclei. Since such systems are generally not possible to solve it is necessary to make assumptions for the potentials. In principle one can assume that the potential functions can be subdivided into three parts:

- *Non-bonded:* Usually non-bonded interactions are Van der Waals interactions and electrostatic interactions.

- *Bonded:* Bonded interactions are necessary for modeling covalent bond-stretching, angle-bending or dihedrals.

- *Restraints:* Special potentials impose restraints on the motion of the system to avoid rearrangements of critical parts during relaxation.

4.1.1 Non-bonded Interactions

Interactions can be simulated by implementing potentials. The potentials evoke a force field which leads to interactions between molecules. Therefore the choice of

the force field is crucial. In general the potential energy U can be described as the sum of n-body potentials u_n:

$$U = \sum_{i}^{N} u_1(\mathbf{r}_i) + \sum_{\substack{i,j \\ i<j}}^{N} u_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{\substack{i,j,k \\ i<j<k}}^{N} u_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \ldots \tag{68}$$

u_1 denotes the one-body term, u_i denotes the i-body term. The one-body term is only meaningful in the presence of a external force field, because without an external force field the energy of the atoms should not depend on the position. Assuming the absence of an external force field the potential depends only of the relative interatomic distances $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$ and the bond angles θ_{ijk}. The potential function reduces to the form:

$$U = \sum_{\substack{i,j \\ i<j}}^{N} u_2(r_{ij}) + \sum_{\substack{i,j,k \\ i<j<k}}^{N} u_3(r_{ij}, r_{ik}, \theta_{ijk}) + \ldots \tag{69}$$

Since three- and higher-body problems are numerically expensive, the potentials have to be approximated as pairwise-additive potentials where many-body potentials are embedded:

$$U = \sum_{\substack{i,j \\ i<j}}^{N} u_2^{eff}(r_{ij}). \tag{70}$$

Such an approximation is often called effective pair potential. A pair potential can be empirical, when it should simply reproduce experimental results. An example for this is the Lennard-Jones potential, or it can be derived from fundamental physical laws like for instance the Coulomb potential.

Lennard-Jones Potential

The Lennard-Jones interaction is an approximation for the interaction between neutral atoms or molecules [78]. It was first proposed by *John Lennard-Jones* in 1924 and its most common form is

$$u^{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{71}$$

where σ_{ij} is the distance at which the inter-particle potential is zero and ϵ_{ij} is the depth of the potential well. The term r^6 describes the attraction at long ranges which are mainly the Van der Waals force and the dispersion force, the r^{-12} term describes the Pauli repulsion due overlapping electron orbitals and is therefore

relevant at short ranges. The potential can be simplified by using Lorentz-Berthelot rules proposed by *H. A. Lorentz* [79] and *D. Berthelot* [80]:

$$\begin{aligned} \sigma_{ij} &= \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \\ \epsilon_{ij} &= \sqrt{\epsilon_{ii}\epsilon_{jj}}. \end{aligned} \tag{72}$$

The L-J potential is limited by two parameters ϵ_{ij} and σ_{ij} which determine the energy and length scale. This has the consequence that the potential is unique and can not be fitted to properties of real materials. Nevertheless the L-J potential is very popular in MD simulations.

The most wide spread alternative is the Buckingham potential [81] proposed by *R. Buckingham*. The repulsive r^{-12} term is replaced by an exponential term Ae^{-Br} which has the advantage that the repulsion can be modeled more precisely. However, if $r \to 0$ the exponential term Ae^{-Br} converges while the r^{-6} diverges which may be problematic if one is dealing with very short range distances.

Coulomb Potential

The Coulomb potential is the central potential for charged particles. It was developed by *Charles Augustin de Coulomb* in 1785 and is fundamental for electrostatics. The potential has the form:

$$u^C(r_{ij}) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r_{ij}}, \tag{73}$$

where ϵ_0 is the electric constant, q_i is the charge of particle i and q_j is the charge of particle j. Since the decay r^{-1} is slow, the error is large when using a cutoff-scheme. It is either necessary to use advanced cutoff schemes like a cutoff in combination with a shift and additionally long-range corrections or it is necessary to introduce a reaction field which is a method to simulate long-range interactions.

4.1.2 Bonded Interactions

Bonded interactions are the interactions within a molecule. They are not only pair interactions but also 3- and 4-body interactions. Bonded interactions can be subdivided into bond stretching interactions which are 2-body interactions, angle-bending interactions which are 3-body interactions and dihedral-angle interactions which are 4-body interactions. In this section the discussion of bond-stretching potentials and angle-bending potentials will be very rudimentary because the used models for water and carbon dioxide are rigid. Neither bond-stretching nor angle-

bending occurs. Since in this work no structures with dihedral properties occur this potential will be neglected.

Bond-stretching Potentials

The simplest bond-stretching potential is a harmonic potential,

$$u^b(r_i j) = \frac{1}{2} k_{ij}^b \left(r_{ij} - b_{ij} \right)^2 \tag{74}$$

where k_{ij}^b is the harmonic force constant and b_{ij} is the equilibrium bond length. Many more potentials, for instance the Morse potential, the cubic potential or the FENE potential can be found in the GROMACS user manual in chapter 4.2 [82].

Angle-bending Potentials

The simplest angle-bending interaction is a harmonic potential and has the same form as equation 74:

$$u^a(\theta_{ijk}) = \frac{1}{2} k_{ijk}^\theta \left(\theta_{ijk} - \theta_{ijk}^0 \right)^2. \tag{75}$$

Again k_{ijk}^θ is a harmonic force constant and θ_{ijk}^0 is the undeflected angle. θ_{ijk} can be computed by

$$\theta_{ijk} = \arccos \left(\frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{r_{ij} r_{jk}} \right) \tag{76}$$

and is hence depending on the positions of three particles.

Alternative angle-bending potentials as the cosine based angle potential or the Urey-Bradley potential can be found in the GROMACS user manual in chapter 4.2 [82].

4.1.3   Restraints

Restraints are special potentials which are used to restrain the motion of the system. This could be necessary to include certain knowledge from experimental results in the system or to avoid disastrous deviations. Typically restraints are implemented by applying an energy penalty if a particle is deviating from a certain value. One has to realize that in GROMACS restraints and constraints are two different concepts. Constraints are needed to fix bond lengths and angles whereas restraints favor certain constellations.

Different restraint algorithms allow to implement restraints for bond lengths, bond angles, dihedrals or orientations of certain vectors. A list of all available restraint algorithms for GROMACS can be found in the user manual in chapter 4.2 [82].

4.2   Water Models

Since H_2O is one of the most important molecules in MD simulations there are dozens of different water models. Classifying all those models is nearly impossible, however, the three main categories to classify them are the number of interaction points (which are called sites), the flexible or rigid nature of the model and the inclusion of polarization effects. In this work the used water model is TIP4P/Ice which is discussed in detail in section 4.2.1.

4.2.1   TIP4P/Ice Potential
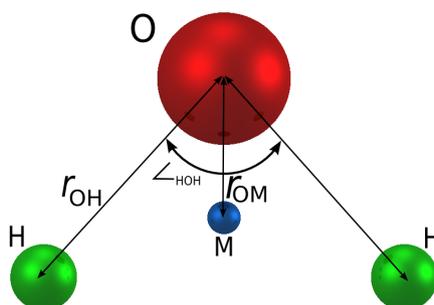
Figure 8: TIP4P model of H_2O molecule. There are four interaction points: one oxygen atom O (red), two hydrogen atoms H (green) and one dummy atom M (blue) which is located near the oxygen atom along the bisector of the HOH angle. Figure from Sklogwiki [83].

The TIP4P/Ice potential (**T**ransferable **I**ntermolecular **P**otential with **4 P**oints) was developed by *J. L. F. Abascal, E. Sanz, R. García Fernández* and *C. Vega* in 2005 to reproduce the coexistence lines of different ice forms [84]. The potential is rigid, planar and has four interaction points: one oxygen atom O, two hydrogen atoms H and one dummy atom M which is located near the oxygen atom along the bisector of the HOH angle (see figure 8). This dummy particle has the purpose of a charge site which improves the electrostatic distribution of the molecule. This setup is based on the Bernal-Fowler model published in 1933 [85]. The used potential is the Lennard-Jones potential combined with the Coulomb potential:

$$U = U^{LJ} + U^C = \sum_{\substack{i,j \\ i<j}}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \right) \tag{77}$$

using the parameters given in table 1.

Table 1: Parameters of TIP4P/Ice water model

	ϵ/k_B (K)	σ (Å)	q (e)	Geometry
O	10.61	3.1668	0.0	O-H: 0.9572 Å
H	0.0	0.0	0.5897	O-M: 0.1577 Å
M	0.0	0.0	-1.1794	H-O-H: 104.5°

The masses are carried by the oxygen and by the hydrogen atoms. The oxygen atom has a mass of 15.9994 u, the hydrogen atom carries a mass of 1.008 u. The phase diagram calculated with this model can be seen in figure 9. The blue stars are the experimental coexistence lines of different ice structures and liquid water. The black dotted line is the phase diagram computed with the TIP4P potential proposed by *Jorgensen et al.* in 1983 [86]. The red line shows the coexistence lines using the TIP4P/Ice potential. The shift of the stability regions of ice II, ice III, ice V and ice VI of approximately 15 K for TIP4P/Ice and approximately 50 K for TIP4P is conspicuous.
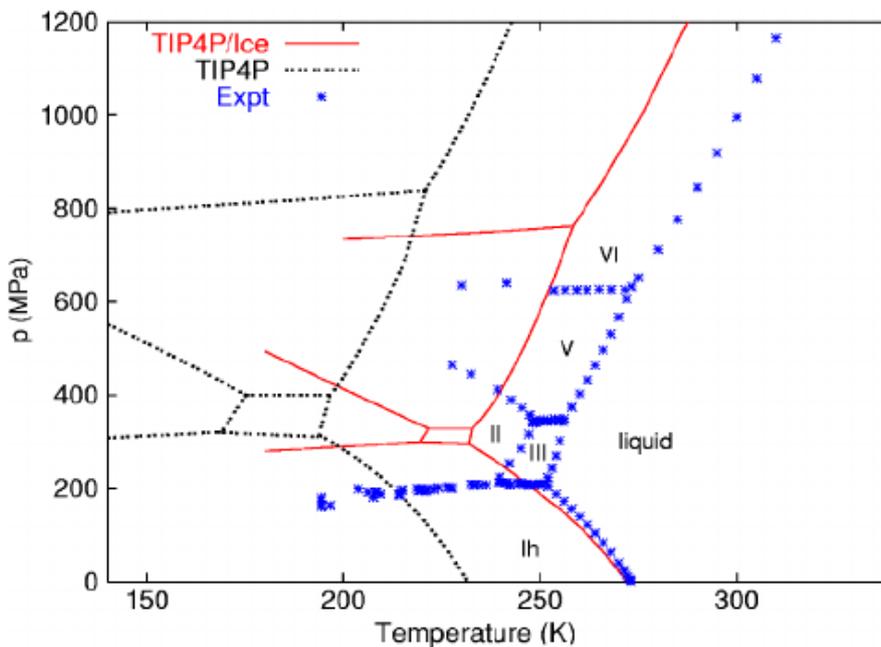
Figure 9: Phase diagram of water using TIP4P (black) and TIP4P/Ice (red) water in comparison to experimental results. Figure from *Abascal et al.*[84].

4.3 CO$_2$ Models

A number of models has been proposed to simulate properties of CO$_2$, but only, the MSM model [87, 88], the EMP2 model by *Harris* and *Yung*, 1995 [89], the exp-

6-variant of this model by *Potoff et al.*, 1999 [90] and the TraPPE model by *Potoff* and *Siepmann*, 2001 [91] are optimized for simulating phase equilibria. Since for the EMP2 model and for the exp-6-variant often special combinig rules are required the chosen potential for this work has been the TraPPE potential. This is done in agreement to other works in the field, e.g.Míguez *et al.* [92] or Sarupria and Debenedetti [93].

4.3.1 TraPPE Potential

The TraPPE potential (**Tra**nsferable **P**otentials for **P**hase **E**quilibria) was developed by *J. J. Potoff* and *J. I. Siepmann* in 2001 for describing the vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen [91]. The molecule is rigid, linear and has 5 sites. The carbon atom and two oxygen atoms are the LJ site and the Coulomb site, these three interaction points are along one line. Along the same line are two virtual particles, each moved outside of the molecule along the line, carrying half of the mass of the molecule (each virtual atom carries a mass of 22.0049 u).
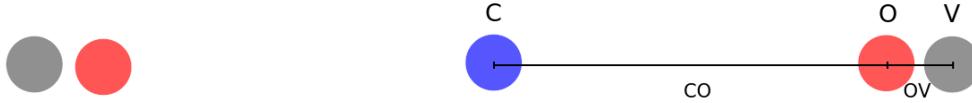
Figure 10: TraPPE model of CO$_2$ molecule: There are five interaction points: one carbon atom (blue), two oxygen atoms O (red) and two virtual sites (gray) which are located next two the oxygen atoms.

The used potential is a combination of LJ-potential and Coulomb potential and has the form

$$U = U^{LJ} + U^C = \sum_{\substack{i,j \\ i<j}}^{N} \left(4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \right). \tag{78}$$

The used parameters are listed in table 2.

Table 2: Parameters of TraPPE CO$_2$ model

	ϵ/k_B (K)	σ (Å)	q (e)	Mass (u)	Geometry
C	27-0	2.80	0.70	0.0	C-O: 1.086378 Å
O	79.0	3.05	0.35	0.0	O-V: 0.1978245 Å
V	0.0	0.0	0.0	22.0049	

In principle the TraPPE potential would also allow to introduce an angle depending harmonic term and a torsion term. Since in this work the TraPPE potential is used

for trapped, single CO_2-molecules the torsion term can be neglected obviously. The angle depending term is neglected for comparability reasons, because other works in the field (e.g.Míguez *et al.* [92] or Sarupria and Debenedetti [93]) neglect it and therefore it is easier to compare results.

5. TWO PHASE COEXISTENCE LINE OF CO$_2$ HYDRATES

In this section the prediction of the two phase coexistence line for both, *sI*-hydrate and χ-hydrate, is presented. There exist several ways for finding the melting temperature of a certain structure. In this work the chosen method was using a two phase system, one phase is the hydrate to observe, the other phase is a mixture of CO$_2$ molecules and H$_2$O molecules. For this systems *NPT* MD simulations along different isobars have been performed. For each pressure the MD simulations have been performed for different temperatures. The evolution of the system for a fixed pressure and a fixed temperature has been analyzed. Depending on whether the potential energy is increasing or decreasing the hydrate is growing or melting. Using bisection the melting temperature for each pressure was found by finding the temperature interval in which on one boundary of the interval the potential energy is increasing while on the other boundary the potential energy is decreasing.

5.1 Initial Configuration

For both hydrates, *sI* and χ, a two phase coexistence system has been prepared. One phase was the hydrate for which the melting temperature should be observed, the other phase was a mixture of CO$_2$ and H$_2$O molecules. Detailed information to the initial configurations is provided in section 5.1.1 and 5.1.2.

5.1.1 *sI*-Hydrate

For *sI* the unit cell was replicated 2 times in each direction. A occupancy rate of 100% was chosen, therefore the unit cell is built of 46 H$_2$O molecules and 8 CO$_2$ molecules, leading to a guest to host ration of 1 : 5.75. Hence the bulk phase contains 368 H$_2$O molecules and 64 CO$_2$ molecules and has a size of $2.324 \times 2.324 \times 2.324$ nm^3. The second phase containing a mixture of CO$_2$ molecules and H$_2$O molecules was built of 368 H$_2$O molecules and 64 CO$_2$ molecules. This phase has been relaxed separately from the hydrate for 500 ps in an *NVT* MD simulation with a fixed volume of $2.324 \times 2.324 \times 2.324$ nm^3 and a temperature of 293 K.

After the relaxation of the CO$_2$-H$_2$O-mixture it was merged along the z-dimension with the hydrate such that the interface of hydrate and mixture was in the x-y plane. The total number of water molecules was 736, the total number of CO$_2$ molecules was 128. The two phase system was relaxed for 500 ps in a *NPT* simulation at a temperature of 263 K and a pressure of 500 bar.

5.1.2 χ-Hydrate

The structure for the χ-hydrate is taken from *Amos et. al.* [1]. The unit cell in the structure of *Amos et. al.* contains 6 H_2O and 6 CO_2 molecules. Since the guest to host ratio is flexible as discussed in section 2.1.2 it was adapted to the ratio of the *sI*-hydrate in section 5.1.1. Therefore five CO_2 molecules have been removed which leads to a guest to host ration of 1:6. This unit cell was duplicated 4 times in each direction leading to 384 H_2O molecules and 64 CO_2 molecules and a box size of $2.51 \times 2.1737 \times 2.519$nm^3. Three additional CO_2 molecules are placed along the chiral channels of the χ-hydrate. This leads to a guest to host ratio of 1:5.73. This hydrate phase was relaxed for 500 ns in a *NVT* MD simulation with a temperature of 23 K and a pressure of 5000 bar.

The liquid phase containing a CO_2-H_2O-mixture was prepared analogously to the way described in section 5.1.1. 368 H_2O molecules and 64 CO_2 molecules are relaxed with an *NVT* MD simulation in a volume of $2.51 \times 2.1737 \times 2.519$nm^3 with a temperature of 293 K and a pressure of 5000 bar. Then the χ-hydrate and the liquid phase were merged along the z direction and relaxed in a *NPT* simulation at a temperature of 250 K and a pressure of 5000 bar.

5.2   Calculation of Liquid-Solid Coexistence Line

For calculating the two phase coexistence line of *sI*-hydrate the initial configurations discussed in section 5.1.1 were used. With these configurations *NPT* MD simulations were performed along different isobars of 10, 20, 40, 100, 200, 400, 500, 600, 700, 800, 900, 1000, 1500 MPa. For each of these pressures two temperatures were determined, one low temperature T_1 for which the hydrate is shrinking and one high temperature T_2 for which the hydrate phase is melting during the time evolution of the system. Whether the hydrate was growing or shrinking could also be determined in the potential energy of the system. If the potential energy was increasing the hydrate was melting, if the potential energy was decreasing the hydrate was growing.

After the determination of the temperature for which the potential energy was increasing T_2 and one for which the potential energy was decreasing T_1 the intermediate temperature $T_3 = \frac{T_1 + T_2}{2}$ was calculated and for this intermediate temperature a *NPT* MD simulation at the same pressure was performed. For this MD simulation again the development of the potential energy was analyzed. If the potential energy was increasing, the intermediate between T_1 and T_3 was chosen as the temperature for the next MD simulation, if the potential energy was increasing the intermediate between T_2 and T_3 was chosen. This bisection was performed until

for every pressure a final equilibrium state with an uncertainty of $\pm 1\,$K was identified (see figure 11a). The growing of the *sI*-hydrate phase can be seen in figure 12.
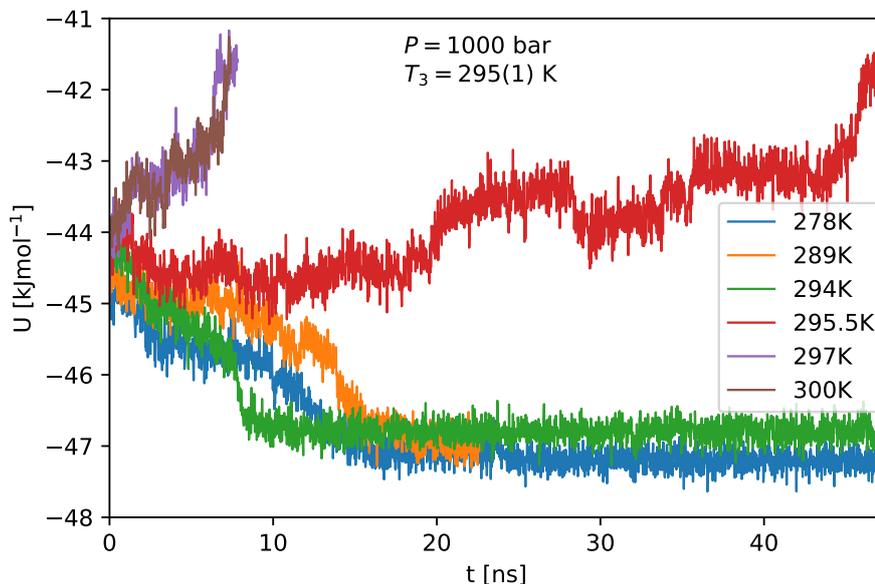
(a) Potential energy of *sI*-phase of CO_2 hydrates
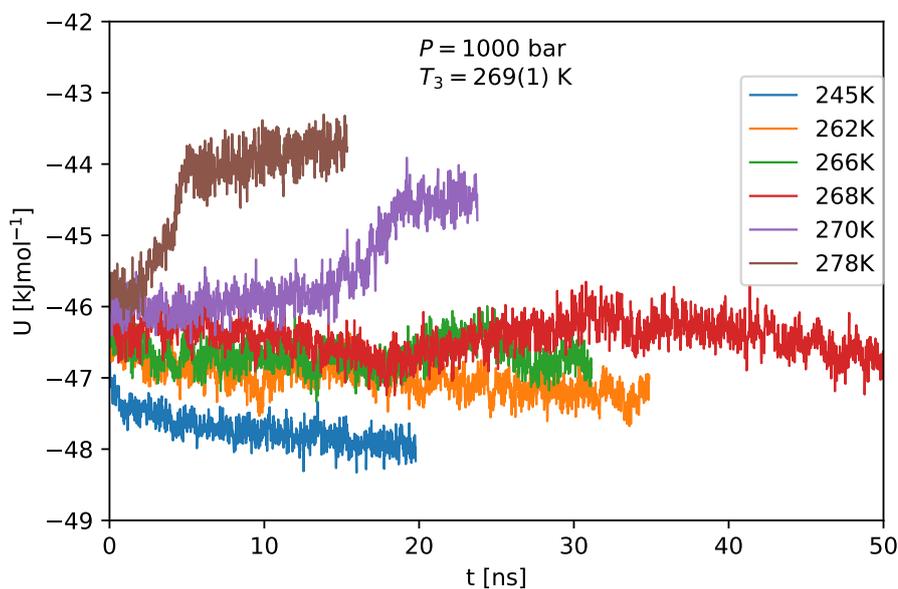
(b) Potential energy of χ-phase of CO_2-hydrates

Figure 11: Evolution of the potential energy as a function of time of the initial two phase configuration. *NPT* MD simulation of: (a) *sI*-hydrate/CO_2-H_2O-mixture at $1000\,$bar and temperatures from $278\,$K to $300\,$K (b) χ-hydrate/CO_2-H_2O-mixture at $1000\,$bar and temperatures from $245\,$K to $278\,$K
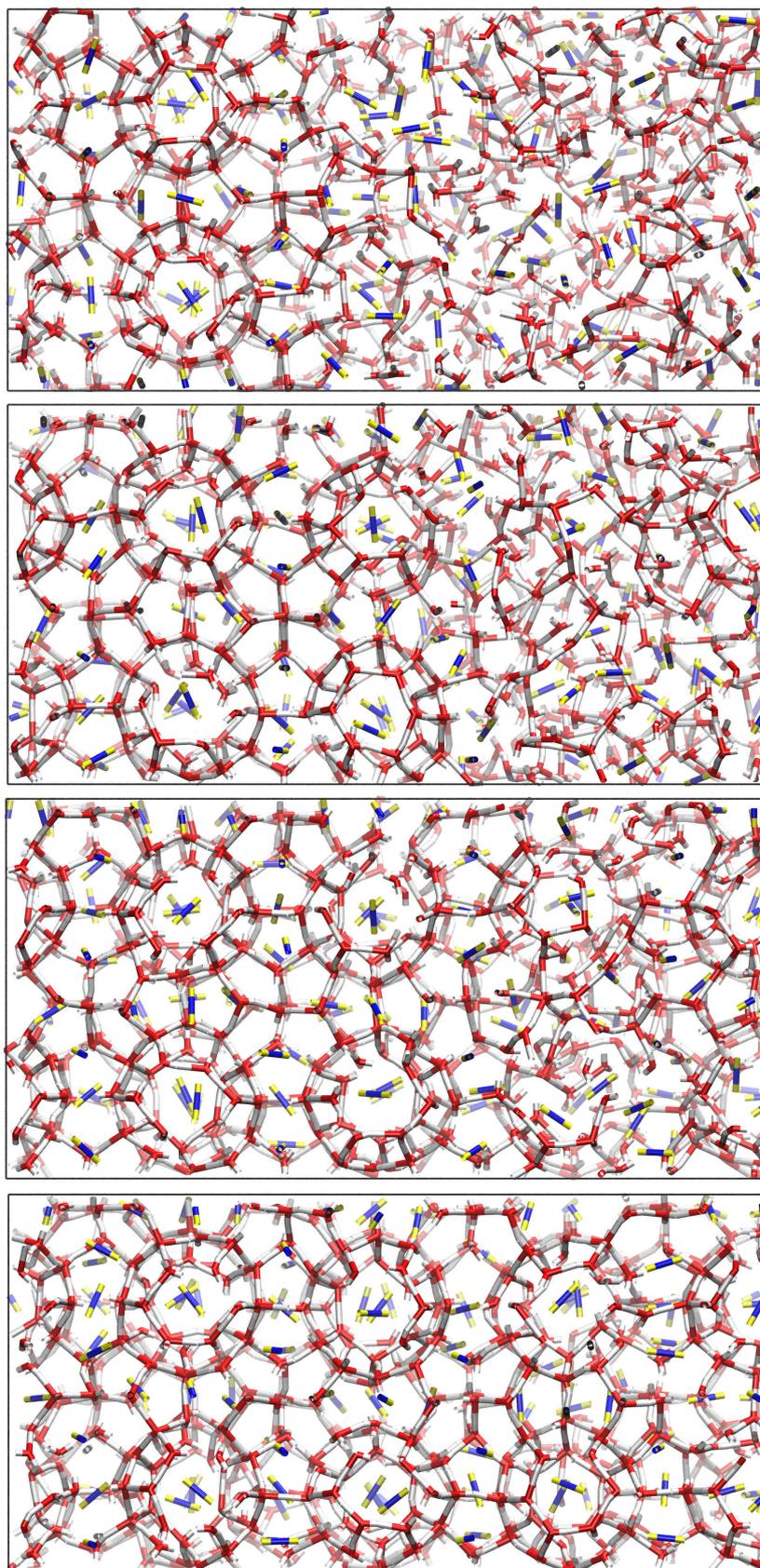
Figure 12: Snapshots of layer by layer *sI*-hydrate growth created with VMD. Beginning at a separated two phase system the hydrate grows in both sides, enabled by using PBC.
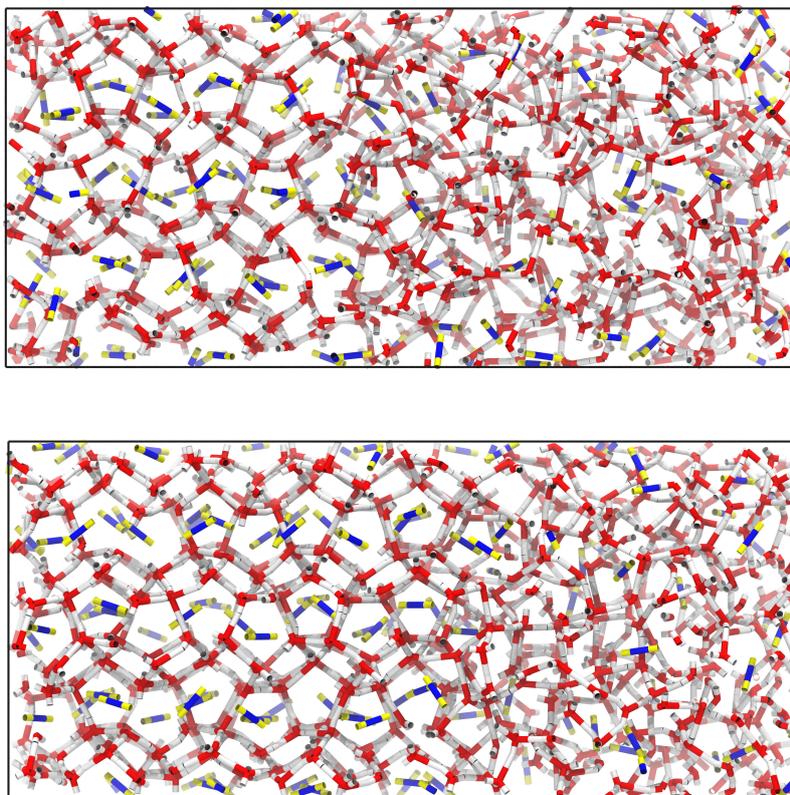
Figure 13: Snapshots of layer by layer χ-hydrate growth created with VMD. Beginning at a separated two phase system the hydrate grows in both sides, enabled by using PBC.

The bisection method was very useful for finding the coexistence line *sI*-hydrate for pressures up to 5000 bar, however for other configurations or higher pressures it reached its limits very fast. For the *sI*-hydrate with pressures higher than 5000 bar the hydrate structure stopped growing, it was either able to stay equal size or to melt. The reason was, that if the temperature was too low, in the initially liquid CO_2-H_2O-mixture the molecules were moving very little or were building an amorphous structure which prevented the hydrate from growing. Since in this process the identification whether the hydrate was growing or keeping equal size was very hard by using the potential energy, instead the density profile was calculated to identify the behavior of the hydrate.

A similar problem occurred for the χ-hydrate system. For low pressures (pressures below 5000 bar) the growing process of the hydrate was very slow. However, at these pressures also the melting process was very quick and so it was possible to use the bisection method between an increasing potential energy and a potential energy which is constant or slightly decreasing. For higher pressures it was again necessary

to use the density profile of the two phase system to identify the behavior of the χ-hydrate.

5.2.1 Density Profile

Whether the hydrate structures were growing or shrinking was also possible to observe in the corresponding density profiles along the z-axis. The calculation of the density profile along the z-axis is straight forward. First the simulation box is subdivided into equally sized slabs with a thickness Δz. The slabs are indexed by an integer b. Furthermore considering PBC the slabs are extended in the x-y-plane. Each slab contains the region $b\Delta z \leq z < (b+1)\Delta z$. This allows defining an indicator function δ_b for the slab b:

$$\delta_b(z) = \begin{cases} 1 & \text{if } b \leq z/\Delta z < b+1 \\ 0 & \text{otherwise} \end{cases} \tag{79}$$

Using the indicator function the density profile can be calculated by

$$\rho_b = (L_x L_y \Delta z)^{-1} = \sum_{i \in \text{ atoms}} \delta_b(z_i) p_i \tag{80}$$

where L_x and L_y are the sides of the periodic cell, z_i is the z-coordinate of the atom i and p is the property for which the density profile should be calculated (for instance the p_i is the mass of atom i if the mass profile is calculated). This algorithms is implemented in the VMD package *density profile tool* by *Toni Giorgino* [94].

The density profiles in figure 14 show the number density profile and are using slab size of $\Delta z = 0.1$ Å. For the number density profile the property p is chosen as $p_i = 1$. The density profile was both calculated for CO$_2$ and H$_2$O separately, the *sI*-hydrate and the χ-hydrate. The initial density functions show the average density profiles of the first 50 frames of the trajectories, the final density profiles show the average density profiles of the last 400 frames.

The density profiles of the *sI*-hydrate system and the χ-hydrate system (figure 14) confirm that the *sI*-hydrate system was able to grow fast (compare the initial state of the *sI*-hydrate/CO$_2$-H$_2$O-mixture in figure 14a with the final state in figure 14b) whereas the growth of the χ-hydrate system was very slow (again compare the initial state of the χ-hydrate/CO$_2$-H$_2$O-mixture in figure 14c with the final state in figure 14d). Due to the slow process it is not possible to reach a pure χ-hydrate system in a reasonable computing time, though it is possible to see that the periodic part in figure 14d is larger than in figure 14c.

(a) Initial state *sI*-hydrate/CO$_2$-H$_2$O-mixture

(b) Final state *sI*-hydrate/CO$_2$-H$_2$O-mixture

(c) Initial state χ-hydrate/CO$_2$-H$_2$O-mixture

(d) Final state χ-hydrate/CO$_2$-H$_2$O-mixture
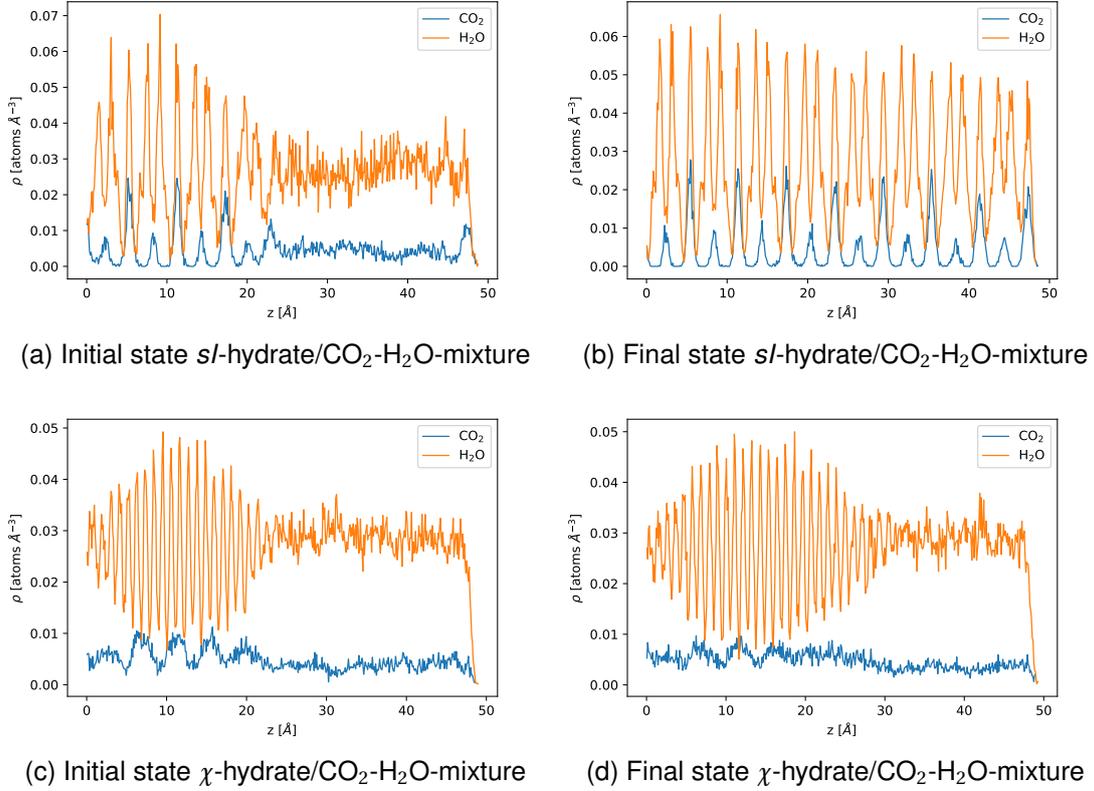
Figure 14: Snapshots of the density profiles along z-axis of (a) initial configuration of *sI*-hydrate/CO$_2$-H$_2$O-mixture at $1000\,\mathrm{bar}$ and $278\,\mathrm{K}$ (b) final state of *sI*-hydrate/CO$_2$-H$_2$O-mixture at $1000\,\mathrm{bar}$ and $278\,\mathrm{K}$ (c) initial configuration of χ-hydrate/CO$_2$-H$_2$O-mixture at $1000\,\mathrm{bar}$ and $262\,\mathrm{K}$ (d) final state of χ-hydrate/CO$_2$-H$_2$O-mixture at $1000\,\mathrm{bar}$ and $262\,\mathrm{K}$

.

5.3   Solid-Liquid Coexistence Line

Using the bisection method described in section 5.2 for each of the pressures 10, 20, 40, 100, 200, 400, 500, 600, 700, 800, 900, 1000, 1500 MPa a corresponding melting temperature was calculated for both, the *sI*-hydrate system and the χ-hydrate system. For the χ-phase additionally the pressures 1100, 1200, 1300, 1400 MPa have been computed. The results are summarized in figure 15. In this figure the orange data points represent the two-phase coexistence of *sI*-hydrate and H$_2$O-CO$_2$-mixture, the green data points represent the two-phase coexistence of χ-hydrate and H$_2$O-CO$_2$-mixture. For pressures up to 900 MPa the error of the method is 2 K. For higher pressures the error is 5 K for the *sI*-hydrate and 3 K for the χ-hydrate. All calculated coexistence temperatures T_3 are listed in table 3. The blue data points are the computed phase diagram of water using TIP4P/ice by Abascal and Vega [84].

The errors of the coexistence temperatures T_3 at high pressures are bigger than at low pressures. This is, because at high pressures the formation of hydrates is very slow
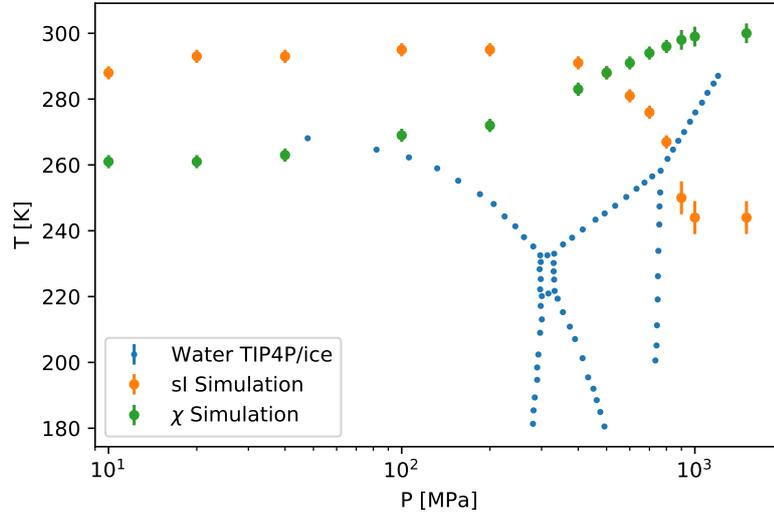
Figure 15: PT projection of the two-phase coexistence line of *sI*-hydrate with H$_2$O-CO$_2$-mixture and χ-hydrate with H$_2$O-CO$_2$-mixture. The orange and green data points are calculated in the way discussed in section 5.2. The blue data points are from the computed phase diagram of water using TIP4P/ice by Abascal and Vega [84].

Table 3: Two-phase coexistence temperatures (T_3) at different pressures obtained using bisection method. The estimated error in T_3 is shown between parentheses.

P [MPa]	T_3^{sI} [K]	T_3^{χ} [K]
10	288(2)	261(2)
20	293(2)	261(2)
40	293(2)	263(2)
100	295(2)	269(2)
200	295(2)	272(2)
400	291(2)	283(2)
500	288(2)	288(2)
600	281(2)	291(2)
700	276(2)	294(2)
800	267(2)	296(2)
900	250(5)	298(3)
1000	244(5)	300(3)
1100	-	303(3)
1200	-	305(3)
1300	-	306(3)
1400	-	305(3)
1500	244(5)	300(3)

or stops completely since the CO$_2$-H$_2$O-mixture becomes less fluid (or is forming an amorphous CO$_2$ filled ice if the temperature is low enough). Therefore close to the two-phase coexistence line the time which is necessary to determine T_3 is so long that it is not within current computing possibilities.

The simulated curve of *sI*-hydrate is in good agreement with the experimental values of *Massani, Mitterdorfer* and *Loerting* [19]. A shift of the coexistence line of approximately $15\,\mathrm{K}$ which occurred in a previous work of *Míguez et al.* [92] does not occur in the simulation method chosen in this work. The reason could be that *Míguez et al.* have chosen to compute the three-phase coexistence line whereas in this work the two-phase coexistence line has been computed. In a three-phase system (*sI*-hydrate-H_2O-CO_2) the CO_2 molecules have to diffuse through the slab of liquid H_2O to reach the *sI*-hydrate surface whereas in a two-phase system the growing process of the hydrate can start immediately.

Nakano et al. showed that the coexistence curve has a maximum temperature of $294.0\,\mathrm{K}$ at a pressure of $328\,\mathrm{MPa}$ [29]. A shifted maximum also occurs in the work of *Míguez et al.* around $90\,\mathrm{MPa}$. This clear maximum can not be seen in this work, since for both, $100\,\mathrm{MPa}$ and $200\,\mathrm{MPa}$ a maximum temperature of $295\,\mathrm{K}$ occurs. Though it is likely that a maximum temperature occurs between $100\,\mathrm{MPa}$ and $200\,\mathrm{MPa}$. Furthermore *Nakano et al.* showed that for pressures lower than the pressure where the maximum temperature is reached the slope $\mathrm{d}p/\mathrm{d}T > 0$ whereas in higher pressure regions $\mathrm{d}p/\mathrm{d}T < 0$ holds. Excluding the point at a pressure of $100\,\mathrm{MPa}$ this relations also hold in this work.

6. SOLID-SOLID COEXISTENCE

In chapter 5 the coexistence line of *sI*-hydrate and a CO_2-H_2O-mixture and the coexistence line of χ-hydrate and a CO_2-H_2O-mixture were computed by using a two phase coexistence. The calculation of the coexistence line of *sI*-hydrate and χ-hydrate can not be done analogously for several reasons. First of all it is not possible for geometric reasons. The unit cell of *sI*-hydrate and χ-hydrate have different dimensions, therefore it is not possible to unify them to a single simulation box. This is even more complicated by the fact that the unit cell of *sI* is a cuboid whereas the unit cell of χ is triclinic. Basically simulations could work by using a thin liquid phase between the two solid phases. However, this simulation would be computationally expensive and therefore alternative concepts are favored.

6.1 Coexistence of *sI*-Hydrate and χ-Hydrate

6.1.1 Theory

To get around the problem of solid-solid interface the coexistence line was determined indirectly by using the state function of the isothermal-isobaric ensemble:

$$\beta G = -\ln(\Delta), \tag{81}$$

where $\beta = \frac{1}{k_B T}$ and Δ_{NpT} is the partition function:

$$\Delta_{NpT} = \frac{\beta p}{N!(2\pi\hbar)^{3N}} \int dV \int d\mathbf{r^N} d\mathbf{p^N} e^{-\beta[H(\mathbf{r^N},\mathbf{p^N})+pV]}. \tag{82}$$

Inserting equation 82 in equation 81 and taking the derivative in direction β one gets:

$$\frac{\partial \beta G}{\partial \beta} = -k_B T - \langle U \rangle_{NpT} - \langle pV \rangle_{NpT}. \tag{83}$$

This can be integrated in the range β_1 to β_2:

$$\beta_2 G_2 - \beta_1 G_1 = \int_{\beta_1}^{\beta_2} \frac{\partial \beta G}{\partial \beta} d\beta =$$

$$= -\int_{\beta_1}^{\beta_2} \left(\langle U \rangle_{NpT} + \langle pV \rangle_{NpT} + k_B T \right) d\beta = -\int_{T_1}^{T_2} \frac{\langle U \rangle_{NpT} + \langle pV \rangle_{NpT} + k_B T}{k_B T^2} dT \tag{84}$$

Considering figure 16 the following equation holds:

$$
\begin{aligned}
0 &= \Delta(\beta G)_{12} + \Delta(\beta G)_{13} - \Delta(\beta G)_{23} \\
&= (\beta G)_{\chi,mix} - (\beta G)_{sI,mix} + (\beta G)_{sI,mix} \\
&\quad - (\beta G)_{sI,\chi} - (\beta G)_{\chi,mix} + (\beta G)_{sI,\chi}.
\end{aligned}
\tag{85}
$$

Here each βG is marked with two indices and each index is representing a phase. The integration of equation 84 is done for a certain system (the sI-phase, the χ-phase or the CO_2-H_2O-mixture) and the limits of the integral are always certain melting temperatures T_1, T_2, T_3. The two indices represent the value of βG evaluated at the melting temperature of the binary system of the two indices. The sum of the integrals 84 of all three systems (the sI-phase, the χ-phase or the CO_2-H_2O-mixture) has to be zero (compare figure 16). The temperatures $T_{sI,mix}$ and $T_{\chi,mix}$ can be computed as discussed in chapter 5. The results can be plugged into equation 85. This results in an equation where only $T_{sI,\chi}$ is unknown and which can be solved for $T_{sI,\chi}$. With this procedure it is possible to find for every fixed pressure the coexistence temperature $T_{sI,\chi}$.
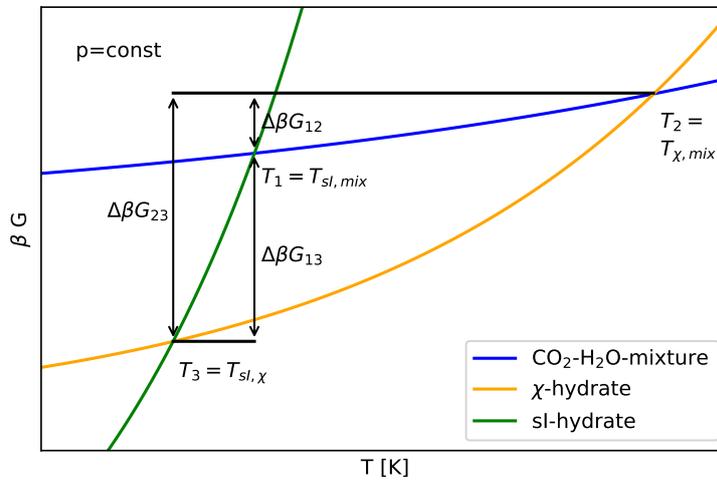
Figure 16: The blue line represents the product of the Gibbs energy and β for a constant pressure p. The green line represents the same product for sI-hydrate, the orange line for χ-hydrate. $T_{sI,mix}$ is the coexistence temperature of sI-hydrate and CO_2-H_2O-mixture at pressure p, $T_{\chi,mix}$ is the coexistence temperature with the χ-hydrate. $T_{sI,\chi}$ is the coexistence temperature of the two hydrates, which should be calculated

The basic structure of the algorithm is summarized in 2.

Algorithm 2 solid-solid-coexistence

1: **for** p in pressures **do**
2:    call coex_sI_mix() $\triangleright$ compute $T_{sI,mix}$
3:    call coex_chi_mix() $\triangleright$ compute $T_{\chi,mix}$
4:    **for** T in temperatures **do**
5:        call E() $\triangleright$ compute $U(T) + pV(T) + k_B T$
6:    **end for**
7:    integrate() $\triangleright$ Integrate equation 84 three times with different limits
8:    solve $\triangleright$ Solve equation 85 for $T_{sI,\chi}$
9: **end for**

6.1.2 Computation of Internal Energy

For finding the coexistence temperature of *sI*-hydrate and χ-hydrate using the way described in section 6.1.1 it is necessary to know the internal energy of *sI*-hydrate, χ-hydrate and the CO_2-H_2O-mixture as a function of pressure and temperature. Therefore each of the three systems was relaxed at each of the pressures 10, 20, 40, 100, 200, 400, 500, 600, 700, 800, 900, 1000, 1200, 1300, 1400, 1500 MPa for 500 ps with constant volume at a temperature of 100 K.

For the *sI*-hydrate a the system had a size of $2 \times 2 \times 2$ cages with an occupancy rate of 100% and a total of 368 H_2O molecules and 64 CO_2 molecules. The χ-hydrate had a size of $4 \times 4 \times 4$ cages, each consisting of 6 H_2O molecules and one CO_2 molecule. Three additional CO_2 molecules were added in the channels of the χ-hydrate such that the ration between CO_2 and H_2O molecules is approximately the same in the *sI*-hydrate and in the χ-hydrate ($1:5.75$ in the *sI*-hydrate and $1:5.73$ in the χ-hydrate). The CO_2-H_2O-mixture was built of 368 H_2O-molecules and 64 CO_2-molecules.

For each of the three systems at each pressure for each temperature between 180 K and 300 K with a step size of 10 K a 2 ns *NPT*-simulation was carried out using a step size of 2 fs. From this simulation the average internal energy for each temperature and pressure combination was calculated by using the internal energy and the volume of every 100th step starting from step 500.000. Since the used systems have different numbers of particles it is furthermore necessary to normalize the internal energies and the volumes by dividing the average internal energy and the volumes by the number of particles.

For each pressure the temperatures had to be increased step by step and the simulations had to be done one after each other, with that, the heating process was very continuously and the relaxation times were short. The result for a pressure of 900 MPa can be seen in figure 17.
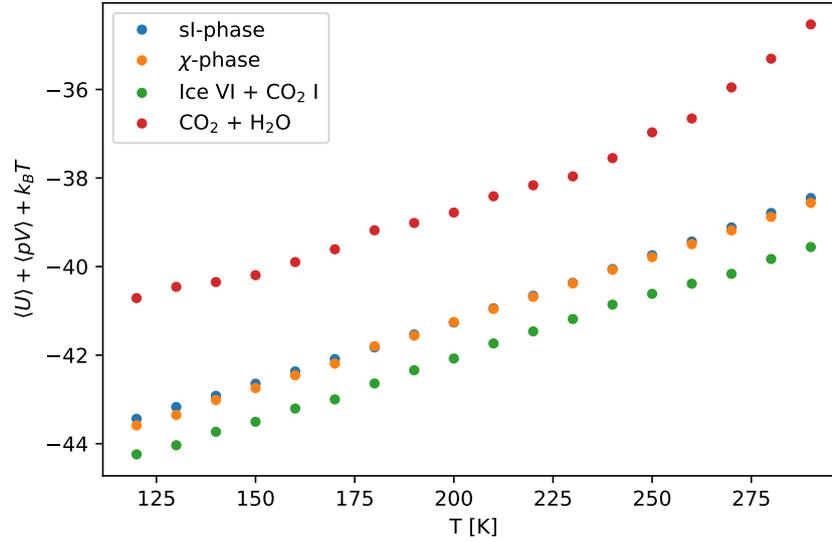
Figure 17: $\langle U \rangle + \langle pV \rangle + k_B T$ as a function of temperature T for a pressure of $900\,\mathrm{MPa}$. The red dots represent the energy of the liquid CO_2-H_2O-mixture, the blue dots represent the *sI*-phase and the orange dots the χ-phase. The green dots show the energy of the separated ice VI + CO_2 I system.

6.1.3   Computation of Coexistence Temperature

After the computation of the average internal energy $\langle U \rangle_{NpT}$ depending on pressure and temperature for *sI*-hydrate, χ-hydrate and the CO_2-H_2O-mixture it was necessary to integrate $-\frac{\langle U \rangle_{NpT} + \langle pV \rangle_{NpT} + k_B T}{k_B T^2}$ as it is shown in equation 84. The integration had to be done three times. The first time it had to be done for the average internal energy of the CO_2-H_2O-mixture in the range $T_{sI,mix}$ to $T_{\chi,mix}$. Secondly it had to be done for the χ-hydrate system using the limits $T_{\chi,mix}$ and χ,sI and thirdly it had to be done for the *sI*-hydrate system in inverse direction using the limits $T_{sI,mix}$ and $T_{sI,\chi}$. Integrating the *sI*-hydrate system in inverse direction was necessary to use $T_{sI,mix}$ as a reference point for the integration. The upper limit $T_{sI,\chi}$ of the last two integrations is still unknown.

By using the three integrals and equation 85 it was then possible to find the unknown coexistence temperature $T_{sI,\chi}$. Therefore all temperatures in the range from $120\,\mathrm{K}$ to $300\,\mathrm{K}$ with a step size of $1\,\mathrm{K}$ were plugged into the upper boundary of the second and the third integral. Then the results of the integrations could be used for equation 85. If a temperature had solved the equation, this temperature would have been the searched coexistence temperature $T_{sI,\chi}$.

Figure 18 shows the result of the procedure for a pressure of $500\,\mathrm{MPa}$ and a pressure of $600\,\mathrm{MPa}$. At the intersection of the orange line which represents the *sI*-hydrate and the green line which represents the χ-hydrate equation 85 is fulfilled and therefore the temperature of the intersection point is the searched coexistence
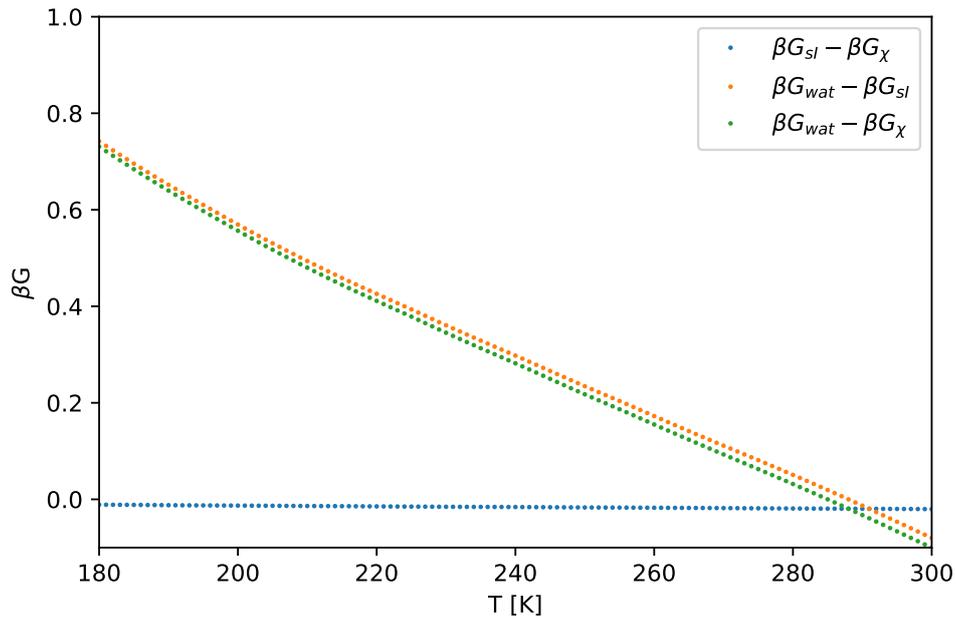
(a) Gibbs free energy for a pressure of $500\,\mathrm{MPa}$
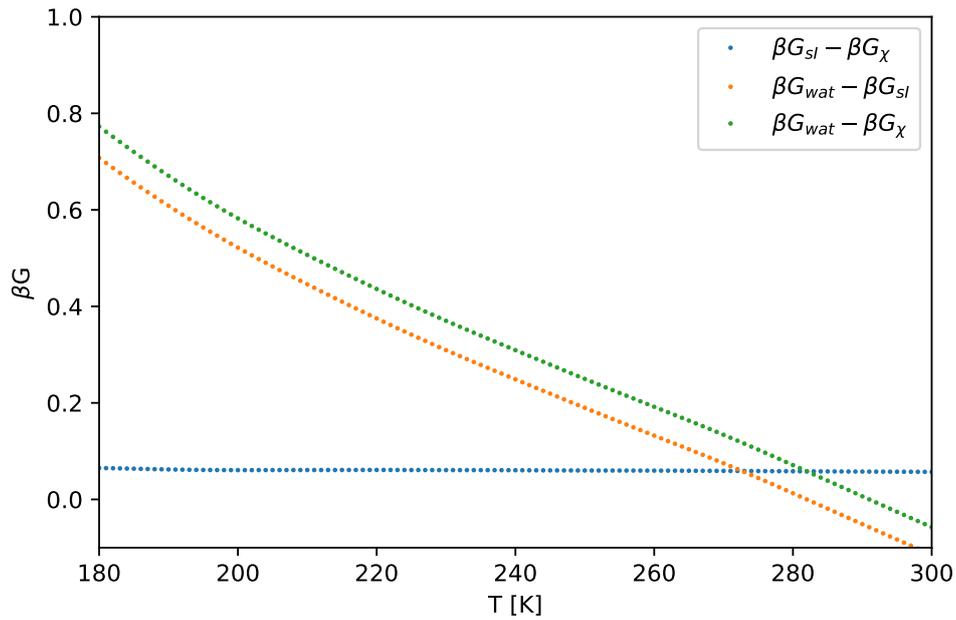
(b) Gibbs free energy for a pressure of $600\,\mathrm{MPa}$

Figure 18: Each line represents the result of the integration in equation 84. The blue line represents the integration of the average internal energy of the CO_2-H_2O-mixture, the orange line of the *sI*-hydrate and the green line of the χ-hydrate. If the green and the orange line would intersect, the temperature at the intersection point would be the coexistence temperature. Since the orange line is above the green line for a pressure of $500\,\mathrm{MPa}$ and below the green line for a pressure of $600\,\mathrm{MPa}$ the pressures at which the two lines intersect are between $500\,\mathrm{MPa}$ and $600\,\mathrm{MPa}$.

temperature $T_{sI,\chi}$. Neither in figure 18a nor in figure 18b there is an intersection of the orange line and the green line. In figure 18a it would be possible, that the intersection point is below 180 K. In figure 18b the orange line and the green line are very parallel, therefore there it is not possible to suspect any intersection point. At low pressures (≤ 500 MPa) the line of the *sI*-hydrate is above the line of the χ-hydrate, at high pressures (≥ 600 MPa) it is reversed. This has the consequence, that only in the region where the line of the *sI*-hydrate and the line of the χ-hydrate are nearly on top of each other a coexistence temperature below the melting temperatures of the hydrates occurs.

The region in which the line of the *sI*-hydrate and the χ-hydrate are on top of each other is between 500 MPa and 600 MPa. In this small region the line of the *sI*-hydrate and the χ-hydrate intersect at temperatures below the melting temperatures of the hydrates. This is also the region in which the melting curves of *sI*-hydrate and χ-hydrate are intersecting).

At pressures below the intersection point the *sI*-hydrate is more stable than the χ-hydrate since the melting curve of *sI*-hydrate is above the melting curve of χ-hydrate. At pressure above approximately 600 MPa the χ-hydrate is more stable than the *sI*-hydrate.

6.2   Coexistence of χ-Hydrate and Ice VI + CO$_2$ I

Experiments suggest, that the χ-hydrate occurs only in a region between 300 MPa and 1000 MPa [19, 20]. At lower pressures the *sI*-hydrate is more stable, which was also verified in section 6.1. At higher pressures solid CO_2 and ice VI are separated and more stable than hydrates [19, 20]. To verify this the whole procedure described in section 6.1 can be done again with solid CO_2 and ice VI instead of *sI*-hydrate.

Again first of all the computation of the internal energy is necessary. This has to be done separately for solid CO_2 and for ice VI. The work of *Datchi et al.* shows, that up to a pressure of 10 GPa the most stable solid CO_2 phase is Phase I [95]. The CO_2 I system was built of $6 \times 6 \times 6$ unit cells with a total of 864 CO_2 molecules. The ice VI system was built of $4 \times 4 \times 4$ unit cells with a total of 640 H_2O molecules. The simulation parameters are the same as used in section 6.1.2. After the computation of the internal energy for all pressures and temperatures a normalization to the average internal energy per particle was necessary, considering that the ratio between CO_2 and H_2O molecules has to be $1 : 5.75$.

After the computation of the average internal energies the integration step was necessary. As limits $T_{\chi,mix}$, $T_{mix,ice\ VI}$ are used, $T_{\chi,ice\ VI}$ should be calculated. The reason for choosing the melting temperature of ice VI and not the melting

temperature of CO_2 1 is, that the coexistence line of CO_2 1 is above the line of ice VI (for pressures higher than 200 MPa), which implicates that ice VI is melting first.

In contrast to the previous calculation of the coexistence line of the *sI*-phase and the χ-phase it was now possible to compute coexistence points and not only the limits in which the coexistence points must occur. It was possible to find six coexistence points in pressures between 700 MPa and 1200 MPa. Figure 19 shows the intersection point for a pressure of 1000 MPa. The coexistence pressures and temperatures can be found in table 4.
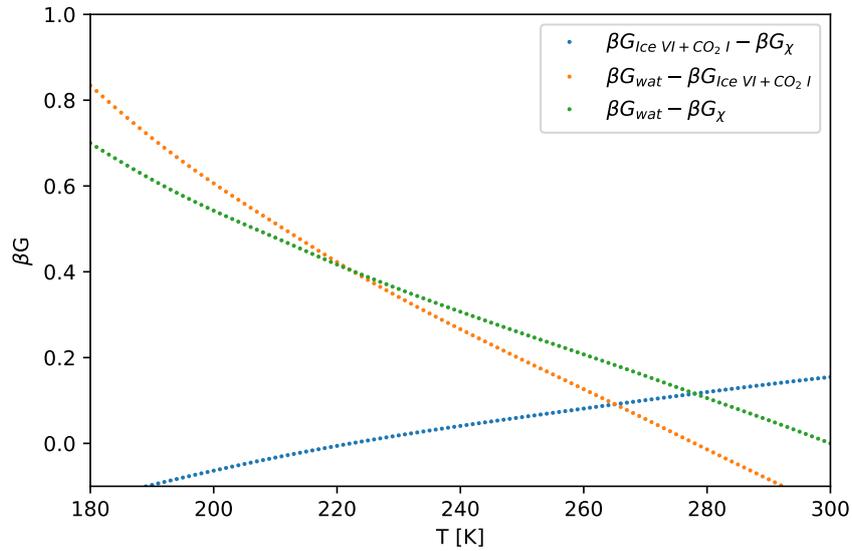
Figure 19: Each line represents the result of the integration in equation 84 for a pressure of 1000 MPa. The blue line represents the difference of the integration of the average internal energy of ice VI + CO_2 I and the χ-phase, the orange line represents the difference of ice VI + CO_2 I and the CO_2-H_2O-mixture and the green line represents the difference of the χ-phase and the CO_2-H_2O-mixture. The intersection temperature of the green line and the orange line is the coexistence temperature of the χ-phase and ice VI + CO_2 I.

Table 4: Six coexistence points between the χ-phase and ice VI + CO_2 I have been found using thermodynamic integration.

P [MPa]	T [K]
700	123(2)
800	165(2)
900	200(2)
1000	222(2)
1100	249(2)
1200	280(2)

Finding melting points at higher pressures (in the used resolution the next higher pressure would have been 1300 MPa was not possible, because at this pressure the melting point would have been above the melting line of ice VI and of the χ-phase.

In principle the thermodynamic integration method could also work for systems above the melting temperature. Here some adjustments of the MD simulation settings would be necessary. The pressure coupling is done with a Parrinello-Rahman barostate, therefore above the melting temperature of the χ-phase or ice VI it would be necessary to use isotropic coupling. The temperature for which isotropic coupling is necessary has to be chosen manually and this has not be done in this work.

6.3   Overview of Coexistence Lines

A summary of the calculated coexistence regions can be seen in figure 20. The red area between 500 MPa and 600 MPa marks the region in which the coexistence line of the *sI*-phase and the χ-phase is located. At lower pressures the most stable hydrate phase is the *sI*-phase. The green points are showing the coexistence line of the χ-phase and ice VI + CO_2 I. Between the coexistence line of the *sI*-phase and the χ-phase (red) and the coexistence line of the χ-phase and ice VI + CO_2 I the most stable phase is the χ-phase. At higher pressures ice VI + CO_2 I is the most stable phase.
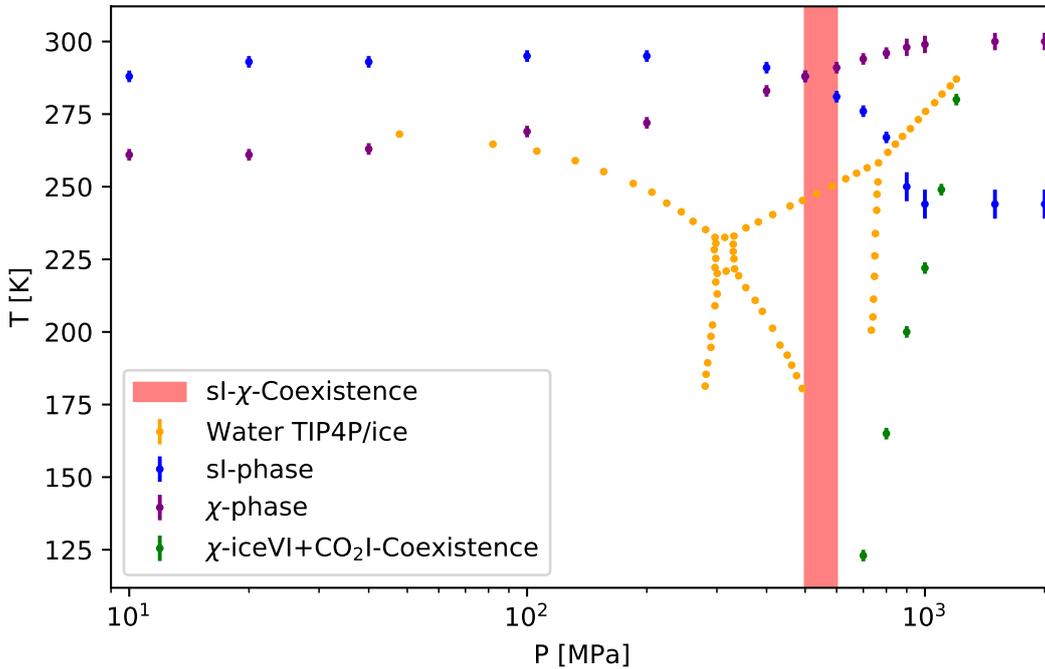
Figure 20: Computed coexistence regions of *sI*-phase, χ-phase and ice VI + CO_2 I. The red region show the limits in which the coexistence pressure of the *sI*-phase and the χ-phase of CO_2 hydrate must be, the green markers shows the coexistence line of the χ-phase and ice VI and CO_2 I. The blue dots show the computed melting curve of the *sI*-phase, the purple dots show the computed curve of the χ-phase.

6.4   Experimental Comparison

This result is also confirmed by different experiments. Massani, Mitterdorfer and Loerting showed the existence of a phase which they called FI (filled ice) in a region between 300 MPa and 1000 GPa[19]. In lower pressure regions they showed the existence of *sI*-hydrate, in higher pressure regions they found experimentally solid CO_2 I and ice VI separated (see figure 21). Massani *et al.* gave no description of the structure which they called FI, but if one assumes that the filled ice structure is the same structure which is in this work called χ-hydrate, then basically the results of Massani *et al.* match with the results calculated in section 6.1 and section 6.2. There are slight differences in the coexistence pressures, in this work the coexistence pressure of *sI*-hydrate and χhydrate is between 500 MPa and 600 MPa. In the work of Massani *et al.* it is about 200 MPa lower. The coexistence line between χ-hydrate and ice VI and CO_2 I separated runs in this work from a pressure of 700 MPa with a assigned temperature of 123 K to a pressure of 1200 MPa with a temperature of 280 K. The curvature of this coexistence line differs from the curvature of the calculations from Massani *et al.* However, despite from the curvature of the coexistence line the region in which the coexistence occurs is very similar and the assumption that the χ-phase is the same as FI in the work of Massani *et al.* seems to be very reasonable.

The gray data points in figure 21 represent experiments of Hirai *et al.* [30] which show a similar region of a hydrate structure as it is shown by Massani *et al.* The gray squares are selected data points of Bollengier *et al.* which represent a phase called high pressure hydrate (HP hydrate). Here it is again possible to suspect that this HP hydrate, for which no structural description is provided, is the same as the χ-hydrate.

In figure 21 there are two regions highlighted which are called "CO_2(s) + LDA" and "CO_2(s) + HDA". This represents CO_2 I and low density amorphous ice and CO_2 and high density amorphous ice respectively. These structures occur at low temperatures (below 150 K respectively 200 K) and a structural description is not provided. In this thesis these structures have been neglected, though in principle it should be possible to compute a solid-solid coexistence exactly the same way as it has been done before, provided that a description of the amorphous structure exists.

Figure 22 is from the work of Bollengier *et al.* In several experiments they showed the stability region of *sI*-hydrate (green diamonds) a phase they called CO_2 high pressure (HP) hydrate (blue diamonds) and the stability region of ice VI and CO_2 I separated (red squares). Additionally they measured the melting curve of ice VI (black pluses) and the melting of ice VI saturated with CO_2 (red crosses). The results are similar to the results of Massani *et al.* and the results of this work. At pressures below 700 MPa they found *sI*-hydrate and at pressures between 700 MPa and 1 GPa they showed the existence of the CO_2 HP hydrate. Since they only examined temperatures above 240 K, these results match with the results of Massani *et al.* if again the HP hydrate
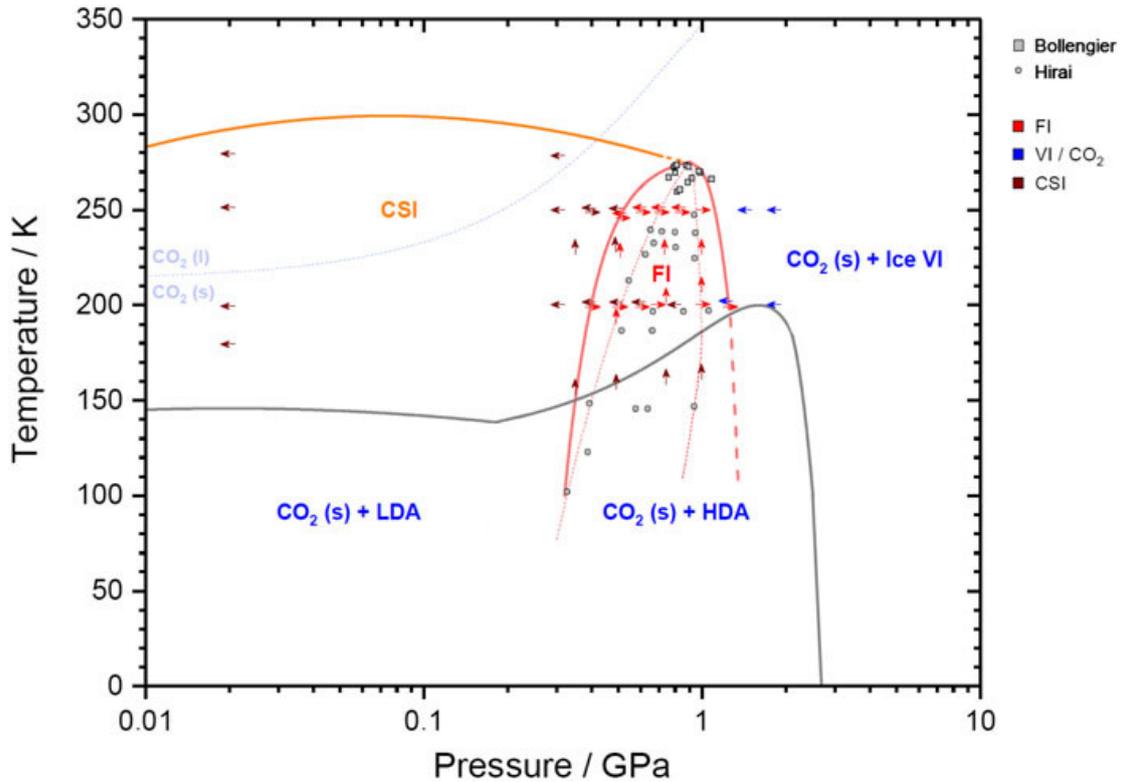
Figure 21: Pressure-temperature phase diagram in a H_2O-CO_2 system. The gray squares represent experiments of Bollengier *et al.*[20], the gray circles represent experiments of Hirai *et al.*[30], Both show the existence of a CO_2-hydrate. The red, blue and brown arrows represents experiments of Massani *et al.* and show the regions of filled ice, CO_2 I + ice VI separated and *sI*-hydrate, respectively [19]. The gray line shows the occurrence of CO_2 I and a low density amorphous and a high density amorphous phase. The light gray dotted curve is the melting curve of CO_2 I. Figure from Massani *et al.*[19].

corresponds to the FI and the χ-phase. A slight difference is, that for temperatures below 260 K a direct coexistence of *sI*-hydrate and ice VI and CO_2 I separated was found. This direct coexistence does not occur in the work of Massani *et al.* and does also not occur in this thesis.

6.5   Shape of Coexistence Lines

The method for computing the solid-solid coexistence pressures used in this work has a resolution of 100 MPa. Even though this resolution allows a wide range of possible shapes of the coexistence line of the *sI*-phase and the χ-phase a big curvature as it appears at high temperatures in the work of Massani *et al.* is out of range.

The region in which the experimental coexistence line of the *sI*-phase and the χ-phase occurs is between approximately 500 MPa and 900 MPa and at a temperature
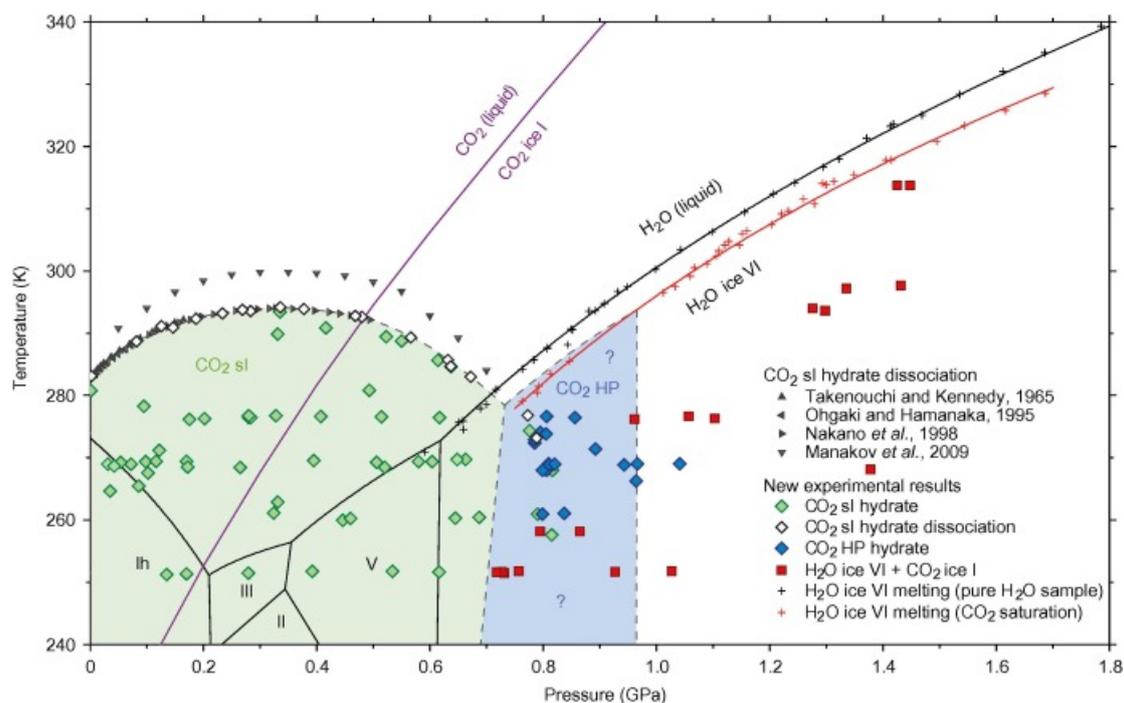
Figure 22: Pressure-temperature phase diagram in a H_2O-CO_2 system. The green diamonds show the stability region of *sI*-hydrate, the white diamond shows the dissociation of *sI*-hydrate, the blue diamond shows the stability region of a CO_2 high pressure hydrate. The red squares represent ice VI and CO_2 I separated, the black and red pluses show the melting curve of ice VI and ice VI saturated with CO_2 respectively. All these data points are measured by *Bollengier et al.*[20]. The upward triangles show the melting curve of *sI*-hydrate measured by *Takenouchi and Kennedy*[28], leftward triangles: *Ohgaki* and *Hamanak*[96], rightward triangles: *Nakano et al.*[29], downward triangles: *Manakov et al.*[97]. The purple line is the melting curve of CO_2 I. The black lines in background show the phase diagram of H_2O. Figure from *Bollengier et al.* [20]

between 230 K and 280 K. In this region there are not only the *sI*-phase and the χ-phase stable but also ice V and CO_2 I. It is possible that there is not a direct transition between the *sI*-phase and the χ-phase but instead a transition from the *sI*-phase to ice V + CO_2 I separated and then to the χ-phase. For finding the stability region for this transition way a solid-solid simulation of the χ-phase and ice V + CO_2 I separated and a solid-solid simulation of the *sI*-phase and ice V + CO_2 I separated would be necessary. This can be done by using the integration procedure introduced at the beginning of this chapter.

A similar problem occurs at slightly lower pressures and temperatures. In the region between 300 MPa and 500 MPa at temperatures between 100 K and 230 K ice II and CO_2 I separated is a third additional stable phase and it is possible that it occurs as intermediate phase in a transition from the *sI*-phase to the χ-phase. Again additional solid-solid simulations would be necessary to understand if a transition with this intermediate product occurs.

The next step to compute a more detailed shape of the transition lines could be by using Gibbs-Duhem integration introduced by David Kofke [98][99]. The Clausius-Clapeyron equation (derived by Rudolf Clausius [100] and Benoît Paul Émile Clapeyron [101])

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{\Delta H}{T \Delta V}$$

where p is the pressure T the temperature holds along the transition line of two phases. ΔV is the volume difference of the two phases and ΔH is the difference in enthalpy. Using a given point on the coexistence line the integration of the Clausius-Clapeyron equation can be started and the transition line can be computed this way. For the given problem of this thesis this procedure can be used using the sI-phase and the χ-phase. As starting point for the integration a point at low temperatures between 500 MPa and 600 MPa can be used. Since the starting point must be very precise, because otherwise the coexistence line would be diverge, it would previously be necessary to use thermodynamic integration in a pressure range between 500 MPa and 600 MPa to find suitable starting pressure and temperature.

Gibbs-Duhem integration could also be used to refine the shape of the transition line between the χ-phase and ice VI + CO_2 I. A suitable starting point would be at a pressure of 700 MPa and a temperature of 123 K.

Using these two methods, the solid-solid integration method with ice V and ice II and the Gibbs-Duhem integration it could be possible to refine the shape of the coexistence line of the different phases.

7. CONCLUSION

This thesis is devoted to study the phase diagram of a binary system containing CO_2 and H_2O molecules. In such binary systems in certain temperature and pressure ranges the molecules can form CO_2-clathrate-hydrates. The focus of this work is to find the stability region of the χ-phase of CO_2-clathrate-hydrates. Molecular Dynamics simulations allow to compute the phase diagram and are therefore an excellent alternative to experiments.

The first step of reproducing certain parts of the phase diagram is to chose suitable models for the occurring molecules in the simulations. TIP4P/ice is commonly used to reproduce different ice phases in the phase diagram of water, therefore it was chosen as model for the hydrate phases which can occur at similar temperatures and pressures. For CO_2 TraPPE was chosen to model the molecules. For a better understanding of the influence of the model it would be possible to redo the calculations using alternative models, for example TIP4P/2005 for H_2O and EPM2[89] or ZD[102] for CO_2. Especially the CO_2 alternatives would be interesting in further investigations because Amos *et al.* showed that the CO_2-molecules in the χ-phase are slightly bended but TraPPE has a fixed angle whereas EMP2 and ZD would allow to reproduce the bending.

With the chosen models the next step was to find the solid-liquid transition temperatures for a range of pressures for both, the *sI*-phase and the χ-phase. This was done by using a two phase model where the hydrate and a mixture of CO_2 and water coexist. The two phase system was relaxed for a range of temperatures and pressures. Knowing that the coexistence temperature is between the lowest temperature at which the hydrate is melting and the highest temperature at which the hydrate is growing, the coexistence temperature can be found using a bisection method.

For a further determination of the stability regions of different hydrates it was necessary to find the solid-solid coexistence line between the *sI*-phase and the χ-phase of the CO_2-hydrates and the χ-phase and ice VI and CO_2 I separated. Since two solid phases can not be simulated in a direct coexistence system a workaround using thermodynamic integration was found. This integration method allowed finding the coexistence temperature between two solids for a certain pressure. Again by scanning a range of pressures, the pressure at which the two solids coexist was found. With this method a coexistence line of the *sI*-phase and the χ-phase was found between 500 MPa and 600 MPa and the coexistence line of the χ-phase and ice VI + CO_2 I was running from (700 MPa, 123 K) to (1200 MPa, 280 K). At low pressures (<500 MPa) the most stable solid phase is the *sI*-phase, at pressures between 600 MPa

and 1200 MPa it is the χ-phase and at higher pressures water and CO_2 are separated and occur as ice VI and CO_2 I.

The stability regions of the *sI*-phase, of the χ-phase and of ice VI + CO_2 I separated approximately match with experimental results of Bollengier et al. [20] and Massani et al. [19]. Despite a slight shift in the pressure of the coexistence lines the fundamental setup of the phase diagram is the same. Compared to experimental results the shape of the coexistence line differs. For more knowledge on the shape of the coexistence lines further simulations using different ice phases would be necessary. Specially the coexistence line of the χ-phase with ice V and CO_2 I separated at higher temperatures (T>200 K) and with ice II and CO_2 I separated at lower temperatures would be interesting.

Bibliography

[1] D. M. Amos, M.-E. Donnelly, P. Teeratchanan, C. L. Bull, A. Falenty, W. F. Kuhs, A. Hermann, and J. S. Loveday, *A Chiral Gas-Hydrate Structure Common to the Carbon Dioxide-Water and Hydrogen–Water Systems*, The Journal of Physical Chemistry Letters **8**, 4295 (2017).

[2] K. A. Kvenvolden, *Methane hydrate - a major reservoir of carbon in the shallow geosphere?*, Chemical Geology **71**, 41 (1988).

[3] C. A. Koh, A. K. Sum, and E. D. Sloan, *State of the art: Natural gas hydrates as a natural resource*, Journal of Natural Gas Science and Engineering **8**, 132 (2012).

[4] R. Boswell and T. S. Collett, *Current perspectives on gas hydrate resources*, Energy & environmental science **4**, 1206 (2011).

[5] F. Ning, Y. Yu, S. Kjelstrup, T. J. Vlugt, and K. Glavatskiy, *Mechanical properties of clathrate hydrates: status and perspectives*, Energy & Environmental Science **5**, 6779 (2012).

[6] Y. MAKOGON, *Gas Hydrates, The State of Study in the USSR and Perspactives for its use*, 3rd Chemical Cong. of North America, Toronto (1988).

[7] M. D. Max, A. H. Johnson, and W. P. Dillon, *Introduction*, Economic Geology of Natural Gas Hydrate (2006).

[8] M. D. Max, *Natural gas hydrate in oceanic and permafrost environments*, vol. 5 (Springer Science & Business Media, 2003).

[9] R. Boswell, *Is gas hydrate energy within reach?*, Science **325**, 957 (2009).

[10] D. Bai, X. Zhang, G. Chen, and W. Wang, *Replacement mechanism of methane hydrate with carbon dioxide from microsecond molecular dynamics simulations*, Energy & Environmental Science **5**, 7033 (2012).

[11] K. Ohgaki, K. Takano, H. Sangawa, T. Matsubara, and S. Nakano, *Methane Exploitation by Carbon Dioxide from Gas Hydrates-Phase Equilibria for CO2-CH4 Mixed Hydrate System*, Journal of Chemical Engineering of Japan **29**, 478 (1996).

[12] R. P. Warzinski and G. D. Holder, *Gas Clathrate Hydrates1*, Energy & Fuels **12**, 189 (1998).

[13] H. Rodhe, *A comparison of the contribution of various gases to the greenhouse effect*, Science **248**, 1217 (1990).

[14] C. Moon, R. Hawtin, and P. M. Rodger, *Nucleation and control of clathrate hydrates: insights from simulation*, Faraday Discussions **136**, 367 (2007).

[15] H. Lee, Y. Seo, Y.-T. Seo, I. L. Moudrakovski, and J. A. Ripmeester, *Recovering methane from solid methane hydrate with carbon dioxide*, Angewandte Chemie **115**, 5202 (2003).

[16] Y. Park, D.-Y. Kim, J.-W. Lee, D.-G. Huh, K.-P. Park, J. Lee, and H. Lee, *Sequestering carbon dioxide into complex structures of naturally occurring gas hydrates*, Proceedings of the National Academy of Sciences **103**, 12690 (2006).

[17] T. Uchida, S. Takeya, T. Ebinuma, and H. Narita, *Replacing methane with CO2 in clathrate hydrate: observations using Raman spectroscopy* (2001).

[18] A. K. Sum, D. T. Wu, and K. Yasuoka, *Energy science of clathrate hydrates: Simulation-based advances*, MRS Bulletin **36**, 205 (2011).

[19] B. Massani, C. Mitterdorfer, and T. Loerting, *Formation and decomposition of CO2-filled ice*, The Journal of Chemical Physics **147**, 134503 (2017).

[20] O. Bollengier, M. Choukroun, O. Grasset, E. Le Menn, G. Bellino, Y. Morizet, L. Bezacier, A. Oancea, C. Taffin, and G. Tobie, *Phase equilibria in the H2O-CO2 system between 250-330 K and 0-1.7 GPa: Stability of the CO2 hydrates and H2O-ice VI at CO2 saturation*, Geochimica et Cosmochimica Acta **119**, 322 (2013).

[21] H. J. Berendsen, D. van der Spoel, and R. van Drunen, *GROMACS: a message-passing parallel molecular dynamics implementation*, Computer Physics Communications **91**, 43 (1995).

[22] H. Davy et al., in *Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London* (The Royal Society, 1832), vol. 1, pp. 385–388.

[23] E. Hammerschmidt, *Formation of gas hydrates in natural gas transmission lines*, Industrial & Engineering Chemistry **26**, 851 (1934).

[24] C. A. Koh, E. D. Sloan, A. K. Sum, and D. T. Wu, *Fundamentals and applications of gas hydrates* (2011).

[25] E. D. Sloan Jr and C. Koh, *Clathrate hydrates of natural gases* (2007).

[26] S. Sarupria and P. G. Debenedetti, *Molecular Dynamics Study of Carbon Dioxide Hydrate Dissociation*, The Journal of Physical Chemistry A **115**, 6102 (2011).

[27] M. H. Waage, T. J. Vlugt, and S. Kjelstrup, *Phase Diagram of Methane and Carbon Dioxide Hydrates Computed by Monte Carlo Simulations*, The Journal of Physical Chemistry B **121**, 7336 (2017).

[28] S. Takenouchi and G. C. Kennedy, *Dissociation Pressures of the Phase CO2· 5 H2O*, The Journal of Geology **73**, 383 (1965).

[29] S. Nakano, M. Moritoki, and K. Ohgaki, *High-pressure phase equilibrium and Raman microprobe spectroscopic studies on the CO2 hydrate system*, Journal of Chemical & Engineering Data **43**, 807 (1998).

[30] H. Hirai, K. Komatsu, M. Honda, T. Kawamura, Y. Yamamoto, and T. Yagi, *Phase changes of CO2 hydrate under high pressure and low temperature*, The Journal of Chemical Physics **133**, 124511 (2010).

[31] J. Loveday and R. Nelmes, *High-pressure gas hydrates*, Physical Chemistry Chemical Physics **10**, 937 (2008).

[32] M. Lauricella, S. Meloni, N. J. English, B. Peters, and G. Ciccotti, *Methane clathrate hydrate nucleation mechanism by advanced molecular simulations*, The Journal of Physical Chemistry C **118**, 22847 (2014).

[33] M. Von Stackelberg, *Feste Gashydrate*, Naturwissenschaften **36**, 327 (1949).

[34] M. Von Stackelberg, *Feste Gashydrate II. Struktur and Raumchemie*, Zeit. F. Elektro. **58**, 25 (1954).

[35] M. v. Stackelberg and H. Müller, *Zur struktur der gashydrate*, Naturwissenschaften **38**, 456 (1951).

[36] W. F. Claussen, *Suggested structures of water in inert gas hydrates*, The Journal of Chemical Physics **19**, 259 (1951).

[37] W. Claussen, *A second water structure for inert gas hydrates*, The Journal of Chemical Physics **19**, 1425 (1951).

[38] L. Pauling, *Pauling, l., and re marsh, 1952, proc. natl. acad. sci. usa 38, 112.*, Proc. Natl. Acad. Sci. USA **38**, 112 (1952).

[39] T. Uchida, *Physical property measurements on CO2 clathrate hydrates. Review of crystallography, hydration number, and mechanical properties*, Waste Management **17**, 343 (1998).

[40] T. Uchida, T. Hondoh, S. Mae, and J. Kawabata, *Physical data of co2 hydrate*, Direct Ocean Disposal of Carbon Dioxide pp. 45–61 (1995).

[41] R. Sun and Z. Duan, *Prediction of CH4 and CO2 hydrate phase equilibrium and cage occupancy from ab initio intermolecular potentials*, Geochimica et Cosmochimica Acta **69**, 4411 (2005).

[42] S. Circone, L. A. Stern, S. H. Kirby, W. B. Durham, B. C. Chakoumakos, C. J. Rawn, A. J. Rondinone, and Y. Ishii, *CO2 hydrate: synthesis, composition, structure, dissociation behavior, and a comparison to structure I CH4 hydrate*, The Journal of Physical Chemistry B **107**, 5529 (2003).

[43] W. Humphrey, A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*, Journal of Molecular Graphics **14**, 33 (1996).

[44] T. A. Strobel, M. Somayazulu, S. V. Sinogeikin, P. Dera, and R. J. Hemley, *Hydrogen-stuffed, quartz-like water ice*, Journal of the American Chemical Society **138**, 13786 (2016).

[45] W. Blase and G. Cordier, *NaGaSn5, eine neue Zintl-Phase mit Ga-Sn-Schrauben/NaGaSn5, a New Zintl Phase Containing Ga-Sn Helices*, Zeitschrift für Naturforschung B **43**, 1017 (1988).

[46] S. Stegmaier, S.-J. Kim, A. Henze, and T. F. Faĺssler, *Tetrahedral Framework Structures: Polymorphic Phase Transition with Reorientation of Hexagonal Helical Channels in the Zintl Compound Na2ZnSn5 and Its Relation to Na5Zn2+ x Sn10-x*, Journal of the American Chemical Society **135**, 10654 (2013).

[47] M. Cogoni, B. D'Aguanno, L. Kuleshova, and D. Hofmann, *A powerful computational crystallography method to study ice polymorphism*, The Journal of Chemical Physics **134**, 204506 (2011).

[48] M. Chaplin, *Water phase diagram* (2018), [Online; accessed 31-August-2018].

[49] B. Kamb, *Structure of ice VI*, Science **150**, 205 (1965).

[50] J.-L. Kuo and W. F. Kuhs, *A first principles study on the structure of ice-VI: Static distortion, molecular geometry, and proton ordering*, The Journal of Physical Chemistry B **110**, 3697 (2006).

[51] E. Almqvist, *History of industrial gases* (2003).

[52] M. Santoro and F. Gorelli, *High pressure solid state chemistry of carbon dioxide*, Chemical Society Reviews **35**, 918 (2006).

[53] S. Walmsley and J. Pople, *Intermolecular vibrations of solid carbon dioxide*, Molecular Physics **8**, 345 (1964).

[54] J. Martin Trusler, *Equation of state for solid phase I of carbon dioxide valid for temperatures up to 800 K and pressures up to 12 GPa*, Journal of Physical and Chemical Reference Data **40**, 043105 (2011).

[55] C.-S. Yoo, A. Sengupta, and M. Kim, *Phase diagram of carbon dioxide: update and challenges*, High Pressure Research **31**, 68 (2011).

[56] J. Li, O. Sode, G. A. Voth, and S. Hirata, *A solid-solid phase transition in carbon dioxide at high pressures and intermediate temperatures*, Nature Communications **6**, 8907 (2015).

[57] K. D. Litasov, A. F. Goncharov, and R. J. Hemley, *Crossover from melting to dissociation of CO2 under pressure: Implications for the lower mantle*, Earth and Planetary Science Letters **309**, 318 (2011).

[58] V. M. Giordano, F. Datchi, and A. Dewaele, *Melting curve and fluid equation of state of carbon dioxide at high pressure and high temperature*, The Journal of Chemical Physics **125**, 054504 (2006).

[59] V. Giordano and F. Datchi, *Molecular carbon dioxide at high pressure and high temperature*, EPL (Europhysics Letters) **77**, 46002 (2007).

[60] V. Iota and C.-S. Yoo, *Phase diagram of carbon dioxide: Evidence for a new associated phase*, Physical Review Letters **86**, 5922 (2001).

[61] V. Iota, C.-S. Yoo, J.-H. Klepeis, Z. Jenei, W. Evans, and H. Cynn, *Six-fold coordinated carbon dioxide VI*, Nature Materials **6**, 34 (2007).

[62] J. W. Gibbs, *Elementary principles in statistical mechanics* (Courier Corporation, 2014).

[63] Y. A. Cengel and M. A. Boles, *Thermodynamics: an engineering approach*, Sea **1000**, 8862 (2002).

[64] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, vol. 1 (Elsevier, 2001).

[65] D. C. Rapaport and D. C. R. Rapaport, *The art of molecular dynamics simulation* (Cambridge University Press, 2004).

[66] E. Feigl, Master's thesis, University of Vienna (2018).

[67] M. O. Steinhauser, in *Molecular Dynamics-Studies of Synthetic and Biological Macromolecules* (InTech, 2012).

[68] T. Darden, D. York, and L. Pedersen, *Particle mesh Ewald: An NâŃĚ log (N) method for Ewald sums in large systems*, The Journal of Chemical Physics **98**, 10089 (1993).

[69] P. P. Ewald, *Die Berechnung optischer und elektrostatischer Gitterpotentiale*, Annalen der Physik **369**, 253 (1921).

[70] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*, Journal of Computational Physics **23**, 327 (1977).

[71] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije, *LINCS: a linear constraint solver for molecular simulations*, Journal of Computational Chemistry **18**, 1463 (1997).

[72] B. Hess, *P-LINCS: A parallel linear constraint solver for molecular simulation*, Journal of Chemical Theory and Computation **4**, 116 (2008).

[73] H. C. Andersen, *Molecular dynamics simulations at constant pressure and/or temperature*, The Journal of Chemical Physics **72**, 2384 (1980).

[74] S. Nosé, *A unified formulation of the constant temperature molecular dynamics methods*, The Journal of Chemical Physics **81**, 511 (1984).

[75] W. G. Hoover, *Canonical dynamics: equilibrium phase-space distributions*, Physical Review A **31**, 1695 (1985).

[76] M. Parrinello and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*, Journal of Applied Physics **52**, 7182 (1981).

[77] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, *GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation*, Journal of Chemical Theory and Computation **4**, 435 (2008).

[78] J. E. Lennard-Jones, *On the determination of molecular fields. II. From the equation of state of gas*, Proc. Roy. Soc. A **106**, 463 (1924).

[79] H. Lorentz, *Ueber die Anwendung des Satzes vom Virial in der kinetischen Theorie der Gase*, Annalen der Physik **248**, 127 (1881).

[80] D. Berthelot, *Sur le mélange des gaz*, Compt. Rendus **126**, 1703 (1898).

[81] R. A. Buckingham, *The classical equation of state of gaseous helium, neon and argon*, Proc. R. Soc. Lond. A **168**, 264 (1938).

[82] D. van der Spoel, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, B. Hess, K. A. Feenstra, E. Lindahl, R. van Drunen, et al., *Gromacs User Manual version 3.1*, Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: http://www.gromacs.org (2001).

[83] McBride, *Sklogwiki, open edit encyclopedia* (2009).

[84] J. Abascal, E. Sanz, R. García Fernández, and C. Vega, *A potential model for the study of ices and amorphous water: TIP4P/Ice*, The Journal of Chemical Physics **122**, 234511 (2005).

[85] J. D. Bernal and R. H. Fowler, *A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions*, The Journal of Chemical Physics **1**, 515 (1933).

[86] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of simple potential functions for simulating liquid water*, The Journal of Chemical Physics **79**, 926 (1983).

[87] C. Murthy, S. O'Shea, and I. McDonald, *Electrostatic interactions in molecular crystals: lattice dynamics of solid nitrogen and carbon dioxide*, Molecular Physics **50**, 531 (1983).

[88] D. Möller and J. Fischer, *Determination of an effective intermolecular potential for carbon dioxide using vapour-liquid phase equilibria from NpT+ test particle simulations*, Fluid Phase Equilibria **100**, 35 (1994).

[89] J. G. Harris and K. H. Yung, *Carbon dioxide's liquid-vapor coexistence curve and critical properties as predicted by a simple molecular model*, The Journal of Physical Chemistry **99**, 12021 (1995).

[90] J. Potoff, J. Errington, and A. Panagiotopoulos, *Molecular simulation of phase equilibria for mixtures of polar and non-polar components*, Molecular Physics **97**, 1073 (1999).

[91] J. J. Potoff and J. I. Siepmann, *Vapor-liquid equilibria of mixtures containing alkanes, carbon dioxide, and nitrogen*, AIChE Journal **47**, 1676 (2001).

[92] J. Míguez, M. Conde, J.-P. Torré, F. Blas, M. Piñeiro, and C. Vega, *Molecular dynamics simulation of CO2 hydrates: Prediction of three phase coexistence line*, The Journal of Chemical Physics **142**, 124505 (2015).

[93] S. Sarupria and P. G. Debenedetti, *Molecular dynamics study of carbon dioxide hydrate dissociation*, The Journal of Physical Chemistry A **115**, 6102 (2011).

[94] T. Giorgino, *Computing 1-D atomic densities in macromolecular simulations: The density profile tool for VMD*, Computer Physics Communications **185**, 317 (2014).

[95] F. Datchi, V. M. Giordano, P. Munsch, and A. M. Saitta, *Structure of carbon dioxide phase IV: Breakdown of the intermediate bonding state scenario*, Physical Review Letters **103**, 185701 (2009).

[96] K. Ohgaki and T. Hamanaka, *Phase-behavoir of co2 hydrate-liquid co2-h2o system at high pressure*, Kagaku Kogaku Ronbun **21**, 800 (1995).

[97] A. Y. Manakov, Y. A. Dyadin, A. G. Ogienko, A. V. Kurnosov, E. Y. Aladko, E. G. Larionov, F. V. Zhurko, V. I. Voronin, I. F. Berger, S. V. Goryainov, et al., *Phase diagram and high-pressure boundary of hydrate formation in the carbon dioxide- water system*, The Journal of Physical Chemistry B **113**, 7257 (2009).

[98] D. A. Kofke, *Gibbs-Duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation*, Molecular Physics **78**, 1331 (1993).

[99] D. A. Kofke, *Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line*, The Journal of Chemical Physics **98**, 4149 (1993).

[100] R. Clausius, *On the motive power of heat, and on the laws which can be deduced from it for the theory of heat* (Dover, 1960).

[101] É. Clapeyron, *Mémoire sur la puissance motrice de la chaleur*, Journal de l'École polytechnique **14**, 153 (1834).

[102] Z. Zhang and Z. Duan, *An optimized molecular potential for carbon dioxide*, The Journal of Chemical Physics **122**, 214507 (2005).