



universität  
wien

# MASTER'S THESIS

Title of the Master's Thesis

**„Structure of the Hydrogen Fluoride Liquid/Vapour  
Interface Investigated Using Molecular Dynamics  
Simulations“**

submitted by

Elija Feigl

in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Vienna, 2018

degree programme code as it appears on  
the student record sheet:

A 066 910

degree programme as it appears on  
the student record sheet:

Computational Science

Supervisor:

Dr. Marcello Sega, Privatdoz.

# Contents

Abstract . . . . .	1
Zusammenfassung . . . . .	2
<b>1. INTRODUCTION</b>	<b>3</b>
<b>2. BACKGROUND – MOLECULAR DYNAMICS OF LIQUIDS</b>	<b>5</b>
2.1. Computer Simulation . . . . .	5
2.2. Statistical Mechanics . . . . .	6
2.2.1. Fundamental Principles . . . . .	6
2.2.2. Averages and Thermodynamics Ensembles . . . . .	8
2.3. Molecular Dynamics Simulation . . . . .	12
2.3.1. Ergodicity . . . . .	12
Temporal Average . . . . .	12
Correlation . . . . .	13
2.3.2. Integrating the Equations of Motion . . . . .	15
2.3.3. Temperature and Pressure Coupling . . . . .	22
2.3.4. Algorithm – Basic Structure of a Molecular Dynamics Simulation	25
2.4. Software . . . . .	26
2.4.1. GROMACS . . . . .	26
<b>3. FORCE-FIELD MODELLING</b>	<b>27</b>
3.1. Terminology & Theory – Modelling . . . . .	27
3.1.1. Empirical Potentials and Pairwise Additive Interactions . . . . .	28
Coulomb . . . . .	28
Lennard-Jones . . . . .	29
3.1.2. Molecular Interactions and Site-Site Force-Fields . . . . .	29
Polarizability . . . . .	30
3.2. Adaptation of Existing Hydrogen Fluoride Models . . . . .	33
3.2.1. Existing Empirical Force-Fields for Hydrogen Fluoride . . . . .	34
3.2.2. Model Adaptation: Polarizable Drude Oscillators Model . . . . .	38
3.2.3. (P)JVP-D Drude Particle Parametrization . . . . .	39
3.3. Simulation of Adapted Models . . . . .	41
3.3.1. Liquid Hydrogen Fluoride in the Isothermal-Isobaric Ensemble	42
Computational Details . . . . .	42
System Description . . . . .	43
Thermodynamic Results . . . . .	44
Structural Results . . . . .	46
3.4. Concluding Remarks for Modelling . . . . .	46
<b>4. LIQUID/VAPOUR INTERFACE</b>	<b>48</b>

---

4.1. Terminology & Theory – Interface Simulation . . . . .	48
4.1.1. Interface Analysis . . . . .	48
4.1.2. Layer-By-Layer Analysis . . . . .	49
4.1.3. Layer Identification Methods . . . . .	50
4.2. Software – Interface Analysis . . . . .	53
4.2.1. PYTIM . . . . .	53
4.2.2. MDAnalysis . . . . .	53
4.2.3. Network Analysis Tool . . . . .	54
4.3. Simulation of Liquid/Vapour Interface of Hydrogen Fluoride . . . . .	54
4.3.1. Liquid/Vapour Coexistence in the Canonical Ensemble . . . . .	54
Computational Details . . . . .	54
System Description . . . . .	55
4.3.2. GITIM Application . . . . .	57
Origin of Voids in the Liquid Phase . . . . .	62
4.3.3. Network Analysis at Liquid/Vapour Coexistence . . . . .	66
Bonds . . . . .	68
Clusters, Branches, Linear Chains and Rings . . . . .	69
4.3.4. Thermodynamic Analysis of the Hydrogen Fluoride Interface . .	74
Vapour Pressure . . . . .	74
Surface Tension . . . . .	75
Density Profiles . . . . .	76
4.4. Concluding Remarks for Interface Analysis . . . . .	81
<b>5. CONCLUSION</b>	<b>83</b>
<b>Bibliography</b>	<b>85</b>

## List of Tables

1.	Model: Parameters . . . . .	37
2.	Liquid Bulk: Energies . . . . .	45
3.	Liquid Bulk: Density . . . . .	45
4.	Liquid/Vapour Coexistence: Structural Data . . . . .	64
	a. Partial Inter-Cluster F-F RDF – $\alpha$ . . . . .	64
	b. Partial H-F RDF and H-F RDF – Cutoff . . . . .	64
5.	Liquid/Vapour Coexistence: Network Properties . . . . .	64
	a. Average Number of Bonds . . . . .	64
	b. Average Cluster Number and Size . . . . .	64

## List of Figures

1.	MD: Statistical Ensembles . . . . .	9
	a. Micro-Canonical . . . . .	9
	b. Canonical . . . . .	9
	c. Isothermal-Isobaric . . . . .	9
2.	MD: Chaotic System and Shadow Hamiltonian . . . . .	20
	a. Chaotic Trajectories . . . . .	20
	b. Shadow Hamiltonian . . . . .	20
3.	MD: Periodic Boundary Conditions . . . . .	21
4.	Model: Site-Site Approximation . . . . .	30
5.	Model: Schematic Representation . . . . .	36
	a. Point-Dipole Model . . . . .	36
	b. Drude Oscillator Model . . . . .	36
6.	Model: Drude Charge Parametrization . . . . .	41
7.	Liquid Bulk: Sample Configuration . . . . .	43
8.	Liquid Bulk: Partial RDF (P)JVP(-D) . . . . .	47
	a. RDF JVP(-D) . . . . .	47
	b. RDF PJVP(-D) . . . . .	47
9.	LJ-Liquid: Sample Configuration – ITIM Layers Coloring . . . . .	50
10.	Algorithm: ITIM and GITIM . . . . .	51
	a. ITIM Schema . . . . .	51
	b. GITIM Schema . . . . .	51
11.	Liquid/Vapour Coexistence: Sample Configuration . . . . .	56
12.	Liquid/Vapour Coexistence: RDF – Inter- and Intra- Cluster . . . . .	59

---

13.	Liquid/Vapour Coexistence: Sample Configuration – GITIM – $\alpha$ . . . .	61
	a. Incorrect Interface . . . . .	61
	b. Correct Interface . . . . .	61
14.	Liquid/Vapour Coexistence: GITIM: Intrinsic Density Profile – $\alpha$ . . . .	61
15.	Liquid/Vapour Coexistence: Sample Configuration – GITIM – Voids . .	63
16.	Liquid/Vapour Coexistence: Sample Configuration – Network Coloring	67
17.	Bonds Distribution . . . . .	68
	a. Number . . . . .	68
	b. Defects . . . . .	68
18.	Cluster Distributions . . . . .	70
	a. Cluster Size Distribution . . . . .	70
	b. Number of Molecules in Cluster Distribution . . . . .	70
19.	Constituent Species Distributions . . . . .	72
	a. Constituent Species Size Distribution . . . . .	72
	b. Number of Molecules in Constituent Species Distribution . . . .	72
20.	Network Composition . . . . .	74
	a. Temperature . . . . .	74
	b. Distribution of Constituent Species . . . . .	74
21.	Thermodynamic Properties . . . . .	76
	a. Vapour Pressure . . . . .	76
	b. Surface Tension . . . . .	76
22.	GITIM: Density Profile – Temperature . . . . .	78
	a. Non- Intrinsic Density Profile – Temperature . . . . .	78
	b. Intrinsic Density Profile – Temperature . . . . .	78
23.	GITIM: Density Profile – Layers . . . . .	80
	a. Regular Density Profile – Layers . . . . .	80
	b. Intrinsic Density Profile – Layers . . . . .	80

## List of Abbreviations

CWT	Capillary Wave Theory
EOM	Equation of Motion
F	Fluoride
H	Hydrogen
HF	Hydrogen Fluoride
H-bond	Hydrogen Bond
LJ	Lennard Jones
MC	Monte Carlo
MD	Molecular Dynamics
PBC	Periodic Boundary Conditions
PME	Particle Mesh Ewald
RDF	Radial Distribution Function
RF	Reaction Field
SCF	Self Consistent Field
SD	Standard Deviation
ITIM	Identification of Truly Interfacial Molecules
GROMACS	GRoningen MACHine for Chemical Simulations
GITIM	Generalized Identification of Truly Interfacial Molecules
PYTIM	PYthon Tool for Intrinsic Molecular analysis
inter-	between members of different groups/molecules/clusters
intra-	between members inside one group/molecule/cluster
JVP	Jedlovsky Vallauri Polarizable
JVP-D	Jedlovsky Vallauri Polarizable - Drude
PJVP	Partay Jedlovsky Vallauri Polarizable
PJVP-D	Partay Jedlovsky Vallauri Polarizable - Drude

## Abstract

Hydrogen fluoride is one of the simplest molecular liquids, however, it is characterised by a linear geometry and a pronounced dipolar nature. Despite the simplicity of its molecular structure, its properties in the liquid phase are different from those of other molecular liquids. These peculiarities are not yet fully understood, also due to technical limitations in its experimental investigation. In this thesis, the liquid/vapour interface of hydrogen fluoride is investigated by means of Molecular Dynamic simulations. To facilitate this, an existing polarizable model with point dipoles is reimplemented using a classical Drude oscillator.

The reported simulation data represent the first microscopic investigation of the liquid/vapour interface of hydrogen fluoride. The model reproduces qualitatively the experimentally known thermodynamic properties of hydrogen fluoride at liquid/vapour coexistence, most notably the slope of the vapour pressure and the surface tension. The extremely low surface tension of hydrogen fluoride results in a highly corrugated interface and in the occurrence of large spontaneously formed void structures in the liquid phase. Intrinsic analysis of the interface shows the presence of well structured successive layers despite a complex morphology and long-wavelength surface fluctuations. A network analysis reveals the presence of long linear chains, as well as heavily branched clusters in the liquid phase, caused by the strong polarity of the model. These structures can protrude from the interface and extend noticeably in the vapour region, generating a unique liquid/vapour morphology, that is not seen in any other molecular liquid.

## Zusammenfassung

Fluorwasserstoff ist eine der einfachsten molekularen Flüssigkeiten, die sich jedoch durch seine lineare Geometrie und seinen ausgeprägten dipolaren Charakter auszeichnet. Trotz der einfachen Molekülstruktur unterscheiden sich die Eigenschaften der flüssigen Phase deutlich von anderen molekularen Flüssigkeiten. Fehlendes Verständnis dieser Besonderheit liegt besonders in der Limitation der experimentellen Untersuchung von Fluorwasserstoff begründet. In dieser Arbeit wird, unter Verwendung von Molekular-Dynamik Simulationen, die Grenzfläche zwischen flüssiger und gasförmiger Phase von Fluorwasserstoff untersucht. Um die Simulation dieses Systems zu ermöglichen, wird ein bestehendes polarisierbares Modell für Fluorwasserstoff mittels eines klassischen Drude Oszillators modifiziert.

Die hier präsentierten Ergebnisse sind die erste mikroskopischen Studie der Grenzfläche zwischen flüssiger und gasförmiger Phase bei Fluorwasserstoff. Das Modell reproduziert qualitativ die experimentell bestimmten thermodynamischen Eigenschaften bei flüssig/gasförmig-Koexistenz, wobei besonders die Steigung der Dampfdruckkurve und die niedrige Oberflächenspannung hervorzuheben sind. Die niedrige Oberflächenspannung von Fluorwasserstoff führt zu einer stark zerfurchten Grenzfläche und der Entstehung von Hohlräumen in der flüssigen Phase. Trotz komplexer Morphologie und langwelliger Fluktuationen der Grenzfläche kann mit Hilfe von Intrinsischer Analyse die Ausprägung stark strukturierter, aufeinanderfolgender Schichten gezeigt werden. Untersuchung der Netzwerkstruktur des simulierten Systems offenbart die Existenz langer, linearer Ketten und stark verzweigter Cluster, die aufgrund der starken Polarität des Modells entstehen. Diese Strukturen können auch aus der Grenzregion herausragen und sich deutlich in die gasförmige Region erstrecken, wodurch eine einzigartige flüssig/gasförmig-Morphologie entsteht, die mit keiner anderen molekularen Flüssigkeit vergleichbar ist.

# 1. INTRODUCTION

Hydrogen fluoride (HF) is the simplest molecule formed by fluorine. In its liquid state, HF is widely used in petrochemical processes and in the ceramics industry. The compound is highly toxic and corrosive. Already at low humidities it converts to hydrofluoric acid, which is capable of dissolving many materials, especially oxide-compounds. As a consequence, few quantitative experiments have been performed on the structural and dynamical properties of liquid HF [1–5].

One of the main characteristics of HF is its propensity to form long molecular chains. Due to its high electronegativity, HF has a large dipole moment (1.83 D [6]), which compares, in the realm of molecular liquids, only to water (1.85 D [7]). In the case of HF, the negative dipole-end is located on the fluorine atom since the polarity is always oriented towards the atom with the larger electronegativity [8]. Consequently, both – the liquid and the vapour phases of HF – are dominated by the Coulomb interaction between the negative end on the fluorine and the positive end on the hydrogen atom. As a result, the molecule has a strong tendency to form hydrogen bonds (H-bond) and the linear character of the molecule results in the appearance of long chains of H-bonded molecules in the liquid phase [2, 9].

However, little is known about the length and lifetime of these chains, due to the lack of experimental data and the difficulties in performing complex experiments with such an aggressive substance as HF. Additionally, theoretical considerations [9, 10] suggest that branched H-bonded clusters of the liquid phase should also be present, although they are subject to strict geometrical constraints since the molecule's linear geometry restricts the number of possible bonds, as well as their orientation [3, 11, 12].

The formation of multiple, small hydrogen-bonded clusters and linear chains in the liquid phase of HF, as well as its thermodynamic properties, underline the exceptional nature of HF. Despite its strong polar nature, in contrast to water, liquid HF has an anomalously low surface tension [13, 14]. Additionally, molecules in the vapour phase of HF also show a strong tendency to form clusters [4]. These atypical features suggest a rather peculiar morphology of the liquid/vapour interface of HF. The low surface tension of the liquid should result in a highly corrugated surface, where dangling chains that extend into the vapour phase could be present. Moreover, small chains and clusters could detach from the liquid phase and persist in the vapour without breaking apart. Yet, experimental studies on the interfacial structure of a system at liquid/vapour coexistence are difficult in general [15, 16] and to the best knowledge of the author, no such study has been reported in the literature for HF.

Using computational resources for simulations and calculations, questions can be addressed which cannot be answered by experimental approaches due to technical and safety restrictions. Efforts in this direction, regarding liquid HF, date back to the 1980s [10, 17–20]. The molecule’s simplicity and its tendency to form H-bonds have made HF one of the most simulated molecular liquids. Nonetheless, accurate modelling of the dipole moment and the molecule’s polarizability present significant challenges for numerical simulations based on empirical potentials. Recent studies stressed the importance of polarizability on the liquid’s structure [2, 3, 21]. Although numerous attempts to implement better and more complex force-fields have been made lately [22–28], three-site models with point dipoles [29, 30] proved to be the most transferable modelling approach so far. These models are especially superior for the simulation of liquid/vapour coexistence over the complete liquid range [12, 29, 30].

For simulation of the liquid/vapour interface of HF an appropriate model, suitable for Molecular Dynamics simulations at liquid/vapour coexistence, is necessary. Popular Molecular Dynamics simulation-packages like GROMACS [31, 32] or NAMD [33] are based on point-charge interactions. Consequently, existing polarizable models [29, 30] have not been used extensively in recent years and simulation of HF at liquid/vapour coexistence has only been performed with inferior, non-polarizable, models [34–37]. Moreover, an algorithm capable of identifying arbitrary interfaces is essential to analyse the complex shape of HF’s liquid/vapour interface. Several methods for the identification of interfacial molecules and the computation of an interfacial surface have been developed in the last decade. Here, we refer to the Identification of Truly Interfacial Molecules (ITIM) algorithm [38], which is efficient, but restricted to interfaces with limited corrugation, and to its generalized variant (GITIM) [39], capable of identifying arbitrarily shaped interfaces.

In this work, the existing polarizable three-site models for HF [29, 30] are modified, by replacing the point-dipole with a Drude oscillator [40], to allow their use in the modern Molecular Dynamics simulation software package GROMACS. Subsequently, simulations of coexisting liquid and vapour phases at multiple temperatures are performed. The structural and topological properties of the interfacial molecules are studied using the ITIM and GITIM algorithms and a network analysis tool, developed specifically to address the characteristics of HF clusters.

To summarise, the main aims of this thesis are (I) to establish and test a point-charge model of HF that is equivalent to the best existing point-dipole models; (II) to use this model to study the thermodynamics, network properties, and interface morphology of HF at the liquid/vapour coexistence.

---

## 2. BACKGROUND – MOLECULAR DYNAMICS OF LIQUIDS

This chapter serves as a brief summary of the basics of computer simulations in general and Molecular Dynamics simulations in particular. Also, the most important definitions and technical terms will be introduced. However, due to the large variety of methods and aspects, a lot of methodology has to be omitted here. For a detailed introduction to statistical mechanics, the reader might refer to *Elementary Principles of Statistical Mechanics* by J. W. Gibbs [8] and for computer simulations of liquids to the books of Allen & Tildesley [41], Rapaport [42], and Frenkel & Smit [43]. The books listed above serve as a reference for the following sections.

### 2.1 Computer Simulation

The Laws of Newtonian Mechanics provide a simple and clear description of the motion of a single particle. Conversely, the involvement of many particles, where many usually means more than two, results in an immediate loss of simplicity. The forces acting on one particle are a result of a potential generated by all other particles and fields in the system and consequently couples the motion of one particle to the motion of all others. Except for a few special cases even the simple Newtonian mechanics of larger systems are expressed by many coupled differential equations that cannot be solved exactly.

Precise theoretical and analytical treatment of many scientific problems fail here. Consequently, most predictions concerning liquids are based on approximate descriptions of materials under certain conditions. Those descriptions often are based on information gathered from experiment or the explicit calculation of drastically simplified systems, which are then adapted to fit different situations. If theory and experiment do not match it is hard and most of the time impossible to correct the theory. Limited knowledge of molecular interactions characterizing the systems behaviour or the complete lack thereof is the main issue. Moreover, not every setup can be realized experimentally and almost every experimental measurement deals with inaccuracy and perturbations making the testing of a theory even harder. Without computer simulations there is no effective way of validation.

Computer simulations provide that missing link, by allowing modelling of those interactions. Although simulations are neither able to provide a theory nor generate exact results, they can solve any modelled system up to a desired accuracy, provided computational resources suffice. Simulations cannot replace theoretical

and experimental evaluations, but complement them. However, the comparison of experiment and simulation gives a clear indication whether or not the modelling of the inter-molecular forces is accurate. Contrarily, comparison of simulation and theory can clearly gauge the correctness of a theory and thereby ultimately link theory and experiment.

There are further benefits accompanying the usage of computer simulations. Once the hardware is acquired and the software developed, the same setup can be applied to multiple different systems and often only small adaptations to algorithms are necessary to expand the usage to a wider field of problems. Contrary to experimental setups, simulations are not influenced by external disturbances like weather or static fields that need to be shielded. Also, no special treatment is necessary to simulate systems that would be hard to set up in an experiment, like extreme temperatures or pressures or experiments that are sensitive to perturbations. Moreover, especially important for the highly corrosive HF, computer simulations do not bear the danger of potential harm on the experimentalist or the environment.

The main simulation technique used in this thesis is called Molecular Dynamics Simulation (MD), which corresponds to computing the time evolution of a many-particle system by solving the equations of motion (EOM) of classical mechanics. The analysis of the time evolution allows the calculation of various thermodynamic properties, like temperature, density, and other equilibrium properties, but also structural observables, like average bond length, bond angles, particle orientation, and radial distribution functions (RDF).

The following sections focus on the basics of statistical mechanics and thermodynamics, that build the foundation for MD simulations, as well as a short overview of the techniques utilized in this thesis.

## 2.2 Statistical Mechanics

### 2.2.1 Fundamental Principles

A complete microscopical description of a system is called its micro-state. For a classical system within the Born-Oppenheimer approximation this includes the exact knowledge of nuclear coordinates  $\mathbf{r}_i$  and momenta  $\mathbf{p}_i$  of all atoms, where the index  $i$  runs over all  $N$  molecules of the system. The electrons' motion is fast compared with that of the nuclei and therefore cancels out on average, justifying this approximation. Following a typical convention for a condensed notation in

Cartesian coordinates, the positions and momenta of all  $N$  atoms in the system are written as

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

Bold letters indicate  $n$ -dimensional vectors, where  $n$  is the space dimension. Consequently, a system of  $N$  particles has  $6N$  degrees of freedom and its Hamiltonian can be expressed as the sum of kinetic and potential energy functions:

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

where  $m_i$  is the atomic mass.

The complex informations about inter-molecular interactions are embedded in the potential energy function  $\mathcal{V}$ . Different approaches have been developed to model these interactions. Although *density functional theory*, the explicit inclusion of electronic degrees of freedom, or the use of *neural networks* are subject to rapid development and investigation, a commonly used approach in computer simulations is the division of the potential energy into terms coupling molecular pairs, triplets and higher order tuples. Many simulations are solely restricted to the use of pairwise additive potentials to model the potential energy of a system since modelling is easy and implementation straightforward and fast. Further detail about the use of pairwise additive potentials in this thesis is given in subsection 3.1.1 of chapter 3. However, this simplification is only justified for specific systems and meaningful effects might be neglected. Polarizability of a molecule like HF is the classical example for a non-negligible many-body effect. Subsection 3.1.2 covers the theoretical background for polarizability and its implications in MD simulations.

A short derivation of the Hamiltonian form of these equations is given in section 2.3.2. Once a suitable expression for the Hamiltonian  $\mathcal{H}$  is derived, the system's time evolution can be calculated by solving its EOM. This approach of explicitly modelling the potential energy clearly displays the strength of MD simulation at validating theoretical concepts since they allow the direct verification whether or not a specific inter-molecular interaction is responsible for a macroscopic effect.

Using the time-evolution of a system, the simulation calculates a series of micro-states accessible by this system. However, due to Ljapunov instability and the finite accuracy of a computer, the trajectory of a single particle as calculated by the computer simulation does not correspond to the actual trajectory of a real particle. A closer analysis of this problem is given in subsection 2.3.2. Fortunately, one specific micro-state and its exact time-evolution is not of importance for statistical

mechanics and thermodynamics since even in experiments the system constantly changes from one micro-state to another. Instead, the statistics of all micro-states accessible by the system and their distribution is essential for the system's behaviour. This representation of the system by the probability  $P_j$  of its micro-states  $j$  is called its macro-state. Thermodynamic observables like temperature or pressure can be derived as averages of a macro-state. In order to accurately specify such a probability, a large number  $M$  of similar systems is necessary from which  $M_j$  are in micro-state  $j$  such that:

$$P_j = \lim_{M \rightarrow \infty} \frac{M_j}{M} \quad (3)$$

The limit can be replaced by the specification of sufficiently large  $M$  and the entity of all similar systems is called statistical ensemble. Whenever a closed System is relaxed without exterior influences, it will reach a specific macro-state called equilibrium at which certain macroscopic quantities like density, temperature, or polarization remain constant over time. This equilibrium state motivates the formulation of the fundamental postulate of statistical mechanics:

*"For an isolated system in equilibrium, the system can be found with equal probability in any micro-state consistent with that macro-state"*

This postulate connects the microscopical properties of a system represented by the micro-state with macroscopic observables of the equilibrium and enables the calculation of these quantities as averages on the ensemble. In MD and Monte Carlo Simulation (MC) averages are computed using this concept. For MC, the ensemble is produced by performing small artificial changes on the system using a stochastic process to sample the correct ensemble. Contrarily, for MD calculating an ensemble average is replaced by an average over a finite time interval. However, the dynamics of the system are merely a convenient way of sampling micro-states, and do not necessarily bear physical meaning. The time average of a system can be calculated as the average over a large number of time steps along the system's trajectory. For an ergodic system, a finite number of micro-states are visited and the time average exactly matches the ensemble average. The following subsection covers this procedure and its restrictions in more detail.

### 2.2.2 Averages and Thermodynamics Ensembles

In conventional MD simulations the volume of the simulation box and the total energy of the system are fixed. While the fixed volume is a direct feature of the MD algorithm, energy conservation is a consequence of Hamiltonian dynamics. In contrast, most experimental studies are performed at constant temperature and

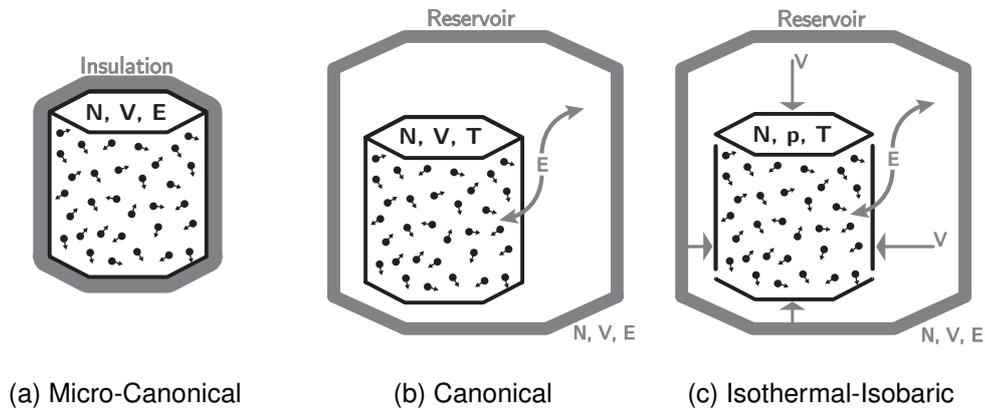


Figure 1: Representation of the three statistical ensembles used in this work

pressure. To account for the different setups, the notion of different thermodynamic ensembles is used in statistical physics. Each setup is represented by different constraints on the system and each combination of constraints results in a specific ensemble. In each ensemble, different observables are permitted to fluctuate, thereby enabling different behaviour.

In the thermodynamic limit, the fluctuation's importance ceases to be significant and the ensembles become mostly indistinguishable. Therefore, the choice of the correct ensemble is only a matter of convenience in most cases. Contrarily, the unconstrained observables can strongly fluctuate for small systems. In computer simulation usually only about  $10^3$  to  $10^5$  molecules are used, due to the limited capacity of computational resources. This mandates, that care has to be taken about the system's specific setup to allow for an accurate computation of the desired quantities in accordance with statistical physics.

In this thesis we make use of three different setups:

- Micro-canonical ensemble ( $N, V, E$ )
- Canonical ensemble ( $N, V, T$ )
- Isothermal-Isobaric ensemble ( $N, p, T$ )

The letters in brackets indicate, which macroscopic variables are constrained to a certain value. For the three ensembles listed above they are: molecule number  $N$ , volume  $V$ , energy  $E$ , temperature  $T$  and pressure  $p$ . Figure 1 shows a visual representation of the three setups. For the sake of completeness it is necessary to emphasize that more thermodynamic ensembles exist and that for each specific algorithms have been designed. However, those ensembles and their implementation will not be covered in this work.

The statistical average, denoted by enclosing angle brackets, of thermodynamic quantities  $A(\mathbf{r}^N, \mathbf{p}^N)$  can be computed by,

$$\langle A \rangle = \int d\mathbf{r}^N d\mathbf{p}^N f(\mathbf{r}^N, \mathbf{p}^N) A(\mathbf{r}^N, \mathbf{p}^N) \quad (4)$$

where  $f(\mathbf{r}^N, \mathbf{p}^N)$  is the probability-density of micro-states specified by  $\{\mathbf{r}^N, \mathbf{p}^N\}$  in the corresponding ensemble.

### Micro-Canonical Ensemble

For the micro-canonical ensemble the system is assumed to have a fixed volume and can neither exchange energy nor particles with the environment. Conventional MD simulations are performed in this ensemble. The constraints on the system imply that the number of possible micro-states of such a system, also called partition function  $\Omega$ , can be expressed as:

$$\Omega = \frac{1}{N!(2\pi\hbar)^{3N}} \int d\mathbf{r}^N d\mathbf{p}^N \delta[\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) - E]. \quad (5)$$

The factor  $\frac{1}{N!}$  accounts for indistinguishable molecules, whereas  $(2\pi\hbar)$  is the phase-space-volume of a micro-state. The expression is basically a sum over all configurations where the Hamiltonian is exactly equal to the energy  $E$ . As a result, the probability density function of the micro-canonical ensemble is:

$$f_{NVE} = \frac{1}{\int d\mathbf{r}^N d\mathbf{p}^N \delta[\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) - E]} \delta[\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) - E] \quad (6)$$

For a system at equilibrium the number of microscopic configurations  $\Omega$  is directly tied to thermodynamics via the entropy  $S$ , which is defined as:

$$S(N, V, E) = k_B \ln \Omega, \quad (7)$$

$k_B$  is the Boltzmann-constant. This formula also implies that at equilibrium, since all micro-states are equally likely, the entropy has to be at its maximum. It is also used to formally define the temperature  $T$  as a thermodynamic property:

$$\frac{1}{T} = \left( \frac{\partial S}{\partial E} \right)_{V, N} \quad (8)$$

Although usually not suitable for experimental comparison, where energy exchange with the environment is generally inevitable, simulations at constant energy are often used for preparing initial configurations.

All other ensembles can be derived from the micro-canonical ensemble using extended Lagrangians. Moreover, it is generally possible to restructure any MD simulation to be performed at fixed volume and energy, but the use of a more appropriate ensemble drastically simplifies the computation.

### Canonical Ensemble

The canonical ensemble describes a system that, like the micro-canonical ensemble, has a fixed volume  $V$  and cannot exchange particles with the environment. However, it is in thermal contact with a heat-bath of constant temperature  $T$  and energy fluctuates. Consequently, the probability to find the system in a micro-state with a given Hamiltonian  $\mathcal{H}$  is given by the Boltzmann-factor  $e^{-\beta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)}$ , with inverse temperature  $\beta = \frac{1}{k_B T}$ . The complete system (simulation box plus heat bath) have fixed volume and energy, hence are in the micro-canonical ensemble.

The number of accessible micro-states  $Q_{NVT}$  can be expressed as:

$$Q_{NVT} = \frac{1}{N!(2\pi\hbar)^{3N}} \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)} \quad (9)$$

and the canonical ensemble's probability density function is,

$$f_{NVT} = \frac{1}{\int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)}} e^{-\beta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)}. \quad (10)$$

The thermodynamic potential that is directly linked to the canonical ensemble is the Helmholtz Free Energy  $F(N, V, T) = E - TS$ . It is related to the canonical partition function by,

$$F(N, V, T) = -k_B T \ln Q_{NVT}. \quad (11)$$

This relation combines the thermodynamic perspective with the microscopical view of statistical physics and can for instance be used to derive the average energy  $E$ . Taking the derivative of the free energy with respect to  $\beta$  in the thermodynamic perspective yields:

$$\frac{\partial(\beta F)}{\partial\beta} = \frac{\partial}{\partial\beta} \left( \beta E - \frac{S}{k_B} \right) = E, \quad (12)$$

whereas in statistical mechanics the derivative is equivalent to:

$$\frac{\partial(\beta F)}{\partial\beta} = -\frac{\partial(\ln Q_{NVT})}{\partial\beta} = \frac{1}{\int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)}} \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)} \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) \quad (13)$$

Using the definition of the ensemble average given in equation 4 these results combine to

$$\langle \mathcal{H} \rangle = E, \quad (14)$$

verifying that the energy of an equilibrated system in the canonical ensemble fluctuates around the average energy  $E$ . Simulations of interfaces of coexisting phases are performed in this ensemble.

### Isothermal-Isobaric Ensemble

The isothermal-isobaric ensemble corresponds to a system in contact with a heat-bath and a volume-reservoir. It cannot exchange particles with the environment. Consequently the system's energy and volume fluctuate around an equilibrium value. Again, the complete system is in the micro-canonical ensemble. This is also the setting at which most experiments for analysing and characterizing fluids are performed. The ensemble is especially useful for the study of thermodynamic and structural properties, and the parametrization and verification of an empirical potential.

The partition function and probability density function for the isothermal-isobaric ensemble are:

$$Q_{NpT} = \frac{\beta p}{N!(2\pi\hbar)^{3N}} \int_0^\infty dV \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta[\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) + pV]} \quad (15)$$

$$f_{NpT} = \frac{1}{\int_0^\infty dV \int d\mathbf{r}^N d\mathbf{p}^N e^{-\beta[\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) + pV]}} e^{-\beta[\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) + pV]}, \quad (16)$$

where  $p$  is the pressure (not to be confused with the momentum  $\mathbf{p}$ ) and  $V$  the volume of the simulated system. The factor  $\beta p$  is added to make the partition function dimensionless and vanishes in the thermodynamic limit.

## 2.3 Molecular Dynamics Simulation

### 2.3.1 Ergodicity

#### Temporal Average

The calculation of averages explicitly using the partition function is only possible for few systems because of the complexity of calculating the partition function itself. For other simulation techniques clever schemes have been designed to avoid the direct calculation. Metropolis MC for instance circumvents the problem by a specific

choice of the acceptance probability of each move. In the case of MD simulations the difficulty is evaded by calculating a temporal average rather than an ensemble-average. This approach also reflects the notion of a real-world measurement, where a macroscopic quantity is measured as its average over the finite duration of the measurement which is usually much larger than the characteristic timescale for the microscopic motion. The temporal-average of a quantity  $A(\mathbf{r}^N, \mathbf{p}^N)$  for a given trajectory is:

$$\bar{A}_t = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t d\tau A(\mathbf{r}^N(\tau), \mathbf{p}^N(\tau)) \quad (17)$$

During its time evolution, a system at equilibrium visits micro-states corresponding to a specific statistic ensemble. If the simulation is set up in such a way that it will eventually visit all accessible micro-states, given long enough simulation time  $t$ , the simulation is called ergodic. Moreover, satisfying ergodicity implies that the temporal average does not depend on the initial conditions of the trajectory. For a simulation that satisfies all these conditions the ergodic hypothesis holds:

$$\bar{A}_t = \langle A \rangle \quad (18)$$

Averages calculated with MD simulations rely on this property, therefore special caution is necessary for ensuring that the simulated system is indeed ergodic. An example for a violation of the ergodic hypothesis is spontaneous symmetry breaking.

### Correlation

The discrete and finite nature of a computer simulation has to be taken in account when thermodynamic properties are calculated. Computer simulations do not provide continuous trajectories but compute the system by solving its EOM using finite difference methods for a given time-step  $\Delta t$ . Additionally, the trajectory of a system can only be calculated for a finite number of time-steps. Consequently, the time average defined above can only be estimated by

$$\bar{A}_t \sim \bar{A} = \frac{1}{M} \sum_{i=1}^M A(\mathbf{r}^M_i), \quad (19)$$

where  $\bar{A}$  denotes the sample-average calculated by the MD simulation of  $M$  time-steps.

The number of steps  $N$  and the step-size  $\Delta t$  have to be chosen such that ergodicity is guaranteed. However, when calculating statistical averages usually not only the correct average value is desired, but also a small value for its variance. The variance is a measure for the deviation from the mean, hence theoretical interpretations of

the result rely on it. Due to the finite sample size of the MD simulation statistical errors are extremely prominent. Moreover, the value of a given quantity  $A(\mathbf{r}^N_i)$  at step  $i$  is not statistically independent of the values at previous time-steps  $A(\mathbf{r}^N_j)$  with  $j < i$ . This dependence is also called correlation. As a result, the sample-average's variance strongly deviates from the variance of the ensemble-average.

The following short calculation illustrates this deviation and motivates the definition of the correlation-function  $C_A(k)$  and the correlation length  $\tau_A$ .

$$\begin{aligned} Var(\bar{A}) &= \langle \bar{A}^2 \rangle - \langle \bar{A} \rangle^2 \\ &= \frac{1}{M^2} \langle \sum_{i=1}^N \sum_{j=1}^N A(\mathbf{r}^N_i) A(\mathbf{r}^N_j) \rangle - \frac{1}{M} \langle \sum_{i=1}^N A(\mathbf{r}^N_i) \rangle \langle \sum_{j=1}^N A(\mathbf{r}^N_j) \rangle \\ &= \frac{1}{M^2} \sum_{i=1}^N \sum_{j=1}^N [\langle A(\mathbf{r}^N_i) A(\mathbf{r}^N_j) \rangle - \langle A(\mathbf{r}^N_i) \rangle \langle A(\mathbf{r}^N_j) \rangle] \end{aligned} \quad (20)$$

For a stationary state the above expression can be rewritten to only depend on the number of steps  $k$  between  $i$  and  $j$ , such that  $i+k=j$ , instead of  $i$  and  $j$  individually. Moreover, for a system at equilibrium the contribution is the same for any  $i$ .

$$\begin{aligned} &= \frac{1}{M^2} \sum_{i=1}^N \sum_{k=-\infty}^{\infty} [\langle A(\mathbf{r}^N_i) A(\mathbf{r}^N_{i+k}) \rangle - \langle A(\mathbf{r}^N_i) \rangle \langle A(\mathbf{r}^N_{i+k}) \rangle] \\ &= \frac{1}{M} \sum_{k=-\infty}^{\infty} [\langle A(\mathbf{r}^N_i) A(\mathbf{r}^N_{i+k}) \rangle - \langle A(\mathbf{r}^N_i) \rangle \langle A(\mathbf{r}^N_{i+k}) \rangle] \end{aligned} \quad (21)$$

By defining the symmetric, normalized auto-correlation-function  $C_A(k)$  as

$$C_A(k) = \frac{\langle A(\mathbf{r}^N_i) A(\mathbf{r}^N_{i+k}) \rangle - \langle A(\mathbf{r}^N_i) \rangle \langle A(\mathbf{r}^N_{i+k}) \rangle}{\langle A^2 \rangle - \langle A \rangle^2} \quad (22)$$

and the correlation length  $\tau_A$  by

$$\tau_A = \sum_{k=0}^{\infty} C_A(k), \quad (23)$$

the final result for the sample average's variance becomes

$$Var(\bar{A}) = \frac{2\tau_A}{M} Var(A). \quad (24)$$

In case of a MD simulation the number of steps corresponds to elapsed time and the correlation function can be reformulated as a time-correlation function.

The above expression illustrates that a large sample size is necessary to achieve statistical accuracy, but it also clarifies that correlations in those samples effectively

reduce the sample size. Hence, for a strongly correlated system a sample size larger by a factor of  $2\tau_A$  is necessary to achieve the same statistical accuracy as for an uncorrelated system. As a result it is always necessary to calculate the correlation-length and hence the auto-correlation-function of a property to allow statistical interpretations. This can either be done exactly using Fourier transform or quantitatively using block-averages.

Although this is clearly inconvenient when sampling statistical properties, linear response theory successfully uses the correlation-function to allow further insight in the behaviour of a system. Applying this theory, so called Green-Kubo relations, which relate the correlation of a property to a transport coefficient, can be derived. The diffusion coefficient  $D$  for example can be calculated as

$$D = \frac{1}{3} \int_0^{\infty} dt \langle \mathbf{v}(t) \mathbf{v}(0) \rangle. \quad (25)$$

### 2.3.2 Integrating the Equations of Motion

Various approaches in defining the classical EOM have been proposed. Although Lagrangian dynamics are often referred to as the most fundamental one and predate the Hamiltonian reformulation by 45 years, this thesis will primarily focus on the latter formulation, due to the explicit notion of momenta and the strong connection of the canonical ensemble to the Hamiltonian formulation.

Using generalized coordinates the Hamiltonian is defined as:

$$\mathcal{H}(\mathbf{p}, \mathbf{q}) = \sum_k \dot{q}_k p_k - \mathcal{K}(\dot{\mathbf{q}}, \mathbf{q}) + \mathcal{V}(\dot{\mathbf{q}}, \mathbf{q}) \quad (26)$$

and the Hamiltonian equations of motion are:

$$\begin{aligned} \dot{q}_k &= \frac{\partial \mathcal{H}}{\partial p_k} \\ \dot{p}_k &= -\frac{\partial \mathcal{H}}{\partial q_k}, \end{aligned} \quad (27)$$

where  $\mathbf{q}$  and  $\mathbf{p}$  are the generalized coordinates and momenta, and  $\mathcal{K}$  and  $\mathcal{V}$  the kinetic and potential energy, respectively.

The Cartesian variant of the Hamiltonian simplifies to

$$\mathcal{H}(\mathbf{r}^N, \mathbf{p}^N) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V}(\mathbf{r}^N) \quad (28)$$

and the Hamiltonian EOM become

$$\begin{aligned}\dot{\mathbf{r}}_i &= \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i &= -\nabla_{\mathbf{r}_i} \mathcal{V} = \mathbf{F}_i,\end{aligned}\tag{29}$$

where  $\mathbf{F}_i$  is the Force exerted on particle  $i$ . For a system of  $N$  rigid molecules this amounts to a system of  $6N$  coupled differential equations of first order, whereas the coupling is realized via the gradient of the potential energy  $\mathcal{V}$ .

As long as no velocity-dependent forces act on the system, the total derivative  $\dot{\mathcal{H}} = \frac{d\mathcal{H}}{dt}$  is zero and therefore the Hamiltonian a constant of the motion. The Hamiltonian EOM also conserve total momentum and, for an isolated system, total angular momentum. However, periodic boundary conditions, as frequently used in MD simulations, break the rotational invariance and angular momentum is not conserved. Furthermore, the motion described by these equations is time-reversible, which means by inverting the momenta the system will retrace its trajectory in the opposite direction back to the initial setting. Additionally, Liouville's theorem holds that the phase space density is conserved under Hamiltonian dynamics.

These properties are of special interest in designing and testing finite integration schemes since any algorithm is required to fulfil the same conditions. Moreover, conserved quantities can easily be used to verify whether an algorithm is correctly implemented or not. Any algorithm that satisfies Liouville's theorem is called symplectic. Often the order of accuracy of an algorithm is of far less importance than the question, whether it is time reversible and symplectic or not. Since it is by concept impossible to calculate the exact trajectory of a system with a computer simulation, it is usually more important to decide how useful the calculated result is rather than how close it is to the exact solution. This problem is investigated more thoroughly in section 2.3.2. Although, many higher order algorithms, like Runge-Kutta [44, 45] and predictor-corrector [46] have been developed, but they are either not symplectic or need multiple force calculations and are thereby computationally expensive.

Two  $2^{nd}$ -order algorithms that fulfil all the above requirements and are simple and straightforward to derive and implement, are the Størmer-Verlet [47] and the velocity Verlet [48]. They also only necessitate a minimal number of force calculations per step, if implemented correctly. Since force calculations are the computationally expensive part of an MD simulation, the two algorithms are fast and only need little memory and are therefore widely applied in MD calculations.

**Størmer-Verlet:**

The algorithm is attributed to *Carl Størmer* around 1907 and was rediscovered and applied to MD simulations about 60 years later by *Loup Verlet* [47]. The algorithm implements an equation to advance the position of a particle by a time-step  $\Delta t$ :

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{F}(t)}{m} \Delta t^2 + \mathcal{O}(\Delta t^4) \quad (30)$$

The expression can easily be derived by adding the equations of the Taylor expansion of  $\mathbf{r}(t + \Delta t)$  and  $\mathbf{r}(t - \Delta t)$ :

$$\begin{aligned} \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m} \Delta t^2 + \mathcal{O}(\Delta t^3) \\ \mathbf{r}(t - \Delta t) &= \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m} \Delta t^2 - \mathcal{O}(\Delta t^3) \end{aligned} \quad (31)$$

Since all terms of odd power are eliminated from equation 30, velocities  $\mathbf{v}$  are not needed to compute the trajectory. Therefore it is easy to verify that the algorithm is time reversible. If needed, velocities can be obtained in hindsight as a central difference:

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2). \quad (32)$$

Adding to the inconvenience of missing velocities, the algorithm is not self-starting, illustrated by the appearance of  $\mathbf{r}(t - \Delta t)$  on the right side of equation 30.

**Velocity Verlet:**

To account for self-starting and the direct calculation of velocities the Størmer-Verlet algorithm can be reformulated in its velocity-dependent form as:

$$\begin{aligned} \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m} \Delta t^2 + \mathcal{O}(\Delta t^3) \\ \mathbf{v}(t + \Delta t) &= \mathbf{v}(t) + \frac{\mathbf{F}(t) + \mathbf{F}(t + \Delta t)}{2m} \Delta t + \mathcal{O}(\Delta t^2) \end{aligned} \quad (33)$$

The position part in equation 33 is the Taylor expansion of  $\mathbf{r}(t + \Delta t)$  and the second can be derived by inserting the first equation in the time-reversed Taylor expansion of  $\mathbf{r}(t)$ :

$$\mathbf{r}(t) = \mathbf{r}(t + \Delta t) - \mathbf{v}(t + \Delta t)\Delta t + \frac{\mathbf{F}(t + \Delta t)}{2m} \Delta t^2 + \mathcal{O}(\Delta t^3) \quad (34)$$

The use of the time reversed Taylor expansion illustrates that the resulting algorithm is, like the Størmer-Verlet, time-reversible. If implemented properly the algorithm does not require more force-calculations than the initial version. Storing both force-terms,  $\mathbf{F}(t + \Delta t)$  and  $\mathbf{F}(t)$ , is also unnecessary since the new velocity can be computed in two steps, hence saving additional memory. First, after calculating the new position using the current force  $\mathbf{F}(t)$ , the new force  $\mathbf{F}(t + \Delta t)$  is calculated and added to the current one to calculate the velocity.

The global error of both these algorithm can be proven to be of second order and both algorithms produce the same trajectory as long as the forces are not velocity-dependent. However, the velocity Verlet cannot be used in the case of velocity-dependent forces. The symplectic nature of both algorithms can be shown by expressing it as a Trotter expansion of the Liouville operator (see chapter 4, *Frenkel & Smit* [43] for details).

A popular variant of the velocity Verlet is the leapfrog [49], where position and velocity are evaluated alternatingly a half-step apart instead of at the same time, leapfrogging each other. The two algorithms share the same properties and their results are equivalent. However, if the velocities are not explicitly needed, leapfrog is often the preferred choice due to its intuitive implementation.

### Instability

The underlying idea for performing MD is solving the EOM of a system in small time-steps  $\Delta t$ , starting from some initial conditions  $\mathbf{s}_0^N = \{\mathbf{r}_0^N; \mathbf{p}_0^N\}$ , and producing a trajectory, namely a sequence of points  $\mathbf{s}_i^N = \{\mathbf{r}_i^N; \mathbf{p}_i^N\}$  in phase-space. Given a sufficiently small time-step and accurate algorithm, this computed trajectory approximates the EOM's exact solution for the given initial conditions  $\mathbf{s}_0^N$ . However, many systems of coupled differential equations are sensitive to small deviations to the initial conditions  $\delta \mathbf{s}_0^N = \{\delta \mathbf{r}_0^N; \delta \mathbf{p}_0^N\}$ . Such systems are called chaotic.

The classification of the behaviour of both these systems can best be described by the notion of Lyapunov exponents  $\lambda$ , defined by:

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{|\delta \mathbf{s}_t^N|}{|\delta \mathbf{s}_0^N|} \quad (35)$$

which can be reformulated as a condition on the perturbation at time  $t$  as:

$$|\delta \mathbf{s}_t^N| \sim |\delta \mathbf{s}_0^N| e^{\lambda t} \quad (36)$$

If the Lyapunov exponent  $\lambda \gg 1$ , the system's behaviour will be chaotic and the calculated trajectory will significantly deviate from the real trajectory. From a solely

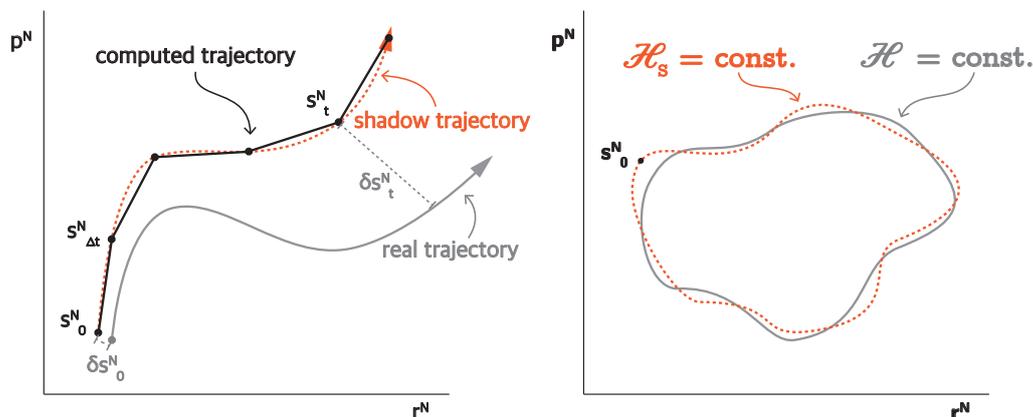
technical point of view, due to the finite memory capacities of computers, it is impossible to design a completely unperturbed system. In a three-dimensional system of  $N$  particles there are  $6N$  coupled ordinary differential equations and therefore  $6N$  Lyapunov exponents are defined. Moreover, the behaviour of the whole system is determined by largest Lyapunov exponent. Real systems are also affected by this so called Lyapunov instability and can become chaotic. However, the problem with MD is, that the perturbation of a simulated system deviates strongly and in an uncontrollable manner from that of a real system. Consequently, the system's trajectory cannot be used to predict the exact motion of a single particle in the system. This fact also resonates in the observation, that simply increasing the order of the integration algorithm is not overly beneficial for most simulations. No matter how accurate the used algorithm is, it will never be able to reproduce the real endpoint of the trajectory of a chaotic system. Without the notion of symplectic algorithms the results of MD simulation would be rendered useless for almost all cases.

It is however possible to prove, that for the trajectory computed by a symplectic algorithm a so called shadow Hamiltonian  $\mathcal{H}_S$  exists, that represents a system for which the computed trajectory is the exact solution. Moreover, the shadow Hamiltonian  $\mathcal{H}_S$  deviates from the original Hamiltonian  $\mathcal{H}$  only by  $\Delta\mathcal{H} = \mathcal{O}(\Delta t^n)$ , where  $n$  is the order of the algorithm used. The maximum size of this deviation does not change over time. Figure 2a shows a comparison of the trajectory computed by the MD simulation, the trajectory of the real system defined by the Hamiltonian  $\mathcal{H}$ , and the trajectory for the shadow Hamiltonian  $\mathcal{H}_S$ . The discrete MD trajectory successfully approximates the shadow system's trajectory, but the real trajectory differs dramatically showing strong sensitivity to small perturbations, hence a chaotic system is present.

As a result it can be confidently concluded, that properties that do not sensitively depend on small changes in the Hamiltonian, for instance the RDF, can be accurately computed by the use of symplectic MD simulations.

### **Simulation Box and Periodic Boundary Conditions**

Due to the limitations of computational resources, MD simulations are restricted to a small number of molecules. It is therefore necessary to confine the particles in a box of length  $L$  and impose conditions for particles at or close to this boundary. In this thesis the simulated systems represent a substance that is several orders of magnitude larger in terms of particle number. Consequently, the boundary conditions should represent the influence of particles in the surrounding, larger substance. Since the larger system cannot be simulated for reasons of limited



(a) Conceptual depiction of the computed trajectory (black dots), the real trajectory (solid grey curve), and the trajectory corresponding to the shadow Hamiltonian  $\mathcal{H}_s$  (dotted orange curve) of a chaotic system. The dots represent the discrete computation for a time-step  $\delta t$  and the black lines, connecting the computed time-steps is a guide for the eye.

(b) Trajectories of constant Hamiltonian  $\mathcal{H}$  (grey, solid line) for the real trajectory and constant shadow Hamiltonian  $\mathcal{H}_s$  (orange, dashed line) for the computed and shadow system.

Figure 2: Molecular Dynamic simulations show a) chaotic behaviour and b) approximate the trajectory of a system with Hamiltonian  $\mathcal{H}_s$ .

computational resources, a commonly applied solution is the use of Periodic Boundary Conditions (PBC).

The concept of PBC is best illustrated by a system that only extends in one spatial dimension. This corresponds to particles positioned along a line. Imposing PBC on this system corresponds to connecting the two loose ends of this line. As a consequence all particles have two neighbours instead of only one neighbour for particles at the loose ends of the line. In two and three dimensions it is usually more convenient to imagine identical copies, called images, of the system surrounding the original box as illustrated for two dimension in figure 3.

However, special care has to be taken when applying PBC. Since the main idea is to reduce the number of simulated particles and thereby be able to represent a large system, it is sensitive to only calculate interactions with the closest copy of each particle. This technique is often referred to as the minimum image convention. Consequently, interactions with periodic images of the particle with itself are automatically excluded. The implication of this technique is the limitation of interactions to contributions within a given cutoff-radius  $r_c$ , where the condition  $2r_c < L$  has to be fulfilled. All interaction within the cutoff are called short-range interaction. The remaining interactions are referred to as long-range. It is possible to take the error produced by the cut-off into account by the use of so called long-range corrections. This approach is, besides other conditions, reasonable

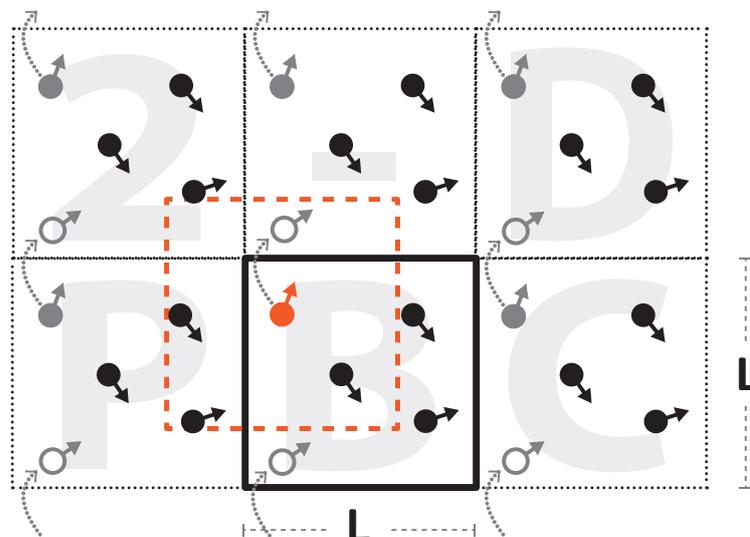


Figure 3: Crossing one of the four box edges of a two dimensional periodic system, molecules leave the box on one side and enter on the opposing side. In a three-dimensional system molecules can cross one of the six cube faces. The orange square (dashed) represent the minimum image of the highlighted (orange) molecule. Consequently, its interactions with the two rightmost molecules in central box *B* are calculated with their periodic image in box *P*.

if the simulation-box is larger than the extension of local structures. Common correction-techniques are tail-correction, Reaction Field method [50–52] (RF), and ewald-summation [53]. The tail-correction is an analytical solution of the integral corresponding to the average contributions of a thermodynamic property beyond the cutoff. However, in three spatial dimensions the integral is only convergent for interactions decreasing faster than  $1/r$  and only analytically solvable if the RDF converges to 1 within the cutoff.

The RF method, assuming a uniform dielectric constant outside the cutoff, is not applicable for biphasic systems. Instead of a direct calculation in real space, Ewald-summation is performed in reciprocal space. Particle-Mesh-Ewald [54, 55] (PME), evaluating the long-range interactions on discrete lattice, is applicable for both slowly decreasing interaction and biphasic systems and is therefore the preferred method for modern MD simulations. Moreover, the PME method's computational complexity is of order  $\mathcal{O}(N \log N)$ . Further details about long-range correction techniques can be found in chapter 2.8 of *Allen & Tildesley* [41]. To avoid unphysical artefacts it is necessary to make sure that no significant correlation effects exceed the simulation box length.

### 2.3.3 Temperature and Pressure Coupling

Applying only the methodology and algorithms discussed in subsection 2.3.2 will result in a simulation in the micro-canonical ensemble. This becomes apparent when re-evaluating the micro-canonical  $(N, V, E)$ -setup, namely constant energy  $E$ , particle number  $N$ , and volume  $V$ . The algorithms described there are specifically designed to integrate the EOM of a fixed number of  $N$  particles while conserving the energy. The constant volume of the simulation box is, if not deliberately changed, an inherent feature of a computer simulation (also see subsection 2.3.2). Hence the conditions of the micro-canonical ensemble are fulfilled.

To perform simulations in the  $(N, V, T)$  or  $(N, P, T)$  ensemble it is necessary to explicitly control temperature and pressure of the system. For simulations in the  $(N, V, T)$ -ensemble a thermostat is needed and the isothermal-isobaric ensemble requires both a thermostat and a barostat.

Two different approaches for temperature and pressure coupling are common for MD simulation. One is the pairing of the principles of MD with stochastic elements and represents rescaling particle velocities and/or coordinates according to the ensemble distribution. The second approach, first introduced by *H.C. Andersen* [56] in 1979 for pressure coupling, is based on the reformulation of the system's dynamics by extending the Hamiltonian – the original approach has been performed by extending the Lagrangian – of the system and performing a (NVE)-simulation for this new system comprised of the actual system and a volume-reservoir. The same technique has been quickly adapted for temperature coupling as well as a large variety of other situations. The basic concepts of the two schemes used in this thesis both follow the second approach and are further discussed in the following sections.

#### Nosé-Hoover Thermostat

The thermostat developed by *Nosé* [57] in 1984 and later improved by *Hoover* [58] consists of a heat-bath represented by an imaginary particle moving along an additional degree of freedom  $s$  with momentum  $p_s$ . The dynamics of this particle are included in the Hamiltonian by

$$\mathcal{H}_{NH}(\mathbf{r}^N, \mathbf{p}^N, p_s, s) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + \mathcal{V}(\mathbf{r}^N) + \frac{p_s^2}{2Q} + gk_B T \ln(s), \quad (37)$$

where  $Q$  has the dimension of  $(energy \cdot time^2)$  and behaves like an imaginary mass. The constant  $g$  is chosen such that the average of this Hamiltonian is equal to the canonical ensemble average at equilibrium.  $T$  is the imposed temperature.

By expressing the coordinates via so called virtual coordinates and introducing a thermodynamic friction coefficient  $\zeta$  as

$$\begin{aligned}\mathbf{r}'_i &= \mathbf{r}_i \\ \mathbf{p}'_i &= \frac{\mathbf{p}_i}{s} \\ t' &= \int^t \frac{dt}{s} \\ \zeta &=: \frac{p_s}{Q}\end{aligned}\tag{38}$$

the EOM for this system can be written as:

$$\begin{aligned}\frac{d\mathbf{r}'_i}{dt'} &= \frac{\mathbf{p}'_i}{m_i} \\ \frac{d\mathbf{p}'_i}{dt'} &= \mathbf{F}_i - \zeta \mathbf{p}'_i \\ \frac{d\zeta}{dt'} &= \frac{1}{Q} \left[ \sum_{i=1}^N \frac{\mathbf{p}_i'^2}{m_i} - 3Nk_B T \right] \\ \frac{d\ln(s)}{dt'} &= \zeta\end{aligned}\tag{39}$$

The constant  $g$  has to be equal to  $3N$ . The first two lines in equation 39 look similar to the original EOM as derived in subsection 2.3.2. However, the momenta-dependent term in the second equation is responsible for coupling the thermostat to the system by either accelerating or decelerating all molecules depending on the sign of  $\zeta$ . The right side of the third equation can be rewritten as  $\frac{2}{Q} [\mathcal{K} - \langle \mathcal{K} \rangle]$ , with  $\mathcal{K}$  the kinetic energy. This reformulation illustrates that  $\zeta$  facilitates a deceleration, if the kinetic energy is above and an acceleration, if it is below its average value.

The thermostat does not react instantaneously to deviations from the target temperature, but rather its time derivative is directly coupled to those deviations. How fast the thermostat reacts is controlled by the parameter  $Q$ , hence it is referred to as a mass. For larger values of  $Q$  the thermostat reacts faster, but disadvantageously this also results in a stronger perturbation of the system's dynamics. A typical method in determining  $Q$ , such that the thermostat reacts as fast as possible without significantly affecting the dynamics, is the comparison of the velocity autocorrelation function for different values of  $Q$ . If the velocity autocorrelation function does not change qualitatively, also the system dynamics is not affected strongly. Consequently, the largest value that does not qualitatively alters the auto-correlation function is the optimum to be picked.

The 4<sup>th</sup> equation, even though not necessary for following the system's dynamics, ensures that the time is scaled correctly. This is essential for sampling the canonical

ensemble. It has been shown by Hoover [58] that the equations 39 are unique. Moreover, the Nosé-Hoover EOM contain a velocity-dependent force-term and are therefore not symplectic. This is caused by the conservation of the extended Hamiltonian  $\mathcal{H}_{NH}(\mathbf{r}^N, \mathbf{p}^N, p_s, s)$  instead of the original one.

### Andersen Barostat

Andersen's [56] variant will be introduced here, although the Parinello-Rahman [59] barostat, which also allows for anisotropic box-scaling, is used in this thesis. The basic concept is the same, but the Andersen thermostat does not necessitate the separate handling of each dimension. Additionally it is reformulated from Lagrangian to Hamiltonian notation in similarity to the Nosé-Hoover thermostat for better comparability.

To simplify rescaling of the simulation box, it is reasonable to switch to scaled coordinates when coupling a barostat to the system. Scaled coordinates and scaled moment are defined as:

$$\begin{aligned}\mathbf{s}_i &= \frac{\mathbf{r}_i}{V^{\frac{1}{3}}} \\ \pi_i &= \mathbf{p}_i V^{\frac{1}{3}}.\end{aligned}\tag{40}$$

For the barostat the volume  $V$  itself becomes a degree of freedom and also a corresponding momentum  $\pi_V$  has to be added to get the Andersen Hamiltonian defined by

$$\mathcal{H}_A(\mathbf{r}^N, \pi^N, \pi_V, V) = \sum_{i=1}^N \frac{\pi_i^2 V^{\frac{2}{3}}}{2m_i} + \mathcal{V}(V^{\frac{1}{3}} \mathbf{s}^N) + \frac{\pi_V^2}{2W} + pV.\tag{41}$$

Here  $W$  represents the mass of an imaginary piston with potential energy  $pV$ , that affects the volume of the simulation box. Again  $p$  is the pressure, not to be confused with the momentum  $\mathbf{p}$  of a particle.

The Andersen EOM in Hamiltonian notation are:

$$\begin{aligned}\frac{d\mathbf{r}_i}{dt} &= \frac{\mathbf{p}_i}{m_i} + \frac{1}{3} \frac{\dot{V}}{V} \mathbf{r}_i \\ \frac{d\mathbf{p}_i}{dt} &= \mathbf{F}_i - \frac{1}{3} \frac{\dot{V}}{V} \mathbf{p}_i \\ \frac{d\pi_V}{dt} &= \frac{1}{3V} \sum_{i=1}^N \left[ \frac{\mathbf{p}_i^2}{m_i} + \mathbf{F}_i \mathbf{r}_i \right] - p \\ \frac{dV}{dt} &= \frac{\pi_V}{W}\end{aligned}\tag{42}$$

Similar to the EOM of the Nosé-Hoover thermostat, the right side of the 3rd line in equation 42 can be interpreted as the difference of the instantaneous pressure  $p_{inst} = \sum_{i=1}^N \left[ \frac{\mathbf{p}_i^2}{m_i} + \mathbf{F}_i \mathbf{r}_i \right]$  and its canonical average  $p = \langle p_{inst} \rangle$ . The 4th equation couples the change of momentum  $\pi_V$  to the change of volume. Consequently, the system undergoes compression, if the instantaneous pressure is too high and expansion if it is too low. The barostat's reaction speed is controlled by  $W$  and like the thermostat it does not react instantaneously to deviation in the pressure, but by the derivative. In similarity to the Nosé-Hoover thermostat, the Andersen EOM contain a velocity-dependent forces  $\frac{1}{3} \frac{\dot{V}}{V} \mathbf{p}_i$  and are not symplectic.

### 2.3.4 Algorithm – Basic Structure of a Molecular Dynamics Simulation

Having discussed the basic concepts of a MD simulation in various ensembles, the basic structure of an MD simulation is depicted in algorithm 1. The algorithm can be separated in two major components. In the first part, the simulated system has to be initialized according to the provided input data. This initialization may include, among other tasks, determining the particle number  $N$ , generating or loading from file a starting configuration, setting up the simulation box, enabling boundary conditions, or removing centre-of-mass motion.

---

#### Algorithm 1: Basic MD structure

---

**Input:** simulation parameters:  $\Delta t$ , *integrator*,  $M$ ,  $N$

**Output:** system trajectory and thermodynamic averages

Initialization();

**for**  $M$  steps **do**

*forces* = **compute\_forces**();

*trajectory* += **integrate\_EOM**( $\Delta t$ , *forces*, *integrator*);

*averages* = **update\_averages**();

**end**

**return** *trajectory*, *averages*

---

The algorithm's second part primarily comprises of the for-loop over  $M$  time-steps of size  $\Delta t$ . For each iteration three essential tasks have to be performed. Firstly, computation of the forces acting on all particles is usually the computationally most expensive part of the simulation and determines the abstraction level or in other words accuracy of the simulation. In general, the computational complexity of calculating the forces for all  $N$  particles is of order  $\mathcal{O}(N^2)$ . Usage of an interaction-distance cutoff in combination with long-range corrections or neighbour-lists, that keep track of molecules close by over multiple time-steps, are two examples

for effective computational schemes that have been developed to reduce the computational effort of force calculation. To actually compute forces, the way in which the simulated particles interact has to be specified. Since force-field design is an essential part of this work, chapter 3 provides a more thorough analysis of this topic. Secondly, the positions of all particles have to be advanced by  $\Delta t$  by solving the EOM using a finite difference integration algorithm. The new positions are then added to the system's trajectory already containing all previous time-steps. Lastly, thermodynamic averages have to be updated according to the new data.

## 2.4 Software

### 2.4.1 GROMACS

Initially developed in 1991 at the University of Groningen, Netherlands, *GRO*ningen *MA*chine for *C*hemical *S*imulations (*GROMACS*) [31, 32] is one of the fastest MD software packages available to date. The software is open-source and released under the GNU General Public License. Due to its versatility in applying different forcefields and file-formats like AMBER, OPLS, or PDB, as well as a large variety of different algorithms and the availability of double precision accuracy, GROMACS is a popular MD package in physical chemistry and the simulation of molecular liquids. The core algorithms are written in C and most of its algorithms can be executed in parallel.

GROMCAS supports Drude oscillators (see subsection 3.1.2), using the SCF method and allowing to set the accuracy of the energy minimization for the Drude particles. Due to the SCF method, simulations using Drude particles cannot be performed in parallel. Other significant features of GROMACS used in this thesis are PME long-range correction for Coulomb and Lennard-Jones, and temperature and pressure coupling using the Nosé-Hoover and Parinello-Rahman methods.

Additionally, the software provides a large selection of analysis tools, for instance for calculating thermodynamic averages and radial distribution functions and is also equipped with a variety of tools to modify topologies. A detailed documentation can be found at <http://manual.gromacs.org/documentation/>. This thesis uses the latest version GROMACS 2016.4. released on September 15, 2017. GROMACS 2018 version is currently available as a beta-release.

## 3. FORCE-FIELD MODELLING

In this chapter the adaptation of existing hydrogen fluoride (HF) models for use in MD simulations and subsequent testing of the newly extended model are described. First, the theoretical background from chapter 2 is extended to basic terminology of force field modelling (section 3.1). Challenges for developing accurate molecular models with a focus on empirical force-fields and the inclusion of polarizability into these model are discussed and therefore can be skipped by the experienced reader. Then, a historical overview of empirical models for HF is presented in section 3.2, followed by a detailed introduction of the models used as a base for this thesis. Finally, the adaptations of those models, by using classical Drude oscillators, are introduced and subsequently the simulation results for a system of liquid HF in a bulk system with the adapted models are evaluated (see sections 3.3 and 3.4). If not explicitly stated, the theoretical sections of this chapter are based on book references [41], [42], and [43].

### 3.1 Terminology & Theory – Modelling

When, modelling the interaction of particles on the microscopic level, it is often necessary to balance accuracy and applicability. Although a description of these interactions should in general be based on the fundamental principles of quantum mechanics, this approach, often called *ab initio* simulation, is highly demanding even for small systems and therefore not widely applicable for the study of most systems. Computational biology, a typical field of application for MD simulations, illustrates this: the folding of proteins might take up to several milliseconds to be completed. However, the number of molecules per protein is of the order of  $10^3$  and the solvent molecules have to be simulated as well. Consequently, if the folding process of a protein should be simulated, the resulting system needs to be extremely large. Computing the trajectory of such a system based on the principles of quantum mechanics is currently infeasible. Although the dynamics of a slow folding protein can be considered an extreme scenario, the same reasoning is true for much smaller systems. In the context of this thesis, several thousands of HF molecules, simulated over a span of multiple nanoseconds, are necessary to accurately compute properties of the liquid-gas interface. In recent years, quantum mechanic force-fields for HF have been successfully applied to systems of up to 300 molecules and a timespan of at most one nanosecond [28, 60]. Since the computational cost of a MD simulation does not scale linearly with system size, it becomes apparent that a simplification of the molecular interactions is necessary.

### 3.1.1 Empirical Potentials and Pairwise Additive Interactions

A typical approach, used to make fast MD simulations of large systems, is a classical representation of atoms or molecules as points of a given mass and charge that interact through forces evoked by a potential energy. In general, this potential energy  $\mathcal{V}$  can then be represented as the sum of  $n$ -body potentials  $v_n$ , depending only on the coordinates of these  $n$  atoms:

$$\mathcal{V} = \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j} v_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \dots \quad (43)$$

The first part  $v_1$  represents the effects of an external field, and the following terms represent atom-atom interactions. However, the computation of three-body potentials and other many-particle effects, summarized under the notion of non-pairwise-additive potentials, is inefficient due to the large computational cost of quantities involving triplets. Non-pairwise-additive interactions are rarely explicitly used for MD simulations.

While some potentials originate from classical physical laws like Coulomb's law, others have been developed to model a specific feature of the liquid and might not bear physical meaning. Often, many-body effects are partially included implicitly in the parametrization of the two-body potential. This approximation, called effective pair potential, results in remarkably good descriptions for most fluid systems, but has the disadvantage of only being applicable for a specific state point. Often two-body potentials are called empirical potentials to emphasize that they should not be taken too literally:

$$\mathcal{V} \approx \sum_i v_1(\mathbf{r}_i) + \sum_i \sum_{j>i} v_2^{eff}(\mathbf{r}_i, \mathbf{r}_j) \quad (44)$$

It is sensible to first try the simplest model possible, since often no direct physical interpretation of a model exists. Two potentials commonly used for MD simulations are the Coulomb interaction representing electrostatics and the Lennard-Jones (12-6) potential which is an empirical description of the interaction of pairs of neutral atoms or molecules. As they are being used throughout this thesis, the two interactions are discussed in more detail:

#### Coulomb

Classical electrostatics derive the two-body potential of a pair of point charges as (in *cgs*-units):

$$v^C(\mathbf{r}_i, \mathbf{r}_j) = \frac{q_i q_j}{r} \quad (45)$$

where  $\mathbf{r} = \mathbf{r}_j - \mathbf{r}_i$  is the separation of the two point charges. The decay of this potential of  $\frac{1}{r}$  is rather slow, resulting in a large error when calculating forces using a small cutoff. Therefore the use of long-range correction techniques is inevitable for liquids with strong Coulomb interactions.

### Lennard-Jones

A simple way of describing the interaction of uncharged particles has been introduced by John Lennard-Jones as early as 1924 [61]. The Lennard-Jones Potential (LJ) combines a long-range attractive part with a short range repulsion and can be parametrized by just two parameters  $\sigma$  and  $\epsilon$ :

$$v^{LJ} = 4\epsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] \quad (46)$$

The attractive  $\frac{1}{r^6}$ -term is modelled after the London-dispersion-forces, arising from induced instantaneous polarization. The repulsive  $\frac{1}{r^{12}}$ -term describes repulsion due to overlapping at distances shorter than  $\sigma$ . The (...)  $^{12}$  term is chosen for convenience and simulation speedup since it is simply the square of the attractive part and is not physically based. The two power functions of order 6 and 12 give the potential its categorization name (12-6). The parameter  $\epsilon$  is the depth of the potential well.

A commonly used alternative is the Buckingham or (exp-6) potential [62], where the repulsive part is modelled as an exponential function. This has the benefit of a "softer" repulsion since, in contrast to LJ, it does not diverge for two atoms closely approaching each other. Historically, usage of the LJ potential was preferred since the calculation of an exponential function provided additional computational cost. Although no longer an issue today, the LJ potential is still used more frequently.

### 3.1.2 Molecular Interactions and Site-Site Force-Fields

In the case of molecular liquids it is also necessary to consider interactions of atoms within the molecule. It is common to treat intra-molecular bonds as rigid or semi-rigid with fixed bond length and bond and torsion angles, to save computational cost since classical treatment of high frequency oscillations would be difficult. The low amplitudes of bond vibrations usually result in only little effect on the dynamics, justifying this simplification.

Interaction of a molecular pair contains contributions from different parts of the molecule. This is often modelled using the site-site approximation. There, each molecule is represented as a collection of sites connected by rigid or semi-rigid bonds as illustrated in figure 4. Each site can be the centre of a specific potential, for instance of a LJ potential, parametrized according to the atom type. The total

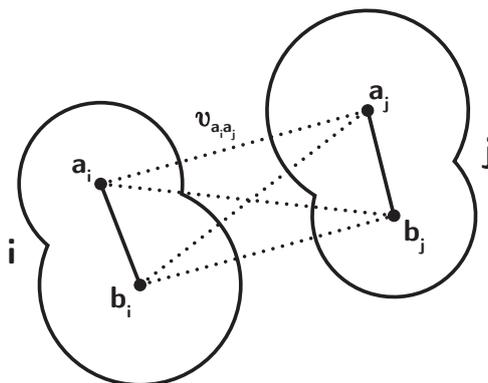


Figure 4: Site-Site approximation of the interaction of two multi-atomic molecules. Points represent interaction sites of a molecules and lines indicate interaction of the two connected sites. Only interactions of sites of different molecules are considered.

interaction of the pair of molecules is the sum of pairwise additive contributions from all distinct sites in the molecules.

In general, sites are located at the positions of atomic nuclei. However, an improvement that is rather efficient is the introduction of additional sites that do not correspond to an atomic position. These so called virtual sites can for instance bear partial charges to multipole-moments of the molecule. Again, the number of additional sites and hence the improvement of the model has to be balanced with the increase in computational cost.

The collection of the location of sites for a molecule and the parametrization of their interaction potentials is usually referred to as the molecule's force-field. Important properties of such force-fields are their applicability to various state-points and situations often referred to as transferability. The improvement as well as the comparison in terms of efficiency and accuracy of different forcefields remains a large field of interest in computational physics and a large variety of different approaches and methods for developing and extending force-fields are present.

### **Polarizability**

Whenever a model proves to be inaccurate, fails to describe specific fluid properties, or is limited in its transferability, an improvement to the model is desirable. A typical approach for doing so is the explicit inclusion of non-additive terms in the force-field.

For fluids composed of molecules with large permanent dipole moments, strong local electric fields can be present without an external field. This results in changes to the charge distribution of a single molecule and can strongly affect molecular interaction. The measure determining the strength of a molecule's reaction to such

a field  $\mathbf{E}$  is called polarizability  $\alpha$ . The change of the charge distribution causes an induced dipole moment  $\mu^{ind}$  defined by:

$$\mu^{ind} = \alpha \mathbf{E} \quad (47)$$

Although not SI,  $\alpha$  is often expressed in  $\text{\AA}^3$  and for an anisotropic medium it is a rank two tensor. This thesis solely focuses on isotropic polarizability, hence  $\alpha$  is a scalar if not explicitly stated otherwise. Due to the induced dipole's dependence on the local electric field, which is in turn generated by the permanent and induced dipole moments of the surrounding molecules, polarizability is a typical example of a non-additive interaction term.

Explicitly including the polarizability into the force-field has proven to be a successful measure for improving empirical forcefields for water [63–65] and HF [29, 30, 66]. A drastic improvement of transferability is often the main improvement of these models. Different strategies to describe polarizability have been successfully used, but not all of them are implemented for every simulation software. Point-dipoles [67], classical Drude oscillators [40], and fluctuating charges [68] are three predominant methods used. The fluctuating charge approach, where charges are allowed to be redistributed between sites at every step, is conceptually different to the two other approaches [69–71]. Therefore, only point-dipoles and Drude oscillator are covered in this thesis.

**Point-Dipole** An induced mathematical point-dipole [67] attached to a single site, characterizes an induced dipole moment  $\mu^{ind}$  expressed as:

$$\mu^{ind} = \alpha_i \left( \mathbf{E}_i^q + \mathbf{E}_i^{ind} \right), \quad (48)$$

where  $\alpha_i$  is the isotropic polarizability,  $\mathbf{E}_i^q$  the electrostatic field created by the charge distribution, and  $\mathbf{E}_i^{ind}$  the field created by the induced dipoles of other atoms acting on atom  $i$ . Since the induced dipoles depend on all other induced dipoles, an iterative procedure, performed until all induced dipoles converge and self-consistency has been reached, is necessary. This approach implies that the implementation of electrostatics has to be augmented to not only include point charges but also dipoles, which has the drawback of hindering development of faster algorithms.

**Drude Oscillator** Drude oscillator (DO) models, alternatively referred to as shell models, represent polarizability by attaching charged but massless particles to a polarizable site using a harmonic spring. This approach has first been used by and consequently named after *Paul Drude* who described it as a technique to model dispersive properties of materials in his book *The Theory of Optics* [40] published

in 1902. The Drude particle, representing the polarizability  $\alpha$  with charge  $q_D$ , is harmonically bound with force constant  $k_D$ . By changing the charge on the site the Drude particle is attached to from  $q$  to  $q - q_D$  the pair maintains an average charge of  $q$  in absence of a field. The Drude particle is oscillating around the atom's position in this case. With an electric field  $\mathbf{E}$ , the Drude particle's displacement  $\mathbf{d}$  and the average induced dipole moment  $\mu^{ind}$  generated by the Drude particle can be written as

$$\mathbf{d} = q_D \frac{\mathbf{E}}{k_D} \quad (49)$$

$$\mu^{ind} = q_D^2 \frac{\mathbf{E}}{k_D} \quad (50)$$

These expressions can be used to parametrize the model for any given polarizability  $\alpha$ , while ensuring that the displacement  $\mathbf{d}$  remains small and the dipole moment comparable to a point-dipole. This likewise ensures that the displacement is smaller than any other inter-atomic distance to avoid artefacts. For any  $\alpha$  the parametrization has to follow the expression:

$$\alpha = \frac{q_D^2}{k_D} \quad (51)$$

In similarity to the point-dipole a DO is usually solved using the Born-Oppenheimer approximation and implementing a self-consistent field (SCF) condition. The Drude particles are relaxed to the position of minimal potential energy using an iterative procedure. The equation defining this minimum can be represented by the force equilibrium of the Drude particle:

$$k_D \mathbf{d} - q_D (\mathbf{E}^q + \mathbf{E}^{ind}) = \mathbf{0} \quad (52)$$

where  $\mathbf{E}^q$  is the electric field of all fixed charges and  $\mathbf{E}^{ind}$  the field generated by all other induced dipoles in the system. The most appealing aspect of this model is its preservation of the Coulomb interaction as the sole interaction necessary to model polarizability of a molecule. This simplification allows a straightforward implementation of the model for most standard MD software packages as well as its compatibility with more advanced algorithms like PME and temperature/pressure coupling [72].

The use of the SCF method for a DO is particularly inefficient, since an iterative energy minimization is necessary at every time step, This reduces the possibility of efficient parallelization [69, 73, 74]. In principle there are faster computational schemes to solve the motion of a Drude particle, however the speedup is usually generated by an estimation of the motion and consequently less accurate than with SCF [75]. However, those alternative methods are not implemented in

the current GROMACS version. The GROMACS 2016.4 implementation, used throughout this thesis, does not support parallel execution of the Drude SCF iteration. From a theoretical standpoint point-dipoles are a limit case of a DO with a stiff, infinitesimally short bond. As a result, DO and point-dipole models are fundamentally equivalent and their results are expected to be qualitatively comparable [21, 69, 73, 75, 76]. This theoretical prediction is also tested for the DO models of HF used in this thesis, which is covered in the next section.

Once all induced dipoles are calculated using one of the above techniques, the polarization energy can be expressed by

$$\mathcal{V}_{pol} = -\frac{1}{2} \sum_i \mu_i \mathbf{E}_i, \quad (53)$$

and the total potential energy of a system without an external field is

$$\mathcal{V} = \mathcal{V}_{pair} + \mathcal{V}_{pol}. \quad (54)$$

In general adding polarizability to a model of a liquid results in the loss of long range structure, caused by an effective screening of electrostatic interactions by the induced dipoles, and moreover, the looser structure results in faster dynamics [21].

## 3.2 Adaptation of Existing Hydrogen Fluoride Models

As a consequence of the highly corrosive and toxic nature of HF only scarce experimental data concerning its structure and thermodynamic properties are available, especially when compared with other molecular fluids. Consequently, the simulation and development of computational models for HF is of special interest for generating thermodynamic properties of the liquid across a wide range of thermodynamic states. Moreover, the simple composition of the molecule of only two atoms and the large polarity has motivated a large number of computer simulation studies. However, the lack of extensive experimental data provided large difficulties for model development. In the following subsection 3.2.1 an overview over existing models for HF, their evolution, and applicability for certain situations is given. Their adaptation of existing models for usage in this thesis is described in subsection 3.2.2. Finally, in subsection 3.3.1, a full thermodynamic and structural comparison of the adapted model with its predecessor are presented.

### 3.2.1 Existing Empirical Force-Fields for Hydrogen Fluoride

The first computational models for simulating liquid HF date back to 1978. *Jorgensen & Cournoyer* [17, 19] used *ab initio* Hartree-Fock theory [77, 78] to compute the binding energies of rigid HF dimers and fit an interaction potential to the calculated results. *Klein & McDonald* [18, 20] parametrized a potential for a flexible two site model by reproducing *ab initio* energies, but solely used inverse power- and exponential-functions for the fit. In 1984 *Cournoyer & Jorgensen* [5] developed a three-site model called HFC that, instead of using complicated functional forms, modelled the non-Coulomb interaction part with a simple Lennard-Jones potential centred at the fluorine site. In this three-site model, the sites of H and F each carry equal positive charge and twice the negative charge is placed in between.

The structure of the two- and three-site models turned out to be rather different, when assessed via the RDF. However, at the time of implementation no experimental data were available. The first neutron diffraction experiment on liquid HF was only reported in 1985 by *Deraman et al.* [11]. The results of this experiment clearly showed that only three-site models can effectively reproduce the structural properties of liquid HF [29, 35, 66]. The inclusion of a third site allowed for accurate modelling of the quadrupole moment of HF and is essential for the correct angle of the H-bond. Since hydrogen bonding has a dominant influence on the structure of liquid HF, a correct bond angle is indispensable and most models developed later are based on three sites [24, 25, 29, 79–81]. Two-site models developed after 1984 [4, 22, 23, 36, 37] usually focused on the study of only a specific aspect of liquid HF instead of the accurate and transferable modelling of HF. A comparison of experimental data with the HFC model showed good agreement and due to the lack of further experimental data no improvements over the HFC model were proposed for over ten years.

After the development of the Reverse Monte Carlo Method [82] and its successful application for deriving arrangements for three-dimensional models from experimentally measured structure-factors, a reparametrization of the HFC model was performed by *Jedlovsky & Vallauri* [79] in 1997. The new model incorporated long-range corrections for the Coulomb interaction in the parametrization using the RF method. Although reaching higher accuracy than any previous model and providing good agreement with *ab initio* simulations [83], the model could not reproduce the temperature dependence of the density of liquid HF and failed to model the HF dimer accurately. A polarizable model called JVP [29] (see subsection 3.2.1) was introduced the same year to account for the importance of cooperative effects in the fluid and further improve the model. It included a point-dipole located at the position of the fluorine atom to explicitly incorporate the molecule's electric polarizability. Although computationally more expensive due to the inclusion of many-body interactions, the model covered a wide range of

thermodynamic states with reasonable accuracy and slightly improved agreement with experimental RDFs. Moreover, it also provided a good representation of the isolated dimer comparable to experimental [84] and *ab initio* [85, 86] data.

*Ab initio* based models [17–20, 80, 83] are parametrized to fit dimer- and oligomer-properties. Therefore, they are generally inferior to effective pair potential models [5, 29, 29] in reproducing the properties of the liquid phase. Consequently, polarizable models are superior for modelling liquid-vapour coexistence [23, 35]. Further structural experiments performed in 2000 by *Pfleiderer et al.* [1] covered a broader range of thermodynamic states and could further manifest the superiority of the JVP model. Additionally, it showed the model's capability to qualitatively reproduce the experimentally observed elongation of the hydrogen bond when transitioning from the liquid to the gas phase [29]. Attempts to surpass the accuracy and transferability of the JVP model using the experimental data provided by *Pfleiderer et al.* [1], as well as the application of Quantum Mechanics/Molecular Mechanics [22–28], have not been successful.

In 2004, more extensive experimental studies, performed by *McLain et al.* [2, 3] were reported. Measuring partial structure factors and computing the three atom-atom partial RDF [3], the new experimental data allow for a more thorough assessment and validation of structural properties for simulations of liquid HF. However, the experiment also provided clear evidence that the length of the HF bond in the JVP model, set to the experimental gas-phase value measured by *Janzen & Bartell* in 1969 [87], is too large [2]. These findings inspired the reparametrization of various HF-models [4, 22, 25, 27, 28, 30]. Even after reparametrization, non-polarizable two-site [22] and three-site [4] models could not match the polarizable models' accuracy. Also, neither *ab initio* [22] nor Quantum Mechanics/Molecular Mechanics [27, 28] were able to reach the same transferability and usability for simulating liquid HF and liquid-vapour coexistence as the JVP model's reparametrization, introduced as PJVP by *Pártay, Jedlowszky & Vallauri* in 2006 [30]. Like its predecessor the PJVP model incorporated polarizability using a point-dipole. However, the length of the HF bond was significantly shortened and all other parameters were recalculated in accordance to the original procedure of the JVP model.

Apart from the accurate modelling of thermodynamic and structural properties of liquid HF by PJVP and JVP, their usability to simulate liquid-vapour coexistence is of great benefit and importance for this thesis, hence these two provide the basis for the models used here. Before the adapted model is introduced in section 3.2.2, a more detailed description of the two point-dipole models is given.

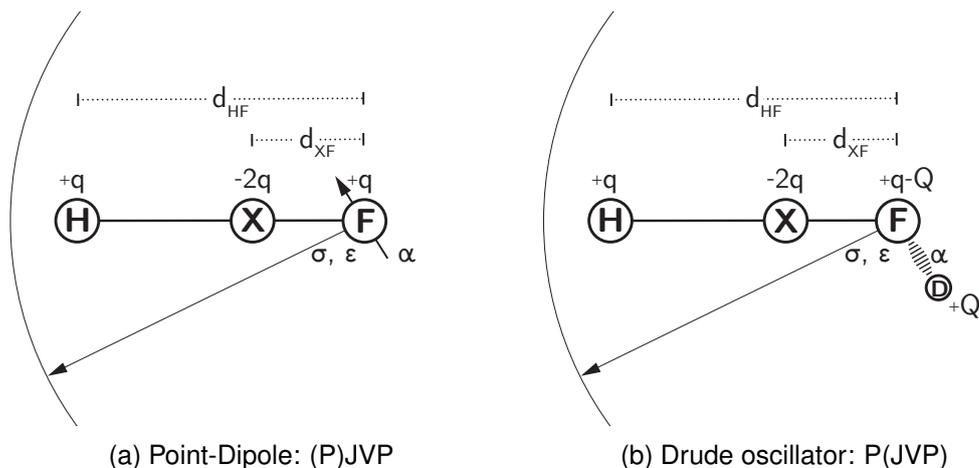


Figure 5: Schematic representation of the geometry of the a) point-dipole and b) Drude oscillator models.

## JVP

Developed in 1997 as an improvement over their previous non-polarizable model [79] the JVP model by *Jedlovsky & Vallauri* [29] is the first HF model to explicitly incorporate polarizability and thereby multi-body effects. As illustrated in figure 5 the potential model is composed of three charge interaction sites in linear alignment.

For description of the Coulomb interaction the H and F sites each carry a positive fractional charge  $+q$  and a charge of  $-2q$  is placed on the auxiliary site X to maintain the molecule's overall electric neutrality. The length of the H-F bond  $d_{HF}$  is set to the gas-phase experimental value of 0.0973 nm determined by *Janzen and Bartell* in 1969 [87]. Reverse Monte Carlo technique [82] was used to determine the value of  $q$  and the position of the X site, such that the experimental values for the dipole moment of 1.83 D [6] and quadrupole moment of 1.83 B [88] of the HF molecule are reproduced by the model. As a result, the X site is placed along the H-F bond at a distance  $d_{XF} = 0.01647$  nm from the F site and a charge  $q = 0.592 e$  is used. Additionally, an induced point-dipole moment located at the F site of the model accounts for the molecule's electric polarizability. The induced dipole is represented by the scalar polarizability  $\alpha$ , which is set to the experimental value for HF of  $0.00083 \text{ nm}^3$  [88]. The non-Coulomb interaction part is modelled using a single Lennard-Jones interaction centre located at the F site. The two parameters  $\sigma = 0.305$  nm and  $\epsilon/k_B = 110$  K have been determined by a series of isothermal-isobaric Monte Carlo simulations to fit the values of internal energy and density at 273 K and  $1 \cdot 10^5$  Pa. Long range corrections have been taken into account via the reaction-field method for these simulations, which has not been the case in preceding models [5, 29]. Model parameters are summarized in table 1.

Table 1: Parameters of point-dipole and Drude oscillator (-D) models

	$d_{HF}$ [nm]	$d_{XF}$ [nm]	$q$ [ $e$ ]	$\sigma$ [nm]	$(\epsilon/k_B)$ [K]	$\alpha$ [nm <sup>3</sup> ]
JVP(-D)	0.0973	0.01647	0.592	0.3050	110.0	0.00083
PJVP(-D)	0.0930	0.01741	0.655	0.3035	105.0	0.00083

In their studies *Jedlovsky & Vallauri* used the model to perform Monte-Carlo simulations of liquid and supercritical HF at a large number of state points [29, 66]. Comparison with different models and experimental results proved, that the inclusion of polarizability into the empirical potential is the superior approach for modelling liquid and supercritical HF. Besides, it provides results for the HF dimer comparable with the accuracy of *ab initio* simulations. Since non-polarizable models are parametrized to account for the molecule's polarization at one specific state point only, JVP was the first model to explicitly account for the change of the molecular dipole moment with varying thermodynamic conditions. Consequently, the elongation of the hydrogen bonds with decreasing density as measured by *Pfleiderer* [1] and the characteristic liquid-vapour coexistence curve are accurately reproduced by the polarizable JVP model [25, 66]. Despite the success in describing the properties of HF in a broad range of thermodynamic states, including liquid-vapour coexistence, significant discrepancies between the JVP model and experimental data for HF persist. The first peaks of the RDF function is shifted to larger distances compared with experimental and *ab initio* results, indicating overestimation of the molecules' close-contact separation. Furthermore, a steeper minimum and broader third peak indicate overestimation of the ordering of distant hydrogen-bonded neighbours. The density of the liquid phase is consistently underestimated by the JVP model, and it has been shown that it is only weakly correlated to the system's internal energy and therefore extremely sensitive to the selection of LJ parameters [30, 66].

## PJVP

Previous studies did not provide consistent values for the bond length of liquid HF [1, 11, 24]. Then, extensive experimental studies of liquid HF published in 2004 by *McLain et al.* [2, 3] provided a value of the bond length consistent with *Pfleiderer et al.* [1] motivating a reparametrization of the JVP model and ample data for its validation. Although the model's main geometry remains unchanged (see figure 5), the HF bond length  $d_{HF}$  was changed to the liquid phase equilibrium value of 0.093 nm [2]. Consequently, all other parameters were recalculated according to the same procedure used for the JVP model [30]. The parameters for the two models are collected in table 1.

A comparison of the reparametrized model PJVP and its predecessor JVP performing MC simulations on various liquid and supercritical state points only revealed significant performance on certain aspects. Thermodynamic properties of the liquid phase improved noticeably. However, the RDFs and therefore the liquid's structure could not be distinctively enhanced by the reparametrization. The calculated data for the coexisting densities of the liquid phase still underestimate the experimental values along the entire coexistence curve, but the discrepancy reduced noticeably with the PJVP model. Besides, the overestimation of molecules' close-contact separation, already present in the JVP model, could not be resolved by the reparametrization. A comparison of the LJ parameters  $\sigma$  and  $\epsilon$  for the two models (see table 1) reveals only a slight modification of their values. Consequently, a substantial improvement of the model might not be achievable by readjusting the parameters, but by changes to the functional form of the interaction potential. *Pártay et al.* suggest that the majority of model's deficits might be related to a drawback of the simplicity of the LJ interaction. Since the repulsive part of this potential is not physically based, it might not be soft enough to allow a close approach of the molecules. A softer short-range repulsion for non-Coulomb interactions might therefore increase the density of the coexisting liquid and provide significant improvements of the model's structural properties [30]. Another possible improvement facilitated by such a change might be the reduced sensitivity of the density with respect to the parametrization.

### 3.2.2 Model Adaptation: Polarizable Drude Oscillators Model

#### The Model

In their respective studies the JVP and PJVP models were used for MC simulations [29, 30]. To determine the magnitude of each induced point-dipole, an iterative procedure was performed for all dipoles in the system until self consistency was reached. This method is especially applicable for MC simulations using single particle moves since the system is perturbed little with each move. Thereby the SCF procedure converges rapidly and few iterations are necessary [30]. However, the usage of induced dipoles necessitates the additional treatment of dipoles that exceeds the interaction of point charges [89, 90]. Popular MD simulation packages like GROMACS, NAMD [33], CHARMM [91], and DL\_POLY [92] are based on point-charge interactions and often do not offer extensions for dipole treatment. Therefore, classical DOs or variations thereof are often used to incorporate polarizability into an MD simulation [73, 75]. The method has been successfully applied for polarizable force-fields of water, acetone, methanol, and ionic liquids [65, 70, 93–98].

The models used for this thesis are modifications of the JVP and PJVP models, which are adapted for usage in simulations with modern MD packages. This necessitated the replacement of the induced dipole with a DO. They will henceforth be referred to as JVP-Drude (JVP-D) and PJVP-Drude (PJVP-D). Figure 5 shows a schematic representation of the adapted potential models featuring an additional interaction site D that is attached to the fluorine atom, instead of the point-dipole.

A charge  $q_D$  is placed on the Drude site D and the charge on the F site is adapted to maintain overall electric neutrality. The Drude charge is attached with a harmonic spring potential characterized by the force constant  $k_D$ . In combination with the fluorine the Drude site forms an oscillating physical dipole [21, 98]. Any pair of parameters for the Drude particle  $\{q_D, k_D\}$  has to fulfil the condition

$$\alpha = \frac{q_D^2}{k_D}. \quad (55)$$

Although geometric and LJ parameters are not changed for the adapted models, there is ambiguity regarding the choice of the Drude parameters [63, 73, 99]. For this study, the Drude charge has been chosen as  $q_D = 4.0 e$  (for detail see subsection 3.2.3).

### 3.2.3 (P)JVP-D Drude Particle Parametrization

In simulation packages polarizability  $\alpha$  and Drude charge  $q_D$  have to be specified and the spring constant  $k_D$  is calculated internally by equation 55. Since the value for  $\alpha$  is provided from experimental data, only the value for the charge  $q_D$  needs to be specified. Equation 56 illustrates that the displacement  $\mathbf{d}$  of the D site is proportional to the local electric field  $\mathbf{E}$ .

$$\mathbf{d} = \alpha \frac{\mathbf{E}}{q_D} \quad (56)$$

The displacement is indirectly proportional to the charge's magnitude, reiterating the importance of the Drude charge parametrization [70, 98].

Several studies suggest that the point-dipoles and classical DO approach produce equivalent results [21, 69, 72, 73, 75, 76]. Nonetheless, a physically sensible range for  $q_D$  exists and structural stability also limits the choice of the Drude charge [63, 99]. Additionally, finite precision of computers, as well as the application of long-range corrections can have a large influence on the optimal value for  $q_D$ . Moreover, restrictions for optimal parameters depend heavily on the simulated system and can significantly change for models with different geometries [73].

Parametrization was performed by evaluating the polarization energy  $\mathcal{V}_{pol}$  of a single configuration for a series of different Drude charges  $q_D$  of the JVP-D model. Both a two-molecule system and an equilibrated bulk configuration with 1000 molecules at ambient conditions were tested. Polarization energy was calculated within GROMACS simulation package version 2016.4. The SCF minimization was stopped when the maximum force surpasses a threshold value of  $1 \cdot 10^{-5} \text{ kJ mol}^{-1} \text{ nm}^{-1}$ , providing the same accuracy for the dipole moment as the point-dipole SCF iteration in reference [29]. Both JVP and PJVP, used as a basis for adaptation, have a point-dipole with equal  $\alpha$  and similar geometry. Consequently, the parametrization of the DO suitable for the JVP-D model is likewise appropriate for the PJVP-D model.

As depicted in figure 6, the polarization energy of the equilibrium configuration diverges for small  $q_D$ . The same is true for the two-molecule system. In accordance with equation 56 small Drude charges result in a considerable displacement and fail to accurately model a point-dipole [21]. Contrarily, the polarization energy converges to the same value for large Drude charges  $q_D$ ; both positive and negative. Furthermore, the polarization energy  $\mathcal{V}_{pol}$  is antisymmetric with respect to the charge's sign, but the sign is merely a consequence of the molecular alignment in the system. For instance, the two-molecule and equilibrium bulk systems show contrary behaviour, likely caused by the arbitrary, parallel arrangement of the two molecules in the small system.

Finite accuracy of computer simulations and constraints of algorithms present additional limitations. The displacement of site D from site F has to exceed a minimum distance to comply with the cutoff for PME long-rang corrections and the system's accuracy. However, a Drude charge  $q_D$  of comparable magnitude as the model's partial charges  $q$  is beneficial with respect to accuracy.

For comparison, the polarization energy  $\mathcal{V}_{pol}$  for the two-molecule system was calculated for the point-dipole model JVP, using the iterative procedure described in the initial study [29]. The convergence limit for the DO's polarization energy and the value computed for the point-dipole differ by 0.15%. Since the SCF minimization enforces a change in dipole moment of less than 0.1% per individual molecule, the point-dipole and the convergence limit of the DO match. Accordingly, the optimal choice for replacing a point-dipole with a DO corresponds to the smallest absolute value  $|q_D|$  for the charge, with  $\mathcal{V}_{pol}$  close to the convergence limit of large charges. This contradicts the proposition, that a negative Drude charge better represents the electronic degrees of freedom [63]. However, for a system of multiple molecules not only accurate representation of the dipole, but also structural stability of the model has to be guaranteed [99]. Other molecules with partial charges of opposite sign as the Drude charge  $q_D$  must not get too close to the Drude site D. Otherwise, the electrostatic attraction can exceed the LJ-repulsion of the molecule and the DO can get trapped inside another molecule.

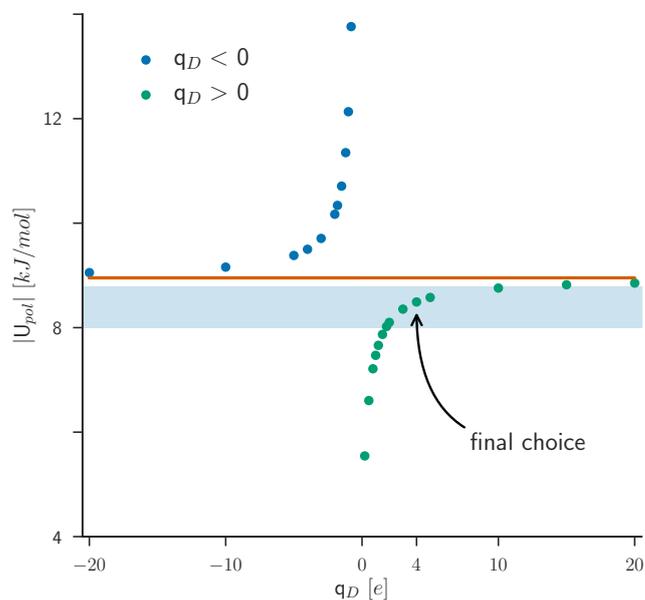


Figure 6: Polarization energy  $\mathcal{V}_{pol}$  of a single equilibrium bulk configuration of 1000 HF molecules for different values of the Drude charge  $q_D$ . The horizontal line represents the convergence limit for charges with large absolute values. The shaded blue band marks the confidence interval (1 standard deviation (SD)) for the polarization energy  $\mathcal{V}_{pol}$  of the point-dipole model [29]. The final choice for the two model is also highlighted.

H-bonding is the most common structural alignment in HF models, despite the spherical appearance of the models ( $\sigma \gg d_{HF}$ ). By definition the charge on the D site is not interacting with the charge on the F site it is attached to. Due to the negative partial charge on the X site, the F site represents the negative end of the permanent dipole of HF. Therefore, a negative Drude charge will on average be displaced in the direction of the positive H site that is hydrogen bonded to the Drude pair's F site. Contrarily, a positive Drude charge will distance itself from the H-bonded H site and towards the molecule's own H site. It thereby repulsively interacts with the negative end of another molecule bonded to this H site further confining the Drude site's position to the interior of the LJ-sphere. Since the arrangement closer to the centre of the LJ-sphere guarantees stability, a positive Drude charge  $q_D$  is the sensible choice for HF models. Consequently, the JVP-D and PJVP-D models were parametrized with a Drude charge of  $q_D = 4.0 e$ . This value corresponds to a deviation of the polarization energy of only 0.05% from the convergence limit.

### 3.3 Simulation of Adapted Models

For extensive comparison of the adapted DO-models with its point-dipole counterpart, MD simulations of liquid HF have been performed. The results

of the JVP study [29] provide the separate energy contributions of polarization energy  $\mathcal{V}_{pol}$ , Coulomb energy  $\mathcal{V}_{qq}$ , and LJ energy  $\mathcal{V}_{lj}$ . The additional information about these quantities allows for a more extensive comparison, therefore the simulations were performed at the conditions employed there. Due to the models' similarity, qualitative results for the JVP-D model also apply for the PJVP-D model. The simulations are restricted to a single state-point since the focus of this study is the agreement of point-dipole and DO-models instead of a complete evaluation [29, 30, 66] of the two models.

### 3.3.1 Liquid Hydrogen Fluoride in the Isothermal-Isobaric Ensemble

#### Computational Details

The MD simulations were performed in the isothermal-isobaric ensemble at temperature of 273 K and pressure of  $1 \cdot 10^5$  Pa. In analogy to references [29, 30], this range covers HF in its liquid state. 1000 HF molecules were placed in a cubic box of size L and regular PBC were enforced. The average temperature was controlled using a Nosé-Hoover thermostat [57, 58] with a relaxation constant of 0.5 ps and the average pressure using an isotropic variant of the Parrinello-Rahman barostat [59] with a relaxation constant of 2 ps and a compressibility of  $5.0 \cdot 10^{-10}$  Pa<sup>-1</sup>. The EOM were integrated every 1 fs using a leapfrog algorithm [49] and particle positions and energies were saved every 1 ps. Moreover, a SCF threshold of  $1 \cdot 10^{-5}$  kJ mol<sup>-1</sup> nm per Drude pair was used. This threshold corresponds to a minimization of the force exerted on the spring and double precision accuracy of the simulation is necessary for this purpose.

An interaction cutoff of 0.87 nm was used for both Coulomb and LJ interaction, however two different long range correction strategies were applied. Although the RF method was used for the initial MC studies, it is not applicable for simulating systems at liquid/vapour coexistence as covered in the next chapter of this thesis. Therefore, the comparison of RF and PME corrections for HF simulations is of great importance since the latter is applicable for systems with a liquid/vapour interface. One setup applied the RF method with an infinite dielectric constant for the Coulomb interaction part and tail-correction for the LJ interaction. 4<sup>th</sup>-order PME for Coulomb and LJ-parts with an isotropic Fourier spacing of 0.09 nm was used for the second setup. Long-range dispersion corrections to energy and pressure were applied in both cases.

Equilibration was performed for 3 ns to ensure relaxation of potential energy and density to their equilibrium values. Thermodynamic averages and particle positions were computed and saved over a period of 1 ns. Simulations were performed with

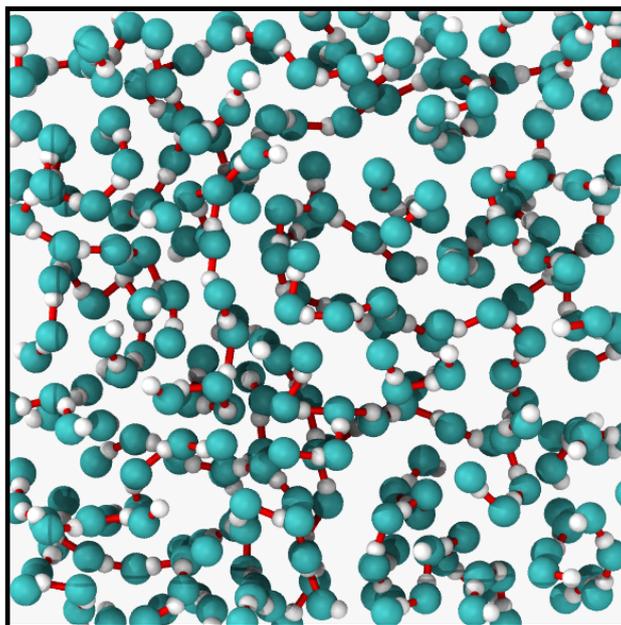


Figure 7: Volume slice ( $L \times L \times \frac{L}{3}$ ) of a single configuration of the equilibrated system at 273 K and  $1 \cdot 10^5$  Pa. Smaller spheres (white) represent hydrogen and larger spheres (turquoise) fluorine atoms. A solid line (red) indicate hydrogen bonded molecules, identified by a distance smaller than first minimum of  $g_{HF}$  at 0.236 nm. The black box represents the simulation box with PBC.

the GROMACS simulation package version 2016.4 [31, 32], compiled with double precision and using the Shell Molecular Dynamics component [75, 100].

### System Description

Figure 7 shows a volume slice of a single configuration of the simulation box. Only one third of the box is displayed to allow for a clearer perspective and the atoms are coloured according to their type. Additionally, lines connect molecular pairs, if their distance is smaller than the first minimum of the H-F partial radial distribution function  $g_{HF}$  (see subsection 3.3.1).

The molecule's tendency to form two bonds and align linearly with only little loose ends and few branches is clearly visible by this representation. Caused by the strong permanent dipole moment of the molecule, the formation of strong hydrogen bonds between the positive end at the H site and the negative end at the F site of two molecules is favourable. Compared with the non-polarizable models the molecules are far less likely to form three or more bonds which has been determined as a main weakness of these models [66]. A result of this mostly linear alignment are fluctuations of the local density that cannot be observed for non-polar liquids like a simple LJ-fluid.

## Thermodynamic Results

The potential energy  $\mathcal{V}$  and density  $\rho$  of the DO models for RF and PME are summarized and compared with experimental data, and data from the two initial studies of the point-dipole models [29, 30] in tables 2 and 3, respectively. Additionally, the different contributions to the total potential energy are also listed, although reference data are only available for the JVP model. In general, data from reference [30] are favoured due to higher accuracy. The Drude charge's contributions to the overall Coulomb energy  $\mathcal{V}_{qq}$  hinder direct comparison with Coulomb energy of the point-dipole models, where point-dipoles energy is treated separately. To account for this difference  $\mathcal{V}_{qq}^*$  is defined and computed as the electrostatic energy of the H, F and X site only. (In GROMACS  $\mathcal{V}_{qq}^*$  can be calculated by splitting the F site into two overlapping sites (rigidly bound with length zero) with charges  $q$  and  $-q_D$  and the use of energy groups). For comparability with the initial JVP study  $\mathcal{V}_{qq}^*$  is reported for the DO-models instead of  $\mathcal{V}_{qq}$ . Consequently, the three contributions of  $\mathcal{V}_{qq}^*$ ,  $\mathcal{V}_{lj}$  and  $\mathcal{V}_{pol}$  do not exactly add up to  $\mathcal{V}$  for the DO-models. A similar discrepancy for the point-dipole model ( $\mathcal{V} \neq \mathcal{V}_{qq}^* + \mathcal{V}_{lj} + \mathcal{V}_{pol}$ ) originates from the different number of samples used for each contribution [29].

The potential energy  $\mathcal{V}$  of the DO-models is 6% (PME) or 6% (RF) higher than for JVP and 6% (PME) or 6% (RF) higher than for PJVP, indicating worse performance compared with the experimental values. In accordance to DO models the potential energy decreases slightly from the JVP-D to the PJVP-D. This effect originates from the H-F-bond shortening and is even more pronounced for higher temperatures [30]. However, both long range correction strategies produce almost identical results for  $\mathcal{V}$ , indicating that both approaches are valid choices for the model.

Contrarily, the separate contributions to the potential energy do not reproduce the DO data (only for JVP) equally good. Moreover, the different long-range correction strategies show different trends regarding those contributions. For PME,  $\mathcal{V}_{qq}^*$  matches the DO data within its accuracy, but the LJ repulsion  $\mathcal{V}_{lj}$  is underestimated by about 16%. For RF, both the LJ and the Coulomb contribution do not match the JVP model. While the deviation of  $\mathcal{V}_{qq}^*$  is comparable to the deviation of the overall energy  $\mathcal{V}$ , the LJ repulsion  $\mathcal{V}_{lj}$  also deviate by 12%. Most importantly, the polarization energy  $\mathcal{V}_{pol}$  of DO and point-dipole model match within accuracy of the data. The separate contributions for the PJVP-D model show a decrease by 4%-10% for Coulomb attraction and polarization energy, compared to the JVP-D model. The LJ centres' non-Coulomb repulsion increases by roughly 30%.

As apparent from table 3, the MD simulations of the Drude model fail to reproduce the correct density. The PME approach is slightly closer to the experimental and MC densities. However, a large sensitivity of the density with respect to small fluctuation

Table 2: Total internal energy  $\mathcal{V}$ , separate contributions  $\mathcal{V}_{qq}^*$ ,  $\mathcal{V}_{lj}$ , and  $\mathcal{V}_{pol}$  of the point-dipole and Drude oscillator (-D) models, and experimental data (Expt.)

	PJVP-D (PME)	PJVP-D (RF)	PJVP <sup>a</sup>	JVP-D (PME)	JVP-D (RF)	JVP <sup>b</sup>	Expt. <sup>c</sup>
$\mathcal{V}$ [kJ mol <sup>-1</sup> ]	-26.877 ± 3e-3	-26.753 ± 5e-3	-28.53 ± 0.37	-26.199 ± 2e-3	-26.038 ± 3e-3	-27.68 ± 0.32	-29.01
$\mathcal{V}_{qq}^*$ [kJ mol <sup>-1</sup> ] <sup>d</sup>	-24.528 ± 5e-3	-23.044 ± 2e-3		-23.500 ± 5e-3	-22.009 ± 4e-3	-23.40 ± 0.40	
$\mathcal{V}_{lj}$ [kJ mol <sup>-1</sup> ]	3.864 ± 5e-3	4.109 ± 3e-3		3.018 ± 6e-3	3.183 ± 5e-3	4.10 ± 0.50	
$\mathcal{V}_{pol}$ [kJ mol <sup>-1</sup> ] <sup>e</sup>	-8.692 ± 3e-3	-8.777 ± 2e-3		-8.047 ± 4e-3	-8.071 ± 4e-3	-8.40 ± 0.40	

<sup>a</sup>Ref. [29]

<sup>b</sup>Ref. [30]

<sup>c</sup>data from Ref. [5] via [30]

<sup>d</sup>see text

<sup>e</sup>GROMACS defines  $\mathcal{V}_{pol}$  as positive. Here the sign is adjusted to fit the definition given in subsection 3.1.2

Table 3: Density  $\rho$  of the point-dipole and Drude oscillator (-D) models, and experimental data (Expt.)

	PJVP-D (PME)	PJVP-D (RF)	PJVP <sup>a</sup>	JVP-D (PME)	JVP-D (RF)	JVP <sup>a</sup>	Expt. <sup>b</sup>
$\rho$ [kgm <sup>-3</sup> ]	949.6 ± 0.4	890.5 ± 0.7	1014.0 ± 28.0	943.7 ± 0.4	893.0 ± 0.4	1000.0 ± 24.0	1015.0

<sup>a</sup>Ref. [30]

<sup>b</sup>Ref. [101]

of the equilibrium energy and LJ parameters  $\sigma$  and  $\epsilon$  was suggested by *Jedlovsky et al.* [79]. The point-dipole model's initial parametrization and its reparametrization are therefore strongly influenced by the system size effects and choice of long-range correction [29, 30]. Comparison of the initial non-polarizable HF model JVNP [79] with its re-evaluation in the PJVP study [30], reveals a deviation in density values of 30% for the exact same model at the equal state point, whereas the system is half as large in the later study (256 PJVP and 512 JVNP). The LJ parametrization of the point-dipole models was also performed for smaller systems (200) than the simulations reported here. In conclusion, simulations of significantly larger systems than the parametrization base without parameter adaptation necessarily results in an incorrect density and the 5%-10% underestimation observed here reflects the fact, that effects of the system size decrease for larger systems.

### Structural Results

For structural comparability of the DO-models, the three partial RDFs  $g_{ij}(r)$  are reported, compared with the results from the initial JVP study [29]. Figure 8 shows the H-F, F-F and H-H RDF at 273 K and  $1 \cdot 10^5$  Pa of the JVP(-D) and PJVP(-D) models. As investigated by *Pártay et al.* [30] the partial RDF of the initial JVP and the reparametrized PJVP models are almost identical for non-critical state points, wherefore an additional comparison with the PJVP data can be omitted.

Firstly,  $g_{ij}(r)$  of the JVP-D Drude models accurately reproduces the position of maxima and minima of the partial RDF point-dipole model. However, a larger first peak in the F-F and H-F correlation can be observed. This is likely related to the underestimated density and even more pronounced for the RF long-range correction strategy (compare with table 3). Additionally, the second maximum of the F-F correlation is less pronounced than for the point-dipole data. This effect is also less dramatic for the PME long range correction. Secondly, the partial RDFs for PJVP-D model are almost identical to the JVP-D model. The largest discrepancy is a minimal increase of the peak position by 0.019 nm for second minimum in the H-F correlation function.

## 3.4 Concluding Remarks for Modelling

The testing of the Drude charge's parametrization performed in subsection 3.2.3 shows the equivalence of a point-dipole model with a DO for a configuration. The subsequent simulation of a system of 1000 HF molecules in the isothermal-isobaric ensemble at 273 K and  $1 \cdot 10^5$  Pa further solidified the correctness of replacing the point-dipole with a DO without significantly changing the model's behaviour. The changes in the potential energy and its three components are within 10% of

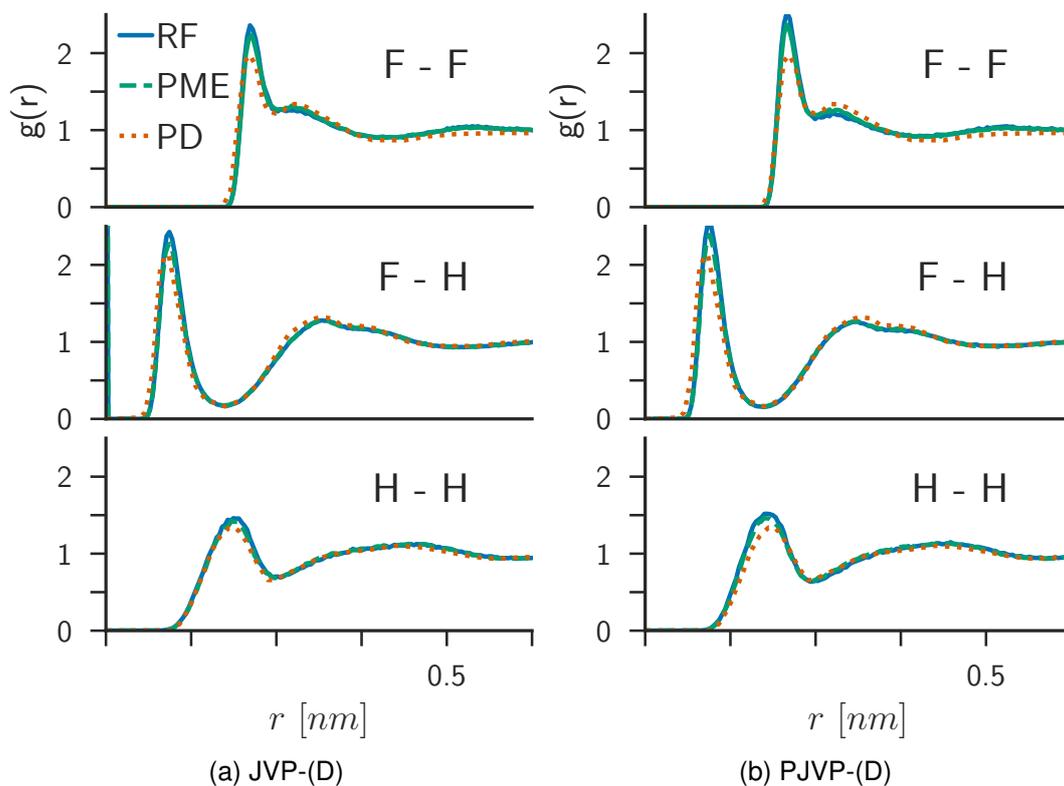


Figure 8: Partial radial distribution functions  $g_{ij}(r)$  of the a) JVP(-D) model and b) PJVP(-D) model. Drude oscillator (RF – solid blue curve, PME – dashed green curve) and point-dipole (PD) models\* (dotted orange curve).

\*data extracted from graph in reference [29](JVP) and [30](PJVP)

previously reported values and mostly originate from different long-range correction methods and the different system sizes. The large deviation in density is most likely solely related to the strong sensitivity with respect to the LJ parameters of the model, whose parametrization is likewise affected by system size. Overall, the absolute value of the density, affected by sensitivity and the large local density fluctuations caused by H-bonded chains, can only be taken qualitatively, regardless of the model-type (non-polarizable, point-dipole, OD).

In conclusion, the JVP-D and PJVP-D models are equivalent to the point-dipole counterparts JVP and PJVP in their ability of reproduce structural and thermodynamic properties of HF. Similar equivalence was successfully shown for various other systems like water [73, 75] in recent years [21, 69, 72, 76]. Moreover, the performed MD simulations show that the PME long-range correction strategy reproduces the thermodynamic and structural properties with the same accuracy as the RF strategy used in the initial studies. Consequently, the use of PME long-range correction instead of the RF method is an appropriate alternative for the presented models.

## 4. LIQUID/VAPOUR INTERFACE

A description of interfaces on an atomistic level, computational methods for its detection in MD simulations, and the application to the liquid/gas interface of HF are covered in this chapter. First, methodology for an atomistic level description of a liquid/gas interface and difficulties with respect to its definition and representation are discussed in section 4.1. Additionally, two algorithms for detecting such interfaces on the molecular level are described in more detail. Section 2.4 introduces a software package implementing these algorithms, as well as further software used to analyse simulation data. The main part of this chapter focuses on the simulation and analysis of the liquid/vapour interface of HF. Details of the computer simulation at various state-points, using the PJVP-D model, introduced in the previous chapter, are reported in section 4.3. Thermodynamic and structural properties, analysed with respect to the interface are also presented there. A critical assessment (section 4.4) of the results concludes the chapter.

### 4.1 Terminology & Theory – Interface Simulation

The description of soft interfaces on a molecular level is severely complicated by the interface's corrugation. Even interfaces that appear to be perfectly planar on a mesoscopic level usually are not flat on the microscopic level. This distortion of the interfacial particle positions can be caused by long ranging thermal waves caused by surface tension, referred to as capillary waves. Steric effects and the formation of clusters and chains formed along the interface by H-bonds or other strong intermolecular interactions strongly impact the interfacial surface. Therefore, the density profile of an interfacial system is smeared out, resulting in its typical sigmoidal shape [8]. Consequently, the description of such a soft, planar interface through a smooth density profile  $\rho(x)$ , where  $x$  is the direction of the interface-normal, is not sufficient for an atomistic analysis [15, 102, 103].

#### 4.1.1 Interface Analysis

In capillary wave theory (CWT) the corrugation caused by fluctuations is explicitly incorporated using a so called intrinsic surface  $x = \xi(\mathbf{R})$ , where  $\mathbf{R}$  is the transverse position along the interface (here  $\mathbf{R} = (y, z)$ ) [102, 104]. Using this concept, an intrinsic density profile  $\tilde{\rho}(x)$  can be defined. This profile depends on the distance to the local position of the interface  $\xi(\mathbf{R})$ , rather than on an another reference position

like the centre of mass, which is the usual choice for regular density profiles. The non-intrinsic and intrinsic density profile can formally be defined by

$$\rho(x) = \frac{1}{A} \left\langle \sum_i \delta(x - x_i) \right\rangle \quad (57)$$

$$\tilde{\rho}(x') = \frac{1}{A} \left\langle \sum_i \delta(x' - x_i + \xi(y_i, z_i)) \right\rangle, \quad (58)$$

where  $A$  is the cross-section area of the simulation box,  $i$  the particle index, and  $\xi(y_i, z_i)$  is the local interface position corresponding to the projected position of particle  $i$  on the surface. The notation  $x'$  indicates the different reference frame for the intrinsic case [105]. Intrinsic profiles are not limited to density, but can be calculated for any thermodynamic property.

Use of intrinsic profiles and the comparison with its non-intrinsic counterpart, allow for distinction of effects caused by corrugation of the interface and effects inherent to features of the liquid along an interface. The *intrinsic sampling method* [15], *Identification of Truly Interfacial Molecules* (ITIM) [38], and its generalization to non-planar interfaces (GITIM) [39] are three examples of methods to identify and calculate interfaces and intrinsic profiles.

A common problem for all three methods (and interface identification methods in general) is the necessity to define a continuous interfacial surface from a discrete set of molecules. Consequently, infinitely many interfaces can be defined, each corresponding to a different choice of at least one free parameter, causing uncertainty on its exact physical interpretation. However, it has been shown that the total intrinsic profile does not react strongly to small changes of this parameter and that a physically sensible range exists. Whether a single, correct, and physically based choice for the interface on the atomistic level exists is a topic of ongoing scientific discussion [15, 38, 39].

The ITIM and GITIM methods, used throughout this thesis, are covered in more detail in subsections 4.1.3.

#### 4.1.2 Layer-By-Layer Analysis

As described in the previous section, explicitly identifying the surface along an interface is reasonable on the atomistic level. A different perspective on an interface is to assign the molecules forming an interface to a set, called the first layer. For most liquids local structure exists in the bulk phase and a local layering structure can also form along the interface. These layers might be smeared out by capillary waves in the non-intrinsic reference frame, but can, if they exist, be easily observed

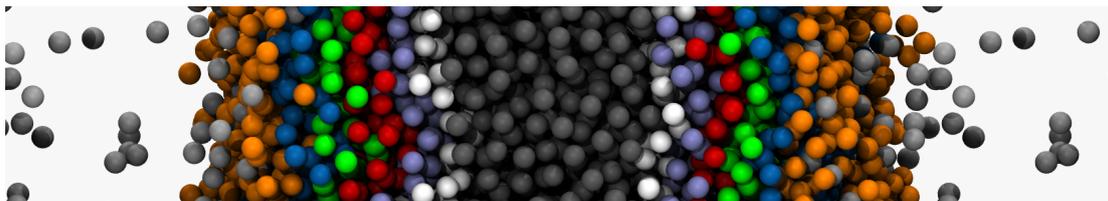


Figure 9: Simulation snapshot of the liquid/vapour interface of a Lennard-Jones fluid. The atoms of the first 6 layers are coloured with respect to their layer (1-orange, 2-blue, 3-green, 4-red, 5-light blue, 6-white). Bulk (grey) and vapour phase atoms (silver) are also visible.

using intrinsic profiles. However, these structural features vanish further away from the interface until the physical meaning of layers is lost in the bulk phase [38, 105].

The easiest way to determine interfacial layers is by repeatedly applying an algorithm like ITIM and removing layers, hence sets of molecules, that have already been identified. The snapshot of a system of a LJ-liquid at liquid/vapour coexistence, coloured according to subsequent layers is shown in figure 9. Contributions of thermodynamic quantities of the subsequent layers can then be determined by sampling solely over the molecules of a given layer. For instance, it has been shown that at least 85% of the surface tension for most molecular liquids is caused by molecules of the first layer, while contributions from the third and subsequent layers are negligible [106].

Nonetheless, special care has to be taken when using computational schemes to identify individual layers. Although the intrinsic profile of the whole system does not react strongly to small changes in the method's free parameter, the same is not true for contributions of different layers along the interface. For ITIM and LJ-liquids a well defined value for the free parameter of the order of the molecular size  $\sigma$  appears to exist [103]. Although this one to one correspondence between the LJ parameters and the parameter determining the interfacial surface does not hold in general, a value corresponding to half of the position of the first peak of the liquids RDF, has proven to be a sensible choice for most small molecular liquids [105–107].

### 4.1.3 Layer Identification Methods

#### Identification of Truly Interfacial Molecules (ITIM)

The ITIM method [38] identifies molecules located at a planar interface and thereby marks the set of molecules forming the first layer of a considered phase. Those molecules are selected by streaming probe spheres of a fixed radius  $R_{ps}$ , along a straight line and perpendicular to the interface, onto the molecules of the liquid phase. As illustrated in figure 10 a molecule is considered to be interfacial, if it is

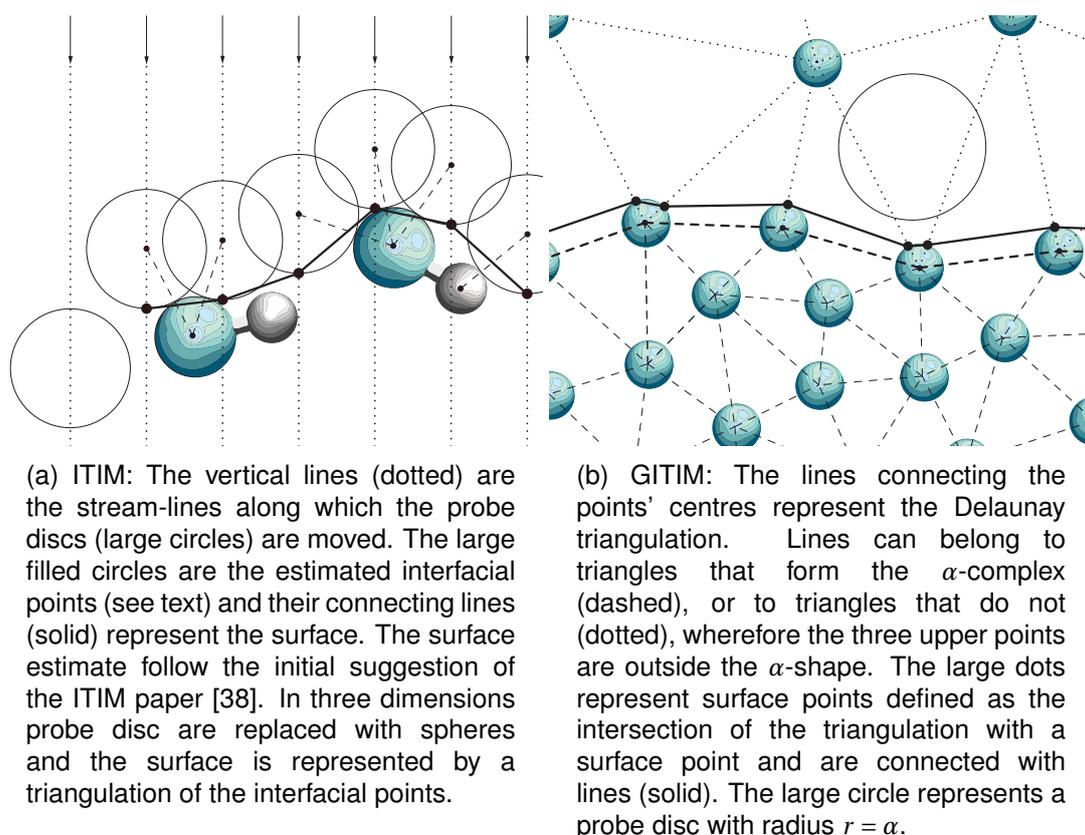


Figure 10: Illustration of the a) ITIM and b) GITIM methods. Also, a surface estimate (solid lines) is presented. However, no general definition for the surface exists and the illustrated surface is just one possibility

the first molecule to be touched by any probe-sphere. A probe sphere touches a molecule, if it touches any of its constituent atoms.

The ITIM approach is only applicable for planar interfaces since the orientation of the lines along which the probe spheres are streamed has to be perpendicular to the interface. Streaming probe spheres at an arbitrary angle introduces artefact and the presence of pockets in the interface cannot be resolved at all. The distribution of the stream-lines has little to no effect on the method's results, as long as the space between two lines is considerably smaller than the probe sphere radius. Although not being uniquely defined, a physically reasonable choice for the probe sphere radius  $R_{ps}$  is the same order of magnitude as the size of the molecules forming the phase. A sensible choice for the molecular radii is the Lennard-Jones radius  $\sigma/2$ , if their non-Coulomb interaction is modelled by a LJ potential. Moreover, results of the method have been shown to be rather insensitive to small changes in  $R_{ps}$  as long as it remains in this range [38, 103].

For interfaces that are formed between two phases of the same substance, for instance a liquid/gas interface, it is necessary to first identify the molecules constituting the considered phase. Multiple different clustering algorithm,

corresponding to different criteria, like local density or connectivity, are used for this purpose. Like the particular value for  $R_{ps}$ , the choice of correct clustering strategy strongly depends on the system it is applied to [108].

Since the interfacial surface is not uniquely defined either, a computationally efficient strategy is preferred. Often, a triangulation of interfacial point is performed, where an interfacial point is defined as the intersection of the touching probe sphere and the line along which it was streamed. The approach combines quick algorithms with the intuitive condition, that all molecular centres are within the liquid phase. Using this definition, the resolution of the interfacial surface can be adjusted easily and chosen such that it balances precision and computational cost [38, 105]. If applicable for analysing an interface, ITIM is regarded as the fastest algorithm for identifying interfacial atoms [107, 109].

### **Generalized Identification of Truly Interfacial Molecules (GITIM)**

The dependence of the ITIM method on the reference frame of streamlines perpendicular to the interface, impedes an appropriate treatment of arbitrarily shaped interfaces. To extend the applicability of the method's approach to non-planar interfaces the GITIM method [39] proposed by *Sega et al.* in 2013, adopts concepts from computational geometry to completely avoid any reference frame.

More specifically, the  $\alpha$ -shape algorithm [110, 111], which was initially designed for digital shape sampling and processing in computer graphics, was used. The algorithm was slightly adapted to provide analogy to the ITIM method, but the general concept and functionality remained unchanged. First a Delaunay triangulation [112] of all atomic positions is performed. The Delaunay triangulation of a set of points is chosen in such a way, that no other point is inside the circumcircle of any triangle in the triangulation. An alternative definition of the Delaunay triangulation as the triangulation that maximizes the smallest angle of all triangles is also commonly used. For the GITIM method, this triangulation forms simplexes that are used to deploy the touching spheres in analogy to the streamed probe-spheres of the ITIM approach. However, instead of streaming along a predefined line, these spheres grow in radius until they touch four molecules. In contrast to the  $\alpha$ -shape algorithm, this mimics the presence of excluded volume, similar to ITIM. This concept is also illustrated in figure 10. Finally, all molecules that are touched by such a sphere, with radius larger than a predefined value  $\alpha$ , form the so-called  $\alpha$ -shape. A recent study [113] suggests, that the excluded volume has to be considered in the triangulation itself. Omission thereof, or the delayed introduction, like in GITIM, can, under certain conditions, lead to artefacts of the inconsistent metric and incorrect interface determination. The lack of fast computational schemes to calculate such triangulations, compared to the quickhull algorithm [114] (scaling

with  $\mathcal{O}(N \log v)$ , where  $N$  and  $v$  are the number of input points and output vertices) used for GITIM, drastically limits their applicability [39].

The  $\alpha$ -complex is in general a concave, topologically disconnected polytope, however for GITIM isolated points or strings of segments are excluded to allow for a reasonable definition of the interfacial surface. An additional benefit over ITIM is the ability to correctly identify pockets and voids in the liquid since an interfacial molecule can be detected at every point in space instead of only at the first molecule along a streamline [39]. Similar to ITIM, the definition of the interfacial surface is independent of the identification of the molecules and the proper definition is still an open question.

## 4.2 Software – Interface Analysis

### 4.2.1 PYTIM

Implemented in Python, PYthon Tool for Intrinsic Molecular analysis (PYTIM) is a collection of several different methods for determining and analysing fluid interfaces. Developed by *Sega et al.*, the software is based on the MDAnalysis package [115, 116] and therefore complies with a large variety of MD simulation software packages, including GROMACS. The current beta-release is available free of charge via the Python Package Index and development is conducted through Github (<https://github.com/Marcello-Sega/pytim>). Available methods are, among others, the interface identification methods: ITIM, GITIM, Chacon-Tarazona [16], and Willard-Chandler [117], as well as the clustering algorithm DBSCAN [118]. Additionally, PYTIM is equipped with a number of analysis tools: for instance for calculation of interfacial surfaces, intrinsic and non-intrinsic profiles for various thermodynamic observables, and structural features like the radial distribution function. All methods can be applied for molecules as well as individual atoms.

### 4.2.2 MDAnalysis

Developed in 2011 (GPL license), MDAnalysis [115, 116] is a toolkit for analysing trajectories generated with the most common MD software packages, among others: NAMD [33], CHARMM [91], and DL\_POLY [92], and the Protein Data Bank PDB format [119]. Implemented in Python, its use of the NumPy package [120] provides easy access to atomic coordinates through arrays and allows application of large number of python data-analysis and data-manipulation tools. The source code is freely accessible on Github (<https://github.com/MDAnalysis/mdanalysis>) and the package can be installed with the Python Package Index.

### 4.2.3 Network Analysis Tool

This tool is an extension on the PYTIM package and freely available on Github (<https://github.com/elija-feigl/pytim-ext-networkHF.git>). The code determines the network structure of liquid HF and similar liquids by identifying linear chains, linear segments of a branched structure, and cycles. First, terminal- and branching-points of chains are identified using the adjacency matrix of the network determined by a cutoff parameter. Then, the algorithm iterates over neighbouring molecules until a second terminal- or branching-point is reached. Distinction between real linear chains, linear segments of the branched structures, and rings is performed afterwards. Rings can either be completely isolated or attached to only one chain, capping it off. Cyclic structures containing more than one branching point are considered as linear segments of branched structures. However, clusters without any terminal point as well as infinite chains caused by PBC are also identified as rings by the algorithm. The tool also calculates the number of terminal- and branching-points, and the number of bonds per molecule. The code is organized such that it mimics the structure of a PYTIM Observable.

## 4.3 Simulation of Liquid/Vapour Interface of Hydrogen Fluoride

### 4.3.1 Liquid/Vapour Coexistence in the Canonical Ensemble

#### Computational Details

MD simulations were performed in the canonical ensemble at 16 different temperatures ranging from 200 to 455 K. The simulations were carried out with 2000 HF molecules, modelled by the PJVP-D model (see previous chapter 3 for detail), using the GROMACS simulation package version 2016.4 [31, 32]. The software was compiled with double precision and SCF Shell Molecular Dynamics [75, 100] were used for treatment of the DOs.

To simulate the liquid/vapour interface, copies of the simulation box of equilibrated liquid bulk HF were inserted at the centre of a  $64.8\sigma \times 10.8\sigma \times 10.8\sigma$  ( $L_x \times L_y \times L_z$ ) box, such that a third of the box in  $x$  direction was filled with a liquid phase spanning the complete  $y$  and  $z$ -direction. Consequently, each periodic simulation cell contains two liquid/vapour interfaces with surface-normal along the  $x$ -direction. Moreover, periodic boundary conditions ensure that the system does not collapse to a single liquid droplet with a spherical interface.

A Nosé-Hoover thermostat [57, 58] with a relaxation constant of 0.5 ps was used to keep the average temperature constant. The equations of motion were integrated every 1 fs with a leapfrog [49] algorithm. The SCF threshold for energy minimization of the Drude particle was set to  $1 \cdot 10^{-3}$  kJ mol<sup>-1</sup> nm. Although this threshold is less strict than initially suggest for the PJVP model [30], a series of short bulk simulations of liquid HF in the canonical ensemble determined the deviation in potential energy  $\mathcal{V}$  as less than 1%, while the simulation time could be reduced by almost 20%. An interaction cutoff of 0.87 nm was used for both Coulomb and LJ interaction. Since RF correction [50] is not suitable for highly inhomogeneous systems, a 4<sup>th</sup>-order PME [54, 55] for Coulomb and LJ interaction with an isotropic Fourier spacing of 0.09 nm was applied. The PME's relative tolerance was set to  $10^{-5}$  and  $10^{-3}$  for Coulomb and LJ, respectively. Again, the parameter was obtained from a series of short canonical MD simulations in the liquid phase.

The different systems were equilibrated for up to 8 ns to ensure relaxation of potential energy  $\mathcal{V}$  and density  $\rho$  to their equilibrium value at liquid/vapour coexistence. Although systems at liquid/vapour coexistence, and interfaces in general, take longer to equilibrate than bulk systems, the inclusion of polarizability to the model loosens the structure in the liquid phase and accelerates the equilibration process compared to non-polarizable models [21]. This feature of the polarizable model is of particular importance since the current version of GROMACS 2016.4 does not include a parallelized algorithm for treatment of Drude particles, wherefore longer simulation runs of systems with 2000 molecules or more would not be feasible. For simulations at temperatures higher than 300 K, multiple liquid droplets formed during the equilibration process, which is likely due to the strong cohesion of chains and clusters formed by hydrogen bonds and the explosive formation of the gas phase that is caused by the initialization process described above. However, these droplets merged into a single liquid phase for the equilibrated systems.

During a run of 3 ns particle position and energy contributions were saved every 0.5 ps. Consequently, a trajectory with 6000 saved configurations was used for structural and interface analysis. After the simulation, the centre of mass of the liquid phase in x-direction has been centred in the middle of the simulation box to facilitate analysis and allow better comparison of the different simulations. Centring was not performed, if no clear distinction between liquid and gas phase was possible.

## System Description

To give an impression on the general behaviour of the simulated system, a volume slice of a single configuration at six selected temperatures is shown in figure 11.

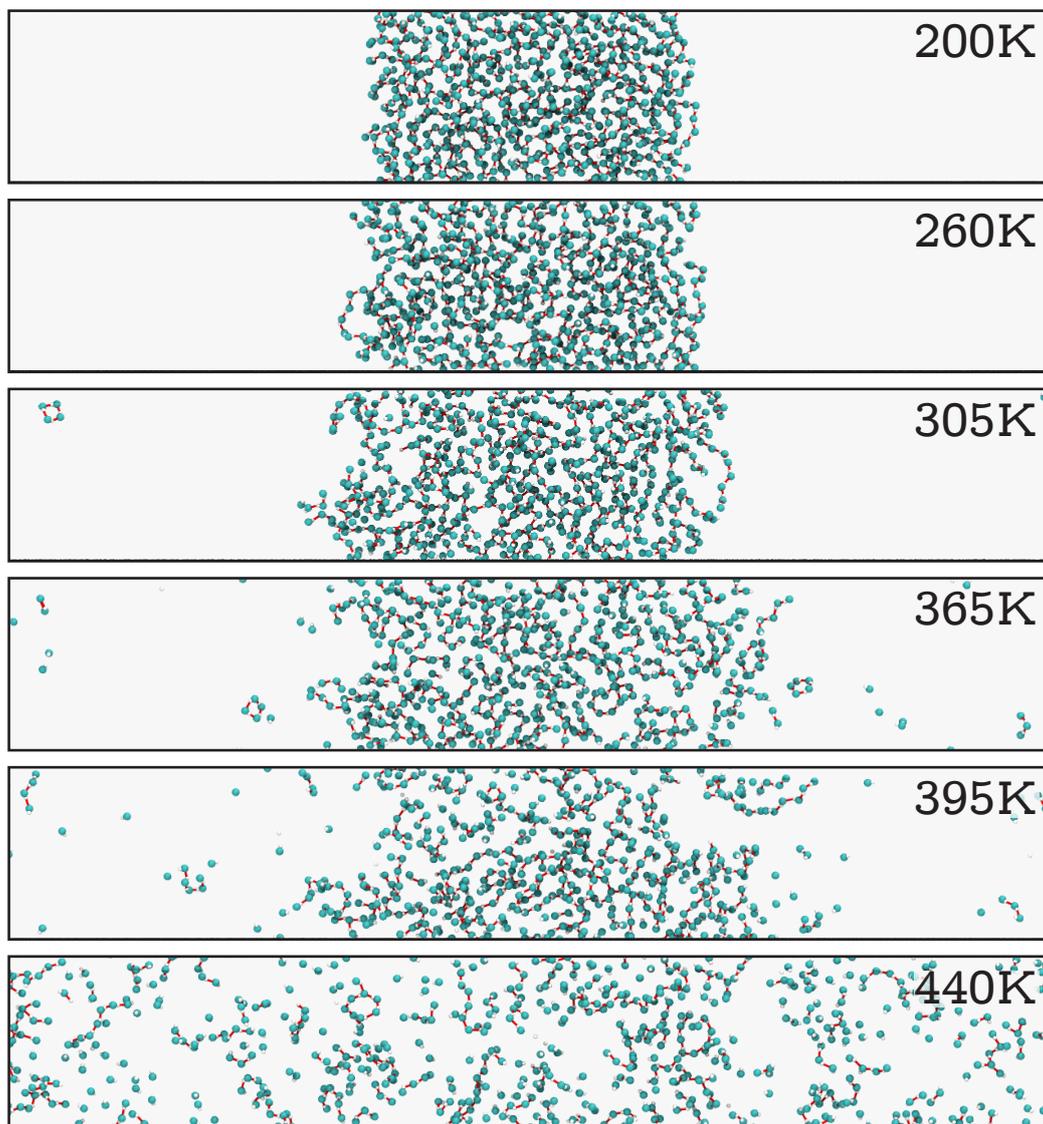


Figure 11: Volume slice ( $L_x \times L_y \times \frac{L_z}{3}$ ) of a single configuration of the equilibrated system for six different temperatures at liquid/vapour coexistence. Smaller spheres (white) represent hydrogen and larger spheres (turquoise) fluorine atoms. A solid line (red) indicates hydrogen bonded molecules, identified by a distance smaller than first minimum of  $g_{HF}$  at 0.236 nm. The increased corrugation of the interface, as well as the formation of pockets and dangling clusters for higher temperatures are clearly visible. At the highest temperature (bottom right) the system is at a super-critical state-point and now separate phases exist.

The volume slice represents only a third of the complete simulation box, to help illustrate the system's nature. All atoms are coloured according to their type, with fluorine in turquoise and hydrogen in white, additionally auxiliary sites are omitted for clarity. Emphasizing the molecule's tendency to form H-bonds, blue lines are drawn between molecular pairs closer than the position of the first minimum of the H-F partial RDF at the given temperature (see table 4b).

Firstly, the expansion of the liquid phase with increasing temperature due to smaller densities can clearly be observed. The liquid phase covers approximately one third of the simulation box for 200 K, but expands until it covers almost the entire box for temperatures close to the critical point. The critical temperature of the PJVP-D model is expected to be close to the critical temperature of the JVP model at 450 K [66], however, no definite value has been calculated [30]. Additionally the simulations presented here are not well suited to accurately represent near-critical behaviour since the finite box size impedes the formation of a coexisting gas-phase for large temperatures. Moreover, the existence of two, macroscopic interfaces is no longer an accurate description for the simulated system close to the critical point. Secondly, a tendency to form clusters, chains, and rings in the gas phase is clearly visible and the increase in gas phase density with increasing temperature is evident. Thirdly, large density fluctuations in the liquid phase can be observed and become more pronounced for higher temperatures. This observation of the occurrence of large voids in the liquid phase is discussed in more detail in subsection 4.3.2. Finally, the snapshots provide a first impression of the highly corrugated nature of a planar HF liquid/gas interface. The formation of pockets and clusters that are dangling from the liquid phase also affects the choice of the appropriate surface identification method.

### 4.3.2 GITIM Application

Before any of the interface detection method presented in subsection 4.1.3 can be applied, all molecules belonging to the liquid-phase have to be identified. A cutoff based cluster search algorithm [121] is used to exclude the vapour phase. Although the first minimum of the H-F partial RDF is the usual choice for the cutoff [30, 105, 108, 121–123], this approach fails to identify the liquid phase for the PJVP model. Stable clusters and chains, that can include a large number of molecules, are present in the liquid. These clusters are held together by the strong dipole-dipole interaction of the PJVP-D model, but do not form a percolating network that includes all molecules. The interaction between clusters appears to be weak compared with the hydrogen bonds forming them and molecules of different clusters are approximately 0.1 nm further apart than H-bonded molecules. Consequently, a larger cutoff criterion is necessary to identify the liquid phase and exclude vapour molecules. Although a more thorough analysis of this finding is certainly necessary, the number of molecules belonging to the liquid phase was found to be rather insensitive to the cutoff criterion of the cluster search. To allow for easy reproduction, the cutoff-distance  $d_{liquid-cluster}$  was chosen as the value where the F-F partial coordination number equals 8.0. This choice corresponds to a range of 0.403 - 0.547 nm for the various temperatures and the exact values are listed in table 4b.

Further consequences of this exceptional behaviour also affect the interface identification. Molecules tend to dissipate from the liquid phase to the gas phase in small clusters instead of individually and therefore can form clusters that extend from the interface of the liquid into the gas phase. These dangling clusters are still connected with the liquid phase and are therefore classified as liquid in this study. Additionally, large pockets can form along the interface when two larger clusters drift apart. Since the ITIM method relies on the streamlines perpendicular to the interface and only detects the first molecule to be touched by a sphere moving along those streamlines, it fails to identify those irregularities and cannot detect the correct interfacial molecules. GITIM on the other hand was specifically designed for this purpose [39]. Although the phenomena described above do not appear for temperatures below 250 K, and application of ITIM would be feasible, GITIM is used for all temperatures to ensure comparability and eliminate potential systematic errors.

Formation of multiple clusters also affects the choice of the GITIM cutoff parameter  $\alpha$ . Previous studies [15, 16, 105, 107, 124] suggested the existence of a physically sensible range for  $\alpha$ . Within this range the shape of the intrinsic profile should not change significantly. Moreover, a more thorough interfacial analysis of liquid argon [103] revealed that the separate contributions of successive layers are far more sensitive to the specific choice of the cutoff parameter. By comparing the peak positions of the overall intrinsic profile and the separate layers, the study revealed the optimal value for  $\alpha$  to precisely match the molecular radius. For most empirical models of small molecules the molecular radius is well-defined as half of the LJ parameter  $\sigma$ . For larger molecules and models with one or more LJ centres, half of the first peak of the intermolecular RDF has proven to be the optimal choice for various liquids [106, 108, 123, 125].

For the PJVP-D model however, the first peak of  $g_{FF}(r)$  for  $\alpha$  does not correspond trivially to  $\sigma$ , due to the presence of strong Coulomb interactions of the H and F sites. Therefore, the minimum energy arrangement of a hydrogen bonded pair is at considerably closer distances than  $\sigma$ . Also, the average distance between two different clusters only has a rather complex relation to  $\sigma$  and exceeds half of the first peak of  $g_{FF}(r)$ . Therefore,  $\alpha = \sigma/2$  is not a suitable choice for HF. Instead, the optimal parameter is very close to half of the first peak of a F-F RDF, that only considers molecules of different clusters. This peak position is subsequently labelled as  $\zeta$ . This contribution to the total RDF is labelled as  $g_{ij}^{soft}(r)$ , emphasizing the weak interaction between clusters. In equation 59, the contribution from molecules within a cluster, formed by the strong H-bond,  $g_{ij}^{hard}(r)$  is the difference of  $g_{ij}^{soft}(r)$  to the overall  $g_{ij}(r)$ .

$$g_{ij}(r) = g_{ij}^{soft}(r) + g_{ij}^{hard}(r) \quad (59)$$

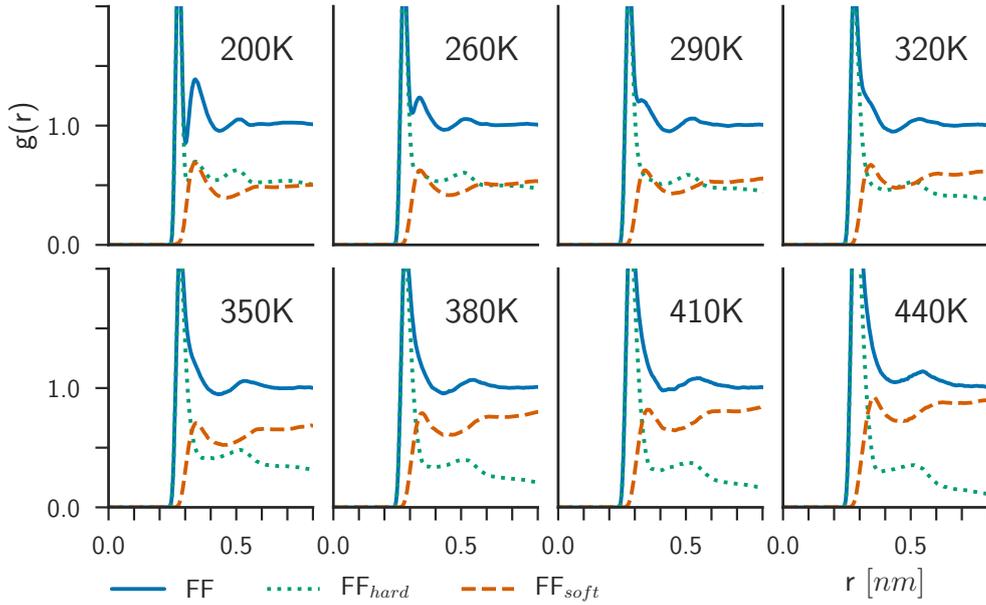


Figure 12: F-F partial radial distribution function  $g_{FF}(r)$  (blue, solid line) and its inter- and intra-cluster contributions  $g_{FF}^{hard}(r)$  (green, dotted line) and  $g_{FF}^{soft}(r)$  (orange, dashed line) at six selected temperatures.

Similar to  $\sigma$  for argon,  $\zeta$  represents the close-contact distance of the largest entities forming the fluid, or in other words the individual clusters' separation. Figure 12 shows  $g_{ij}(r)$  for six selected temperatures and its separate contributions from  $g_{FF}^{hard}(r)$  and  $g_{FF}^{soft}(r)$ . Radial distribution functions in this section are computed for the system at liquid/vapour coexistence, although only being properly defined for a bulk system. Since the vapour density is low, deviations from the correct bulk RDF are expected to be small. Moreover, the normalization was adjusted to mimic the convergence to 1 at large radii  $r$  for liquid bulk systems. Clusters were computed by applying the same cutoff based cluster search algorithm [121] used to identify the liquid phase, with a cutoff corresponding to the first minimum of  $g_{HF}(r)$  (see  $d_{HF}$  in table 4b).

The complete  $g_{FF}(r)$  for small temperatures clearly shows a second maximum at around 0.33 nm, roughly 9% larger than  $\sigma$ . It disappears for temperatures higher than 310 K, merging with the first maximum. Along the complete temperature range,  $g_{FF}^{soft}(r)$  has a clear maximum that corresponds to the parameter  $\zeta$ . This inter-cluster maximum is at the same position as the second maximum in  $g_{FF}(r)$  (if the latter is visible). The parameter  $\zeta$  identifies the average distance of molecules of neighbouring clusters as the origin of the second peak in  $g_{FF}(r)$ . The peak position slowly increases to 0.347 nm at 455 K and its height decreases indicating a looser inter-cluster structure for higher temperatures. Nonetheless, the peak is not vanishing at the critical point, and it persists even at slightly higher temperatures, suggesting interaction of different clusters even at supercritical states. Additionally,

the broad first minimum in  $g_{FF}^{soft}(r)$  appears to shift the second minimum of the complete  $g_{FF}(r)$  to larger values. Similar phenomena have also been reported for the Oxygen-Oxygen partial RDF  $g_{OO}^{H_2O}(r)$  of high density water and amorphous ice [126–130]. Recent studies [131, 132] point to the presence of interstitial water, non-H-bonded molecules located between the first and second coordination shell, as the origin of a peak or shoulder around 0.35 nm in  $g_{OO}^{H_2O}(r)$ . Besides the similarities, non-H-bonded neighbours in water occur less frequently than HF, due to the structural differences of the liquid. Therefore, identification of a causal relationship of non-H-bonded molecules and their effects on the RDF are difficult and certainly need further investigation.

Apart from the contribution of the first maximum in  $g_{FF}^{soft}(r)$ , the overall F-F correlation and thereby the structure of liquid HF is dominated by contributions from the the intra-cluster  $g_{ij}^{hard}(r)$ . Pronounced first and second maxima correspond to the first and second hydrogen bonded neighbours inside a cluster. This intra-cluster order is extremely stable for all temperatures and emphasizes the importance of the H-bond on the local structure in the liquid. Moreover, a peak at 0.33 nm for 200 K  $g_{ij}^{hard}(r)$  and the broadening of the first peak for larger temperatures can be attributed to branching inside a single cluster.

If solely molecules of linear cluster segments and chains are considered to define these RDFs, the peak at 0.35 nm in  $g_{FF}^{hard}(r)$  disappears. The difference allows to assign this feature of  $g_{ij}^{hard}(r)$  to the presence of branching inside a single cluster. Consequently, two neighbouring branches of the same clusters exhibit the same feature in the RDF as two neighbouring clusters. However, the small number of molecules per linear segments strongly impedes proper sampling, wherefore the initial definition of  $g_{FF}^{soft}(r)$  is favourable. The normalization and slope for large values of  $r$  of both inter- and intra-cluster RDF solely indicate the sampling imbalance and reflect the liquid's network properties. The steep decrease of  $g_{FF}^{hard}(r)$  at near-critical temperatures for example, is a consequence of the absence of large linear chains and clusters. Therefore, few molecules have H-bonded  $3^{rd}$  or  $4^{th}$  neighbours and the intra-cluster RDF converges to zero. Contrarily, all molecules at larger distances are part of a different cluster and consequently contribute to  $g_{FF}^{soft}(r)$ . A detailed description of the network properties and the identification of linear chains and linear cluster segment is given in subsection 4.3.3.

Nonetheless, application of GITIM with a parameter choice of  $\alpha = \zeta/2$  fails to identify any reasonable layers. Sample configurations at 200 K and 365 K, as depicted in figure 13 a, reveal that the algorithm detects numerous voids with a equal or larger than  $\zeta$  inside the liquid phase. Voids become even more frequent and get larger for higher temperatures. This finding suggests, that half of the first peak of the inter-entity RDF [103] is not the universal choice for  $\alpha$  for all liquids and HF is certainly an exception. Simply increasing  $\alpha$  until all voids inside the liquid phase disappear is

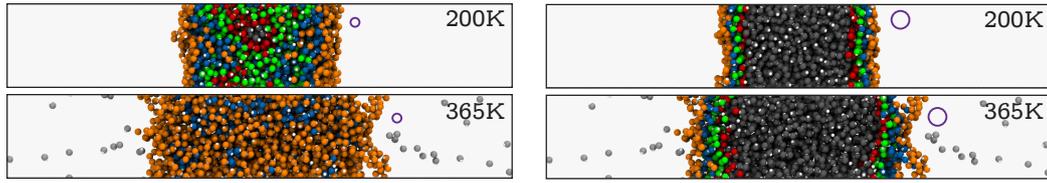
(a)  $\alpha = \zeta/2$ : incorrect interface detection(b)  $\alpha = \zeta$ : correct interface detection, voids can still be present at high temperatures

Figure 13: Sample configuration of the system at 200 K and 365 K. Two different values for parameter  $\alpha$  ( a)  $\zeta/2$  and b)  $\zeta/2$ ) were used and molecules of the first 4 layers are coloured with respect to their layer (1-orange, 2-blue, 3-green, 4-red). Bulk (grey) and vapour phase atoms (silver) are also visible. The purple circle represents a single probe sphere of radius  $\alpha$ , illustrating the GITIM identification. The black box represents the simulation box with PBC.

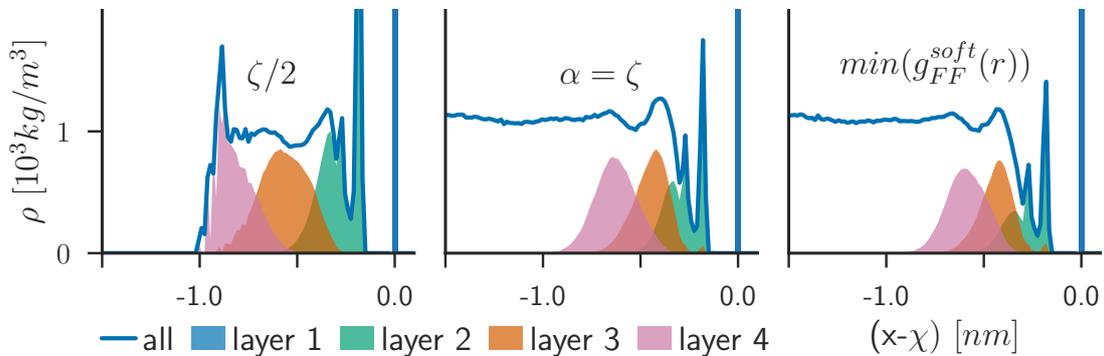


Figure 14: Liquid part of the intrinsic density profile (solid line) for three different values of  $\alpha$  ( $\zeta/2$ ,  $\zeta/2$ , and  $\min(g_{FF}^{soft})$ ). The separate contributions of the first (blue) second (green), third (orange), and fourth (purple) layer are shown, although per definition the first layer only contributes to the delta peak at 0.

not a suitable strategy either since, especially at higher temperatures, the diameter of single voids can become much larger than the interface's characteristic features. Although GITIM provides an option to exclude small voids by only picking the largest cluster of molecules as interfacial, this approach also fails for the simulated system. This failure suggests, that no lateral percolating network along the interface exists, as has been reported for other highly polar liquids like water [38, 121, 133, 134]. The presence of voids in the liquid phase is discussed in more detail in subsection 4.3.2.

Consequently, a more thorough determination of a proper choice for  $\alpha$  is necessary. Since the choice  $\alpha = \zeta/2$  apparently is too small to identify interfacial layers, larger values for  $\alpha$  were tested and evaluated according to the scheme described for argon and the ITIM method [103]. This scheme implies calculating the intrinsic density profile for varying  $\alpha$  and comparing peak positions with contributions of individual layers.

The intrinsic density profiles for three different values of  $\alpha$  are shown in figure 14. These three choices correspond to  $\zeta/2$ ,  $\zeta$ , and the position of the first minimum

of  $g_{FF}^{soft}(r)$  - labelled  $min(g_{FF}^{soft})$ . These values for all 16 different temperatures are collected in table 4a and are chosen for illustration because of their easy reproducibility. A more thorough analysis of the intrinsic density profile, also discussing the atypical form of the second layer, is given in subsection 4.3.4. As already suggested, the layer-by-layer contributions of the intrinsic density profile reveal, that the value of  $\zeta/2$  is certainly too small. The individual layers' peak positions are shifted further away from the interface than the corresponding maxima of the intrinsic density profile, similar to  $\alpha < \sigma$  for argon [103]. The opposite is true for the minimum of the F-F RDF  $min(g_{FF}^{soft})$ , with the peaks too close to the interface. Instead choosing  $\alpha = \zeta$ , hence twice as large as originally anticipated, produces the most accurate results in matching the peaks and maxima. However, no values close to  $\zeta$  have been tested and compared, therefore  $\alpha = \zeta$  might only be close to and not the optimal value.

Although clearly superior in terms of the intrinsic profile, a large number of voids is still identified by GITIM when using  $\zeta$  as the cutoff parameter. To account for this problem, results of the interfacial and layer-by-layer analysis will also be reported for  $\alpha = min(g_{FF}^{soft})$  alongside the primary choice of  $\alpha = \zeta$ . Since for  $min(g_{FF}^{soft}(r))$  only few voids are identified in the liquid phase, comparison of the two results can be used to assess the influence of the voids on the overall results. To add perspective, figure 13 b shows a configuration coloured according to interfacial layers identified by the GITIM method with  $\alpha = \zeta$ .

### Origin of Voids in the Liquid Phase

By applying the GITIM algorithm with  $\alpha = \zeta$  voids are detected inside the liquid phase for temperatures higher than 260 K. These voids are generally spherical and not stable, usually disappearing within 1 ps. The voids' diameter as well as their numbers increase with temperature and above 320 K voids also appear for the larger parameter choice  $\alpha = min(g_{FF}^{soft})$ . For temperatures above 410 K extremely large and stable voids can be observed and the liquid phase already starts to dissipate along the interface. Due to the large diameter of some voids, examples given in figure 15, their detection is certainly not an artefact (compare [113]) of the GITIM algorithm and its parametrization strategy.

Evidence of large voids in sub-critical liquids is rare in the literature. In a study from 2000 Jedlowszky [135] reports the presence of voids equivalent to the size of a cluster composed of up to 11 HF molecules for liquid bulk HF modelled with JVP at 288 K and  $1 \cdot 10^5$  Pa. The voids are described as being primarily located between long, linear hydrogen bonded chains. Complexly shaped clusters, caused by the (almost) tetrahedral angle of the H-bond and an average number of bonds per molecule of around 2.0 are the likely cause of the liquids

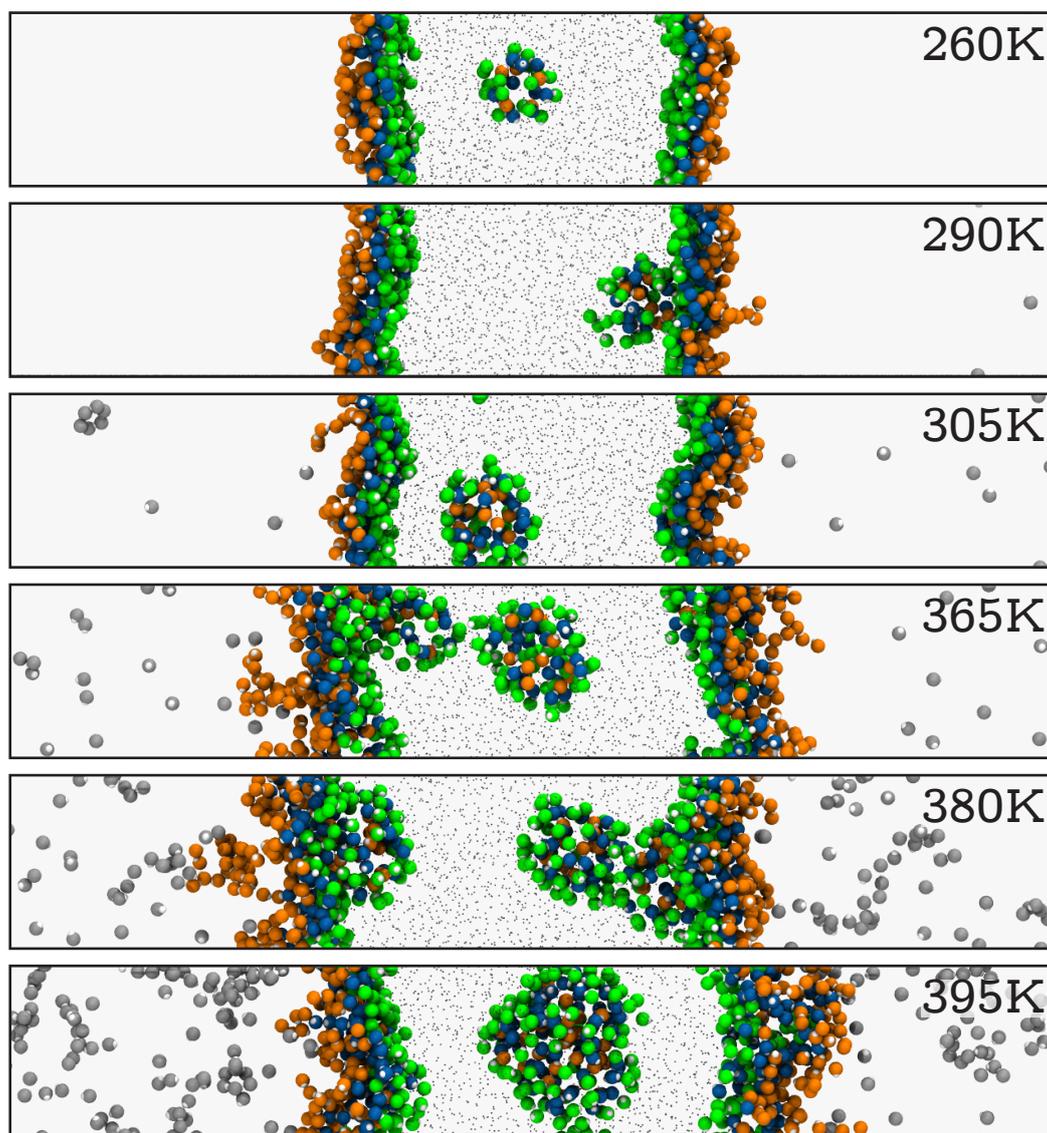


Figure 15: Sample configurations of the system at different temperatures containing voids, pockets, or large dangling clusters. The first three layers are coloured with respect to their layer (1-orange, 2-blue, 3-green), as determined by GITIM using  $\alpha = \zeta$  (ranging from 0.327 nm at 260 K to 3.344 nm at 395 K). The vapour phase molecules (silver) are visible, but molecules of the 4<sup>th</sup> layer, and liquid bulk molecules are excluded to allow for better identification of the peculiar structures. The black box represents the simulation box with PBC.

inability to form a space-filling structure. However, the average temperature is close to the experimental boiling point of HF at 292.7 K [136] and approximately 5 K above the boiling point of the PJVP-D model (see subsection 4.3.4). The JVP model's boiling point is expected to be similar to the PJVP-D model and therefore the reported system is likely above the boiling point [135]. Another recent study [37] about a simple dipolar fluid model with variable dispersion interaction reports thermodynamic properties comparable to HF, modelled with

Table 4: Liquid/vapour coexistence: Structural Data for the PJVP-D model for 16 different temperatures.

(a) Positions of the first peak ( $\zeta$ ) first minimum ( $\min(g_{FF}^{soft})$ ) of the inter-cluster partial F-F radial distribution function $g_{FF}^{soft}(r)$																
[nm]	200 K	230 K	260 K	275 K	290 K	305 K	320 K	335 K	350 K	365 K	380 K	395 K	410 K	425 K	440 K	455 K
$\zeta$	0.327	0.334	0.327	0.334	0.329	0.334	0.330	0.334	0.337	0.334	0.340	0.344	0.340	0.347	0.347	0.347
$\min(g_{FF}^{soft})$	0.438	4.400	0.439	0.439	0.438	0.447	0.438	0.437	0.441	0.438	0.439	0.440	0.441	0.442	0.444	0.446

(b) Cutoff distance for H-bond criterion ( $d_{H-bond}$ ) and determination of the liquid cluster ( $d_{liquid-cluster}$ )																
[nm]	200 K	230 K	260 K	275 K	290 K	305 K	320 K	335 K	350 K	365 K	380 K	395 K	410 K	425 K	440 K	455 K
$d_{H-bond}$	0.228	0.232	0.235	0.236	0.238	0.240	0.242	0.244	0.245	0.246	0.247	0.249	0.251	0.253	0.254	0.255
$d_{liquid-cluster}$	0.403	0.413	0.421	0.426	0.429	0.434	0.440	0.445	0.452	0.459	0.470	0.480	0.498	0.520	0.533	0.547

Table 5: Liquid/vapour coexistence: Network Properties for the PJVP-D model for 16 different temperatures

(a) Average number of bonds for all temperatures																
	200 K	230 K	260 K	275 K	290 K	305 K	320 K	335 K	350 K	365 K	380 K	395 K	410 K	425 K	440 K	455 K
$\langle N_{HB} \rangle$	2.27	2.32	2.33	2.33	2.33	2.33	2.33	2.32	2.28	2.23	2.17	2.13	2.07	1.99	1.89	1.83

(b) Average number and average size of clusters ( $\langle N_C \rangle, \langle S_C \rangle$ ); linear chains ( $\langle N_L \rangle, \langle S_L \rangle$ ); linear cluster segments ( $\langle N_B \rangle, \langle S_B \rangle$ ), and rings ( $\langle N_R \rangle, \langle S_R \rangle$ ).																
	200 K	230 K	260 K	275 K	290 K	305 K	320 K	335 K	350 K	365 K	380 K	395 K	410 K	425 K	440 K	455 K
$\langle N_C \rangle$	13.59	19.07	29.74	38.30	46.00	53.48	62.39	72.30	90.78	112.44	139.35	160.98	191.90	232.63	270.17	290.41
$\langle S_C \rangle$	147.09	104.72	67.00	51.90	43.07	36.91	31.48	26.98	21.29	17.00	13.53	11.53	9.52	7.67	6.45	5.88
$\langle N_L \rangle$	3.44	7.85	16.32	22.71	28.79	34.72	42.17	49.85	63.07	78.76	98.90	114.75	138.48	171.00	202.02	219.95
$\langle S_L \rangle$	50.29	22.93	12.80	10.27	8.59	7.35	6.47	5.77	5.34	4.92	4.61	4.37	4.15	4.01	3.86	3.70
$\langle N_B \rangle$	25.94	59.91	110.86	139.88	168.71	198.55	227.07	252.22	271.0	286.34	293.34	300.39	298.23	280.82	265.34	260.45
$\langle S_B \rangle$	52.72	23.70	13.48	11.04	9.23	7.93	7.03	6.33	5.92	5.52	5.18	4.86	4.56	4.26	3.97	3.77
$\langle N_R \rangle$	52.04	44.82	34.95	28.15	24.82	22.35	18.65	16.37	11.59	8.95	7.22	6.15	5.46	4.85	4.49	4.29
$\langle S_R \rangle$	10.46	13.14	16.00	17.05	19.65	22.57	24.69	27.69	27.16	27.70	27.29	28.10	28.98	28.43	26.67	25.96

PJVP-D, for weak dispersion interaction. However, these models are incapable of reproducing structural properties, due to the missing quadrupole-moment of the two-site approach (compare section 3.2) and liquid/vapour coexistence could not be observed. Although no voids were reported for this model, the study further solidifies the claim that voids might originate from steric restrictions of the H-bonded clusters since the absence of quadrupolar interactions results in linear H-bonds and reduces the complexity of the cluster-shape.

Similar phenomena concerning atypical liquid-vapour coexistence can be found in studies performed on dipolar hard spheres [137–140] and patchy particles [141–143]. These studies suggest, that the liquid-vapour coexistence might be replaced by a defect-driven phase separation [140] or the existence of an additional critical point caused by the preferred formation of rings and associated entropy reduction in the vapour phase [143]. HF is often referred to as the real fluid closest resembling dipolar hard spheres and the presence of rings in its gas phase is well documented [34, 83] and also reproduced by the PJVP-D model. The formation of rings in the vapour impedes ideal gas behaviour in the gas phase, resulting in the atypical slope for the vapour pressure's temperature dependence (subsection 4.3.4)

For the PJVP-D model, the critical point of a potential defect-driven phase separation [140] – corresponding to the temperature, where the total number of defects exceeds the number of molecules with two neighbours – at a temperature of about 295 K, coincides with the occurrence of larger voids in the liquid phase. Furthermore, in similarity to patchy particles with dissimilar patches [143], the PJVP-D model's structure is determined by two independent interaction schemes. Molecules forming a cluster are bound by Coulomb interaction (the LJ interaction is repulsive for an H-bond), while the inter-cluster interaction is a combination of weaker interactions, like LJ attraction and alignment of dipolar chains. Consequently, the PJVP-D model's similarity to these systems, despite the clear differences, can potentially help to obtain a deeper analysis of the origin of voids as is presented here.

Several auxiliary simulations have been performed to ensure, that the presence of voids is neither an artefact of the model – the polarizable model in general or the DO in particular –, insufficient equilibration, nor a finite size effect, caused by inappropriate choice of the simulated system. A MD simulation of the JVNP model [79], the non-polarizable predecessor of the JVP model, at liquid/vapour coexistence was carried out at temperatures of 260 K and 365 K. Since the non-polarizable model does not suffer the same drawback with regards to parallelization as the DO model, the system was equilibrated for as long as 20 ns to ensure that the voids are no metastable artefact of a non-equilibrium state. Interface detection with GITIM revealed the presence of voids at both simulated temperatures, similar to the PJVP-D model. Although the voids generally appear to be slightly smaller

on average than for the polarizable model, the overestimation of the number of hydrogen bonds and the tendency to form heavily branched structures is a reported failure of the JVNP model [79, 135]. Furthermore, liquid bulk simulations in the isothermal-isobaric ensemble (similar: subsection 3.3.1) along the isotherm at 260 K (thereby below the model's boiling point) also showed large fluctuations in the local density and the presence of voids when analysed with the GITIM method.

In summary, the presence of voids in the liquid phase of the system at liquid/vapour coexistence is neither an artefact of the polarizable model, insufficient equilibration, nor exclusively caused by the presence of an interface. Conversely, the interface's independence and the distribution of voids cannot be concluded from these data and a more detailed analysis of the size distribution of voids and their dependence on temperature, pressure, and the proximity to interfaces is necessary.

### 4.3.3 Network Analysis at Liquid/Vapour Coexistence

The formation of hydrogen bonded chains in liquid HF has been subject to many computational and experimental studies [2, 3, 11, 12, 24, 27, 83, 135, 144, 145]. Due to the various approaches, either using geometric criteria [24, 27, 83] or an energetic argument [12, 135], to identify hydrogen bonded networks in computer simulations of HF, comparability of different studies suffers strong limitations. Here, a geometric criterion involving the cutoff distance  $d_{H-bond}$  is used. Two HF molecules are identified as H-bonded, if the H atom of one molecule is closer to the F atom of the other than the first minimum of  $g_{HF}(r)$ . No angular criterion for the H-bond is applied. Additionally, the information provided by application of GITIM is used to describe the number of bonds per molecule and the average cluster size, with respect to individual layers. A cluster is determined to belong to a layer, if any of its molecules is identified as part of the examined layer.

Network properties, including identification of all substructures of branched clusters, were determined using the code described in section 4.2. Clusters as a whole entity were determined by a cutoff based cluster search algorithm [121] implemented in PYTIM.

Figure 16 illustrates the system at various temperatures, coloured according to the molecule's affiliation to monomers, linear chains, branched clusters, or rings. The extension of large, dangling clusters into the gas phase along the interface is clearly visible (compare figure 15). This structural phenomenon is unique for HF and in clear contrast to other strongly dipolar liquids like water [38, 39, 73, 106, 107, 121, 122, 133]. Additionally, the presence of all four different structural variants throughout the liquid phase and the formation of chains and rings in the vapour phase [37, 83] can be observed. The existence of multiple separate clusters and chains, as well as the large number of molecules forming branched structures

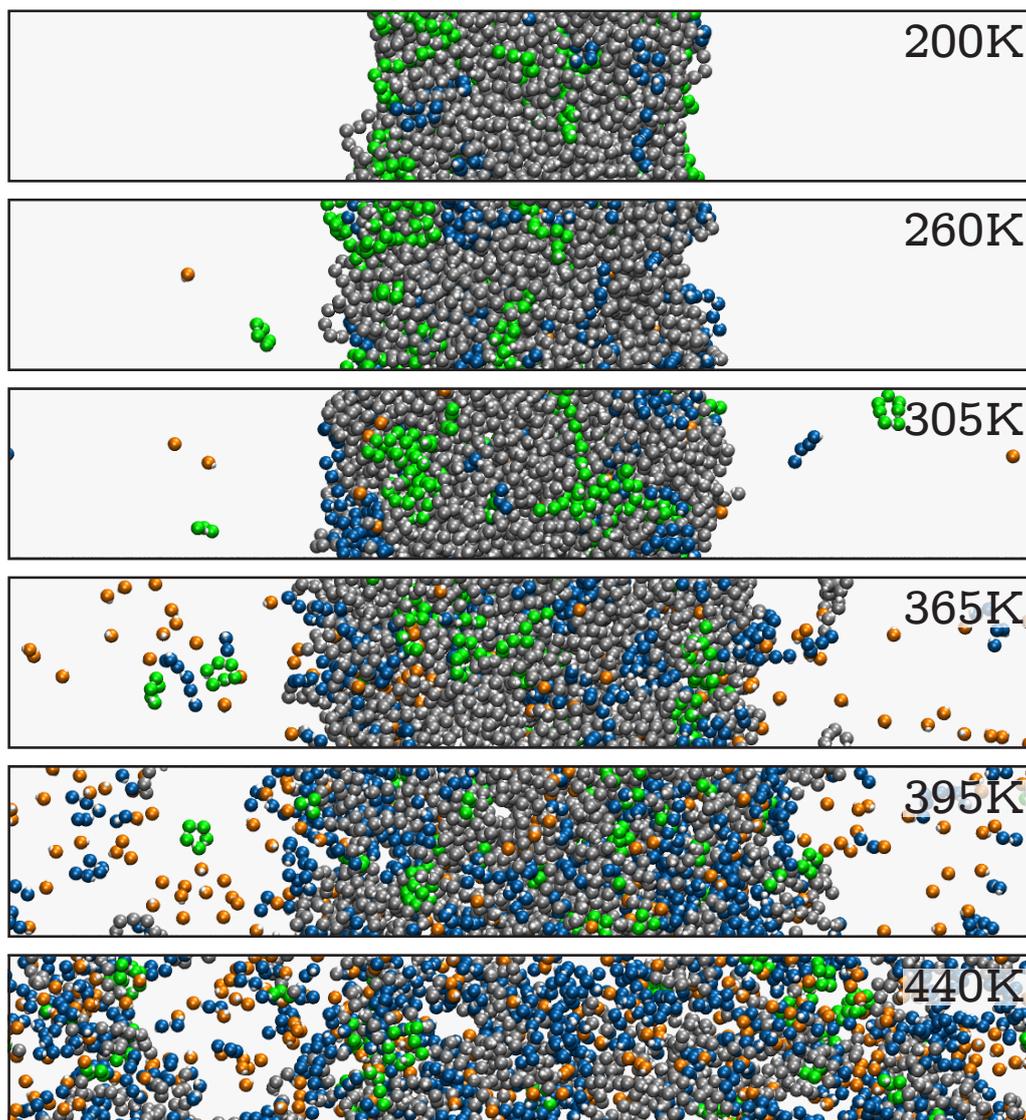
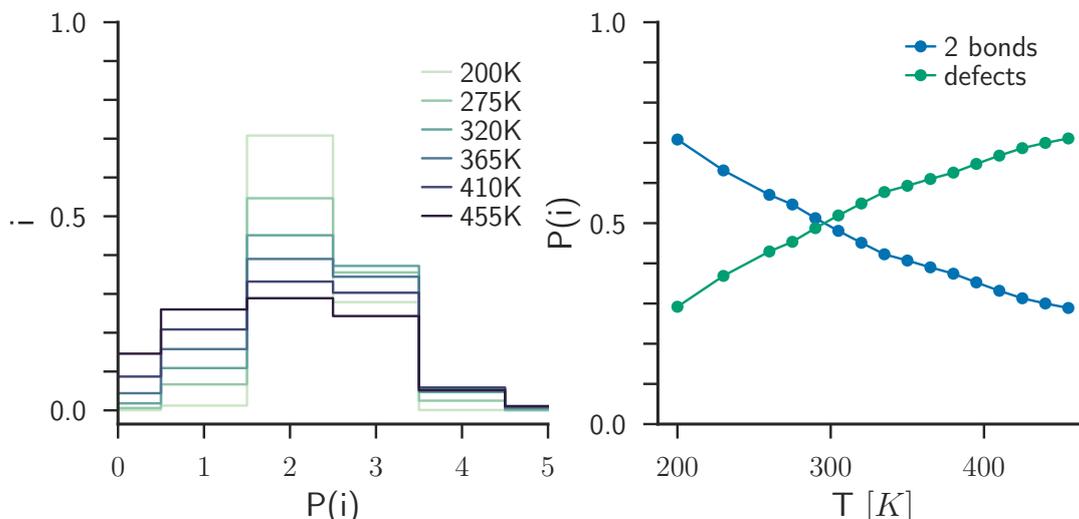


Figure 16: Sample configuration of the system at six different temperatures. The molecules are coloured according to the type of structure – monomers (orange), linear chains (blue), branched clusters (grey), and rings (green) – they belong to. The black box represents the simulation box with PBC.

is also visible. In previous studies branching was generally considered as a side effect [12, 24] or simply neglected [83, 146]. However, the results presented here suggest that the importance of branching to the structure of the liquid phase is more significant than previously suggested.



(a) Distribution of the hydrogen-bond number per molecule for six different temperatures

(b) Percentage of defects (blue) and molecules with two H-bonds (green).

Figure 17: Hydrogen-bond per molecule

## Bonds

The distribution of the number of H-bonds per molecule for multiple temperatures is shown in figure 17a and the average number of H-bonds per molecule  $\langle N_{HB} \rangle$  is listed in table 5a. For all simulated temperatures at least one quarter of the molecules forms exactly two bonds, therefore rendering the formation of linear segments in the liquid phase an important structural feature. This general trend for forming linear segments is consistent with previous studies [12, 27, 83, 135]. The fraction of molecules with two bonded neighbours varies from 70.8% at 200 K to only 28.8% near the critical point at 455 K. At 305 K already only 48.1% of the molecules have two bonded neighbours and 41.0% contribute as branching points by forming three or more H-bonds. Furthermore, the concentration of three bonds has a maximum at intermediate temperatures. It first increases from 28.9% at 200 K to a maximum of 37.4% at 335 K before decreasing again to a minimum of 24.3% at 455 K. The number of molecules with 4 hydrogen bonded neighbours reaches a plateau at around 6% for temperatures higher than 305 K and is negligible for low temperatures. Moreover, the number of 5 bonds per molecule never exceeds 1% and is probably an artefact of sensitivity of the cutoff-parameter or a consequence of the simple bonding criterion that certainly overestimates the number of bonded neighbours by neglecting angular dependence. More than 4 bonds would also not reflect the physical properties of HF, since in general every fluorine atom can bind up to three hydrogen atoms, resulting in a maximum of 4 H-bonds per individual HF molecule [12].

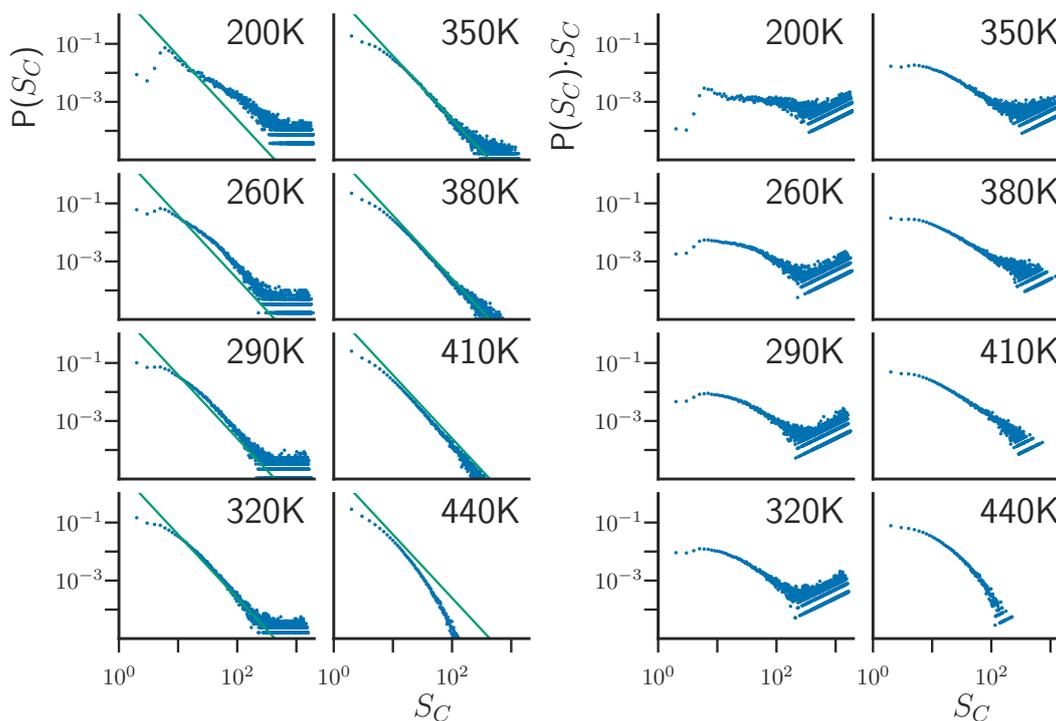
In addition to multiple bonds, the number of terminal chain molecules – molecules with only one bonding neighbour and therefore at the end of a chain – significantly increases to up to 25% for higher temperatures. If both terminal molecules and branches are considered as defects from the energetic optimum of two hydrogen-bonds [140, 143], the total number of defects exceeds the number of molecules with two neighbours for temperatures higher than 295 K, coinciding with the occurrence of larger voids in the liquid phase. The temperature dependence of defects and molecules with exactly two bonds are shown in figure 17b.

A layer-by-layer perspective shows that the percentage of molecules forming 4 bonds is approximately half as large in the first layer as in any other layer. For temperatures below 320 K no other significant trend regarding the distribution of H-bonds per molecule in individual layers can be observed. Conversely, the percentage of loose ends and the number of monomers is considerably larger in the first layer than in all subsequent layers for temperatures above 320 K. However, the large number of monomers in the first layer might be a result of the large cutoff  $d_{liquid-cluster}$  necessary to identify the liquid layer, as gas-phase monomers that are close to the interface can be falsely tagged as liquid. The difference between the first and subsequent layers increases monotonic, with 25% more chain ends and monomers at 410 K. Although the layer-by-layer contributions of the intrinsic density profile are rather sensitive to the choice of the GITIM parameter  $\alpha$ , the distribution of the hydrogen bond number per molecule for each individual layer is not.

### Clusters, Branches, Linear Chains and Rings

A group of HF molecules, where every molecule is H-bonded to at least one other molecule in the group is called a cluster and the cluster size  $N_C$  is the number of its constituting molecules. The distribution of cluster-size and the distribution of the number of molecules to clusters of a given size, referred to here as the cluster and molecular size distributions, are shown in figure 18.

In a three dimensional system at the percolation threshold the cluster size distribution follows a power law:  $P(N_C) \approx N_C^{-\tau}$ , with an exponent of  $\tau = 2.19$  [147–149]. Similarly, the molecular size distribution follows the same power law but with exponent  $\tau' = 1.19$ . In general, a system is percolating, if the distribution is slightly below the critical line at intermediate  $N_C$  and strongly exceeds the critical line for large  $N_C$ . However, finite size errors due the overestimation of clusters containing almost all molecules in the system, resulting in a hump in the distribution at large cluster sizes even in a non-percolating system, impedes accurate detection of the percolation threshold. Moreover, poor statistics represented by the large statistical noise of the distribution make an observation of the drop at intermediate cluster sizes impossible. Nonetheless, the power law is a great reference for describing the



(a) Cluster-size distribution of the H-bonded clusters (dots). The solid lines indicate the predicted distribution at the percolation threshold.

(b) Molecular-size distribution of the H-bonded clusters (dots). The solid lines indicate the predicted cluster size distribution at the percolation threshold, scaled by the cluster size.

Figure 18: Distributions of a) clusters-size and b) number of molecules in a clusters of a given size, at eight selected temperatures, plotted on a logarithmic scale.

cluster size distribution, and can be used to identify percolating and non-percolating systems far from the percolation limit. The distributions for the PJVP-D model show, that the system is certainly percolating at temperatures below 320 K and contrarily not percolating above. An exact determination of the percolation threshold in the remaining temperature range is not possible by only using these distributions.

Although the distribution at small  $N_C$  is only slightly below the general trend for larger temperatures, clusters containing four or fewer molecules are far less frequent below 320 K and especially clusters containing three molecules are rare. In contrast, the number of clusters with five to eight molecules is almost constant for all temperatures, with significant deviations only at 200 K, where clusters with six molecules are slightly more frequent and the number of clusters containing five molecules is slightly lower than on average.

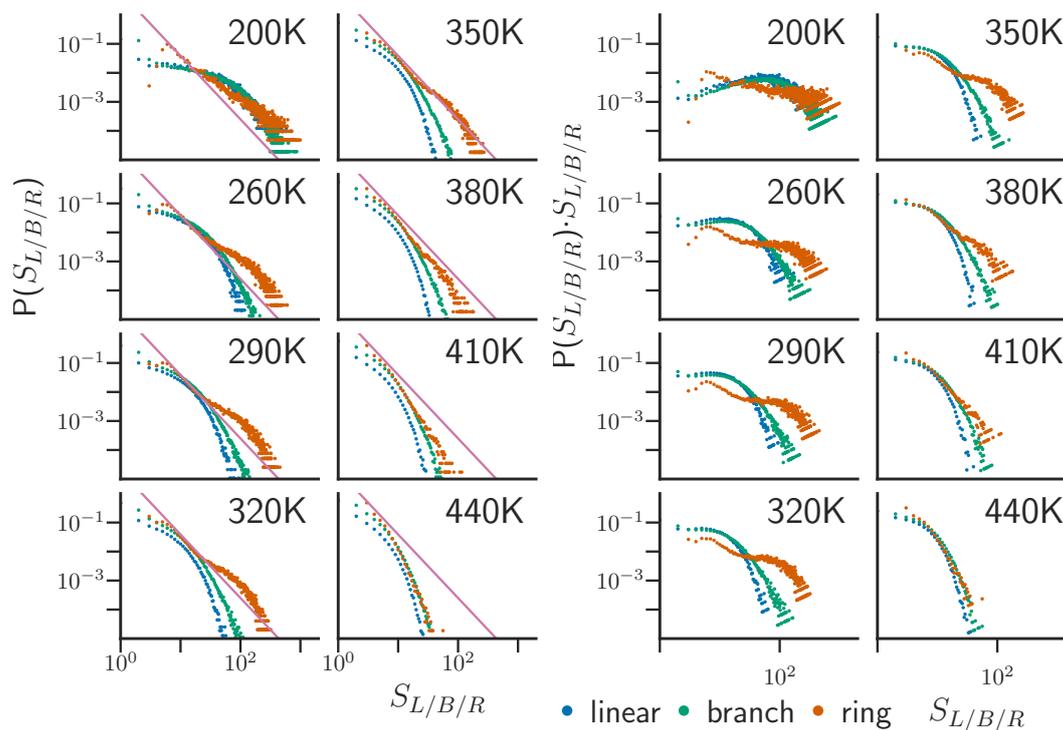
Especially at low temperatures, few molecules are part of dimers and trimers, pointing to a reduced stability of these clusters in the gas phase. In contrast to this dependence for smaller cluster sizes, the molecular size distribution over the

complete range of  $N_C$  shows, that the exponential trend predicted by percolation theory is only correct for temperatures above 380 K. An almost uniform cluster size distribution for HF at ambient conditions has been reported elsewhere [27]. Such a distribution cannot be observed at coexistence and similarity to a uniform distribution is only partially given at 200 K. However, the two simulations cannot be compared definitively due to the different pressure. in the bulk and at coexistence. The data provided here are in far better agreement with results provided for the cluster-size distribution of the non-polarizable JVNP model [12]. Below the critical point the probability of a single molecule to form a cluster of size  $N_C$  has a maximum of around 6. An additional maximum of the molecular size distribution for  $N_C$  larger than 500 is a consequence of the overpopulation of clusters containing almost all molecules caused by the finite system size.

For a temperature of 200 K, the probability of a molecule to be part of a cluster containing fewer than 900 molecules (45%) in the simulation box is about as large as the probability for cluster with more than 900 molecules. At 365 K this percentage already exceeds 90% and at 455 K the probability to be part of a cluster with fewer than 100 molecules already rises to about 85%. This data illustrates the structural change of the liquid phase from few, large and heavily branched clusters to multiple, smaller and less branched clusters. However, layer-by-layer analysis of the cluster analysis shows little conclusive trends.

In table 5b the average number of clusters  $\langle N_C \rangle$  and the average cluster size  $\langle S_C \rangle$  are summarized. Notably, those numbers merely reflect the finite system size and do not bear any direct physical meaning. The average cluster size  $\langle S_C \rangle$  decreases rapidly from 147.0 at 200 K to only 5.9 at 455 K. An opposing dependence on temperature is true for the average number of clusters  $\langle N_C \rangle$ . Monomers are not included in the definition of a cluster, hence the average number of monomers can be extracted as the difference of the number of molecules in the box and the product of  $\langle N_C \rangle$  and  $\langle S_C \rangle$ . The number of monomers increases more than linearly from 0% at 200 K to 14.7% at 455 K.

Breaking down the clusters into constituent species allows for further analysis of the network properties of HF. Since linear chains and rings are referred to as the dominant structures in HF [2, 3, 11, 12, 24, 27, 83, 135, 144, 145] clusters are categorized as linear chains, rings, and linear branches. Linear chains are completely isolated chains without any branching point, monomers are not included. Rings can either be completely isolated or attached to a linear segment. Infinite chains caused by PBC are also classified as rings and the minimal number of molecules necessary to form an infinite chain through the simulation box is roughly 10. Linear branches are the building blocks of branched structures and can be attached to rings.



(a) Size distribution of the H-bonded structure. The three considered structures are linear chains (blue dots), linear branches (green dots), and rings (orange dots). The solid lines indicate the predicted distribution of the cluster size at the percolation threshold.

(b) Distribution of number of molecules of the H-bonded structure. The three considered structures are linear chains (blue dots), linear branches (green dots), and rings (orange dots).

Figure 19: Distributions of a) structure-size and b) number of molecules in a structure of a given size, at eight selected temperatures, plotted on a logarithmic scale.

The size and molecular size distribution of linear chains, branches, and rings, in equivalence to the two distributions for clusters, are shown in figure 19. The hump in the distribution of rings at large  $S_R$  is a result of the definition of infinite chains as rings. Consequently linear chains and branches show no such feature.

However, the molecular distribution of both linear chains and branches have a maximum at intermediate structure sizes for temperatures below 320 K. Contrarily, the molecular distribution of rings has a maximum at far smaller values. Previous studies suggest that rings containing three [4] or six to eight [24, 83] molecules are especially stable and rings with sizes up to 13 occur in isolation [83]. The data obtained for the system at liquid/vapour coexistence show good agreement with these studies at temperatures below 335 K. The anomaly of increased occurrence of rings with 13 molecules reported in reference [83] is also reproduced, although at a different state-point (liquid/vapour coexistence at 200 K). Since at this temperature not a single molecule is in the gas phase, the increased stability of  $S_R = 13$

likely originates from a dense arrangement inside a cluster. Additionally, very few molecules form rings containing three to five molecules at 200 K, indicating that these arrangements are predominantly formed in the gas phase. For temperatures above 335 K rings with 3 molecules become the most likely arrangement and the probability for larger rings decreases slightly faster than exponential, with rings of size  $S_R = 4$  being slightly less likely. The second hump in the distribution of rings, located at  $S_R > 30$  reflects the presence of spanning chains. Its effect on the cluster size distribution is clearly visible, especially when compared to the distribution of linear chains and branched structures. Per definition, these structures cannot cross the periodic boundaries and their distributions are therefore always below the percolation threshold and strictly monotonically decreasing.

Again, layer-by-layer analysis of the linear chain, branch and ring distribution shows little clear tendencies. Rings with 4 to 8 molecules are more common in layers close to the interface and the opposite is true for rings containing 3 molecules. Short linear chains are less frequent in the first layer than in any other layer for temperatures above 290 K. Additionally, small branches are far less probable in the first layer for temperatures below 365 K. However, these trends might be an effect of the clustering procedure necessary to identify the liquid phase, as has already been pointed out for the increased number of monomers in the first layer. Nonetheless, the dominance of monomers, small unbranched linear chains and rings, especially with  $S_R = 3$ , in the gas phase at all temperatures as reported in earlier studies [4, 24, 34, 83] is apparent from the data.

The average number and average size of linear chains ( $\langle N_L \rangle, \langle S_L \rangle$ ), linear cluster segments ( $\langle N_B \rangle, \langle S_B \rangle$ ) and rings ( $\langle N_R \rangle, \langle S_R \rangle$ ) are listed in table 5b. The average size of linear chains  $\langle S_L \rangle$  and linear segments  $\langle S_B \rangle$  both decrease from around 51 molecules on average at 200 K to only little below 4 at 455 K. Contrarily, the average number of linear chains  $\langle N_L \rangle$  constantly increases with temperature whereas the average number of branches  $\langle N_B \rangle$  has a maximum at 396 K. The average number of rings increases with higher temperatures to a maximum of 28.98 at 410 K before decreasing to 25.97 at 455 K. The percentage of molecules in rings is almost constant for temperatures below 335 K and decreases for higher temperatures.

Additionally, the percentages of molecules belonging to a monomer, a cluster, or one of its constituent forms are depicted in figure 20b as a function of temperature. The number of H-bonds per molecule as a function of temperature is shown for reference in figure 20a.

If considering the percentage of molecules, the maximum for linear branches shifts to 350 K. This temperature is comparable with the maximum for three H-bonds per molecule, indicating that the number of branching points per cluster drastically decreases with temperature. Consequently, the structure of the liquid phase changes from a phase of few heavily branched, large clusters at low temperatures

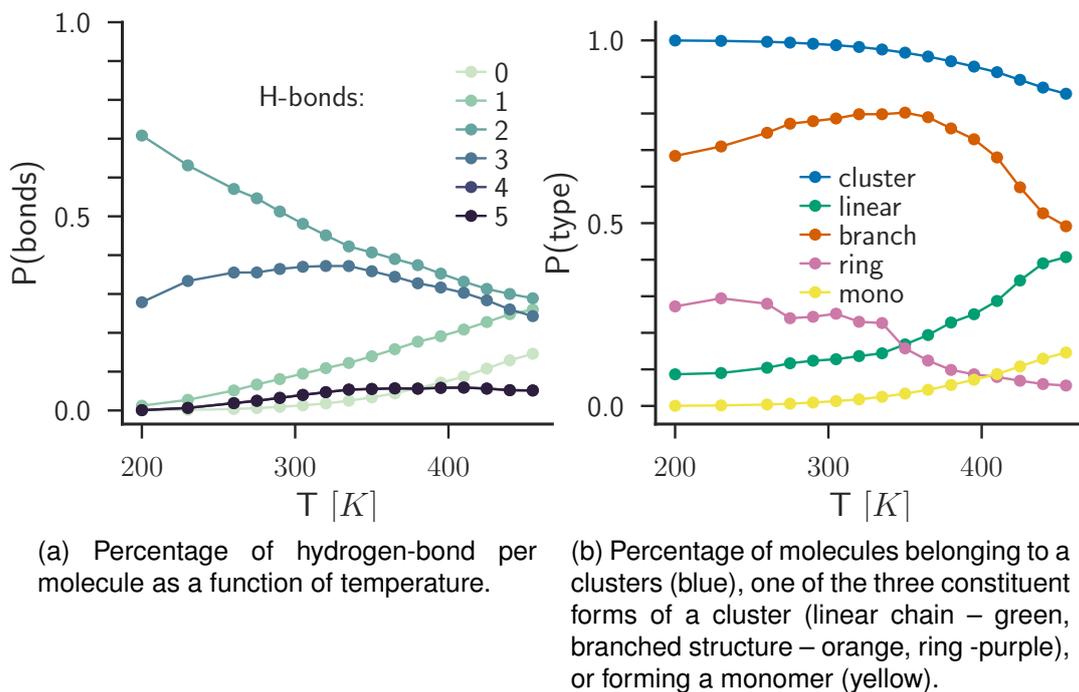


Figure 20: Network composition properties as a function of temperature.

to many small clusters with little branching at higher temperature. The presence of a maximum in the average number of branched structure suggests that at high temperatures ( $T > 350$  K) branching progressively loses importance as a mechanism to minimize the system's free energy.

#### 4.3.4 Thermodynamic Analysis of the Hydrogen Fluoride Interface

As motivated in chapter 3, the absolute value of the density as calculated by the polarizable models depends strongly on the system size. However, the temperature and pressure dependence of the density are the same for both the point dipole, as well as the DO models. The coexistence curve for both point dipole models has already been reported by *Pártay et al.* [30]. Consequently, the coexistence densities and the liquid-vapour coexistence curve are not reported here.

#### Vapour Pressure

The vapour pressure  $p_{vap}$  of the system is provided in GROMACS by the normal component of the pressure tensor in direction of the interface normal, hence  $p_{xx} = p_{vap}$ . In figure 21a the vapour pressure is shown and compared with experimental data of HF [150]. Although the PJVP-D model slightly overestimates the experimental slope in the vapour pressure, it shows a slight upward bent shape also present in the experimental data [35, 150]. A possible cause is the formation of rings in

the gas phase and the consequent deviation from ideal gas behaviour [37]. This atypical feature of HF [35] could not be reproduced by non-polarizable models like HFC [5] and consequently solidifies the superiority of the polarizable models for simulating liquid/gas coexistence. Contrarily, the overestimation of the vapour pressure, especially close to the critical temperature is a reported failure of the JVP and PJVP model [30]. As a consequence, the model's boiling point, estimated by an exponential fit of the computed data, is 9.6 K lower than the experimental value [101] at 283.1 K.

### Surface Tension

The surface tension  $\gamma$  of an interface can be calculated as the average difference between components of the pressure tensor normal and tangential with respect to the interface normal, normalized by the length of the simulation box and the number of interfaces present in the box [37, 105, 106]. These components are calculated from the virial within GROMACS and long-range contributions to the pressure tensor are accounted for via the PME correction of the pressure.

The surface tension of other strongly dipolar fluids like water has been subject to multiple studies and is available for the complete temperature range [13]. In contrast, experimental data on the surface tension of HF [14] are only available for temperatures between 190 K and 290 K. Computational studies on the coexistence properties of HF [34–36] suggest that the surface tension of HF is significantly different from that other dipolar fluids. The predominant linear arrangement in hydrogen bonded clusters and the high critical density of HF, are suspected to affect the lighter slope and far smaller absolute value of the surface tension compared to water [36]. The surface tension of HF calculated for the PJVP-D model, the experimental results for HF [14], and the experimental surface tension of water [13] are compared in figure 21b. Although the results show a clear difference between the surface tension of the PJVP-D model and water, the latter having a far steeper slope and being about 5 to 10 times higher, the experimental values for HF do not match with the PJVP-D model. The computed slope for HF is slightly steeper than the experimental one and the computed values are approximately  $7.0 \text{ N m}^{-1}$  higher than the experimental surface tension. Considering the models' initial purpose of simulating the liquid bulk phase of HF, its ability to qualitatively reproduce to the extremely low surface tension of HF is surprising. Moreover, the surface tension calculated here is only an estimate for a planar interface of a liquid phase without significant voids and a more accurate calculation is expected to result in an even lower value. Analysing the models ability to reproduce the experimental surface tension therefore necessitates further investigation.

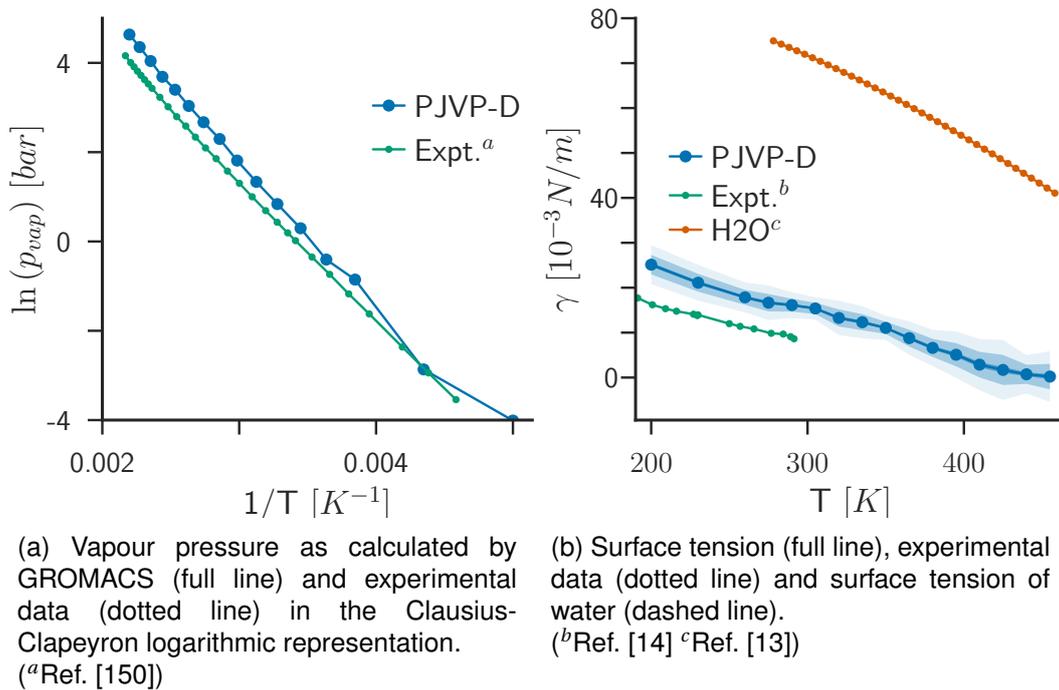


Figure 21: Thermodynamic properties of the PJVP-D model as a function of temperature. Errors are illustrated by the shaded halo and represent deviations corresponding to  $[\pm 1SD, \pm 5SD, \pm 10SD]$ . Error for the vapour-pressure do not exceed the symbol size and are omitted.

## Density Profiles

The regular density profile  $\rho(x)$  can be obtained by dividing the simulation box into slabs of equal volume and equal extent in direction of the interface normal. The number density in each slab is then obtained as the time-average of the number of particles in the respective slab divided by its volume. Contrarily, the intrinsic density profile  $\tilde{\rho}(x')$  is calculated by computing the distance of each molecule in the box to the interface. Since the interface is not planar, especially for high temperatures, distance calculation within PYTIM was performed using the distance measure for a generic surface. In the current implementation, the distance of a single molecule to the interface is calculated as the absolute distance, called intrinsic distance here, to the closest atom of the first layer. A molecule is inside the liquid phase if a sphere of radius  $\alpha$  around its F site contains another molecule and in the gas phase otherwise. The intrinsic density profile  $\tilde{\rho}(x')$  is obtained by calculating the number of molecules within intervals of this calculated intrinsic distance, normalized by the volume representing a single interval.

In figure 22 both intrinsic and regular density profile are reported for six selected temperatures. By construction, the first layer contributes to the intrinsic profile as a delta function and by definition liquid molecules have a negative distance to the interface. Both profiles show the decrease in density of the liquid phase with

temperature, represented by the maximum of the sigmoidal shape for the non-intrinsic and the convergence value for the intrinsic profile, but their representation of the interface differs drastically. The regular profile  $\rho(x)$  shows a broadening of both the liquid phase and the interfacial region for higher temperatures. Moreover, it becomes an almost uniform distribution for 440 K, which is to be expected at or close to the critical point. Although numerous voids are present in the liquid and long chains extend into the gas phase, the regular density profile  $\rho(x)$  does not differ significantly from the profile of a non-dipolar liquid like argon [105].

Contrarily, the intrinsic density profile  $\tilde{\rho}(x')$  features significant differences to the profile of simpler liquids. Instead of a structured profile consisting of contributions of successive layers [105], two narrow peak are present close to the interface. Those two peaks broaden slightly with increasing temperature, but do not get smaller with respect to the liquid bulk density. Further away from the first layer two additional maxima are visible for temperatures below 290 K. Although the latter broadens and disappears at around 290 K the first of these two peaks and its subsequent minimum are identifiable up to 380 K. Additional maxima farther away from the interface cannot be identified, indicating that the liquid/vapour interface of HF exhibits less distinguishable layers than for instance water or argon [105, 151]. At higher temperatures all but the two narrow peaks close to the interface have deteriorated into a shape similar to the sigmoidal shape of the regular density profile. Even at 440 K the first two peaks are still pronounced in the intrinsic profile, besides the almost uniform non-intrinsic density profile. This suggests the presence of larger clusters at the critical point and beyond. The existence of these clusters even at the critical point is also apparent from the network analysis data in the previous chapter. Due to the present implementation of the intrinsic distance, influence of the dangling clusters on the intrinsic density profile is drastically reduced since another molecule, that is also part of the 1<sup>st</sup> layer, but not part of such a cluster, is usually closer to the considered molecule.

The intrinsic density profile's shape is certainly related to the strong tendency to form linear chains and clusters in liquid HF. The layer-by layer contributions to intrinsic and regular density profiles are displayed in figure 23. Although the layer-wise contributions of the non-intrinsic density profile  $\rho(x)$  show no distinct feature that indicates any difference of the interface structure compared to non-polar liquids [39, 103, 105], the presence of large voids inside the liquid phase is clearly visible due to a non vanishing contribution to all layers for temperatures above 335 K. Using the larger GITIM parameter choice of  $\min(g_{FF}^{soft}(r))$  the effect can be eliminated for temperatures below 395 K solidifying the claim that voids are the origin of this atypical feature in the regular density profile. Moreover, the uniform contribution of these voids to the density profile shows, that the distribution of voids in the system is at equilibrium.

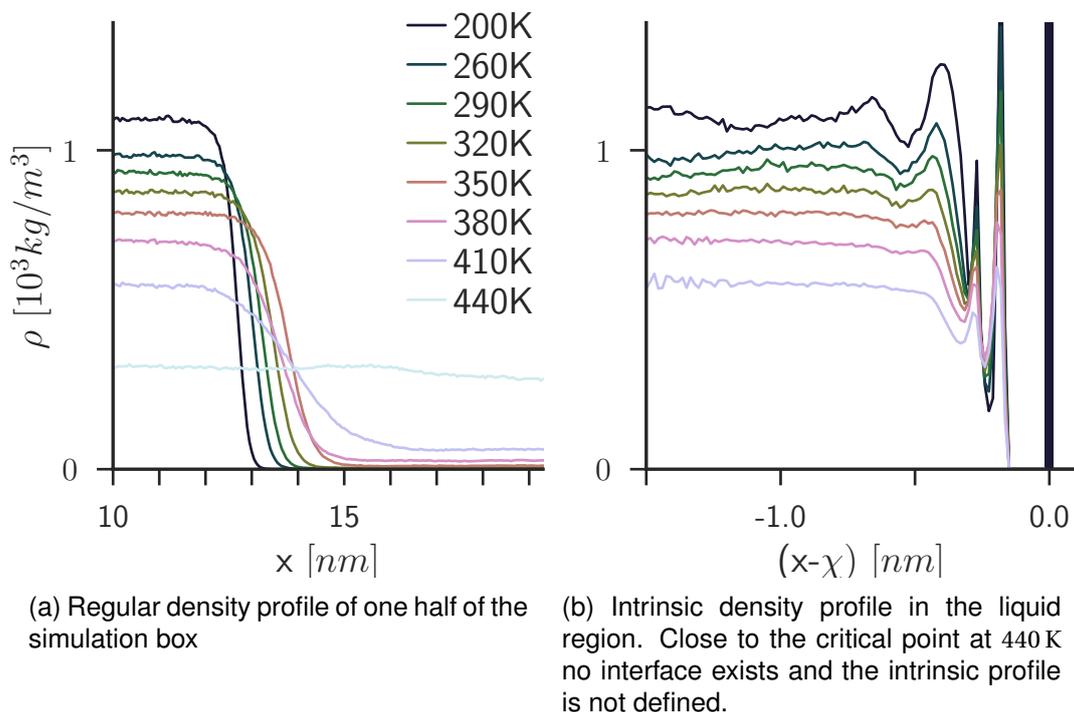
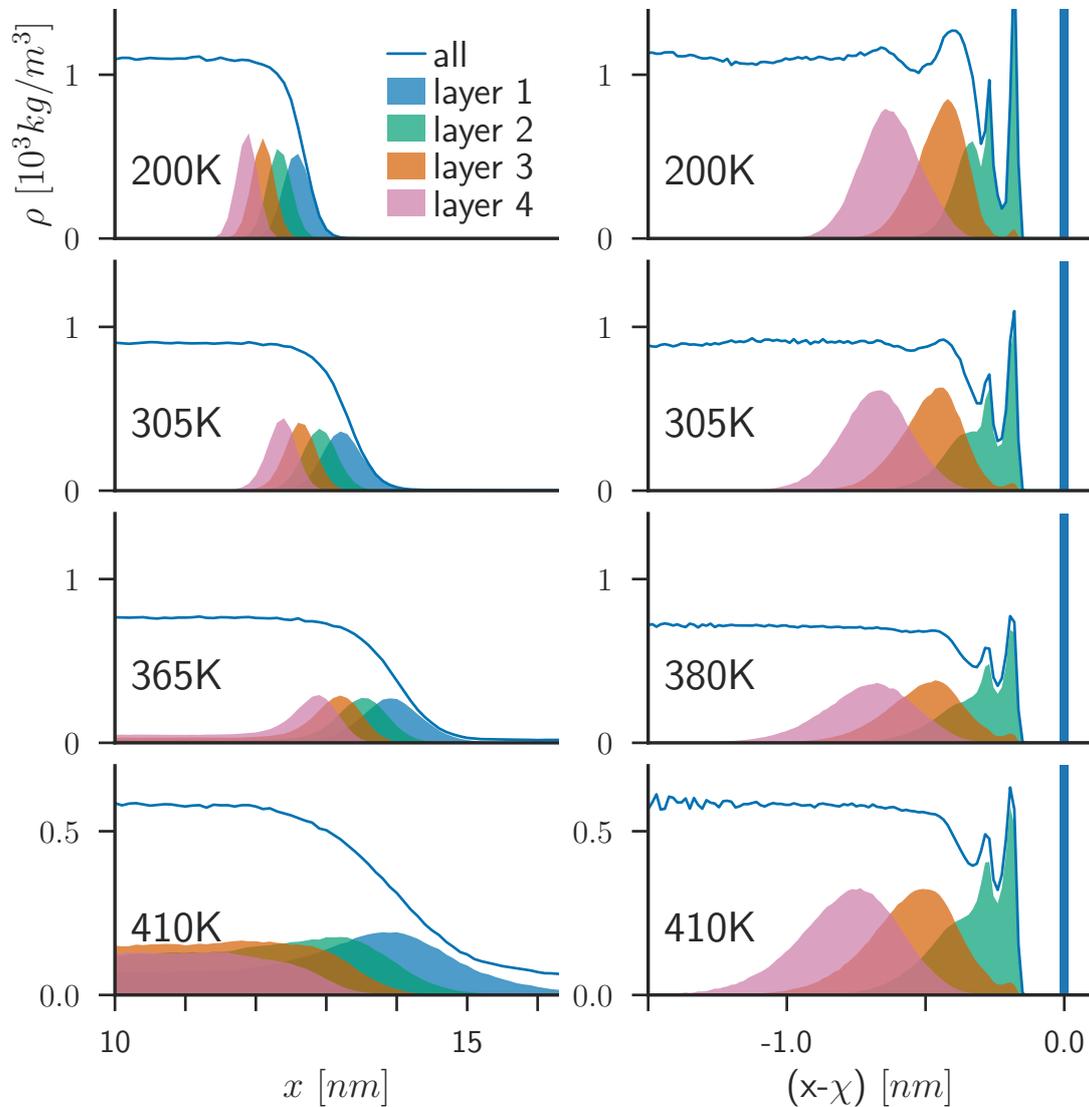


Figure 22: Both a) regular density profile and b) intrinsic density profile as function of temperature for 8 (respectively 7) selected temperatures.

Contrarily, individual layers of the intrinsic density profile  $\bar{\rho}(x')$  suggest that the existence of multiple H-bond between the first two layers cause the two, sharp spikes close to the interface. Both peaks, as well as an additional third peak that does not result in an individual peak in the overall profile, are attributed to the third layer. The positions of these three peaks coincide with the first peak of the  $g_{HF}(r)$ ,  $g_{FF}(r)$  and  $g_{FF}^{soft}(r)$ , hence the position of the first H and F atom of H-bonded neighbours inside a cluster and the closest molecule of a different cluster. The third of the second layer peaks is by far the broadest, resulting from the weaker inter-cluster bond. This effect correlates with the corresponding peak's width in the RDF. Since the third layer is close to the last peak of the second layer (they are only 0.06 nm apart), they are not separable in the overall intrinsic density profile. The the third layer's position also further confirms the presence of numerous chains passing through multiple layers, by coinciding with the molecule's distance to the second neighbour along a linear chain. Furthermore, it is slightly asymmetric, which is attributed to the packing of molecules near the interface [105] and most likely incorporates a contribution of inter-cluster bonds to this layer at around 0.55 nm (compare second maximum of  $g_{FF}^{soft}(r)$ ), similar to the third peak of the second layer. The third layer also has a small contribution at the position of the first minimum  $g_{HF}(r)$ . This contribution close to the interface is present for all temperatures, as well as any other tested choice of  $\alpha$ , which suggests that it is not an artefact of a faulty parameter choice. A probable source of the peak is a favourable, but rare arrangement of a branch in the first

---

layer, where each of the three molecules forming the branch are assigned to different layers. Finally, the fourth layer shows no significant features in the intrinsic density profile and starts to become strongly asymmetric in the regular density profile  $\rho(x)$  for temperatures above 395 K. It is still reasonably shaped in the intrinsic density profile  $\tilde{\rho}(x')$  for temperatures up to 410 K, although the corresponding peak in the overall profile disappears earlier.



(a) Intrinsic and non-intrinsic density profile and layer contribution at 3 selected temperatures

(b) Liquid part of the intrinsic density profile (solid line) for three different temperatures. The separate contributions of the first (blue) second (green), third (orange), and fourth (purple) layer are shown, although per definition the first layer only contributes to the delta peak at 0.

Figure 23: Density profile a) regular and b) and layer contribution at four selected temperatures. Note the different vertical scaling of the lowest graph.

## 4.4 Concluding Remarks for Interface Analysis

In conclusion, the presented results show great qualitative agreement of the PJVP-D model with HF at liquid/vapour coexistence and represent the largest MD simulation of a polarizable HF model reported to date. The use of modern MD simulation packages was made possible by the adaptation of the polarizable PJVP-D model and its applicability for this setup. The model is able to qualitatively reproduce the extremely low surface tension and the atypical slight concave-up bent of the vapour pressure curve of HF. This is especially notable due to the model's origin as a liquid bulk model [29]. Moreover, the simulations reveal the fascinating morphology of the liquid gas interface of the PJVP-D model and help to better understand the structure of the liquid phase and its composition of multiple branched and unbranched clusters.

However, several difficulties impede a definitive interpretation of the presented results and further data are certainly necessary to allow for accurate descriptions of the liquid/gas interface of HF. The major insights regarding the adversities present for interfacial analysis are summarized here.

As noted in previous studies [37, 152], a large extension of the simulation box in direction of the interface normal is necessary to allow for accurate simulation of liquid/vapour interfaces close to the critical point. Consequently, the box size used here for the system might have been too small to allow for two stable, independent interfaces for temperatures above 425 K. Moreover, the box size in the other two dimensions might have considerable effects, due to the presence of long chains and clusters. Additionally, the equilibration time used for this system is considerably shorter than for the comparable systems used in reference [37]. Shorter equilibration times can certainly be expected for the polarizable model [21] and the presence of voids in the liquid phase is (as discussed in subsection 4.3.2) an inherent feature of the simulated system. However, it is still unclear whether long equilibration times might eventually decrease the number and size of the voids, especially at higher temperatures. Both a larger system and longer simulation times are necessary to identify and eventually eliminate artefacts related to those shortcomings.

Further research on the presence of voids, in liquid HF in general and how they need to be taken into account for interfacial analysis in particular, is certainly necessary. Especially their effect on the calculation of the surface tension of the liquid phase remain an open point. The proper choice for the parameter  $\alpha$  with HF and the differences to other liquids also necessitates in-depth analysis and the discussion presented here is incomplete. The peak- and layer-matching approach presented here might only be valid for apolar liquids and was only performed at a single temperature. However, the importance of inter-cluster distances and voids on the parameter choice suggests that the ideal choice of  $\alpha$  depends on the state point

and is not universal for each fluid. Supplementary investigations about the origin of voids in liquid HF and the similarities of HF to the extraordinary critical behaviour of dipolar hard spheres [137–140] and simple dipolar models [36, 37, 141–143] could help to better understand the surface properties of HF.

Lastly, the intrinsic density profile and its layer-by-layer contributions have been shown to be atypical compared to simpler liquids [39, 105]. However, the highly corrugated nature of the interface and large number of isolated chains and clusters extending into the gas-phase certainly affect the intrinsic representation. The increase of such effects with higher temperatures results in a profile, strongly smeared out beyond the two inter-cluster peaks, for one third of the temperature range of liquid HF. Moreover, the current implementation of the GITIM method, only considering the closest interfacial molecule, eliminates the region containing the dangling chains from the intrinsic profile. To also account for effects caused by those objects a different approach for treatment of interfaces of that kind is necessary.

## 5. CONCLUSION

This thesis is devoted to the study of the structural and topological properties of HF along the liquid/vapour coexistence line. Molecular Dynamics simulation was chosen as the preferred tool for this investigation, because the relevant observables, in case of HF, are not easily accessible experimentally.

The first part of this work is focused on the development of a computationally efficient model for the simulation of HF, while the second part describes its application in the actual investigation of the properties of its liquid/vapour interface.

Liquid HF is characterised by the presence of long molecular chains, as well as by an extremely low surface tension and, therefore, pronounced long-wavelength interfacial fluctuations. These peculiarities require the usage of a relatively large simulation box (2000 molecules), especially considering the necessity of employing computationally expensive polarizable models. In order to handle these large interfacial systems with efficient simulation codes, it was necessary to reimplement the polarizable PJVP model [30] by exchanging the point-dipole, responsible for modelling polarizability, with a Drude oscillator [40]. This adaptation allowed to run the simulations using the GROMACS [31, 32] simulation package. The Drude oscillator model has been tested by performing isothermal-isobaric simulations in the liquid bulk at ambient conditions, and reproduced most structural and thermodynamic properties of the point-dipole model [12, 29, 30, 66]. However, closer inspection revealed that the density of both polarizable models and their non-polarizable counterpart is extremely sensitive to small changes of the Lennard Jones parameters [79], likely due to large local density fluctuations and the presence of spanning clusters. Despite this conceptual problem in reproducing the density of the point-dipole model, equivalence with respect to structural and thermodynamic properties of point-dipole and Drude oscillator could clearly be demonstrated for the HF models [21, 69, 72, 76].

With a Drude oscillator model at hand, it has been possible to perform molecular dynamics simulations of systems with 2000 HF molecules along the complete liquid/vapour coexistence temperature range. The intrinsic and layer-by-layer analysis of the liquid/vapour interface revealed large qualitative and quantitative differences with respect to other polar liquids, further illustrating the exceptional properties of HF.

Firstly, the simulations revealed that large voids, comparable to the size of a clusters with up to 10 molecules, are present in the liquid phase of HF. Although, they certainly occur more often at liquid/vapour coexistence, similar voids could also be identified for various liquid bulk simulations. A network analysis performed on the liquid/vapour system identified multiple independent H-bonded clusters

in the liquid phase. The clusters can either be linear chains, rings or strongly branched structures and the large density fluctuations and voids are possibly a result of the cluster's inability to form a space-filling arrangement [135]. Separate evaluation of contributions from molecules within the same cluster and molecules of different clusters to the F-F radial distribution function, shows that the second peak of the RDF at temperatures below 305 K originates from the contribution of inter-cluster bonds. Moreover, the same contribution can be identified over the complete temperature range, although no independent peak is visible in the overall radial distribution function, due to the dominance of the first peak. The weaker bond (compared to the intra-cluster hydrogen bond) between two clusters is certainly connected to the presence of voids. The presence of voids and whether they need to be included as part of the interface or not remains an open question. Additionally, the occurrence of voids might have gone unnoticed for simulations of other molecular liquids at coexistence since they can only be detected by algorithms for arbitrary interfaces like GITIM [39] and not by its predecessors for planar interfaces.

Secondly, the simulations reveal the highly corrugated nature of the liquid/vapour interface of HF, especially at higher temperatures. Large dangling clusters that extend into the gas phase and large pockets along the interface impose high demands on computational schemes for interface identification. Furthermore, the network properties of HF along the liquid/vapour coexistence show that multiple features of the liquid phase change at temperatures around the percolation threshold temperature  $T_p \approx 350$  K. Beyond  $T_p$ , not only the number of branched structures, but also the number of molecules with three H-bonded neighbours decreases and the number and size of voids increases.

Thirdly, it could be shown that the adapted model is capable of reproducing the thermodynamic properties of HF at coexistence, most notably the peculiar slope of the vapour pressure and the extremely low surface tension. The qualitative agreement with experimental coexistence data underlines the transferability of the the model and its applicability for both liquid and vapor phases [29, 30]. Thereby, these results reaffirm the importance of including polarizability into an empirical, three-site, bulk model for HF. The overestimation of the surface tension is likely caused by an overly simplistic numerical calculation and the disregarded contributions of voids and strong surface fluctuations. Even better agreement with experimental data is expected for a more accurate calculation of the surface tension.

Finally, analysis of the density profile of the liquid/vapour interface of HF displays significant differences from apolar molecular liquids and water [39, 105]. The shape of the intrinsic density profile reveals clear layers at the interface despite its high corrugation. However, individual layers at the interface are not equally as pronounced as in other molecular liquids [103, 105, 151]. At most four distinctive

layers (at the lowest evaluated temperature) could be identified. Moreover, no percolating lateral network, as is present for water [38, 121, 133, 134], could be identified in those layers. The absence of a clear layering structure beyond the second layer for higher temperatures is likely caused by clusters passing through multiple layers and the strong corrugation of the interface. The threefold contribution to the third molecular layer indicates H-bonded clusters spread both parallel and perpendicular to the interface, possibly smearing out the layering. None of these features is visible in the regular density profile. The failure of the regular density profile to identify the unique structure of the liquid/vapour interface clearly shows the importance of intrinsic analysis for the proper description of soft interfaces.

In conclusion, this thesis represents the first investigation on the microscopic structure of the liquid/vapour interface of HF, which turned out to have an extremely complex morphology. These findings reaffirmed the unique nature of HF among molecular liquids, but also showed its importance as a model to help improve algorithms and the general understanding of the intrinsic properties of fluid interfaces.

## Bibliography

- [1] T. Pfleiderer, I. Waldner, H. Bertagnolli, K. Tödheide, and H. E. Fischer, *The structure of liquid and supercritical deuterium fluoride from neutron scattering using high-pressure techniques*, J. Chem. Phys. **113**, 3690 (2000), ISSN 00219606.
- [2] S. E. McLain, C. J. Benmore, J. E. Siewenie, J. Urquidi, and J. F. Turner, *On the structure of liquid hydrogen fluoride*, Angew. Chemie - Int. Ed. **43**, 1952 (2004), ISSN 14337851.
- [3] S. E. McLain, C. J. Benmore, J. E. Siewenie, J. J. Molaison, and J. F. Turner, *On the variation of the structure of liquid deuterium fluoride with temperature*, J. Chem. Phys. **121**, 6448 (2004), ISSN 00219606.
- [4] M. Kreitmeir, G. Heusel, H. Bertagnolli, K. Tödheide, C. J. Mundy, and G. J. Cuello, *Structure of dense hydrogen fluoride gas from neutron diffraction and molecular dynamics simulations*, J. Chem. Phys. **122** (2005), ISSN 00219606.
- [5] M. E. Cournoyer and W. L. Jorgensen, *An improved intermolecular potential function for simulations of liquid hydrogen fluoride*, Mol. Phys. **51**, 119 (1984), ISSN 13623028.
- [6] J. S. Muentzer and W. Klemperer, *Hyperfine Structure Constants of HF and DF*, J. Chem. Phys. **52**, 6033 (1970), ISSN 0021-9606.
- [7] S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman, *Dipole moment of water from Stark measurements of H<sub>2</sub>O, HDO, and D<sub>2</sub>O*, J. Chem. Phys. **59**, 2254 (1973), ISSN 0021-9606.
- [8] J. W. Gibbs, *Elementary Principles of Statistical Mechanics*, vol. 1 (Scribner, C., New York, 1902), ISBN 978-0486789958.
- [9] B. Desbat, *Structure of liquid hydrogen fluoride studied by infrared and Raman spectroscopy*, J. Chem. Phys. **78**, 6377 (1983), ISSN 0021-9606.
- [10] W. L. Jorgensen, *Monte Carlo Simulations of Liquid Hydrogen Fluoride*, J. Am. Chem. Soc. **100**, 7824 (1978), ISSN 15205126.
- [11] M. Deraman, J. Dore, J. Powles, J. Holloway, and P. Chieux, *Structural studies of liquid hydrogen fluoride by neutron diffraction*, Mol. Phys. **55**, 1351 (1985), ISSN 0026-8976.
- [12] R. Vallauri and P. Jedlovsky, *Structural properties of liquid HF: a computer simulation investigation*, Mol. Phys. **93**, 15 (1998), ISSN 0026-8976.

- 
- [13] N. B. Vargaftik, B. N. Volkov, and L. D. Voljak, *International Tables of the Surface Tension of Water*, J. Phys. Chem. Ref. Data **12**, 817 (1983), ISSN 0047-2689.
- [14] J. H. Simons and J. W. Bouknight, *The Density and Surface Tension of Liquid Hydrogen Fluoride*, J. Am. Chem. Soc. **54**, 129 (1932), ISSN 0002-7863.
- [15] E. Chacón and P. Tarazona, *Intrinsic profiles beyond the capillary wave theory: a Monte Carlo study.*, Phys. Rev. Lett. **91**, 166103 (2003), ISSN 0031-9007.
- [16] E. Chacón and P. Tarazona, *Characterization of the intrinsic density profiles for liquid surfaces*, J. Phys. Condens. Matter **17**, S3493 (2005), ISSN 0953-8984.
- [17] W. L. Jorgensen and M. E. Cournoyer, *Quantum and statistical studies of liquids. 1. An intermolecular potential function for the hydrogen fluoride dimer from ab initio 6-31G computations*, J. Am. Chem. Soc. **100**, 4942 (1978), ISSN 0002-7863.
- [18] M. L. Klein, I. R. McDonald, and S. F. O'Shea, *An intermolecular force model for  $\{(HF)_2\}$* , J. Chem. Phys. **69**, 63 (1978).
- [19] W. L. Jorgensen, *Basis set dependence of the structure and properties of liquid hydrogen fluoride*, J. Chem. Phys. **70**, 5888 (1979).
- [20] M. L. Klein and I. R. McDonald, *Structure and dynamics of associated molecular systems. I. Computer simulation of liquid hydrogen fluoride*, J. Chem. Phys. **71**, 298 (1979), ISSN 0021-9606.
- [21] M. Schmollngruber, V. Lesch, C. Schröder, A. Heuer, and O. Steinhauser, *Comparing induced point-dipoles and Drude oscillators*, Phys. Chem. Chem. Phys. **17**, 14297 (2015), ISSN 1463-9076.
- [22] S. Izvekov and G. A. Voth, *Effective Force Field for Liquid Hydrogen Fluoride from Ab Initio Molecular Dynamics Simulation Using the Force-Matching Method*, J. Phys. Chem. B **109**, 6573 (2005), ISSN 1520-6106.
- [23] S. J. Wierzchowski and D. A. Kofke, *Hydrogen fluoride phase behavior and molecular structure: Ab initio derived potential models*, J. Chem. Phys. **119**, 6092 (2003), ISSN 00219606.
- [24] M. Kreitmair, H. Bertagnolli, J. J. Mortensen, and M. Parrinello, *Ab initio molecular dynamics simulation of hydrogen fluoride at several thermodynamic states*, J. Chem. Phys. **118**, 3639 (2003), ISSN 00219606.
- [25] S. J. Wierzchowski, Z. H. Fang, D. A. Kofke, and J. L. Tilson, *Three-body effects in hydrogen fluoride: survey of potential energy surfaces*, Mol. Phys. **104**, 503 (2006), ISSN 0026-8976.

- [26] M. J. McGrath, J. N. Ghogomu, C. J. Mundy, I.-F. W. Kuo, and J. I. Siepmann, *First principles Monte Carlo simulations of aggregation in the vapor phase of hydrogen fluoride*, Phys. Chem. Chem. Phys. **12**, 7678 (2010), ISSN 1463-9076.
- [27] M. J. McGrath, I.-F. W. Kuo, and J. I. Siepmann, *Liquid structures of water, methanol, and hydrogen fluoride at ambient conditions from first principles molecular dynamics simulations with a dispersion corrected density functional*, Phys. Chem. Chem. Phys. **13**, 19943 (2011), ISSN 1463-9076.
- [28] M. J. M. Mazack and J. Gao, *Quantum mechanical force field for hydrogen fluoride with explicit electronic polarization*, J. Chem. Phys. **140**, 204501 (2014), ISSN 00219606.
- [29] P. Jedlovsky and R. Vallauri, *Computer simulations of liquid HF by a newly developed polarizable potential model*, J. Chem. Phys. **107**, 10166 (1997), ISSN 0021-9606.
- [30] L. B. Pártay, P. Jedlovsky, and R. Vallauri, *Development of a new polarizable potential model of hydrogen fluoride and comparison with other effective models in liquid and supercritical states*, J. Chem. Phys. **124**, 184504 (2006), ISSN 00219606.
- [31] H. Berendsen, D. van Der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation*, Comput. Phys. Commun. **91**, 43 (1995), ISSN 00104655.
- [32] S. Páll, M. Abraham, C. Kutzner, B. Hess, and E. Lindahl, *Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS*, vol. 8759 (2015), ISBN 9783319159751.
- [33] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, *Scalable molecular dynamics with NAMD*, J. Comput. Chem. **26**, 1781 (2005), ISSN 0192-8651.
- [34] D. P. Visco and D. A. Kofke, *Vapor-liquid equilibria and heat effects of hydrogen fluoride from molecular simulation*, J. Chem. Phys. **109**, 4015 (1998), ISSN 0021-9606.
- [35] D. P. Visco and D. A. Kofke, *A comparison of molecular-based models to determine vapor-liquid phase coexistence in hydrogen fluoride*, Fluid Phase Equilib. **158-160**, 37 (1999), ISSN 03783812.
- [36] V. C. Weiss and W. Schröer, *Anomalous corresponding-states surface tension of hydrogen fluoride and of the Onsager model*, J. Chem. Phys. **122**, 084705 (2005), ISSN 0021-9606.

- [37] V. C. Weiss and F. Leroy, *Corresponding-states behavior of a dipolar model fluid with variable dispersion interactions and its relevance to the anomalies of hydrogen fluoride*, J. Chem. Phys. **144**, 224501 (2016), ISSN 0021-9606.
- [38] L. B. Pártay, G. Hantal, P. Jedlovszky, Á. Vincze, and G. Horvai, *A new method for determining the interfacial molecules and characterizing the surface roughness in computer simulations. Application to the liquid-vapor interface of water*, J. Comput. Chem. **29**, 945 (2008), ISSN 01928651.
- [39] M. Sega, S. S. Kantorovich, P. Jedlovszky, and M. Jorge, *The generalized identification of truly interfacial molecules (ITIM) algorithm for nonplanar interfaces*, J. Chem. Phys. **138**, 044110 (2013), ISSN 00219606.
- [40] P. Drude, *The Theory of Optics* (Kessinger Publishing, New York, New York, USA, 1902), ISBN 9780548647752.
- [41] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, vol. 38 (Oxford University Press, New York, 1988), ISBN 0198556454.
- [42] D. C. Rapaport, *The Art of Molecular Dynamics Simulation* (Cambridge University Press, Cambridge, 2004), 2nd ed., ISBN 9780511816581.
- [43] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (1996).
- [44] C. Runge, *Ueber die numerische Auflösung von Differentialgleichungen*, Math. Ann. **46**, 167 (1895), ISSN 0025-5831.
- [45] M. W. Kutta, *Beitrag zur näherungsweise Integration totaler Differentialgleichungen*, Zeitschrift für Math. und Phys. **46**, 435 (1901).
- [46] J. D. Lambert, *Numerical methods for ordinary differential systems: the initial value problem* (John Wiley & Sons, Inc., 1991).
- [47] L. Verlet, *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*, Phys. Rev. **159**, 98 (1967), ISSN 0031-899X.
- [48] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters*, J. Chem. Phys. **76**, 637 (1982), ISSN 0021-9606.
- [49] O. Buneman, *Time-reversible difference procedures*, J. Comput. Phys. **1**, 517 (1967), ISSN 00219991.
- [50] L. Onsager, *Electric Moments of Molecules in Liquids*, J. Am. Chem. Soc. **58**, 1486 (1936), ISSN 0002-7863.

- [51] M. Neumann, *The dielectric constant of water. Computer simulations with the MCY potential*, J. Chem. Phys. **82**, 5663 (1985), ISSN 0021-9606.
- [52] J. Barker and R. O. Watts, *Monte Carlo studies of the dielectric properties of water-like models*, Mol. Phys. **26**, 789 (1973), ISSN 0026-8976.
- [53] P. P. Ewald, *Die Berechnung optischer und elektrostatischer Gitterpotentiale*, Ann. Phys. **369**, 253 (1921), ISSN 00033804.
- [54] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *A smooth particle mesh Ewald method*, J. Chem. Phys. **103**, 8577 (1995), ISSN 0021-9606.
- [55] T. Darden, D. York, and L. G. Pedersen, *Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems*, J. Chem. Phys. **98**, 10089 (1993), ISSN 0021-9606.
- [56] H. C. Andersen, *Molecular dynamics simulations at constant pressure and/or temperature*, J. Chem. Phys. **72**, 2384 (1980), ISSN 0021-9606.
- [57] S. Nosé, *A unified formulation of the constant temperature molecular dynamics methods*, J. Chem. Phys. **81**, 511 (1984), ISSN 0021-9606.
- [58] W. G. Hoover, *Canonical dynamics: Equilibrium phase-space distributions*, Phys. Rev. A **31**, 1695 (1985), ISSN 0556-2791.
- [59] M. Parrinello and A. Rahman, *Polymorphic transitions in single crystals: A new molecular dynamics method*, J. Appl. Phys. **52**, 7182 (1981), ISSN 0021-8979.
- [60] K. Tsumuraya, Y. Ohde, and T. Oshimi, *Short-Lived HF Molecules in Superionic Hydrogen Fluoride at Extreme Conditions*, J. Phys. Soc. Japan **84**, 024711 (2015), ISSN 0031-9015.
- [61] J. E. Lennard-Jones, *On the Determination of Molecular Fields. II. From the Equation of State of a Gas*, Proc. R. Soc. A Math. Phys. Eng. Sci. **106**, 463 (1924), ISSN 1364-5021.
- [62] R. A. Buckingham, *The Classical Equation of State of Gaseous Helium, Neon and Argon*, Proc. R. Soc. A Math. Phys. Eng. Sci. **168**, 264 (1938), ISSN 1364-5021.
- [63] G. Lamoureux, E. Harder, I. V. Vorobyov, B. Roux, and A. D. MacKerell, *A polarizable model of water for molecular dynamics simulations of biomolecules*, Chem. Phys. Lett. **418**, 245 (2006), ISSN 00092614.
- [64] P. J. van Maaren and D. van Der Spoel, *Molecular dynamics simulations of a water with a novel shell-model potential*, J. Phys. Chem. B **105**, 2618 (2001).

- [65] W. Yu, P. E. M. Lopes, B. Roux, and A. D. MacKerell, *Six-site polarizable model of water based on the classical Drude oscillator*, J. Chem. Phys. **138**, 034508 (2013), ISSN 00219606.
- [66] P. Jedlovszky, M. Mezei, and R. Vallauri, *Comparison of polarizable and nonpolarizable models of hydrogen fluoride in liquid and supercritical states: A Monte Carlo simulation study*, J. Chem. Phys. **115**, 9883 (2001), ISSN 00219606.
- [67] U. Niesar, G. Corongiu, E. Clementi, G. R. Kneller, and D. K. Bhattacharya, *Molecular dynamics simulations of liquid water using the NCC ab initio potential*, J. Phys. Chem. **94**, 7949 (1990), ISSN 0022-3654.
- [68] S. B. Zhu, S. Yao, J. B. Zhu, S. Singh, and G. W. Robinson, *A flexible/polarizable simple point charge water model*, J. Phys. Chem. **95**, 6211 (1991), ISSN 0022-3654.
- [69] H. Wang and W. Yang, *Determining polarizable force fields with electrostatic potentials from quantum mechanical linear response theory*, J. Chem. Phys. **144**, 224107 (2016), ISSN 0021-9606.
- [70] P. E. M. Lopes, B. Roux, and A. D. MacKerell, *Molecular modeling and dynamics studies with explicit inclusion of electronic polarizability: theory and applications*, Theor. Chem. Acc. **124**, 11 (2009), ISSN 1432-881X.
- [71] S. W. Rick, S. J. Stuart, and B. J. Berne, *Dynamical fluctuating charge force fields: Application to liquid water*, J. Chem. Phys. **101**, 6141 (1994), ISSN 0021-9606.
- [72] A. Li, A. Voronin, A. T. Fenley, and M. K. Gilson, *Evaluation of Representations and Response Models for Polarizable Force Fields*, J. Phys. Chem. B **120**, 8668 (2016), ISSN 1520-6106.
- [73] A. A. Chialvo, F. Moucka, L. Vlcek, and I. Nezbeda, *Vapor-Liquid Equilibrium and Polarization Behavior of the GCP Water Model: Gaussian Charge-on-Spring versus Dipole Self-Consistent Field Approaches to Induced Polarization*, J. Phys. Chem. B **119**, 5010 (2015), ISSN 1520-6106.
- [74] M. Sprik and M. L. Klein, *A polarizable model for water using distributed charge sites*, J. Chem. Phys. **89**, 7556 (1988), ISSN 0021-9606.
- [75] J. A. Lemkul, J. Huang, B. Roux, and A. D. MacKerell, *An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications*, Chem. Rev. **116**, 4983 (2016), ISSN 0009-2665.
- [76] J. Huang, A. C. Simmonett, F. C. Pickard, A. D. MacKerell, and B. R. Brooks, *Mapping the Drude polarizable force field onto a multipole and induced dipole model*, J. Chem. Phys. **147**, 161702 (2017), ISSN 0021-9606.

- [77] D. R. Hartree, *The Wave Mechanics of an Atom with a non-Coulomb Central Field. Part III. Term Values and Intensities in Series in Optical Spectra*, Math. Proc. Cambridge Philos. Soc. **24**, 426 (1928), ISSN 0305-0041.
- [78] V. Fock, *Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems*, Zeitschrift für Phys. **61**, 126 (1930), ISSN 1434-6001.
- [79] P. Jedlovsky and R. Vallauri, *Computer simulation study of liquid HF with a new effective pair potential model*, Mol. Phys. **92**, 331 (1997), ISSN 0026-8976.
- [80] R. G. Della Valle and D. Gazzillo, *Towards an effective potential for the monomer, dimer, hexamer, solid, and liquid forms of hydrogen fluoride*, Phys. Rev. B **59**, 13699 (1999), ISSN 01631829.
- [81] A. Muñoz-Losa, I. Fdez.-Galván, M. E. Martín, and M. A. Aguilar, *Theoretical Study of Liquid Hydrogen Fluoride. Application of the Averaged Solvent Electrostatic Potential/Molecular Dynamics Method*, J. Phys. Chem. B **107**, 5043 (2003), ISSN 1520-6106.
- [82] R. L. McGreevy and L. Pusztai, *Reverse Monte Carlo Simulation: A New Technique for the Determination of Disordered Structures*, Mol. Simul. **1**, 359 (1988), ISSN 0892-7022.
- [83] U. Rothlisberger and M. Parrinello, *Ab Initio Molecular Dynamics Simulation of Liquid Hydrogen Fluoride*, J. Chem. Phys. **106**, 4658 (1997).
- [84] B. J. Howard, T. R. Dyke, and W. Klemperer, *The molecular beam spectrum and the structure of the hydrogen fluoride dimer*, J. Chem. Phys. **81**, 5417 (1984).
- [85] W. Klopper, M. Quack, and M. A. Suhm, *A new ab initio based six-dimensional semi-empirical pair interaction potential for HF*, Chem. Phys. Lett. **261**, 35 (1996), ISSN 00092614.
- [86] W. Klopper, M. Quack, and M. A. Suhm, *HF dimer: Empirically refined analytical potential energy and dipole hypersurfaces from ab initio calculations*, J. Chem. Phys. **108**, 10096 (1998), ISSN 0021-9606.
- [87] J. Janzen, *Electron-Diffraction Structural Study of Polymeric Gaseous Hydrogen Fluoride*, J. Chem. Phys. **50**, 3611 (1969), ISSN 00219606.
- [88] F. de Leluw and A. Dymanus, *Magnetic properties and molecular quadrupole moment of HF and HCl by molecular-beam electric-resonance spectroscopy*, J. Mol. Spectrosc. **48**, 427 (1973), ISSN 00222852.
- [89] H. Kornfeld, *Die Berechnung elektrostatischer Potentiale und der Energie von Dipol- und Quadrupolgittern*, Zeitschrift für Phys. **22**, 27 (1924), ISSN 1434-6001.

- [90] D. Adams and G. Dubey, *Taming the Ewald sum in the computer simulation of charged systems*, J. Comput. Phys. **72**, 156 (1987), ISSN 00219991.
- [91] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., *CHARMM: The biomolecular simulation program*, J. Comput. Chem. **30**, 1545 (2009), ISSN 01928651.
- [92] I. T. Todorov, W. Smith, K. Trachenko, and M. T. Dove, *DL\_POLY\_3: new dimensions in molecular dynamics simulations via massive parallelism*, J. Mater. Chem. **16**, 1911 (2006), ISSN 0959-9428.
- [93] H. Yu, T. Hansson, and W. F. van Gunsteren, *Development of a simple, self-consistent polarizable model for liquid water*, J. Chem. Phys. **118**, 221 (2003), ISSN 0021-9606.
- [94] E. Harder, V. M. Anisimov, I. V. Vorobyov, P. E. M. Lopes, S. Y. Noskov, A. D. MacKerell, and B. Roux, *Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator*, J. Chem. Theory Comput. **2**, 1587 (2006), ISSN 1549-9618.
- [95] D. P. Geerke and W. F. van Gunsteren, *On the Calculation of Atomic Forces in Classical Simulation Using the Charge-on-Spring Method To Explicitly Treat Electronic Polarization*, J. Chem. Theory Comput. **3**, 2128 (2007), ISSN 1549-9618.
- [96] C. Schröder and O. Steinhauser, *Simulating polarizable molecular ionic liquids with Drude oscillators*, J. Chem. Phys. **133**, 154511 (2010), ISSN 0021-9606.
- [97] C. Schröder, *Comparing reduced partial charge models with polarizable simulations of ionic liquids*, Phys. Chem. Chem. Phys. **14**, 3089 (2012), ISSN 1463-9076.
- [98] G. Lamoureux and B. Roux, *Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm*, J. Chem. Phys. **119**, 3025 (2003), ISSN 00219606.
- [99] H. Yu, T. W. Whitfield, E. Harder, G. Lamoureux, I. V. Vorobyov, V. M. Anisimov, A. D. MacKerell, and B. Roux, *Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field*, J. Chem. Theory Comput. **6**, 774 (2010), ISSN 1549-9618.
- [100] J. A. Lemkul, B. Roux, D. van Der Spoel, and A. D. Mackerell, *Implementation of extended Lagrangian dynamics in GROMACS for polarizable simulations using the classical Drude oscillator model*, J. Comput. Chem. **36**, 1473 (2015), ISSN 1096987X.

- [101] I. Sheft, A. J. Perkins, and H. H. Hyman, *Anhydrous hydrogen fluoride: Vapor pressure and liquid density*, J. Inorg. Nucl. Chem. **35**, 3677 (1973), ISSN 00221902.
- [102] F. P. Buff, R. A. Lovett, and F. H. Stillinger, *Interfacial Density Profile for Fluids in the Critical Region*, Phys. Rev. Lett. **15**, 621 (1965), ISSN 0031-9007.
- [103] M. Sega, *The role of a small-scale cutoff in determining molecular layers at fluid interfaces*, Phys. Chem. Chem. Phys. **18**, 23354 (2016), ISSN 1463-9076.
- [104] J. S. Rowlinson and B. Widom, *Molecular theory of capillarity*, International series of monographs on chemistry (Clarendon Press, 1982), ISBN 9780198556428.
- [105] M. Sega, B. Fábián, and P. Jedlovszky, *Layer-by-layer and intrinsic analysis of molecular and thermodynamic properties across soft interfaces*, J. Chem. Phys. **143**, 114709 (2015), ISSN 00219606.
- [106] M. Sega, B. Fábián, G. Horvai, and P. Jedlovszky, *How is the surface tension of various liquids distributed along the interface normal?*, J. Phys. Chem. C **120**, 27468 (2016), ISSN 19327455.
- [107] M. Jorge, G. Hantal, P. Jedlovszky, and M. N. D. Cordeiro, *A Critical Assessment of Methods for the Intrinsic Analysis of Liquid Interfaces: 2. Density Profiles*, J. Phys. Chem. C **114**, 18656 (2010), ISSN 1932-7447.
- [108] M. Sega and G. Hantal, *Phase and interface determination in computer simulations of liquid mixtures with high partial miscibility*, Phys. Chem. Chem. Phys. **19**, 18968 (2017), ISSN 14639076.
- [109] M. Jorge, G. Hantal, P. Jedlovszky, and M. N. D. S. Cordeiro, *A Critical Assessment of Methods for the Intrinsic Analysis of Liquid Interfaces: 2. Density Profiles*, J. Phys. Chem. C **114**, 18656 (2010), ISSN 1932-7447.
- [110] H. Edelsbrunner, D. Kirkpatrick, and R. Seidel, *On the shape of a set of points in the plane*, IEEE Trans. Inf. Theory **29**, 551 (1983), ISSN 0018-9448.
- [111] H. Edelsbrunner and E. P. Mücke, *Three-dimensional alpha shapes*, ACM Trans. Graph. **13**, 43 (1994), ISSN 07300301.
- [112] B. Delaunay, *Sur la sphère vide*, Bull. l'Academie des Sci. l'URSS pp. 793–800 (1934).
- [113] J. Skvor, J. Skvára, J. Jirsák, and I. Nezbeda, *A general method for determining molecular interfaces and layers*, J. Mol. Graph. Model. **76**, 17 (2017), ISSN 10933263.

- [114] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, *The quickhull algorithm for convex hulls*, ACM Trans. Math. Softw. **22**, 469 (1996), ISSN 00983500.
- [115] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, *MDAnalysis: A toolkit for the analysis of molecular dynamics simulations*, J. Comput. Chem. **32**, 2319 (2011), ISSN 01928651.
- [116] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney, et al., *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations*, Proc. 15th Python Sci. Conf. pp. 98–105 (2016).
- [117] A. P. Willard and D. Chandler, *Instantaneous liquid interfaces*, J. Phys. Chem. B **114**, 1954 (2010), ISSN 15206106.
- [118] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*, Proc. 2nd Int. Conf. Knowl. Discov. Data Min. pp. 226–231 (1996), ISSN 09758887.
- [119] H. M. Berman, *The Protein Data Bank*, Nucleic Acids Res. **28**, 235 (2000), ISSN 13624962.
- [120] S. van der Walt, S. C. Colbert, and G. Varoquaux, *The NumPy Array: A Structure for Efficient Numerical Computation*, Comput. Sci. Eng. **13**, 22 (2011), ISSN 1521-9615.
- [121] L. B. Pártay, G. Horvai, and P. Jedlovszky, *Temperature and Pressure Dependence of the Properties of the Liquid-Liquid Interface. A Computer Simulation and Identification of the Truly Interfacial Molecules Investigation of the Water-Benzene System*, J. Phys. Chem. C **114**, 21681 (2010), ISSN 1932-7447.
- [122] M. Sega, B. Fábián, and P. Jedlovszky, *Pressure Profile Calculation with Mesh Ewald Methods*, J. Chem. Theory Comput. **12**, 4509 (2016), ISSN 1549-9618.
- [123] M. Sega, B. Fábián, and P. Jedlovszky, *Nonzero Ideal Gas Contribution to the Surface Tension of Water*, J. Phys. Chem. Lett. **8**, 2608 (2017), ISSN 19487185.
- [124] M. Jorge and M. N. D. Cordeiro, *Intrinsic structure and dynamics of the water/nitrobenzene interface*, J. Phys. Chem. C **111**, 17612 (2007), ISSN 19327447.
- [125] M. Sega and C. Dellago, *Long-Range Dispersion Effects on the Water/Vapor Interface Simulated Using the Most Common Models*, J. Phys. Chem. B **121**, 3798 (2017), ISSN 1520-6106.
- [126] A. H. Narten, *Liquid Water: Atom Pair Correlation Functions from Neutron and X-Ray Diffraction*, J. Chem. Phys. **56**, 5681 (1972), ISSN 0021-9606.

- [127] G. Jancso, P. Bopp, and K. Heinzinger, *Molecular dynamics study of high-density liquid water using a modified central-force potential*, Chem. Phys. **85**, 377 (1984), ISSN 03010104.
- [128] L. Bosio, G. P. Johari, and J. Teixeira, *X-ray study of high-density amorphous water*, Phys. Rev. Lett. **56**, 460 (1986), ISSN 0031-9007.
- [129] A. K. Soper, F. Bruni, and M. A. Ricci, *Site-site pair correlation functions of water from 25 to 400 C: Revised analysis of new and old diffraction data*, J. Chem. Phys. **106**, 247 (1997), ISSN 0021-9606.
- [130] J. L. Finney, A. Hallbrucker, I. Kohl, A. K. Soper, and D. T. Bowron, *Structures of High and Low Density Amorphous Ice by Neutron Diffraction*, Phys. Rev. Lett. **88**, 225503 (2002), ISSN 0031-9007.
- [131] A. M. Saitta and F. Datchi, *Structure and phase diagram of high-density water: The role of interstitial molecules*, Phys. Rev. E **67**, 020201 (2003), ISSN 1063-651X.
- [132] S. Biswas, D. Chakraborty, and B. S. Mallik, *Interstitial Voids and Resultant Density of Liquid Water: A First-Principles Molecular Dynamics Study*, ACS Omega **3**, 2010 (2018), ISSN 2470-1343.
- [133] L. B. Pártay, P. Jedlovszky, and G. Horvai, *Structure of the Liquid-Vapor Interface of Water-Acetonitrile Mixtures As Seen from Molecular Dynamics Simulations and Identification of Truly Interfacial Molecules Analysis*, J. Phys. Chem. C **113**, 18173 (2009), ISSN 1932-7447.
- [134] G. Hantal, M. Darvas, L. B. Pártay, G. Horvai, and P. Jedlovszky, *Molecular level properties of the free water surface and different organic liquid/water interfaces, as seen from ITIM analysis of computer simulation results*, J. Phys. Condens. Matter **22**, 284112 (2010).
- [135] P. Jedlovszky, *The local structure of various hydrogen bonded liquids: Voronoi polyhedra analysis of water, methanol, and HF*, J. Chem. Phys. **113**, 9113 (2000), ISSN 0021-9606.
- [136] A. G. Streng, *Miscibility and compatibility of some liquefied and solidified gases at low temperatures*, J. Chem. Eng. Data **16**, 357 (1971), ISSN 0021-9568.
- [137] M. E. van Leeuwen and B. Smit, *What makes a polar liquid a liquid?*, Phys. Rev. Lett. **71**, 3991 (1993), ISSN 0031-9007.
- [138] J. J. Weis and D. Levesque, *Chain formation in low density dipolar hard spheres: A Monte Carlo study*, Phys. Rev. Lett. **71**, 2729 (1993), ISSN 0031-9007.

- [139] J.-M. Caillol, *Search of the gas-liquid transition of dipolar hard spheres*, J. Chem. Phys. **98**, 9835 (1993), ISSN 0021-9606.
- [140] T. Tlusty and S. A. Safran, *Defect-Induced Phase Separation in Dipolar Fluids*, Science (80-. ). **290**, 1328 (2000), ISSN 00368075.
- [141] L. Rovigatti, J. Russo, and F. Sciortino, *No Evidence of Gas-Liquid Coexistence in Dipolar Hard Spheres*, Phys. Rev. Lett. **107**, 237801 (2011), ISSN 0031-9007.
- [142] L. Rovigatti, J. Russo, and F. Sciortino, *Structural properties of the dipolar hard-sphere fluid at low temperatures and densities*, Soft Matter **8**, 6310 (2012), ISSN 1744-683X.
- [143] L. Rovigatti, J. M. Tavares, and F. Sciortino, *Self-Assembly in Chains, Rings, and Branches: A Single Component System with Two Critical Points*, Phys. Rev. Lett. **111**, 168302 (2013), ISSN 0031-9007.
- [144] J. W. Ring, *Vibrational modes of H-bonded chains in liquid and solid HF*, J. Chem. Phys. **68**, 2911 (1978), ISSN 00219606.
- [145] J. Shamir and A. Netzer, *N.M.R. Studies of Anhydrous Hydrogen Fluoride Solutions*, Can. J. Chem. **51**, 2676 (1973), ISSN 0008-4042.
- [146] C. Spickermann, E. Perlt, M. von Domaros, M. Roatsch, J. Friedrich, and B. Kirchner, *Coupled cluster in condensed phase. part II: Liquid hydrogen fluoride from quantum cluster equilibrium theory*, J. Chem. Theory Comput. **7**, 868 (2011), ISSN 15499618.
- [147] P. L. Leath, *Cluster size and boundary distribution near percolation threshold*, Phys. Rev. B **14**, 5046 (1976), ISSN 0556-2805.
- [148] D. Stauffer, *Scaling theory of percolation clusters*, Phys. Rep. **54**, 1 (1979), ISSN 03701573.
- [149] P. Jedlovsky, I. Brovchenko, and A. Oleinikova, *Percolation Transition in Supercritical Water: A Monte Carlo Simulation Study*, J. Phys. Chem. B **111**, 7603 (2007), ISSN 1520-6106.
- [150] C.-P. Chai Kao, M. Paulaitis, G. Sweany, and M. Yokozeki, *An equation of state/chemical association model for fluorinated hydrocarbons and HF*, Fluid Phase Equilib. **108**, 27 (1995), ISSN 03783812.
- [151] J. Chowdhary and B. M. Ladanyi, *Water/hydrocarbon interfaces: Effect of hydrocarbon branching on single-molecule relaxation*, J. Phys. Chem. B **112**, 6259 (2008), ISSN 15206106.
- [152] P. Orea, J. López-Lemus, and J. Alejandro, *Oscillatory surface tension due to finite-size effects*, J. Chem. Phys. **123**, 114702 (2005), ISSN 0021-9606.