Between the Scales: Water from different Perspectives

Dissertation

zur Erlangung des Grades

“Doktor der Naturwissenschaften”

am Fachbereich Physik
der Johannes Gutenberg-Universität Mainz

Christoph Junghans
geb. in Merseburg

Max-Planck-Institut für Polymerforschung

Mainz, 9. August 2010

Tag der mündlichen Prüfung: 29. Okt. 2010

Abstract

Water is one of the most frequently studied fluids on earth. In this thesis, water was investigated at two resolutions using multi-scale computer simulation techniques. First, the atomistic and coarse-grained resolutions were studied separately. In the atomistic resolution, a water molecule is described chemically by three atoms, while in the coarse-grained case, a molecule is modeled by a sphere.

In this work, various coarse-grained models have been developed using different coarse-graining techniques, mainly iterative Boltzmann inversion and iterative inverse Monte Carlo, which are structure-based approaches that aim to reproduce distributions, such as the pair distribution functions, of the atomistic model. In this context the Versatile Object-oriented Toolkit for Coarse-graining Applications (VOTCA) was developed.

It was studied to which extent the coarse-grained models can simultaneously reproduce several properties of the underlying atomistic model, such as thermodynamic properties like pressure and compressibility or structural properties, which have not been used in the coarse-graining process, e. g. the tetrahedral packing behavior, which is responsible for many special properties of water.

Subsequently, these two resolutions were combined using the adaptive resolution scheme, which combines the advantage of atomistic details in a small cavity of high resolution with the computational efficiency of the coarse-grained model in order to access larger time and length scales. In this scheme, the introduced coarse-grained models were used to study the influence of the hydrogen bonds on the hydration of small fullerenes. It was found that the interface structure is more dependent on the nature of the interaction between the solute and water molecules than on the presence of the hydrogen bond network.

Zusammenfassung

In der vorliegenden Arbeit werden verschiedene Wassermodelle in sogenannten Multiskalen-Computersimulationen mit zwei Auflösungen untersucht, in atomistischer Auflösung und in einer vergröberten Auflösung, die als "coarse-grained" bezeichnet wird. In der atomistischen Auflösung wird ein Wassermolekül, entsprechend seiner chemischen Struktur, durch drei Atome beschrieben, im Gegensatz dazu wird ein Molekül in der coarse-grained Auflösung durch eine Kugel dargestellt.

Die coarse-grained Modelle, die in dieser Arbeit vorgestellt werden, werden mit verschiedenen coarse-graining Methoden entwickelt. Hierbei kommen hauptsächlich die "iterative Boltzmann Inversion" und die "iterative Monte Carlo Inversion" zum Einsatz. Beides sind struktur-basierte Ansätze, die darauf abzielen bestimmte strukturelle Eigenschaften, wie etwa die Paarverteilungsfunktionen, des zugrundeliegenden atomistischen Systems zu reproduzieren. Zur automatisierten Anwendung dieser Methoden wurde das Softwarepaket "Versatile Object-oriented Toolkit for Coarse-Graining Applications" (VOTCA) entwickelt.

Es wird untersucht, in welchem Maße coarse-grained Modelle mehrere Eigenschaften des zugrundeliegenden atomistischen Modells gleichzeitig reproduzieren können, z.B. thermodynamische Eigenschaften wie Druck und Kompressibilität oder strukturelle Eigenschaften, die nicht zur Modellbildung verwendet wurden, z.B. das tetraedrische Packungsverhalten, welches für viele spezielle Eigenschaft von Wasser verantwortlich ist.

Mit Hilfe des "Adaptive Resolution Schemes" werden beide Auflösungen in einer Simulation kombiniert. Dabei profitiert man von den Vorteilen beider Modelle: Von der detaillierten Darstellung eines räumlich kleinen Bereichs in atomistischer Auflösung und von der rechnerischen Effizienz des coarse-grained Modells, die den Bereich simulierbarer Zeit- und Längenskalen vergrössert.

In diesen Simulationen kann der Einfluss des Wasserstoffbrückenbindungsnetzwerks auf die Hydration von Fullerenen untersucht werden. Es zeigt sich, dass die Struktur der Wassermoleküle an der Oberfläche hauptsächlich von der Art der Wechselwirkung zwischen dem Fulleren und Wasser und weniger von dem Wasserstoffbrückenbindungsnetzwerk dominiert wird.

Contents

List of Figures

List of Tables

Overview

The main topics of this thesis are simulation models of water studied at different resolutions. Several publications have been produced during that time, which can be found on the next page.

The work started with Ref. [1], in which an extension of the DPD thermostat is discussed and coarse-grained water is used as an example in Ref. [2]. With this the interest to understand the fundamentals behind coarse-grained water models increased. Hence, different models were compared in detail in Ref. [3]. In this comparison only one coarse-graining technique was used, for that reason a systematic study of 3 different techniques was done in Ref. [4] producing a coarse-graining toolkit as a useful side effect.

With this zoo of coarse-grained water models adaptive resolution simulations, which are described in detail in Ref. [5], have been done. To this end, this scheme was implemented in two simulation packages, whose technical questions have been addressed in Ref. [6]. Finally, the adaptive resolution scheme was used to study the solvation of fullerenes in Ref. [7].

This thesis tells the details of the story. After a short introduction to multi-scale techniques, the adaptive resolution scheme and coarse-graining in chapter 1, theoretical backgrounds are explained in chapter 2 starting with molecular dynamics and Fokker-Planck theory followed by details about different coarse-graining techniques and the adaptive resolution scheme. After that the results of the studies described above are explained in detail in the last two chapters, first from the coarse-grained perspective in chapter 3 and then from the multi-scale view obtained by the adaptive resolution scheme in chapter 4. The conclusion and an outlook can be found in chapter 5.

Related Publications

[1] **C. Junghans**, M. Praprotnik, and K. Kremer. Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat. *Soft Matter*, 4(1):156–161, 2008.

[2] M. Praprotnik, **C. Junghans**, L. Delle Site, and K. Kremer. Simulation approaches to soft matter: Generic statistical properties vs. chemical details. *Computer Physics Communications*, 179(1-3):51–60, 2008.

[3] H. Wang, **C. Junghans**, and K. Kremer. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *The European Physical Journal E: Soft Matter and Biological Physics*, 28(2):221–229, 2009.

[4] V. Rühle, **C. Junghans**, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile Object-Oriented toolkit for Coarse-Graining applications. *Journal of Chemical Theory and Computation*, 5(12):3211–3223, 2009.

[5] **C. Junghans**, M. Praprotnik, and L. Delle Site. Adaptive resolution schemes. In J. Grotendorst, N. Attig, S. Blügel, and D. Marx, editors, *Multiscale Simulation Methods in Molecular Sciences: Lecture Notes*, page 359. Forschungszentrum Jülich, 2009.

[6] **C. Junghans** and S. Poblete. A reference implementation of the adaptive resolution scheme in ESPResSo. *Computer Physics Communications*, 181(8):1449–1454, 2010.

[7] B. P. Lambeth, **C. Junghans**, K. Kremer, C. Clementi, and L. Delle Site. Communication: On the locality of hydrogen bond networks at hydrophobic interfaces. *The Journal of Chemical Physics*, 133(22):221101, 2010.

Chapter 1

Introduction

Water is one of the most studied liquids due to its common appearance and general importance, but despite its chemical simplicity, it is still not adequately understood. Over the years, simulations have gained increasing attention in the attempt to provide a better understanding of water as a liquid as well as a solvent [8, 9]. Many different classical simulation models have been developed, which enable different features of water to be captured. Density functional theory can partially reproduce electronic structures of water in bulk and close to certain metal surfaces [10]. Molecular dynamics (MD) simulations can give further insight, e. g. into the role of water during protein folding [11]. On a much larger, not solely particle-based scale, approaches like the lattice Boltzmann method [12] or dissipative particle dynamics (DPD) simulations [13] are employed to describe hydrodynamic effects. DPD simulations are used to give qualitative insight into the behavior of clusters of particles, so called DPD particles. The lattice Boltzmann method discretizes the fluid to the nodes of a lattice and hence the lattice constant determines the modeled length scale. These descriptions of water take place on significantly different scales, so called resolutions, which are described by different physical laws.

Nevertheless, simulations on different scales, not only for water but also for systems in condensed matter, chemistry and material science, are becoming a standard procedure. The fast progress of computer technology and the concurrent development of novel, powerful simulation methods has strongly contributed to this expansion. This has led to the situation that detailed sequential studies from the electronic scale to the mesoscopic scale and even to the continuum can be performed nowadays. However, sequential approaches do not couple scales in a direct way. Their central idea is to employ results from one scale to build simplified models in a physically consistent fashion, keeping the modeling approximations under control as much as possible. And

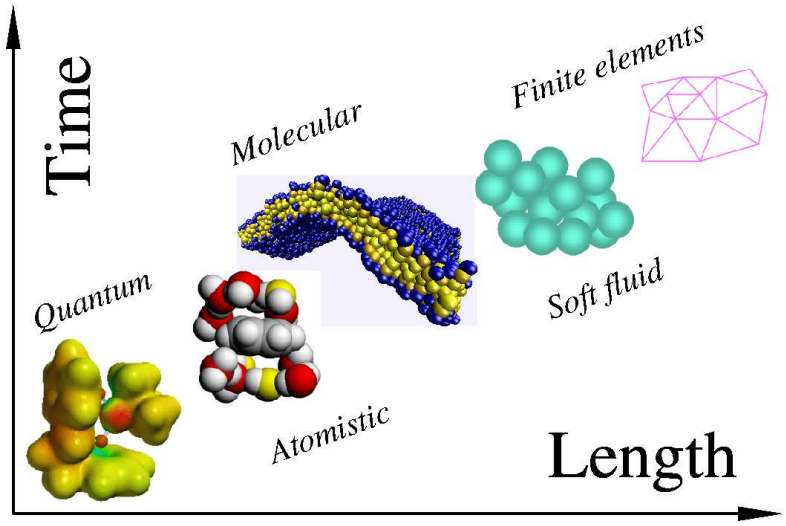

Figure 1.1: **Modeling on different scales**: An illustration of simulation models on different scales using different level of resolution. Picture taken from Ref. [14].

then using this simplified model in a separate stage on much bigger systems. An example of a hierarchy of different simulation models can be found in fig. 1.1. This is where multi-scale techniques enter the game.

1.1 Multi-scale techniques

A step beyond these sequential schemes is represented by approaches, where the scales are coupled in a concurrent fashion within a unified computational scheme. For problems like edge dislocation in metals, cracks in solid materials or simulations of organic electronics, the local chemistry affects large scale material properties and vice versa. These are typical questions to which the idea of concurrent scale methods has been applied. In these cases, quantum based methods are interfaced with classical atomistic and continuum approaches within a single computational scheme [15, 16, 17].

A further example is the Quantum Mechanics - Molecular Mechanics scheme (QMMM) [18], which is mainly used for soft matter systems. It is based on the idea that a fixed subsystem is described with a quantum resolution while the remainder of the system is treated at the classical atomistic level. A typical example of an application of the QMMM method is the study of the solvation process of large molecules. For this specific example the interesting chemistry happens locally within the region defined by a few solvation shells. Thus, it is treated at a quantum level while the sta-

tistical/thermodynamical effect of the fluctuating environment (solvent) far from the molecules is treated in a rather efficient way at a classical level. There are several more examples in the same fashion [19, 20].

All of these methods, although computationally robust, are characterized by a non-trivial conceptual limitation, i. e. the region of high resolution is fixed and thus the exchange of particles among the different regions is not allowed. While this may not be crucial for hard matter, it is certainly a strong limitation for soft matter, i. e. complex fluids, since relevant density fluctuations are suppressed. The natural step forward to overcome this problem is to design adaptive resolution methods, which allow for the exchange of particles among regions of different resolution.

In general, in such a scheme a molecule moving from a high resolution region to a lower one would gradually lose some degrees of freedom until the lower resolution is reached and yet the statistical equilibrium among the two different regions is maintained at any instant. Recently some schemes based on this idea, for classical MD, have been presented in literature, where forces [21], potentials [22] or Lagrangians [23] are interpolated between the two resolutions. They are based on different conceptual approaches regarding the way the scales are coupled and the way the equilibrium of the overall system is assured, their details will be discussed in sec. 2.4. Now the adaptive resolution scheme (AdResS), which interpolates the forces, will be introduced.

1.2 Adaptive Resolution Scheme

As stated before, many problems in condensed matter, material science and chemistry are multi-scale in nature, meaning that the interplay between effects on different scales plays the fundamental role for the understanding of relevant properties. An exhaustive description of the related physical phenomena requires in principle the simultaneous treatment of all the scales involved. This is not only a prohibitive task because of the limited computational resources but above all because the large amount of produced data would mostly contain information not essential to the problem analyzed and may overshadow the underlying fundamental physics or chemistry of the system.

A solution to this problem is to treat only those degrees of freedom in a simulation, which are strictly required by the problem. In the AdResS method the combination of all-atom classical MD and coarse-grained MD leads to a hybrid scheme where the molecule can adapt its resolution, passing from an all-atom to a coarse-grained representation when going from the high

resolution region to the lower one and vice versa. Thus, the method allows the molecules to change their number of degrees of freedom on the fly in a continuous manner. In this way the limitation to small time and length scales in the all-atom approach is overcome due to the fact that only a limited region is treated with atomistic degrees of freedom. This region, which can vary in space and time, is treated with atomistic degrees of freedom, while the remaining part of the system is treated in the coarse-grained representation. Thus, the atomistic details are lost in the coarse-grained part, but those degrees of freedom relevant to the particular property under investigation are retained. In this way one can reach much larger length and time scales and still describe the system with high resolution, where strictly required.

The drawback of the method is that one has to develop a coarse-grained model first, which on one hand has to be simpler and more efficient, but on the other hand has to retain some of the properties of the atomistic system. However, in some cases a very simple coarse-grained model, e. g. a Lennard-Jones fluid, is already sufficient.

1.3 Coarse-Graining

Bottom-up coarse-graining [24] is used to provide such a systematic link between these levels of description. Bottom-up means that knowledge from the more detailed system is used to develop the coarse-grained model. In this thesis, different coarse-grained models will be studied, which conserve first the structural properties, namely the radial distribution function $g(r)$, and second, other properties such as pressure, compressibility or tetrahedral packing.

For MD simulations on the coarse-grained level alone, so called coarse-grained MD [25] can be performed, where each molecule is represented by a set of coarse-grained beads, and can reproduce parts of the structure. Of course, some properties of the atomistic system are lost, but a significant speed up can be achieved due to a smaller number of degrees of freedom, smoother potentials and larger time steps. Many examples for such models, from polymer melts over water to lipid bilayers and small peptides can be found in the literature [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37].

Chapter 2

Theory & Simulation Methods

In this chapter, the simulation methods employed and their theoretical background are described. As the main aim of this work is to understand properties of water on different scales, one first considers the molecular dynamics and the thermostats, due to the fact that these methods are used for both, atomistic and coarse-grained simulations. The section about thermostats will be more detailed to allow for the introduction and understanding of a novel thermostat.

Subsequently, some attention is given to the derivation of coarse-grained force fields on the basis of a reference system, which is an atomistic system here. With this knowledge, one can then discuss the adaptive resolution scheme, which couples the atomistic and the coarse-grained representations within one simulation. With these methods at hand we will be able to understand some effects of water on two different scales of resolution even when these are coupled.

However, when writing this thesis, already around 300 books [38] with "computer simulations" and 450 books [39] with "molecular dynamics" in their titles have been composed. Therefore, the theory in this chapter is reduced to a bare minimum, essential to understand the context and the novel facts.

2.1 Molecular Dynamics

Basically, molecular dynamics is an algorithm to integrate Newton's equation of motion. In this section, the velocity Verlet algorithm, which is the most commonly used integration scheme in molecular dynamics, is derived by expansion of this equation. Later, we will see that one can derive the same algorithm from a Fokker-Planck picture.

2.1.1 Newton's Equation of Motion

The fundamental equation in classical mechanics is

$$m\ddot{\vec{r}} = \vec{F} \ , \tag{2.1}$$

which is known as Newton's equation of motion, with $\vec{r}$ being the position of a particle of mass m, and $\vec{F}$ the force acting on that particle. Difficulty arises when there is more than one particle present and the force must be determined by all the surrounding particles:

$$m_i\ddot{\vec{r}}_i = \vec{F}_i(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_{N-1}, \vec{r}_N) \tag{2.2}$$

For convenience, one can write this as two coupled first-order differential equations:

$$\dot{\vec{r}}_i = \frac{\vec{p}_i}{m_i} \tag{2.3}$$

$$\dot{\vec{p}}_i = \vec{F}_i \tag{2.4}$$

If the force is integrable, i. e. there is a potential, Newton's Equation conserves energy and can be rewritten as the *Hamilton equations*:

$$\dot{\vec{r}}_i = \frac{\partial \mathcal{H}}{\partial \vec{p}_i} \ , \tag{2.5}$$

$$\dot{\vec{p}}_i = -\frac{\partial \mathcal{H}}{\partial \vec{r}_i} \ , \tag{2.6}$$

where $\mathcal{H}$ is the *Hamiltonian*, which is in most cases identical to the energy [40]. These equations generate a micro-canonical ensemble (NVE).

2.1.2 Velocity Verlet Algorithm

The velocity Verlet algorithm is an integration algorithm, which is accurate up to cubic order in the time step Δt.

A Taylor expansion of eq. 2.3 and eq. 2.4 will lead to

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \frac{\Delta t}{m_i}\vec{p}_i(t) + \frac{\Delta t^2}{2m_i}\vec{F}_i(t) + \mathcal{O}(\Delta t^3) \ , \tag{2.7}$$

$$\vec{p}_i(t + \Delta t) = \vec{p}_i(t) + \Delta t \ \vec{F}_i(t) + \frac{\Delta t^2}{2}\dot{\vec{F}}_i(t) + \mathcal{O}(\Delta t^3) \ . \tag{2.8}$$

As $\dot{\vec{F}}_i(t)$ is unknown one expands:

$$\vec{F}_i(t + \Delta t) = \vec{F}_i(t) + \Delta t \ \dot{\vec{F}}_i(t) + \mathcal{O}(\Delta t^2) \ , \tag{2.9}$$

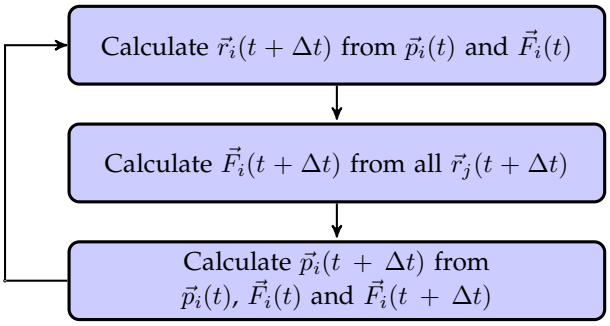

Figure 2.1: **Velocity Verlet algorithm**: The computational steps of the algorithm, using equations 2.11 and 2.12. The third step can become tedious if the forces depend on the velocities as well.

and arrives at:

$$\Delta t^2 \dot{\vec{F}}_i(t) = \left(\vec{F}(t+\Delta t) - \vec{F}_i(t)\right) \Delta t + \mathcal{O}(\Delta t^3) \tag{2.10}$$

With this the *velocity Verlet algorithm* results in:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \frac{\Delta t}{m_i}\vec{p}_i(t) + \frac{\Delta t^2}{2m_i}\vec{F}_i(t) + \mathcal{O}(\Delta t^3) \ , \tag{2.11}$$

$$\vec{p}_i(t+\Delta t) = \vec{p}_i(t) + \frac{\Delta t}{2}\left(\vec{F}_i(t) + \vec{F}_i(t+\Delta t)\right) + \mathcal{O}(\Delta t^3) \tag{2.12}$$

It is important to note that eq. 2.12 has to be solved in an iterative manner if $\vec{F}_i$ depends on the velocities as well.

From this derivation one can see why the velocity Verlet algorithm is correct up to order Δt^2, but one can not prove easily why this algorithm is symplectic. A symplectic integrator is defined as being time-reversible and phase-space volume conservative, where the latter property is more demanding to prove and we will not show it at this point as this follows for free from the derivation done in sec. 2.2.2.1.

2.2 Thermostats

In the previous section, the fundamental equations of molecular dynamics have been explained and we have shown how the velocity Verlet algorithm follows from Newton's equations. Basically, this provides the possibility of simulating a micro-canonical ensemble.

In this section, we will derive the same algorithm from a different perspective, the Fokker-Planck picture, which gives understanding to why it is phase space volume conserving and time-reversible, which makes it a symplectic integrator. In this picture, it is also easy to switch to different equations of motion, which allow the sampling at different ensembles. These additions to Newton's equations are called *thermostats*. It is important to note that not all thermostats generate a canonical ensemble, but a comparable useful ensemble such as an iso-kinetic ensemble. First we will derive the Fokker-Planck equation from general principles and then we will add stochastic terms to the equations of motion to introduce Langevin thermostats.

2.2.1 Fokker-Planck Picture

Let us consider a process defined on a state space Γ with state points $\vec{X}$. These state points can be related to the common mechanical phase space $\{\vec{x}_i, \vec{p}_i\}$ or to the phase space $\{\vec{x}_i\}$ of an over-damped Brownian motion. Assuming we know some points $\vec{X}(t_i) = \vec{X}_i$ of the trajectory of this process, one could ask for the conditional probability density to arrive at point $\vec{X}_n$ under the condition that the process has passed through $\vec{X}_j$ with $j = 1 \ldots n-1$:

$$P(\vec{X}_n, t_n, \vec{X}_{n-1}, t_{n-1}, \ldots, \vec{X}_2, t_2, \vec{X}_1, t_1) \tag{2.13}$$

It is clear that in the case of a deterministic process the probability density will always be a δ-function of the deterministic trajectory, but we also have to keep random processes in mind. Under the assumption that the process is Markovian, the conditional probability simplifies to

$$P(\vec{X}_n, t_n | \vec{X}_{n-1}, t_{n-1}) \ , \tag{2.14}$$

due to the fact that the Markov processes have "no memory". This assumption is reasonable as long as there are no long correlation times involved.

The conditional probability has certain natural properties:

- Normalization
$$\int \mathrm{d}\vec{X} \ P(\vec{X}, t | \vec{X}_0, t_0) = 1 \tag{2.15}$$
- Initial condition
$$P(\vec{X}, t_0 | \vec{X}_0, t_0) = \delta(\vec{X} - \vec{X}_0) \tag{2.16}$$
- Decomposability (*Chapman-Kolmogorov equation*)
$$P(\vec{X}_3, t_3 | \vec{X}_1, t_1) = \int \mathrm{d}\vec{X}_2 \ P(\vec{X}_3, t | \vec{X}_2, t_2) P(\vec{X}_2, t_2 | \vec{X}_1, t_1) \tag{2.17}$$

All information about the conditional probability is basically stored in its moments:

$$\mu_\nu(t, \vec{X}_0, t_0) = \int \mathrm{d}\vec{X}\ (\vec{X} - \vec{X}_0)^\nu P(\vec{X}, t|\vec{X}_0, t_0) \tag{2.18}$$

If all moments exist and are not infinite, one can invert this equation with the help of a Fourier transformation [41]:

$$P(\vec{X}, t|\vec{X}_0, t_0) = \sum_{\nu=1}^{\infty} \frac{1}{\nu!} \mu_\nu(t, \vec{X}_0, t_0) \left(-\frac{\partial}{\partial \vec{X}}\right)^\nu \delta(\vec{X} - \vec{X}_0) \tag{2.19}$$

After inserting this into eq. 2.17 and differentiating with respect to t, one obtains the *Generalized Fokker-Planck equation*

$$\frac{\partial}{\partial t} P(\vec{X}, t|\vec{X}_0, t_0) = \sum_{\nu=1}^{\infty} \left(-\frac{\partial}{\partial \vec{X}}\right)^\nu D^\nu(\vec{X}, t) P(\vec{X}, t|\vec{X}_0, t_0)\ , \tag{2.20}$$

where $D^\nu(\vec{X}, t)$ are the *Kramers-Moyal coefficients*

$$\begin{aligned} D^\nu(\vec{X}, t) = \frac{1}{\nu!} \frac{\partial}{\partial \tau} \mu_\nu(t+\tau, \vec{X}, t)\Big|_{\tau=0} &= \lim_{\tau \to 0} \frac{< (\vec{X} - \vec{X}_0)^\nu >}{\tau \nu!} \\ &= \lim_{\tau \to 0} \frac{< (\Delta \vec{X})^\nu >}{\tau \nu!}\ , \end{aligned} \tag{2.21}$$

where $< >$ is the thermodynamic average defined by eq. 2.18. A much more detailed derivation can be found in the book of Risken [41], which only has the Fokker-Planck equation as its topic.

The first coefficient ($\nu = 1$) is called the drift coefficient, while the second coefficient ($\nu = 2$) is known as the diffusion coefficient. As the time change τ tends to 0, these coefficients are given by the short-time behavior of the system, i. e. short in comparison to the slow degrees of freedom of the system. This formula should not be confused with the diffusion constant determined from the long-time behavior (see eq. 2.79), where the time change Δt is large in comparison to the fast degrees of freedom that are not explicitly taken into account.

For convenience one can rewrite the Fokker-Planck equation (see eq. 2.20) as:

$$\frac{\partial}{\partial t} P(\vec{X}, t|\vec{X}_0, t_0) = \mathrm{i}\mathcal{L} P(\vec{X}, t|\vec{X}_0, t_0)\ , \tag{2.22}$$

which defines the *Fokker-Planck operator* $\mathcal{L}$. In case of a homogeneous process, i. e. when D^ν does not depend on t, the formal solution to the Fokker-Planck equation is simply:

$$P(\vec{X}, t|\vec{X}_0, t_0) = \mathrm{e}^{\mathrm{i}\mathcal{L}(t-t_0)} \delta(x - x_0) = \mathcal{T} \delta(x - x_0) \tag{2.23}$$

Here, we basically moved all of the difficulty into the *time-evolution operator* $\mathcal{T}$. Due to its exponential form, this operator has in the case of Hamiltonian dynamics several beneficial properties:

- Time-reversibility: $\Delta t = (t - t_0) \rightarrow -\Delta t$: $\mathcal{T} \rightarrow \mathcal{T}^{-1}$
- Conservation of phase space volume: $\mathcal{T}$ is unitary[1]

Therefore, a operator splitting of $\mathcal{T}$ will always lead to a symplectic integrator due to the fact that time-reversibility and conservation of phase space volume is defining symplecticity. We will show later that such a splitting of the deterministic Fokker-Planck operator will lead to the velocity Verlet algorithm.

2.2.1.1 Pawula Theorem

At this point, there is only one important connection between the Kramers-Moyal coefficients missing, namely the *Pawula Theorem*:

> If for some ν, one has $D^\nu = D^{\nu+1} = \ldots = 0$, then one also has $D^3 = D^4 = \ldots = 0$

The proof is not very difficult and simply uses the fact that the averaging $< >$ in the Kramers-Moyal coefficients (see eq. 2.21) can be used as a weighted scalar product of the $\mathcal{L}_2$ Hilbert space with P as weighting function. Therefore, one can make use of the Cauchy-Schwarz inequality to derive a recursion for the Kramers-Moyal coefficients $D^\nu(\vec{X}, t)$.

However, the conclusions one can draw are very important. Basically, there exist only *four* kind of Fokker-Planck processes:

- No dynamics: $D^1 = D^2 = \ldots = 0$
- Deterministic dynamics: $D^2 = D^3 = \ldots = 0$
- Standard Fokker-Planck process: $D^3 = D^4 = \ldots = 0$
- Non-trivial dynamics: all $D^k \neq 0$

The last case applies for all *real* world examples, but is not considered here as it can be approximated in most cases as a standard Fokker-Planck process.

[1] This is not completely trivial, one still has to show that $\mathcal{L}$ is self-adjoint in the Hilbert space, where the averaging $< >$ is used as a weighted scalar product.

2.2.2 Deterministic Dynamics

It is not obvious that the case $D^2 = D^3 = \ldots = 0$ is really a deterministic case that is why we look into that in detail for Hamiltonian dynamics. First we calculate the drift coefficient of Newton's equations, eq. 2.3 and eq. 2.4

$$D^1(\vec{X},t) = \lim_{\tau\to 0} \frac{<\Delta\vec{X}>}{\tau} = \lim_{\tau\to 0} \frac{\dot{\vec{X}}(t)\tau + \mathcal{O}(\tau^2)}{\tau} = \dot{\vec{X}}(t) \ , \tag{2.24}$$

which are just the velocity and the acceleration of all particles written as one vector $\dot{\vec{X}} = \{\dot{\vec{x}}_i, \dot{\vec{p}}_i\}$. Furthermore, it is known that a deterministic process has the probability distribution

$$P(\vec{X},t|\vec{X}_0,t_0) = \delta(\vec{Y}(t) - \vec{X}) \ , \tag{2.25}$$

where $\vec{Y}(t)$ is the trajectory with the initial condition $\vec{Y}(t_0) = \vec{X}_0$, which is generated by Newton's equations of motion. Finally, one can show that this solution obeys the Fokker-Planck equation:

$$\begin{aligned}
\frac{\partial}{\partial t}P(\vec{X},t|\vec{X}_0,t_0) &= \frac{\partial}{\partial t}\delta(\vec{Y}(t) - \vec{X}) \\
&= -\dot{\vec{Y}}(t)\cdot\frac{\partial}{\partial\vec{X}}\delta(\vec{Y}(t) - \vec{X}) \\
&= -\frac{\partial}{\partial\vec{X}}\cdot\left[\dot{\vec{Y}}(t)\delta(\vec{Y}(t) - \vec{X})\right] \\
&= -\frac{\partial}{\partial\vec{X}}\cdot\left[D^1(\vec{Y}(t),t)\delta(\vec{Y}(t) - \vec{X})\right] \\
&= -\frac{\partial}{\partial\vec{X}}\cdot\left[D^1(\vec{X},t)\delta(\vec{Y}(t) - \vec{X})\right]
\end{aligned} \tag{2.26}$$

After this proof, we can write down the Fokker-Planck operator for this process

$$\mathrm{i}\mathcal{L}_\mathrm{D} = -\sum_i \left(\frac{\partial}{\partial\vec{r}_i}\cdot\dot{\vec{r}}_i + \frac{\partial}{\partial\vec{p}_i}\cdot\dot{\vec{p}}_i\right) = -\sum_i \left(\frac{\vec{p}_i}{m_i}\cdot\frac{\partial}{\partial\vec{r}_i} + \vec{F}_i\cdot\frac{\partial}{\partial\vec{p}_i}\right) \ , \tag{2.27}$$

where the subindex D stands for deterministic. This operator is also known as the *Liouville operator*.

2.2.2.1 Velocity Verlet Algorithm

The deterministic Fokker-Planck operator can be split into a position and a momentum part, $\mathcal{L}_r$ and $\mathcal{L}_p$. Under the assumption of having a homogeneous process, which is true as long as the forces do not depend explicitly on

the time, one can split the deterministic time-evolution operator (see eq. 2.23) using a Trotter expansion:

$$\mathcal{T}_{\mathrm{D}} = \mathrm{e}^{\mathrm{i}\mathcal{L}_{\mathrm{D}}\Delta t} = \mathrm{e}^{\mathrm{i}(\mathcal{L}_{\mathrm{r}}+\mathcal{L}_{\mathrm{p}})\Delta t} = \mathrm{e}^{\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\mathrm{e}^{\mathrm{i}\mathcal{L}_{\mathrm{r}}\Delta t}\mathrm{e}^{\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2} + \mathcal{O}(\Delta t^3) \tag{2.28}$$

Hence, the position and momentum propagation, which are generated by the inverse time-evolution operator $\mathcal{T}_{\mathrm{D}}^{-1}$, decouple as well:

$$\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\vec{r}_i = \vec{r}_i \tag{2.29}$$

$$\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\vec{p}_i = (1 - \mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2)\,\vec{p}_i + 0 = \vec{p}_i + \frac{\Delta t}{2}\vec{F}_i \tag{2.30}$$

$$\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{r}}\Delta t}\vec{p}_i = \vec{p}_i \tag{2.31}$$

$$\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{r}}\Delta t}\vec{r}_i = (1 - \mathrm{i}\mathcal{L}_{\mathrm{r}}\Delta t)\,\vec{r}_i + 0 = \vec{r}_i + \Delta t\frac{\vec{p}_i}{m_i} \tag{2.32}$$

Thus, one immediately obtains the *velocity Verlet algorithm* (see eq. 2.11 and eq. 2.12):

$$\begin{aligned}
\vec{r}_i(t+\Delta t) &= \mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{r}}\Delta t}\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\vec{r}_i(t) + \mathcal{O}(\Delta t^3) \\
&= \mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\left[\vec{r}_i + \Delta t\frac{\vec{p}_i(t)}{m_i}\right] + \mathcal{O}(\Delta t^3) \\
&= \vec{r}_i + \frac{\Delta t}{m_i}\vec{p}_i(t) + \frac{\Delta t^2}{2m_i}\vec{F}_i(t) + \mathcal{O}(\Delta t^3)\ ,
\end{aligned} \tag{2.33}$$

$$\begin{aligned}
\vec{p}_i(t+\Delta t) &= \mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{r}}\Delta t}\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\vec{p}_i(t) + \mathcal{O}(\Delta t^3) \\
&= \mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{r}}\Delta t}\left[\vec{p}_i(t) + \frac{\Delta t}{2}F_i(\{\vec{r}_j(t)\})\right] + \mathcal{O}(\Delta t^3) \\
&= \mathrm{e}^{-\mathrm{i}\mathcal{L}_{\mathrm{p}}\Delta t/2}\left[\vec{p}_i(t) + \frac{\Delta t}{2}\vec{F}_i(\{\vec{r}_j(t+\Delta t)\})\right] + \mathcal{O}(\Delta t^3) \\
&= \vec{p}_i(t) + \frac{\Delta t}{2}\left(\vec{F}_i(t) + \vec{F}_i(t+\Delta t)\right) + \mathcal{O}(\Delta t^3)
\end{aligned} \tag{2.34}$$

Here, we used the assumption that $\vec{F}_i$ depends on the position, but not on the velocities. After this derivation one can basically understand why the velocity Verlet algorithm is a *symplectic integrator*.

The advantage of this derivation compared to the one presented before is that the calculation can be done for any order in Δt using a higher Trotter expansion and for other equations of motion, like the stochastic ones in the next section. In addition, a different splitting, instead of position and momentum, of the Fokker-Planck operator will lead to a different integrator, which is used in the case of thermostats.

2.2.3 Canonical ensemble

In the previous part of the chapter, deterministic dynamics were considered, which basically describe a micro-canonical ensemble (NVE). However, most of the time, one is more interested in a canonical ensemble (NVT) to compare with the experiment, but also to stabilize the system to overcome the errors made by the integrator. In a canonical ensemble the system is in a thermodynamic equilibrium with a heat-bath instead of being isolated without energy exchange.

There are certain additions to Newton's equations of motion to generate a canonical ensemble, which are called thermostats. They can act in a local or a global fashion. A local thermostat acts on degrees of freedom, which only depend on local properties of the system, e. g. particle velocity, relative velocity between two neighboring particles or a random force. In contrast, a global thermostat acts on a degree of freedom of the whole system, e. g. the kinetic energy itself or an additional virtual mass of the system.

2.2.3.1 Rescaling Thermostats

The simplest way to adjust the temperature to a certain value T_0 is to rescale all velocities with a factor

$$\lambda = \sqrt{T/T_0} \,, \tag{2.35}$$

where T is the actual temperature of the system. Two widely used examples for rescaling thermostats are the Berendsen thermostat [42] and the stochastic velocity rescaling thermostat [43]. The Berendsen thermostat rescales the kinetic energy K with a constant rate to the desired mean value.

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \frac{K(T) - K(T_0)}{\tau} \,, \tag{2.36}$$

where τ is the coupling constant of the Berendsen thermostat. This procedure will lead to the right mean, but not to the correct fluctuations of the temperature. Therefore, the Berendsen thermostat does not generate a "real" canonical ensemble, but an iso-kinetic ensemble. To overcome this problem the velocity rescaling thermostat of Bussi et al. [43] adds a stochastic term to the kinetic energy.

$$\frac{\mathrm{d}K}{\mathrm{d}t} = \frac{K(T) - K(T_0)}{\tau} + 2\sqrt{\frac{K(T)K(T_0)}{\tau}}\eta \,, \tag{2.37}$$

where η is white noise with a certain fixed variance. However, both thermostats act in a global manner, which can lead to unphysical situations, due to the fact that the distribution of the kinetic energy is correct, but the

distributions of individual degrees of freedom are wrong, e. g. the so-called *flying ice cube*, where all bonds are frozen and the center of system moves with a very high velocity.

2.2.4 Langevin Thermostats

The basic idea behind Langevin Thermostats [44] is to couple a molecule to a background fluid, which is actually not simulated. This fluid and the particles rub against each other, so the particles are damped according to their velocity by friction and get randomly moved by the background fluid. The correct ratio of friction and random noise ensures the right temperature.

Due to this randomness in the motion one calls this propagation *Stochastic dynamics*, whose equation of motion is given by:

$$m_i \ddot{\vec{r}}_i = \vec{F}_i^{\mathrm{C}} + \vec{F}_i^{\mathrm{D}} + \vec{F}_i^{\mathrm{R}} = \vec{F}_i - \xi_i \vec{v}_i + \sigma_i \vec{\eta}_i \ , \tag{2.38}$$

where $\vec{F}_i^{\mathrm{C}}$, $\vec{F}_i^{\mathrm{D}}$ and $\vec{F}_i^{\mathrm{R}}$ stand for conservative, damping and random force, respectively.

The damping force $\vec{F}_i^{\mathrm{D}}$ is a Stokes-like friction force, which acts against the direction of the velocity. The random force $\vec{F}_i^{\mathrm{R}}$ acts in a random direction and is completely local, due to the fact that the background fluid is idealized and there are no interactions of the fluid with itself. This locality leads to independent noise, $\vec{\eta}_i$, in space, time and particles. The noise and the friction pre-factors, σ_i and η_i, are particle type dependent, because different particles can have different sizes and surface roughness.

The ratio between ξ_i and η_i can be derived from the Fokker-Planck equation (see eq. 2.22). In the Fokker-Planck picture, having the correct distribution means that the stationary solution of the Fokker-Planck equation

$$\mathrm{i}\mathcal{L}P = \frac{\partial}{\partial t} P = 0 \tag{2.39}$$

is the canonical distribution. This simplifies to

$$\mathrm{i}\mathcal{L}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} = 0 \ , \tag{2.40}$$

ignoring the normalization factor of the canonical distribution. Due to the fact that this equation of motion is nothing else than Newton's equation of motion with some additions one can split the Fokker-Planck operator in a deterministic part $\mathcal{L}_{\mathrm{D}}$ and a stochastic dynamics part $\mathcal{L}_{\mathrm{SD}}$:

$$\mathrm{i}\mathcal{L} = \mathrm{i}\mathcal{L}_{\mathrm{D}} + \mathrm{i}\mathcal{L}_{\mathrm{SD}} \ . \tag{2.41}$$

We will first discuss the deterministic part.

2.2.4.1 Deterministic Part

We already know the deterministic Fokker-Planck operator from above (see sec. 2.2.2):

$$\mathrm{i}\mathcal{L}_\mathrm{D} = -\sum_i \left(\frac{\vec{p}_i}{m_i} \cdot \frac{\partial}{\partial \vec{r}_i} + \vec{F}_i \cdot \frac{\partial}{\partial \vec{p}_i} \right) \tag{2.42}$$

So one can calculate:

$$\begin{aligned}
\mathrm{i}\mathcal{L}_\mathrm{D} \mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} &= \mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \frac{-1}{\mathrm{k_B}T} \mathrm{i}\mathcal{L}_\mathrm{D}\mathcal{H} \\
&= \mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \frac{1}{\mathrm{k_B}T} \sum_i \left(\frac{\vec{p}_i}{m_i} \cdot \frac{\partial}{\partial \vec{r}_i} + \vec{F}_i \cdot \frac{\partial}{\partial \vec{p}_i} \right) \mathcal{H} \\
&= \mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \frac{1}{\mathrm{k_B}T} \sum_i \left(\frac{\vec{p}_i}{m_i} \cdot \frac{\partial \mathcal{H}}{\partial \vec{r}_i} + \vec{F}_i \cdot \frac{\partial \mathcal{H}}{\partial \vec{p}_i} \right) \\
&= \mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \frac{1}{\mathrm{k_B}T} \sum_i \left(\frac{\partial \mathcal{H}}{\partial \vec{p}_i} \cdot \frac{\partial \mathcal{H}}{\partial \vec{r}_i} - \frac{\partial \mathcal{H}}{\partial \vec{r}_i} \cdot \frac{\partial \mathcal{H}}{\partial \vec{p}_i} \right) \\
&= 0
\end{aligned} \tag{2.43}$$

Hence, the deterministic Fokker-Planck operator accepts $\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T}$ as an equilibrium distribution.

2.2.4.2 Stochastic Part

It is left to show that:

$$\mathrm{i}\mathcal{L}_\mathrm{SD} \mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} = 0 \tag{2.44}$$

For this, we have to derive $\mathrm{i}\mathcal{L}_\mathrm{SD}$, which depends on the Kramers-Moyal coefficients. The coefficients in $\vec{r}_i$ are zero due to the fact that Newton's equation in $\vec{r}_i$ (see eq. 2.3) stays unchanged in eq. 2.38. For the drift coefficients in $\vec{p}_i$ one obtains:

$$\begin{aligned}
D_i^1(\vec{p},t) &= \lim_{\tau \to 0} \frac{< \Delta \vec{p}_i >}{\tau} \\
&= \lim_{\tau \to 0} \frac{< \Delta(-\xi_i \vec{v}_i + \sigma_i \vec{\eta}_i >}{\tau} \\
&= \lim_{\tau \to 0} \frac{-\xi_i \vec{v}_i(t)\tau + \sigma_i \int_t^{t+\tau} \mathrm{d}t' \ < \vec{\eta}_i(t') > + \mathcal{O}(\tau^2)}{\tau} \\
&= -\xi_i \vec{v}_i(t) \ ,
\end{aligned} \tag{2.45}$$

where we made use of the fact that the random noise is homogenous in time, $<\vec{\eta}_i>=0$. The corresponding diffusion coefficients are given by:

$$\begin{aligned} D_{ij}^2(\vec{p},t) &= \lim_{\tau\to 0}\frac{<\Delta\vec{p}_i\otimes\Delta\vec{p}_j>}{2\tau} \\ &= \lim_{\tau\to 0}\frac{\xi_i\vec{v}_i\otimes\xi_j\vec{v}_j\tau^2+\int_t^{t+\tau}\mathrm{d}t'\ \int_t^{t+\tau}\mathrm{d}t''\ \sigma_i\sigma_j<\vec{\eta}_i(t')\otimes\vec{\eta}_j(t'')>}{2\tau} \\ &\quad -\lim_{\tau\to 0}\frac{\xi_i\vec{v}_i\tau\otimes\sigma_j\int_t^{t+\tau}\mathrm{d}t'\ <\vec{\eta}_j(t')>}{2\tau} \\ &\quad -\lim_{\tau\to 0}\frac{\xi_j\vec{v}_j\tau\otimes\sigma_i\int_t^{t+\tau}\mathrm{d}t'\ <\vec{\eta}_i(t')>}{2\tau}+\lim_{\tau\to 0}\frac{\mathcal{O}(\tau^2)}{2\tau} \\ &= \lim_{\tau\to 0}\frac{\sigma_i\sigma_j\delta_{ij}c\mathbb{1}\int_t^{t+\tau}\mathrm{d}t'\ \int_t^{t+\tau}\mathrm{d}t''\ \delta(t'-t'')+\mathcal{O}(\tau^2)}{2\tau} \\ &= \frac{\sigma_i^2}{2}\delta_{ij}c\mathbb{1}\ , \end{aligned} \tag{2.46}$$

where c is a constant, which determines the variance of the noise, which we chose to be 2 for convenience. Again, one relies on the fact that the noise is independent in space and time. For now we assume, that all higher Kramers-Moyal coefficients vanish due to the fact that this is a standard Fokker-Planck process (see the Pawula Theorem in sec. 2.2.1). However, one can also obtain such vanishing coefficients by requiring some special noise, i. e. Gaussian white noise. It is important to note that the terms in the first order of τ actually determine the Kramers-Moyal coefficients.

From eq. 2.45 and eq. 2.46 we see that the stochastic part of the Fokker-Planck operator is:

$$\mathrm{i}\mathcal{L}_{\mathrm{SD}}=\sum_i\frac{\partial}{\partial\vec{p}_i}\cdot\left(\xi_i\vec{v}_i+\sigma_i^2\frac{\partial}{\partial\vec{p}_i}\right) \tag{2.47}$$

Finally one can determine the ratio between noise and friction:

$$\begin{aligned} \mathrm{i}\mathcal{L}_{\mathrm{SD}}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} &= \sum_i\frac{\partial}{\partial\vec{p}_i}\cdot\left(\xi_i\vec{v}_i+\sigma_i^2\frac{\partial}{\partial\vec{p}_i}\right)\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \\ &= \sum_{i\alpha}\frac{\partial}{\partial p_i^\alpha}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T}\left(\xi_i v_i^\alpha-\frac{\sigma_i^2}{\mathrm{k_B}T}\frac{\partial\mathcal{H}}{\partial p_i^\alpha}\right) \\ &= \sum_{i\alpha}\frac{\partial}{\partial p_i^\alpha}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T}\frac{\partial\mathcal{H}}{\partial p_i^\alpha}\left(\xi_i-\frac{\sigma_i^2}{\mathrm{k_B}T}\right) \end{aligned} \tag{2.48}$$

This can be read as the *Fluctuation-Dissipation theorem*:

$$\mathrm{i}\mathcal{L}_{\mathrm{SD}}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T}=0\leftrightarrow\xi_i\mathrm{k_B}T=\sigma_i^2 \tag{2.49}$$

The left side is the requirement that $e^{-\mathcal{H}/k_B T}$ is the equilibrium distribution, which was proven to be equivalent to having the right ratio between noise and friction, $\sigma_i^2/\xi_i = k_B T$.

2.2.4.3 Summary

If one has a Langevin-type equation of motion

$$m_i \ddot{\vec{r}}_i = \vec{F}_i^{\mathrm{C}} + \vec{F}_i^{\mathrm{D}} + \vec{F}_i^{\mathrm{R}} = \vec{F}_i - \xi_i \vec{v}_i + \sigma_i \vec{\eta}_i \ , \tag{2.50}$$

then one generates a canonical ensemble if the *Fluctuation-Dissipation theorem* is fulfilled:

$$\xi_i k_B T = \sigma_i^2 \tag{2.51}$$

The noise has to have certain properties:

- Homogeneity: $< \vec{\eta}_i >= 0$
- Independency in time and space: $< \vec{\eta}_i(t) \otimes \vec{\eta}_j(t') >= 2\delta_{ij}\delta(t-t')\mathbb{1}$

Above, we have assumed that this motion is a standard Fokker-Planck process to save the calculation of the higher Kramers-Moyal coefficients. Actually, one would have to prove that all higher moments in $\Delta\vec{p}_i$ have an order above τ^1, which is the case for Gaussian white noise.

However, for integrators, which are only first order accurate in Δt, one can use random numbers from any symmetric distribution, as long as the second moment has the correct variance [45]. Basically, the error made in drawing the random number is of the same order as the error made by the integrator. Thus, there is nothing to gain or loose by using box-shaped noise instead of Gaussian white noise. For second order integrators the 4th moment has to be correct as well.

The Langevin thermostat is one of the standard local stochastic thermostats, which generates a canonical ensemble. In addition these local thermostats stabilize the system, which global thermostats usually do not do. Also the rescaling thermostat of Bussi et al. [43] can be viewed as a Langevin thermostat however applied to the whole kinetic energy as a degree of freedom. Nevertheless, all Langevin thermostats discussed so far do not conserve momentum, which is critical from the point of view of hydrodynamics. Therefore, the Dissipative Particle Dynamics Thermostat, which does conserve the momentum, will be discussed in the next section

2.2.5 Dissipative Particle Dynamics Thermostat

The initial idea of *dissipative particle dynamics* (DPD) was to merge several individual particles to a larger particle that interact via a much smoother interaction, basically a weak harmonic interaction. So weak that these fluffy particles can even interpenetrate each other and hence, the dynamics of the particles is not really physical. However, for the point of statistic averages DPD can serve as a very good model for fluids.

As in the case of stochastic dynamics a damping and a random force are added to generate a canonical ensemble. Combining the damping and random parts of the DPD with a common force field results in a *DPD thermostat* [46]. The DPD thermostat is known to satisfy Newton's third law by construction. Due to mass, momentum and temperature conservation, hydrodynamics is also correctly reproduced [47]. This will be very useful for both, coarse-grained and adaptive resolution, simulations of soft matter.

In Langevin dynamics, two additional forces are introduced, a damping and a random force, whose ratio defines the temperature. For DPD, the approach is similar [13, 48]:

$$\ddot{\vec{r}}_i = \vec{F}_i^{\mathrm{C}} + \vec{F}_i^{\mathrm{D}} + \vec{F}_i^{\mathrm{R}} \ , \tag{2.52}$$

where $\vec{F}_i^{\mathrm{C}}$ denotes the conservative force on the ith particle. The damping and random forces can be split up in particle pair forces

$$\vec{F}_i^{\mathrm{D}} = \sum_{j \neq i} \vec{F}_{ij}^{\mathrm{D}} \ , \tag{2.53}$$

$$\vec{F}_i^{\mathrm{R}} = \sum_{j \neq i} \vec{F}_{ij}^{\mathrm{R}} \ , \tag{2.54}$$

where the damping force reads as

$$\vec{F}_{ij}^{\mathrm{D}} = -\xi_{ij}^{\parallel} w^{\mathrm{D}}(r_{ij})(\hat{r}_{ij} \cdot \vec{v}_{ij})\hat{r}_{ij}, \tag{2.55}$$

and the random force is given by

$$\vec{F}_{ij}^{\mathrm{R}} = \sigma_{ij}^{\parallel} w^{\mathrm{R}}(r_{ij})\eta_{ij}\hat{r}_{ij} \ . \tag{2.56}$$

In these equations the relative velocity $\vec{v}_{ij} = \vec{v}_i - \vec{v}_j$ between the ith and jth particle is introduced, while $\hat{r}_{ij}$ denotes the unit vector of the interatomic axis $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$.

It is important to note that both equations satisfy Newton's third law, which leads to the correct hydrodynamics. $\xi_{ij}^{\parallel}$ is the friction constant and $\sigma_{ij}^{\parallel}$

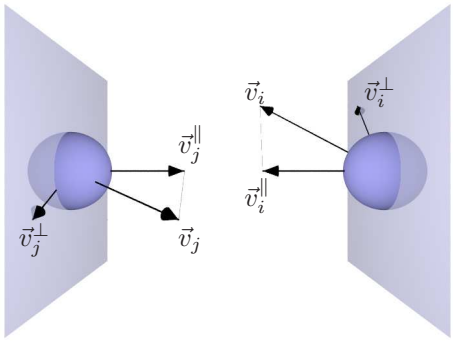

Figure 2.2: **Decomposition of the relative velocity**: The decomposition of the relative velocity between two particles, in the components on the interatomic axis and perpendicular to this space, leads in a natural way to the standard and transverse DPD thermostat. DPD thermostat

is the noise strength, which depend on the particle types. $w^{\mathrm{D}}(r_{ij})$ and $w^{\mathrm{R}}(r_{ij})$ are r-dependent weight functions. In sec. 2.2.8 we will see that the coupling to the thermostat is weak and hence the dynamics is not influenced very much [46]. These parameters are connected by the fluctuation-dissipation theorem, which has a similar form as in the case of stochastic dynamics (see eq. 2.51). We will prove this in the remainder of this section.

2.2.5.1 Noise Properties

The noise properties of stochastic dynamics have to be slightly generalized due to the fact that in the case of DPD we have a random force acting on pairs of particles. The noise variable η_{ij} is:

- Symmetric: $\eta_{ij} = \eta_{ji}$
- Homogenous: $\langle \eta_{ij}(t) \rangle = 0$
- Independent in time and particle pairs, symmetric in pairs (Newton's third law):

$$\langle \eta_{ij}(t)\eta_{kl}(t') \rangle = 2(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\delta(t - t')$$

With this knowledge about the noise, we can now derive the Fokker-Planck operator for DPD.

2.2.5.2 DPD Fokker-Planck operator

As Newton's equation in $\vec{r}_i$ (see eq. 2.3) stays unchanged, one can apply the same separation ansatz for the Fokker-Planck operator as for stochastic dynamics (see eq. 2.41) and split the operator into two parts:

$$\mathrm{i}\mathcal{L} = \mathrm{i}\mathcal{L}_{\mathrm{D}} + \mathrm{i}\mathcal{L}_{\mathrm{DPD}} \tag{2.57}$$

For the deterministic part, it is known that (see eq. 2.43):

$$\mathrm{i}\mathcal{L}_{\mathrm{D}}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} = 0 \tag{2.58}$$

To show the equivalent for the DPD part, $\mathcal{L}_{\mathrm{DPD}}$, we first have to calculate the first Kramers-Moyal coefficients

$$\begin{aligned} D_i^1(\vec{p},t) &= \lim_{\tau\to 0}\frac{<\Delta\vec{p}_i(t)>}{\tau} \\ &= \lim_{\tau\to 0}\sum_j \frac{\xi_{ij}^{\parallel} w^{\mathrm{D}}(r_{ij}(t))(\hat{r}_{ij}(t)\cdot\vec{v}_{ij}(t))\hat{r}_{ij}(t)\tau}{\tau} \\ &\quad + \lim_{\tau\to 0}\sum_j \frac{\sigma_{ij}^{\parallel} w^{\mathrm{R}}(r_{ij}(t))\hat{r}_{ij}(t)\int_t^{t+\tau}\mathrm{d}t' \ <\eta_{ij}(t')> +\mathcal{O}(\tau^2)}{\tau} \\ &= \sum_j \left(\xi_{ij}^{\parallel} w^{\mathrm{D}}(r_{ij}(t))(\hat{r}_{ij}(t)\cdot\vec{v}_{ij}(t))\hat{r}_{ij}(t)\right) \end{aligned} \tag{2.59}$$

and the second Kramers-Moyal coefficients

$$
\begin{aligned}
D^2_{ij}(\vec{p},t) &= \lim_{\tau\to 0}\frac{<\Delta\vec{p}_i\otimes\Delta\vec{p}_j>}{2\tau} \\
&= \lim_{\tau\to 0}\sum_{i'j'}\frac{\dots\int_t^{t+\tau}\mathrm{d}t' \ <\eta_{ii'}(t')>+\dots\int_t^{t+\tau}\mathrm{d}t' \ <\eta_{jj'}(t')>}{2\tau} \\
&\quad +\lim_{\tau\to 0}\sum_{i'j'}\frac{\sigma^{\parallel}_{ii'}w^{\mathrm{R}}(r_{ii'}(t))\sigma^{\parallel}_{jj'}w^{\mathrm{R}}(r_{jj'}(t))\hat{r}_{ii'}(t)\otimes\hat{r}_{jj'}(t)+\mathcal{O}(\tau)}{2\tau}\cdot \\
&\quad \cdot\int_t^{t+\tau}\mathrm{d}t'\int_t^{t+\tau}\mathrm{d}t'' \ <\eta_{ii'}(t')\eta_{jj'}(t'')> \\
&\quad +\lim_{\tau\to 0}\frac{\mathcal{O}(\tau^2)}{2\tau} \\
&= \lim_{\tau\to 0}\sum_{i'j'}\frac{\sigma^{\parallel}_{ii'}w^{\mathrm{R}}(r_{ii'}(t))\sigma^{\parallel}_{jj'}w^{\mathrm{R}}(r_{jj'}(t))\hat{r}_{ii'}(t)\otimes\hat{r}_{jj'}(t)+\mathcal{O}(\tau)}{2\tau}\cdot \\
&\quad \cdot 2(\delta_{ij}\delta_{i'j'}+\delta_{ij'}\delta_{i'j})\tau \\
&= \delta_{ij}\sum_{i'}\sigma^{\parallel}_{ii'}w^{\mathrm{R}}(r_{ii'}(t))\sigma^{\parallel}_{ji'}w^{\mathrm{R}}(r_{ji'}(t))\hat{r}_{ii'}(t)\otimes\hat{r}_{ji'}(t) \\
&\quad +\sigma^{\parallel}_{ij}w^{\mathrm{R}}(r_{ij}(t))\sigma^{\parallel}_{ji}w^{\mathrm{R}}(r_{ji}(t))\hat{r}_{ij}(t)\otimes\hat{r}_{ji}(t) \\
&= \delta_{ij}\sum_{i'}[\sigma^{\parallel}_{ii'}w^{\mathrm{R}}(r_{ii'}(t))]^2\hat{r}_{ii'}(t)\otimes\hat{r}_{ii'}(t) \\
&\quad -[\sigma^{\parallel}_{ij}w^{\mathrm{R}}(r_{ij}(t))]^2\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t)\ , \qquad (2.60)
\end{aligned}
$$

where the term $i = j$ was added to the sums, which can be done because it is zero.

The calculations are very similar to those for stochastic dynamics (see eq. 2.45 and eq. 2.46). Due to the limit in τ one only has to determine the term in first order of τ.

Putting it all together, the Fokker-Planck operator of dissipative particle

dynamics is:

$$\begin{aligned}
\mathrm{i}\mathcal{L}_{\mathrm{DPD}} &= -\sum_i \frac{\partial}{\partial \vec{p}_i} D_i^1(\vec{p},t) \\
&\quad + \sum_{ij} \frac{\partial}{\partial \vec{p}_i}\frac{\partial}{\partial \vec{p}_j} D_{ij}^2(\vec{p},t) \\
&= \sum_{ij\alpha} \frac{\partial}{\partial p_i^\alpha} \xi_{ij}^{\parallel} w^{\mathrm{D}}(r_{ij}(t))(\hat{r}_{ij}(t)\cdot\vec{v}_{ij}(t))\hat{r}_{ij}^\alpha(t) \\
&\quad + \sum_{ii'\alpha\beta} \frac{\partial}{\partial p_i^\alpha}\frac{\partial}{\partial p_i^\beta}[\sigma_{ii'}^{\parallel} w^{\mathrm{R}}(r_{ii'}(t))]^2 \left(\hat{r}_{ii'}(t)\otimes\hat{r}_{ii'}(t)\right)^{\alpha\beta} \\
&\quad - \sum_{ij\alpha\beta} \frac{\partial}{\partial p_i^\alpha}\frac{\partial}{\partial p_j^\beta}[\sigma_{ij}^{\parallel} w^{\mathrm{R}}(r_{ij}(t))]^2 \left(\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t)\right)^{\alpha\beta} \\
&= \sum_{ij} \xi_{ij}^{\parallel} w^{\mathrm{D}}(r_{ij}(t)) \frac{\partial}{\partial \vec{p}_i}\cdot(\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t))\cdot\vec{v}_{ij}(t) \\
&\quad + \sum_{ij} [\sigma_{ij}^{\parallel} w^{\mathrm{R}}(r_{ij}(t))]^2 \frac{\partial}{\partial \vec{p}_i}\cdot(\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t))\cdot\left[\frac{\partial}{\partial \vec{p}_i} - \frac{\partial}{\partial \vec{p}_j}\right] ,
\end{aligned} \tag{2.61}$$

where $\hat{r}_{ij}\otimes\hat{r}_{ij}$ is the projector on the interatomic axis between particle i and particle j.

In the literature [44, 47] the Fokker-Planck operator is then further simplified to

$$\begin{aligned}
\mathrm{i}\mathcal{L}_{\mathrm{DPD}} &= \sum_{ij} \xi_{ij}^{\parallel} w^{\mathrm{D}}(r_{ij}(t))\hat{r}_{ij}(t)\cdot\frac{\partial}{\partial \vec{p}_i}\hat{r}_{ij}(t)\cdot\vec{v}_{ij}(t) \\
&\quad + \sum_{ij} [\sigma_{ij}^{\parallel} w^{\mathrm{R}}(r_{ij}(t))]^2 \hat{r}_{ij}(t)\cdot\frac{\partial}{\partial \vec{p}_i}\hat{r}_{ij}(t)\cdot\left[\frac{\partial}{\partial \vec{p}_i} - \frac{\partial}{\partial \vec{p}_j}\right] ,
\end{aligned} \tag{2.62}$$

but without that simplification one can identify the projector $\hat{r}_{ij}\otimes\hat{r}_{ij}$ in the Fokker-Planck operator, which later allow us to generalize this calculation for any projector.

Finally one can determine how the DPD Fokker-Planck operator acts on

the equilibrium distribution:

$$\begin{aligned}
\mathrm{i}\mathcal{L}_{\mathrm{DPD}}P &= \mathrm{i}\mathcal{L}_{\mathrm{DPD}}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \\
&= \sum_{ij}\xi_{ij}^{\parallel}w^{\mathrm{D}}(r_{ij}(t))\frac{\partial}{\partial\vec{p_i}}\cdot(\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t))\cdot\vec{v}_{ij}(t)\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \\
&\quad -\sum_{ij}\frac{[\sigma_{ij}^{\parallel}w^{\mathrm{R}}(r_{ij}(t))]^2}{\mathrm{k_B}T}\times \\
&\quad \times\frac{\partial}{\partial\vec{p_i}}\cdot(\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t))\cdot\left[\frac{\partial\mathcal{H}}{\partial\vec{p_i}}-\frac{\partial\mathcal{H}}{\partial\vec{p_j}}\right]\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T} \\
&= \sum_{ij}\left[\xi_{ij}^{\parallel}w^{\mathrm{D}}(r_{ij}(t))-\frac{[\sigma_{ij}^{\parallel}w^{\mathrm{R}}(r_{ij}(t))]^2}{\mathrm{k_B}T}\right]\times \\
&\quad \times\frac{\partial}{\partial\vec{p_i}}\cdot(\hat{r}_{ij}(t)\otimes\hat{r}_{ij}(t))\cdot\vec{v}_{ij}(t)\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T}
\end{aligned} \tag{2.63}$$

So the fluctuation-dissipation theorem [47] reads

$$\mathrm{i}\mathcal{L}_{\mathrm{DPD}}\mathrm{e}^{-\mathcal{H}/\mathrm{k_B}T}=0\leftrightarrow\xi_{ij}^{\parallel}w^{\mathrm{D}}(r_{ij})\mathrm{k_B}T=[\sigma_{ij}^{\parallel}w^{\mathrm{R}}(r_{ij})]^2\ , \tag{2.64}$$

which for convenience can be split in a normalized distance dependent part

$$\left(w^{\mathrm{R}}(r)\right)^2=w^{\mathrm{D}}(r) \tag{2.65}$$

and a part depending on the strength of noise and friction

$$\xi_{ij}^{\parallel}\mathrm{k_B}T=(\sigma_{ij}^{\parallel})^2\ . \tag{2.66}$$

It is important to note that the second equation has the same form as in the stochastic dynamic case (see eq. 2.51), but the freedom is much larger. In addition to the friction constant, $\xi_{ij}^{\parallel}$, there is a function $w_{ij}^{\mathrm{R}}(r)$ to choose. Usually w_{ij}^{R} is called weighting function and has the form:

$$w_{ij}^{\mathrm{R}}(r)=(1-r/r_{\mathrm{cut}})^n,\ \ n=1,2,3,\ldots,\infty \tag{2.67}$$

With increasing n, the function tends towards the Heaviside function. A detailed study about the influence of the weighting function can be found in [49].

The above DPD equations conserve the total momentum and correctly reproduce the hydrodynamic interactions in the system.

However, as we will validate for a Lennard-Jones fluid later in sec. 2.2.8 and as previous studies [46, 50] have shown, the strength of the friction $\zeta^{\parallel}$ does not influence the viscosity in linear order. To be able to tune the value of the viscosity of the system while preserving all the virtues of the standard DPD thermostat presented above, its extension named *Transversal DPD thermostat* is introduced in the next subsection.

2.2.6 Transversal DPD Thermostat

As it is known in the literature [46] and will be shown later for a simple liquid (see sec. 2.2.8), the DPD thermostat in the above form is not capable of controlling liquid properties such as viscosity and the diffusion constant. Hence, we extend the DPD equations in such a way that these quantities may be controlled by changing the parameters of the thermostat. A similar approach coming from the viewing of smoothed particle hydrodynamics was taken in Ref. [51].

We consider a more general version of DPD and study the terms, which are not used in the standard DPD approach, i. e. the damping of the perpendicular components of the relative particle velocities.

The motivation for this is that the perpendicular components should be the slow degrees of freedom, which stay untouched in the case of standard DPD thermostat, while the fast degrees of freedoms are the relative velocities on the interatomic axis, which are influenced by the standard DPD thermostat.

Therefore, we generalize eq. 2.55 and eq. 2.56 as

$$\vec{F}_{ij}^{\mathrm{D}} = -\xi_{ij} w^{\mathrm{D}}(r_{ij}) \mathbb{P}_{ij}(\vec{r}_{ij}) \cdot \vec{v}_{ij} \tag{2.68}$$

and

$$\vec{F}_{ij}^{\mathrm{R}} = \sigma_{ij} w^{\mathrm{R}}(r_{ij}) \mathbb{P}_{ij}(\vec{r}_{ij}) \cdot \vec{\eta}_{ij} \ , \tag{2.69}$$

where ξ and σ are the friction constant and the noise strength of the generalized thermostat, respectively. $\mathbb{P}_{ij}(\vec{r}_{ij})$ is a projection operator

$$\mathbb{P} = \mathbb{P}^{\mathrm{T}} = \mathbb{P}^2, \tag{2.70}$$

which is symmetric in the particle indices ($\mathbb{P}_{ij} = \mathbb{P}_{ji}$). The scalar noise (see page 23) is replaced by a noise vector $\vec{\eta}_{ij}$

$$\langle \vec{\eta}_{ij}(t) \otimes \vec{\eta}_{kl}(t') \rangle = 2\mathbb{1}(\delta_{ik}\delta_{jl} - \delta_{il}\delta_{jk})\delta(t - t') \ , \tag{2.71}$$

which is antisymmetric in the particle indices ($\vec{\eta}_{ij} = -\vec{\eta}_{ji}$) due to the symmetry of the projection operator and the antisymmetry of the pair force (Newton's third law).

The corresponding Fokker-Planck operator, $\mathcal{L}_{\mathrm{DPD}}$, is a sum of two parts:

$$\begin{aligned} \mathrm{i}\mathcal{L}_{\mathrm{DPD}} \ = \ & \sum_{ij} \xi_{ij} w^{\mathrm{D}}(r_{ij}(t)) \frac{\partial}{\partial \vec{p}_i} \cdot \mathbb{P}_{ij}(\vec{r}_{ij}(t)) \cdot \vec{v}_{ij}(t) \\ & + \sum_{ij} [\sigma_{ij} w^{\mathrm{R}}(r_{ij}(t))]^2 \frac{\partial}{\partial \vec{p}_i} \cdot \mathbb{P}_{ij}((\vec{r}_{ij}(t)) \cdot \left[\frac{\partial}{\partial \vec{p}_i} - \frac{\partial}{\partial \vec{p}_j} \right] \end{aligned} \tag{2.72}$$

The equilibrium condition $\mathcal{L}_{\mathrm{DPD}} e^{-\mathcal{H}/k_B T} = 0$ then yields the fluctuation-dissipation theorem in the same form as given by eq. 2.65 and eq. 2.66.

For the case of the projector chosen along the interatomic axis between particle i and j

$$\mathbb{P}_{ij}(\vec{r}_{ij}) = \hat{r}_{ij} \otimes \hat{r}_{ij}, \tag{2.73}$$

we retain the standard DPD thermostat (see sec. 2.2.5) including the same Fokker-Planck operator and the same noise properties.

Alternatively, one can project on the plane perpendicular to the interatomic axis

$$\mathbb{P}_{ij}(\vec{r}_{ij}) = \mathbb{1} - \hat{r}_{ij} \otimes \hat{r}_{ij} \ . \tag{2.74}$$

The space defined by this projector is orthogonal to the one of the standard DPD and defines an extension of the DPD thermostat named the *Transverse DPD thermostat*. For illustration of the decomposition of relative velocity see fig. 2.2 on page 23. Note that due to this orthogonality, the new thermostat can be used in combination with the standard one. This enables us to control at the same time two friction constants, $\xi^{\parallel}$ and $\xi^{\perp}$, for the standard and the Transverse DPD thermostat, respectively. Galilei invariance and hydrodynamic conservation laws remain valid by construction.

It is important to note that due to the fact that the Transverse DPD thermostat forces between molecules are not central, the total angular momentum of the system is not a conserved quantity anymore. Nevertheless, these perturbations happen only on a local scale and are isotropically in space, and therefore average out [51]. In the meanwhile an extension to Transverse DPD thermostat was published, which repairs this violation of angular momentum by introducing a four body-term [52].

The viscosity is very sensitive to the damping perpendicular to the interatomic axis. This damping mimics the shear of those degrees of freedom that were integrated out in the coarse-graining procedure (see sec. 2.3). In a system with two particles, the stochastic forces of the Transverse DPD thermostat act in the same direction as the shear forces. The mean force acting on a particle in a sheared system with more than two particles is hence a sum of two contributions: a force coming from the Transverse DPD thermostat and another one originating from the shearing of the probe. Therefore, the shear viscosity in a simulation with the Transverse DPD thermostat is always higher than with the standard one.

Viewing this from the Green-Kubo picture [53, 54], which relates the viscosity to the correlation functions of the stress tensor, this additional viscosity arises from the projected stress-stress autocorrelation function, which can be derived by the Mori-Zwanzig formalism [55, 56, 57]. The full derivation is cumbersome and still under discussion [58].

After discussing how the DPD thermostat is technically implemented in the open-source package ESPResSo in the next section, we will demonstrate its ability to control transport properties by the molecular dynamics simulation of a simple Lennard-Jones liquid in sec. 2.2.8 and of atomistic TIP3P water in sec. 3.4.

2.2.7 Implementation in ESPResSo

ESPResSo [59] is a software for soft matter simulations. A particular advantage is its expandability and modularity that simplifies development and testing of new methods.

One can view the DPD thermostat as a usual pair interaction. A general difficulty is that one has to ensure that random numbers are drawn in the correct way, so that Newton's third law, $\vec{F}_{ij} = -\vec{F}_{ji}$, is satisfied, e. g. both particles have exactly the same absolute value of the force. In ESPResSo this problem is avoided by calculating the DPD forces for pairs of particles only [46]. In the case of a parallelized simulation, where the system is divided in domains, which are distributed among different processes, the pair forces across the process boundaries are only calculated once and are then communicated to ensure that the same random number is used for both particles.

For convenience, in ESPResSo the friction constant is scaled with the reduced mass of the particles

$$\xi_{ij} \to \frac{2m_i m_j}{m_i + m_j} \hat{\xi}_{ij} \ . \tag{2.75}$$

This draws a parallel to the common implementation of stochastic dynamics, where the friction constant also has the unit 1/time. All derivations above are still valid due to the fact that the fluctuation-dissipation theorem (see eq. 2.66) holds for all values of ξ_{ij}.

The DPD feature of ESPResSo can be enabled by adding:

```
#define DPD
```

to the ESPResSo configuration file, e. g. `myconfig.h`. The following `TCL` command will then switch on the thermostat in the simulation:

```
thermostat dpd $temp $friction $r_cut \
                     $friction2 $r_cut2 $wf $wf2
```

where `temp` determines the temperature, `friction`, `r_cut` gives the friction constants $\xi^{\parallel}$ and the cut-offs of the DPD thermostat, while `friction2` and `r_cut2` gives the friction constants $\xi^{\perp}$ and the cut-offs of the transverse DPD thermostat. `wf` and `wf2` is an integer n defining the weighting function according to eq. 2.67.

2.2.8 Numerical Validation on Lennard-Jones Particles

As simulation of particles interacting via Lennard-Jones interaction is one of the simplest liquids one can imagine, we will validate the ability of the transverse DPD thermostat by changing the diffusion constant and the viscosity of this liquid.

All simulations of the Lennard-Jones (LJ) liquid are performed using the ESPResSo package [59].

We used the repulsive Weeks-Chandler-Andersen (WCA) potential

$$U_{\mathrm{LJ}}(r) = 4\epsilon \left(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^{6}}{r^{6}} + \frac{1}{4} \right) \tag{2.76}$$

with the cut-off at $r_{\mathrm{cut}} = 2^{1/6}\sigma$, σ and ϵ being the standard LJ parameters of length and energy.

A step function is chosen as the weight function (see eq. 2.67) for both thermostats

$$w^{\mathrm{D}}(r) = \begin{cases} 1, & r < r_{\mathrm{c}} \\ 0, & r \geq r_{\mathrm{c}} \end{cases} \quad . \tag{2.77}$$

The simulations were carried out with a system consisting of $N_{\mathrm{part}} =$ 1000, 2000 and 4000 LJ particles at a temperature $\mathrm{k_B}T = 1.2\epsilon$ and density $\rho = N_{\mathrm{part}}/V = 1/(1.05\sigma)^3$ in a cubic box with periodic boundary conditions due to the fact that we have reference data using only a different thermostat for this simulation setup (T,ρ) [60].

First, we checked the dependency of pressure and temperature on the strength of the friction in an equilibrium simulation. We set the reference temperature to $1.2\epsilon/k_B$ and measured the instantaneous temperature, defined as

$$T = \frac{2E_{\mathrm{kin}}}{3N_{\mathrm{part}}}, \tag{2.78}$$

where E_{kin} and N_{part} are the kinetic energy and the number of particles of the system, respectively. The relative deviation between the measured and the reference temperature was smaller than 1.2% for all strengths of friction and all combinations of thermostats. The mean pressure at that temperature turned out to be $(9.8 \pm 0.2)\,\epsilon/\sigma^3$, which is in perfect agreement with the results of previous studies [60].

Next, we studied the dependency of the liquid transport properties, i. e. diffusion constant and shear viscosity, on the friction constants, $\xi^{\parallel}$ and $\xi^{\perp}$, for both, the standard and the transverse DPD thermostats.

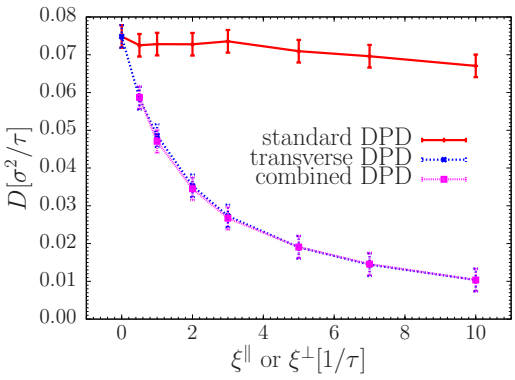

Figure 2.3: **Diffusive dynamics of a LJ system**: Diffusion constant (4000 LJ particles) as a function of the friction measured in equilibrium for different thermostats. In the case of the combined DPD thermostat only the strength of the friction parameter $\xi^{\perp}$ was varied while the friction for the standard one was held constant at the value $\xi^{\parallel} = 1.0$. The errors were obtained by averaging over several runs.

2.2.8.1 Diffusion Constant

The diffusion constant was computed from the particle displacements using the Einstein relation

$$D = \lim_{t\to\infty} \frac{< (\vec{r}(t) - \vec{r}(0))^2 >}{6t}. \tag{2.79}$$

Only a small influence on this constant from the standard DPD thermostat [46] is expected, but a considerable one from the new transverse DPD thermostat. In fig. 2.3 one sees that previous results for the standard DPD thermostat could be confirmed, the diffusion constant stayed nearly unchanged. The value of the diffusion constant approaches the Newtonian dynamics value ($D = 0.08\sigma^2/\tau$) for vanishing friction of the Transverse DPD thermostat.

The diffusion constant D is very sensitive to the friction $\xi^{\perp}$ for the Transverse DPD thermostat. By changing $\xi^{\perp}$ it is therefore possible to adjust the diffusion constant.

2.2.8.2 Shear Viscosity

The shear viscosity was measured in a non-equilibrium molecular dynamics (NEMD) simulation by shearing the system with a constant shear rate in the y-direction [46]

$$\dot{\gamma} = \frac{\partial u_x}{\partial y} \ . \tag{2.80}$$

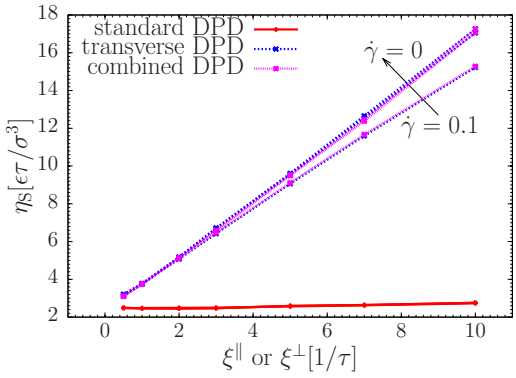

Figure 2.4: **Shear viscosity of a LJ system**: Shear viscosity measured with the NEMD algorithm for different thermostats (4000 LJ particles) and different shear rates (0.1 and 0.01 shown), which are then extrapolated to vanishing shear rates. The errors are obtained by Jackknife analysis [61].

The viscosity can then be determined according to

$$\eta_S = \frac{F}{\dot{\gamma} L^2} \ , \tag{2.81}$$

where F is the mean force (momentum transfer per time unit) and L^2 is the area of the slab. In a NEMD simulation the equilibrium distribution is not exactly the canonical distribution, but can be approximated in first order by one, so that the DPD thermostat, which generates a canonical ensemble, can still be used.

In fig. 2.4 one sees that the apparent shear viscosity, η_S, measured in a NEMD simulation approaches the zero-shear viscosity with decreasing shear rate. We found nearly no dependency on the strength of the friction $\xi^{\parallel}$ for the standard DPD thermostat, (see also Ref. [46]). In contrast, the friction $\xi^{\perp}$ for the transverse DPD thermostat gives a very sensitive switch of controlling the viscosity (see fig. 2.4). In the case of the combination of both thermostats the shear viscosity is mostly controlled by the Transverse DPD thermostat. In the limit of a vanishing shear rate a value of $2.45 \pm 0.07 \epsilon\tau/\sigma^3$ was extrapolated, which matches former results [60]. This extrapolation has to be done in order to obtain the equilibrium viscosity of the LJ system. Additionally, we also checked the correlation functions of the pressure tensor in an equilibrium simulation to show that these are in accordance with the expected Green-Kubo picture: all (non-auto) off-diagonal - (off-)diagonal elements are uncorrelated (60 of 81 possible elements are zero [2]).

[2] All(9×9) minus autocorrelation terms(9) minus diagonal-diagonal(3×2) terms minus

For higher values of $\xi^{\perp}$ we ended up measuring the "viscosity of the thermostat". In this regime also the apparent viscosity becomes increasingly dependent on the shear rate. This is due to the fact that the dynamics is dominated by the thermostat forces, which are linear in $\xi^{\perp}$ and the response to the shear forces is not linear anymore. Hence the linear approximation in eq. 2.81 breaks down and gives an inaccurate value for the viscosity.

2.2.8.3 Summary

Above, it was shown that the thermostat is working for a reference liquid. One can assume that this will work for any pair-wise interaction, but here we will only show that for a coarse-grained water model (see sec. 3.4). To conclude, we have introduced an extension of the DPD thermostat, which on the one hand inherits the good hydrodynamic properties, and on the other hand allows to adjust the dynamics.

2.2.9 Thermostats for Molecules

When describing a small molecule in two different resolutions, e. g. atomistic and coarse-grained resolution, the question is how this geometrical transformation maps down to the thermostat?

In this section we will discuss the special case of center of mass mapping, but more general mapping schemes can be found in sec. 2.3.1. The force acting on a center-of-mass of the αth molecule is just:

$$\vec{F}_\alpha = \sum_{i\in\alpha} \vec{f}_i \ , \tag{2.82}$$

where i labels all the atoms in the molecule. For the damping force of stochastic dynamics (see sec. 2.2.4, eq. 2.38) this means:

$$\vec{F}^{\mathrm{D}}_\alpha = \sum_{i\in\alpha} \vec{f}^{\mathrm{D}}_i = -\sum_{i\in\alpha} \xi_i \vec{v}_i = -\xi \sum_{i\in\alpha} m_i \vec{v}_i = -\xi M_\alpha \vec{V}_\alpha \tag{2.83}$$

So the structure of the damping force is retained if $\xi_i = \xi m_i$, which is a common convention in standard molecular dynamics to ensure that the heavier particles experience higher damping. The summed up random force is:

$$\vec{F}^{\mathrm{R}}_\alpha = \sum_{i\in\alpha} \vec{f}^{\mathrm{R}}_i = \sum_{i\in\alpha} \sigma_i \vec{\eta}_i = \sqrt{\mathrm{k_B} T \xi M_\alpha} \sum_{i\in\alpha} \sqrt{\frac{m_i}{M_\alpha}} \vec{\eta}_i = \hat{\sigma}_\alpha \hat{\vec{\eta}}_\alpha \ , \tag{2.84}$$

symmetric terms(6×1)

where we have defined $\hat{\sigma}_\alpha = \sqrt{\mathrm{k_B}T\xi M_\alpha}$. The only thing left to do is to prove that $\hat{\vec{\eta}}_\alpha$ has the right noise properties (see page 21). Homogeneity is trivial due to the fact that $\vec{\eta}_i$ is homogeneous:

$$< \hat{\vec{\eta}}_\alpha >= \sum_{i\in\alpha} \sqrt{\frac{m_i}{M_\alpha}} < \vec{\eta}_i >= \sum_{i\in\alpha} \sqrt{\frac{m_i}{M_\alpha}} 0 = 0 \tag{2.85}$$

To prove the independence in time and space one calculates:

$$\begin{aligned} < \hat{\vec{\eta}}_\alpha \otimes \hat{\vec{\eta}}_\beta > &= \sum_{i\in\alpha}\sum_{j\in\beta} \frac{\sqrt{m_j m_i}}{M_\alpha} < \vec{\eta}_i(t) \otimes \vec{\eta}_j(t') > \\ &= \sum_{i\in\alpha}\sum_{j\in\beta} \frac{m_i}{M_\alpha} 2\delta_{ij}\delta(t-t')\mathbb{1} \\ &= 2\delta_{\alpha\beta}\delta(t-t')\mathbb{1}\sum_{i\in\alpha}\frac{m_i}{M_\alpha} = 2\delta_{\alpha\beta}\delta(t-t')\mathbb{1} \end{aligned} \tag{2.86}$$

This shows that all noise properties are retained. The thermostat acting on the center of mass sites also obeys the fluctuation-dissipation theorem, if the friction constants of the atoms are scaled with their masses. This basically means that the sum of Langevin thermostats is again a Langevin thermostat, as long as the friction constants are chosen correctly.

This can be understood due to the fact that the momentum autocorrelation function decays with a correlation rate proportional to ξ/m and hence the choice, $\xi_i = \xi m_i$, makes the autocorrelation functions of all atoms decay with the same correlation time.

2.3 Coarse-Graining

After discussing the molecular dynamics method and thermostats, the focus will now be shifted to different coarse-graining methods, which have been used to obtain force fields for coarser representations of the studied systems. Later these coarse-grained force fields will be used in the adaptive resolution scheme to couple these two resolutions. But let us first take a look at the general ideas behind different methods to understand their advantages and disadvantages.

2.3.1 Mappings

The mapping scheme relates two representations of a system. The scheme is usually based on intuition, which comes from structural properties on which

the coarse-graining is based or from knowledge about the internal chemical structure.

Here the focus will lie on coarse-graining techniques that link two *particle-based* descriptions with a different number of degrees of freedom. The system with the larger number of degrees of freedom is denoted as the *reference* system. The system with the reduced number of the degrees of freedom is referred to as the *coarse-grained* system. An example is an all-atom (reference) and a united atom (coarse-grained) molecular representation, where the number of degrees of freedom is reduced by embedding hydrogens into heavier atoms. Another example, which is analyzed in detail here, is an all-atom (three sites) and a single site model of water. Many examples of such models can be readily found in the literature [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37].

We assume that the following pre-requisites should be satisfied:

- Both the reference and the coarse-grained descriptions are represented by a set of point sites, $\vec{r} = \{\vec{r}_i\}$, $i = 1, 2, \ldots, n$ in case of the reference system, and $\vec{R} = \{\vec{R}_j\}$, $j = 1, 2, \ldots, N$ in case of the coarse-grained system. The coordinates $\{\vec{R}_j\}$, which are obtained from an atomistic trajectory shall not be confused with the coordinates of a trajectory obtained from coarse-grained simulations.

- A mapping scheme, i. e. a relation between $\vec{r}$ and $\vec{R}$, can be expressed as $\vec{R} = \hat{M}\vec{r}$, where $\hat{M}$ is an $n \times N$ matrix, which is a special case of a linear relation between $\vec{r}$ and $\vec{R}$. Non-linear mapping can be useful in some special case, but will not be considered here. The mapping matrix $\hat{M}$ is a block-diagonal matrix and to construct it, it is enough to specify building blocks for each molecule type, e. g. for polymers it is enough to specify $\hat{M}$ for one repeat unit only.

- For the reference system, the coordinates and forces of a trajectory that samples a canonical ensemble or at least the part of it, in which one is interested to reproduce it on a coarse-grained level, are required.

The prime task of systematic coarse-graining is to devise a process for the coarse-grained system. In many cases this process can be simply described by its potential energy function $U(\vec{R})$, which we will assume here. To do this, one can use several coarse-graining approaches.

2.3.2 Coarse-Graining Approaches

From the point of implementation, these approaches can be divided in *iterative* and *non-iterative* methods. *Boltzmann inversion* is a typical example of

a non-iterative method [26]. In this method, which is exact for independent degrees of freedom, coarse-grained interaction potentials are calculated by inverting the distribution functions of the coarse-grained system. Another example of a non-iterative method is *force matching*, where the coarse-grained potential is chosen in such a way that it reproduces the forces on the coarse-grained beads [62, 30]. Configurational sampling [63], which matches the potential of mean force, also belongs to this category. Boltzmann inversion and force matching only require a trajectory of a reference system. Often this is a "special" trajectory, which is designed to decouple the degrees of freedom of interest, e. g. a single polymer chain in vacuum with appropriate exclusions [26]. Once this is known, coarse-grained potentials can be calculated for any mapping matrix $\hat{M}$.

Iterative methods refine the coarse-grained potential $U(\vec{R})$ by re-iterating coarse-grained simulations and calculating corrections to the potential on the basis of the reference and coarse-grained observables (e. g. radial distribution function or pressure). The simplest example is the iterative Boltzmann inversion method [64], which is an iterative analogue of the Boltzmann inversion method. A more sophisticated update function is the inverse Monte Carlo approach [65].

One can also classify systematic coarse-graining approaches by micro- and macroscopic observables used to derive the coarse-grained potential, such as structure-based [65, 66, 26], force-based [62, 67, 30], and potential-based approaches [68], where the name identifies the observable used for coarse-graining. Note that hybrids of these methods are also possible [28, 37].

In this thesis we will mainly focus on structure-based approaches, nevertheless we will also discuss some of the other methods in order to develop a foundation for the coarse-graining studies performed here.

2.3.3 Simple Boltzmann Inversion

Boltzmann inversion is the simplest method one can use to obtain coarse-grained potentials [26]. It is mostly used for *bonded* potentials, such as bonds, angles, and torsions. Boltzmann inversion is structure-based and only requires positions of atoms.

The idea of Boltzmann inversion stems from the fact that in a canonical ensemble independent degrees of freedom q obey the Boltzmann distribution, i. e.

$$P(q) = Z^{-1} \mathrm{e}^{-U(q)/\mathrm{k_B}T} \ , \tag{2.87}$$

where $Z = \int \mathrm{d}q \ \mathrm{e}^{-U(q)/\mathrm{k_B}T}$ is the partition function. Once $P(q)$ is known one can invert eq. 2.87 and obtain the coarse-grained potential, which in this case

is a potential of mean force

$$U(q) = -\mathrm{k_B}T \ln P(q) \ . \tag{2.88}$$

Note that the normalization factor Z is not important since it would only enter the coarse-grained potential $U(q)$ as an irrelevant additive constant.

In practice, $P(q)$ is computed from the trajectory of the reference system, which is sampled either by Monte Carlo, molecular dynamics, stochastic dynamics, or any other integrator that ensures a canonical distribution of states.

Boltzmann inversion is simple to implement, however one has to be careful with the rescaling of the probability P due to orientational entropy as well as computational issues. The probability rescaling can be explained on a particular example of coarse-graining of a single polymer chain by beads with bond, angle and torsion potentials. In this case the coarse-grained potential U depends on three variables: bond length r, angle θ and torsion angle φ.

Assuming, as before, a canonical distribution and the independence of the coarse-grained degrees of freedom, we can write

$$P\left(r, \theta, \varphi\right) = \mathrm{e}^{-U(r,\theta,\varphi)} = P_r(r) P_\theta(\theta) P_\varphi(\varphi) \ . \tag{2.89}$$

If we now compute the histograms for the bonds $H_r(r)$, angle $H_\theta(\theta)$, and torsion angle $H_\varphi(\varphi)$ we must rescale them in order to obtain the volume normalized distribution functions. Note that in all the discussion, as before, the irrelevant normalization prefactor Z has been ignored.

$$P_r(r) = \frac{H_r(r)}{4\pi r^2} \ , P_\theta(\theta) = \frac{H_\theta(\theta)}{\sin\theta} \ , P_\varphi(\varphi) = H_\varphi(\varphi) \ . \tag{2.90}$$

The coarse-grained potential can then be calculated by Boltzmann inversion of the distribution functions

$$U(r, \theta, \varphi) = U_r(r) + U_\theta(\theta) + U_\varphi(\varphi) \ , \tag{2.91}$$

with

$$U_q(q) = -\mathrm{k_B}T \ln P_q(q), \ q = r, \theta, \varphi \ . \tag{2.92}$$

On the technical side, the implementation of the Boltzmann inversion method requires *smoothing* of $U(q)$ to provide a continuous force. Splines can be used for this purpose. Poorly and unsampled regions, namely regions with high $U(q)$, shall be *extrapolated*. Since the contribution of these regions to the canonical density of states is small the exact shape of the extrapolation is less important.

Another crucial issue is the cross-correlation of the coarse-grained degrees of freedom. Independence of the coarse-grained degrees of freedom is the main assumption that allows factorization of the probability distribution, eq. 2.89, and the potential, eq. 2.92, hence, one has to carefully check whether this assumption holds in practice. This can be done by performing coarse-grained simulations and comparing cross-correlations for all pairs of degrees of freedom in atomistic and coarse-grained resolution, e. g. using a two-dimensional histogram, analogous to a Ramachandran plot known from protein simulations [69].

Checking the linear correlation coefficient would be another possibility, however it does not guarantee statistical independence of variables, e. g. the correlation of x with x^2 is 0 if x has a symmetric probability density $P(x) = P(-x)$. This case is often encountered in systems used for coarse-graining [70, 71]. The concept of a two-dimensional histogram is illustrated in sec. 3.3.3 for liquid propane and a single molecule of hexane in sec. 3.3.4.

2.3.4 Iterative Methods

Iterative Boltzmann inversion (IBI) is a natural extension of the Boltzmann inversion method. Since the goal of the coarse-grained model is to reproduce the distribution functions of the reference system as accurately as possible, one can also iteratively refine the coarse-grained potentials using a numerical scheme. Depending on the update function, this can be done either by using the iterative Boltzmann inversion [64] or the inverse Monte Carlo [65, 66] method. The iterative Boltzmann inversion method will be discussed first.

2.3.4.1 Iterative Boltzmann Inversion

In the iterative Boltzmann inversion, the coarse-grained potential is refined according to the following scheme:

$$U^{(n+1)} = U^{(n)} + \Delta U^{(n)} \ , \tag{2.93}$$

with

$$\Delta U^{(n)} = \mathrm{k_B}T \ln \frac{P^{(n)}}{P_{\mathrm{ref}}} = U^{\mathrm{ref}}_{\mathrm{PMF}} - U^{(n)}_{\mathrm{PMF}} \ . \tag{2.94}$$

One can easily see that convergence is reached as soon as the distribution function $P^{(n)}$ matches the reference distribution function P_{ref}, or, in other words, the potential of mean force, $U^{(n)}_{\mathrm{PMF}}$ converges to the reference potential of mean force.

Note that this is nothing else but a numerical scheme that allows one to match the coarse-grained and the reference distribution functions. It can be

seen as a first-order correction to the interaction potential with respect to a gas of non-interacting particles. Indeed, in an ideal gas, the probability of finding two particles at a distance r is

$$P^{(0)}(r) = 4\pi r^2 / V(r_{\max}) \ , \tag{2.95}$$

which is equivalent to the statement that the radial distribution function of an ideal gas is 1. Substituting $P^{(0)}(r)$ into eq. 2.94 we obtain the first iteration

$$U^{(1)}(r) = -k_{\mathrm{B}} T \ln(P_{\mathrm{ref}}/4\pi r^2) \ , \tag{2.96}$$

which is the potential of mean force, eq. 2.88.

IBI can be used to refine both bonded and non-bonded potentials. It is primarily used for simple fluids with the aim of reproducing the radial distribution function of the reference system in order to obtain non-bonded interactions [64]. It can have convergence problems for multicomponent systems, even if we never experienced them. The reason is simply that it does not account for cross-correlation correction terms, that is, the updates for P_{AA}, P_{AB}, and P_{BB} are not coupled (the subscript denotes a single component in a multicomponent system). For such systems, the inverse Monte Carlo method may work better. The scheme can be stabilized by multiplying the update function, $\Delta U^{(n)}$, by a factor $\eta \in [0..1]$.

On the implementation side, IBI has the same issues as the inverse Boltzmann method, i. e. smoothing and extrapolation of the potential must be implemented.

One should also note that according to the Henderson theorem [72, 73], which is a classical analogue of the Hohenberg-Kohn theorem [74], the pairwise coarse-grained potential $U(r)$ is unique up to an additive constant and exists [75, 76]. This theorem, in principle, states that all structure-based iterative methods must converge to the same coarse-grained potential, provided that their aim is to exactly reproduce pair correlation functions of the reference system. As we will see later, this is often not the case in practice, since small changes in the radial distribution function often lead to big changes in the pair potential, i. e. it is difficult to control systematic errors during the calculation of the potential update.

2.3.4.2 Pressure Correction

Another issue of coarse-graining is that coarse-grained models can not reproduce all the statistical or thermodynamic properties of the reference system. Pressure, compressibility (see tab. 3.3 in sec. 3.2.2), or viscosity [77] are often very different from those of the reference system, which is in general not a

problem. But in some cases, one needs to correct some of these. For example, the viscosity can be adjusted by tuning the parameters of the DPD thermostat (see sec. 3.4) and the pressure can be corrected iteratively by adding a linear term to the non-bonded potential

$$\Delta U^{\text{pressure}}(r) = A\left(1 - \frac{r}{r_{\text{cut}}}\right), \tag{2.97}$$

where A is either a constant, e. g. $-0.1\,\mathrm{k_B}T$ [64] or can be estimated in a different way. The virial expression of the pressure P_i is given by:

$$P_i V = N\mathrm{k_B}T - \frac{2}{3}\pi N\rho \int_0^{+\infty} \mathrm{d}r\; r^3 \frac{\mathrm{d}V_i(r)}{\mathrm{d}r} g_i(r) \tag{2.98}$$

where $V_i(r)$ and $g_i(r)$ are the potential and RDF of the i-th step the iteration, respectively. The corrected potential is ought to match the correct pressure P^{target}, that is

$$P^{\text{target}} V \approx N\mathrm{k_B}T - \frac{2}{3}\pi N\rho \int_0^{+\infty} \mathrm{d}r\; r^3 \frac{\mathrm{d}}{\mathrm{d}r}\left[V_i(r) + \Delta V_i(r)\right] g_i(r)\ . \tag{2.99}$$

The approximation appears due to the fact that $g_i(r)$ is the RDF of the uncorrected system. Therefore, the unknown A_i satisfies

$$-\left[\frac{2\pi N\rho}{3r_{cut}} \int_0^{r_{cut}} \mathrm{d}r\; r^3 g_i(r)\right] A_i \approx (P - P_{\text{target}})V\ . \tag{2.100}$$

Usually one can not reach the target pressure by one step of the pressure-correction due to stabilization issues. Therefore, the pressure-correction has to be applied in a different way depending on the coarse-grained system, e. g. every second step of iterative Boltzmann inversion or even after the inversion already converged.

It should however be mentioned at this point that the pressure-correction leads to a potential, where the isothermal compressibility deviates significantly from that of the reference system. For water we have shown that compressibility and pressure can not be corrected simultaneously (see sec. 3.2.2).

2.3.4.3 Inverse Monte Carlo

Inverse Monte Carlo (IMC) is another iterative procedure that refines the coarse-grained potentials until the coarse-grained model reproduces a set of reference distribution functions. It is very similar to IBI except that the update of the potential, ΔU, is calculated using rigorous thermodynamic arguments.

The name "inverse Monte Carlo" is somehow confusing and is due to the fact that the original algorithm was combined with Monte Carlo sampling of the phase space [65]. However, practically any sampling method can be used (e. g. molecular dynamics or stochastic dynamics) as long as it provides a canonical sampling of the phase space.

A detailed derivation of the IMC method can be found in Ref. [65]. Here, we briefly recapitulate a more compact version for non-bonded interactions, which is outlined in Ref. [73], emphasizing technical problems encountered during implementation and application of the method.

The idea of IMC is to express the potential update ΔU in a thermodynamically consistent way in terms of measurable statistical properties, e. g. the radial distribution function $g(r)$. Considering a single-component system as an example the Hamiltonian of the system can be written as

$$\mathcal{H} = \sum_{i,j} U(r_{ij}) \ , \tag{2.101}$$

where $U(r_{ij})$ is the pair potential and it is assumed that all interactions depend only on the distance r_{ij}, between particles i and j. Further it is assumed that this potential is short-ranged, i. e. $U(r_{ij}) = 0$ if $r_{ij} \geq r_{\text{cut}}$.

The next step is to tabulate the potential $U(r)$ on a grid of M points, $r_\alpha = \alpha \Delta r$, where $\alpha = 0, 1, \ldots, M$, and $\Delta r = r_{\text{cut}}/M$ is the grid spacing. Then the Hamiltonian, eq. 2.101, can be rewritten as

$$\mathcal{H} = \sum_{\alpha} U_\alpha S_\alpha \ , \tag{2.102}$$

where S_α is the number of particle pairs with interparticle distances $r_{ij} = r_\alpha$, which correspond to the tabulated value of the potential U_α.

On one hand, the average value of S_α is related to the radial distribution function $g(r)$

$$\langle S_\alpha \rangle = \frac{N(N-1)}{2} \frac{4\pi r_\alpha^2 \Delta r}{V} g(r_\alpha) \ , \tag{2.103}$$

where N is the number of atoms in the system, hence $\frac{1}{2}N(N-1)$ is then the number of all pairs and $\Delta r = r_{\text{cut}}/M$ is the grid spacing, V is the total volume of the system.

On the other hand, $\langle S_\alpha \rangle$ is a function of the potential U_α and hence can be expanded in a Taylor series with respect to changes in the potential, ΔU_γ

$$\langle S_\alpha \rangle_0 + \Delta \langle S_\alpha \rangle = \langle S_\alpha \rangle_0 + \sum_{\gamma} \frac{\partial \langle S_\alpha \rangle}{\partial U_\gamma} \Delta U_\gamma + \mathcal{O}(\Delta U^2) \ . \tag{2.104}$$

The derivatives $\partial \langle S_\alpha \rangle / \partial U_\gamma$ can be obtained from the correlation of $\langle S_\alpha \rangle$

$$\begin{aligned} A_{\alpha\gamma} &= \frac{\partial \langle S_\alpha \rangle}{\partial U_\gamma} \\ &= \frac{\partial}{\partial U_\gamma} \frac{\int \mathrm{d}q \; S_\alpha(q) \mathrm{e}^{-\beta \sum_\lambda U_\lambda S_\lambda(q)}}{\int \mathrm{d}q \; \mathrm{e}^{-\beta \sum_\lambda U_\lambda S_\lambda(q)}} \\ &= \beta \left(\langle S_\alpha \rangle \langle S_\gamma \rangle - \langle S_\alpha S_\gamma \rangle \right) \; . \end{aligned} \tag{2.105}$$

Equations 2.103, 2.104, and 2.105 allow calculation of the correction for the potential by solving a set of linear equations

$$\langle S_\alpha \rangle - S_\alpha^{\mathrm{ref}} = A_{\alpha\gamma} \Delta U_\gamma \; , \tag{2.106}$$

where S_α^{ref} is given by the target radial distribution function. The procedure is then repeated until convergence is reached.

A clear advantage of the IMC compared to the IBI method is that the update of the potential is rigorously derived using statistical mechanics and hence the iterative procedure shall converge faster with the IMC update than with the empirical IBI update.

Another advantage is that, in the case of multicomponent mixtures, IMC takes into account correlations of observables, that is updates for U_{AA}, U_{AB}, and U_{BB} are interdependent, where A and B denote different particle types. In comparison, in the IBI method these updates are always independent due to the fact that they only depend on the individual distributions. This can lead to convergence problems in IBI for multicomponent systems.

The advantages come, of course, at a computational cost. As it is clear from eq. 2.105, one has to calculate cross-correlations of S_α. This requires much longer runs to get statistics that is good enough to calculate the potential update to a similar accuracy as IBI. The accuracies of the update functions of IMC and IBI methods are compared in sec. 3.2 for the case of a coarse-grained model of water.

Another issue of the IMC method is the stability of the scheme. Several factors can influence it: The first, and rather technical, point is that the radial distribution function of the reference system $g^{\mathrm{ref}}(r_\alpha)$ has to be calculated using exactly the same convention for the grid as S_α (e. g. the function value should be assigned to the middle of the interval), otherwise the scheme becomes unstable. Second, inversion of $A_{\alpha\gamma}$ requires that it shall be well defined. This means that one has to remove or handle with extra care the regions, which are not sampled, such as those at the beginning of the radial distribution function.

The convergence can be significantly improved if a smoothing of the potential update ΔU is used. Note that it is better to do smoothing of the update function, not the potential itself, since the latter has more features, which can be lost due to too aggressive smoothing. The convergence can also be improved by introducing a multiplicative pre-factor n for the update function in the case of iterative Boltzmann inversion or using a regularization procedure by adding thermodynamic constraints [78].

Finally, we have also noticed that the systematic error in $\langle S_\alpha S_\gamma \rangle$ is always higher in the vicinity of the cut-off, which leads to a shift in the tail of the interaction potential and, as a result, to a large offset of pressure. The cross-correlation term $\langle S_\alpha S_\gamma \rangle$ is also very sensitive to the box size, and special care must be taken in order to converge the results with respect to system size. Finite size effects are discussed in detail in sec. 3.3.2, where liquid methanol was coarse-grained.

2.3.5 Force Matching

Force matching (FM) is another approach to evaluate coarse-grained potentials [62, 30, 79]. In contrast to the structure-based approaches, its aim is not to reproduce various distribution functions, but instead tries to match *forces* on coarse-grained beads as closely as possible. A detailed formal statistical mechanical framework of force matching applied to a liquid state, or a multi-scale coarse-graining method, is given in Ref. [80, 81, 82, 83]. FM is a non-iterative method and hence is less computationally demanding.

The method works as follows: We first assume that the coarse-grained force-field (and therefore the forces) depends on M parameters $g_1, \ldots, g_M$. These parameters can be pre-factors of analytical functions, tabulated values of the interaction potentials, or coefficients of splines used to describe these potentials.

In order to determine these parameters, the reference forces on coarse-grained beads are calculated by properly re-weighting the forces on the atoms

$$\vec{f}_i^{\,\mathrm{ref}} = M_i \sum_\alpha \frac{w_\alpha \vec{f}_\alpha}{m_\alpha} \ , \tag{2.107}$$

where $M_i = \left(\sum_\alpha w_\alpha^2/m_\alpha\right)^{-1}$ is the mass of the bead i, index α numbers all atoms belonging to this bead, $\vec{f}_\alpha$ is the force on the atom α, m_α is its mass and w_α are mapping coefficients used to obtain the position of the coarse-grained bead, $\vec{R}_i = \sum_\alpha w_\alpha \vec{r}_\alpha$. If the center of mass is used in the mapping, eq. 2.107 simplifies to the sum of the forces.

By calculating the reference forces for L snapshots we can write down $N \times L$ equations

$$\vec{f}_{il}^{\,\mathrm{cg}}(g_1, \ldots, g_M) = \vec{f}_{il}^{\,\mathrm{ref}},\ i = 1, \ldots, N, l = 1, \ldots, L\ . \tag{2.108}$$

Here $\vec{f}_{il}^{\,\mathrm{ref}}$ is the force on the bead i, $\vec{f}_{il}^{\,\mathrm{cg}}$ is the coarse-grained representation of this force. Index l enumerates snapshots picked for coarse-graining. By running the simulations long enough one can always ensure that $M < N \times L$. In this case the set of equations 2.108 is overdetermined and can be solved in a least-squares sense.

Though the underlying idea of FM is very simple, implementation wise it is the most complicated method. Going back to the set of equations 2.108 one can see that $\vec{f}_{il}^{\,\mathrm{cg}}$ is, in principle, a non-linear function of its parameters $\{g_i\}$. Therefore, it is useful to represent the coarse-grained force-field in such a way that equations 2.108 become linear functions of $\{g_i\}$. This can be done using splines to describe the functional form of the forces [30].

An adequate sampling of the system requires a large number of snapshots L. Hence, the applicability of the method is often constrained by the amount of available memory. To remedy the situation, one can split the trajectory into blocks, find the coarse-grained potential for each block and then perform averaging over the blocks.

2.3.6 Implementation (VOTCA)

With a rich zoo of methods plus their combinations at hand, it is natural to ask for an optimal method for a specific class of systems. On a more fundamental level one might question whether the different methods provide the same coarse-grained potential and whether it is possible to formulate a set of (even empirical) rules favoring one method with respect to another. It is obvious this is a difficult task to be treated analytically, especially for realistic systems. Often it is not clear if a certain assumption, which was made by a particular coarse-graining technique, is still fulfilled, hence one needs to apply all available methods to the specific system and then compare and quantify the degree of discrepancy between the coarse-grained and reference descriptions.

This is, however, cumbersome due to the absence of a single package where all these methods are implemented with the same accuracy and the same level of technical detail. That is why we have started writing such a coarse-graining package: the *Versatile Object-oriented Toolkit for Coarse-graining Applications* (VOTCA) [5]. It is available as open source [84] under Apache license [85]. The package was designed in collaboration with V.

Rühle [86] and A. Lukyanov. The focus of the author of this thesis was placed on the design of the iterative framework described in sec. 2.3.6.2.

Above we have described the basic ideas behind each method, paying special attention to the technical issues one has to overcome when implementing them. Later in the results chapter (see sec. 3) we then illustrate the applicability of these methods by coarse-graining different systems with the help of VOTCA. But first we will discuss the different parts of VOTCA in more detailed fashion.

2.3.6.1 Coarse-Graining Engine

In a nutshell, coarse-graining is nothing more than an analysis of the canonical ensemble of a reference (high resolution) system. In addition to this analysis, iterative methods require canonical sampling of the coarse-grained system, which can be done using either molecular dynamics (MD), stochastic dynamics (SD), or Monte Carlo (MC) techniques. The latter are implemented in many standard simulation packages. Rather than implementing its own MD/SD/MC modules, the VOTCA toolkit allows swift and flexible integration of existing programs in such a way that sampling is performed in the program of choice. Only the analysis needed for systematic coarse-graining is done using the package tools.

First of all the mapping engine converts the trajectory of the reference system to the trajectory of the coarse-grained system, which then can be stored or analyzed. The analysis tools include calculations of probability distributions of bonded and non-bonded interactions, correlation and auto-correlation functions, and updates for the coarse-grained pair potential. In that sense also force matching is a kind of advanced analysis. Analysis tools of the MD package can also be integrated into the coarse-graining work-flow, if needed.

The package offers a flexible framework for reading, manipulating and analyzing of MD/SD/MC topologies and trajectories. Its core is modular and new file formats can be integrated without changing the existing code. An interface for GroMaCS [87] and ESPResSo [59] topologies and trajectories is provided. An interface to ESPResSo++ [88] is planned.

The coarse-graining procedure itself is controlled by several Extensible Markup Language (XML) input files, which contain mapping and other options required for the work-flow control. In the mapping it is possible to select groups of interactions that will be used for coarse-graining or analysis.

2.3.6.2 Iterative Framework

For iterative processes, like iterative Boltzmann inversion and iterative inverse Monte Carlo, we have implemented a framework, which allows easy controlling of work-flow. The work-flow chart is shown in fig. 2.5. The work-flow is implemented as a shell script, which can, in principle, be run on all available operating systems and provides the flexibility needed to call external (or overload existing) scripts and programs written in other programming languages. An interface to read values from the steering XML files in C++, Perl and shell is also provided.

During the global initialization the initial guess for the coarse-grained potential is calculated from the reference radial distribution function or converted from a given potential guess to the internal format. The actual iterative step starts with an iteration initialization. It searches for possible checkpoints, converts and copies files from the previous step and the base directory. Then the simulation run is prepared by converting potentials to the format required by the external sampling program and the actual sampling is performed. Currently, an interface with GroMaCS [87] is implemented and extension to other packages is straightforward. After sampling the phase space, the potential update ΔU is calculated. Often the update requires post-processing, such as smoothing, interpolation, extrapolation or fitting to an analytical form. A simple pressure correction [64] can also be seen as a post-processing of ΔU, due to the fact that it only adds a linear inter-particle separation function. Finally, the new potential is determined and post-processed. If the iterative process continues, the next iterative step starts to initialize.

The iterative framework does all these step in an automated fashion, which allows to iterate as long as needed. An iteration with more than 10000 steps would barely be possible without a framework like this.

The iterative framework is flexible enough to implement tasks, which are only remotely related coarse-graining, e. g. determination of the thermodynamic force for AdResS (see sec. 2.4.6.10) or automated fitting of the viscosity (see sec. 3.4). These examples show the power of the framework and future extension, like simplex [89, 64] are in preparation.

2.3.6.3 Summary

VOTCA is an open source package under Apache license, which allows free distribution and open development, but also proprietary developments for industry are possible. The package provides an unified interface to the coarse-graining methods introduced above and enables the user to do coarse-graining

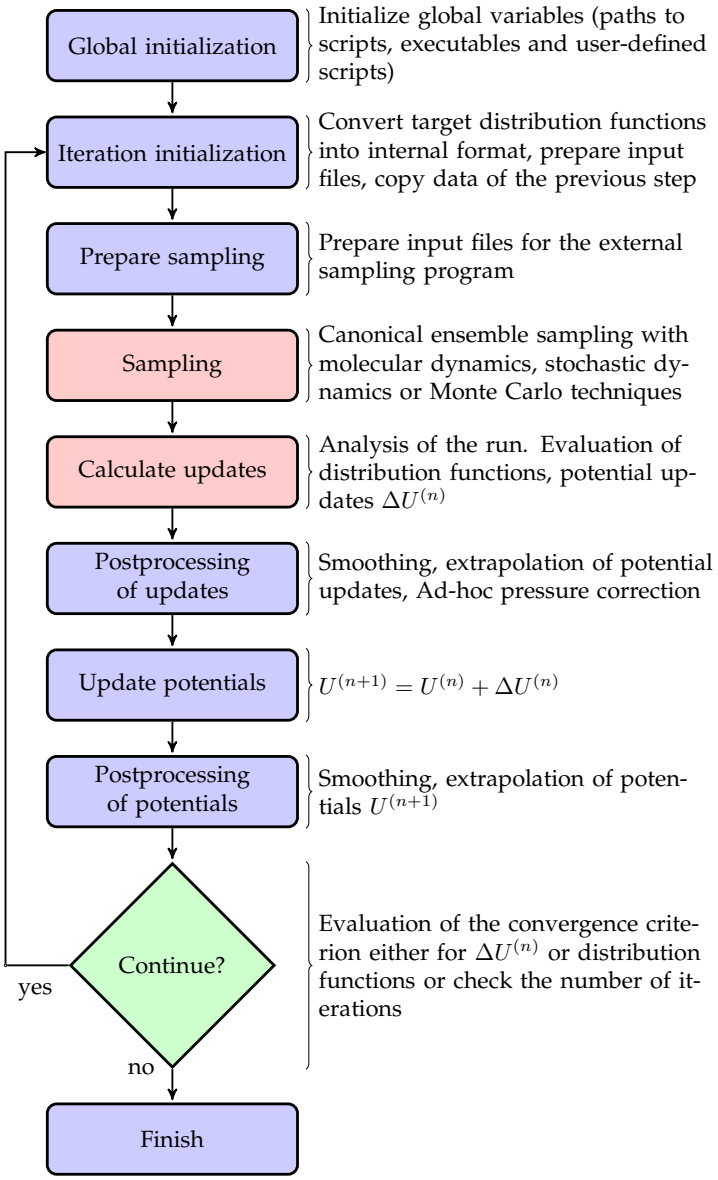

Figure 2.5: **Block-scheme of the iterative framework**: Block-scheme of the workflow control for the iterative methods. The most time-consuming parts are marked in red.

in an automated fashion.

2.3.7 Potential of Mean Force using Constraints

For some systems it is often impossible to measure the full distribution of a degree of freedom, which can be due to e. g. sampling problems and finite size. For this reason we will describe another method to determine the potential of mean force (and also the distribution).

Usually a potential of mean force is obtained by inverting the distribution (see sec. 2.3.3). However, this method is based on the fact that the full range of the degree of freedom q was sampled, which can be tedious. For example in the case that q is the distance r between two objects, the distribution of short distances can not be measured due to the energy barriers of the repulsion and distribution for long distances due to the finite size of the simulation box.

Nevertheless one can also measure the potential of mean force by constraining the distance and measuring the average constrained force $< \vec{f}_{\mathrm{c}} >$ to keep this distance [90]. The mean force is then the negative of this constrained force. So integration of this force

$$V(r_2) - V(r_1) = \int_{r_1}^{r_2} \mathrm{d}r \ < -\vec{f} > (r) = \int_{r_1}^{r_2} \mathrm{d}r \ < \vec{f}_{\mathrm{c}} > (r) \ , \qquad (2.109)$$

will give the free energy difference. However, the above equation misses an entropic part, which even pulls apart noninteracting particles. The entropic force is given by [91]:

$$-\frac{\mathrm{d}}{\mathrm{d}r}\left[-\mathrm{k_B}T \log\, 4\pi \left(\frac{r}{r_0}\right)^2\right] = \frac{2\mathrm{k_B}T}{r} \ , \qquad (2.110)$$

where $\mathrm{k_B} \log\, 4\pi(r/r_0)^2$ is just the entropy of the phase space volume $4\pi r^2$. This yields the potential of mean force:

$$V_{\mathrm{MF}}(r) = \int_{r_{\mathrm{cut}}}^{r} \mathrm{d}r' \ < \vec{f}_{\mathrm{c}} > (r') + \frac{2\mathrm{k_B}T}{r'} + C \ , \qquad (2.111)$$

where C is a constant that can be chosen in such a way that $V_{\mathrm{MF}}(r)$ is zero at the cut-off r_{cut}.

With the help of this method it is possible to measure the potential of mean force and radial distribution functions of parts of molecules or objects, e. g. ions, which appear in a very low concentration.

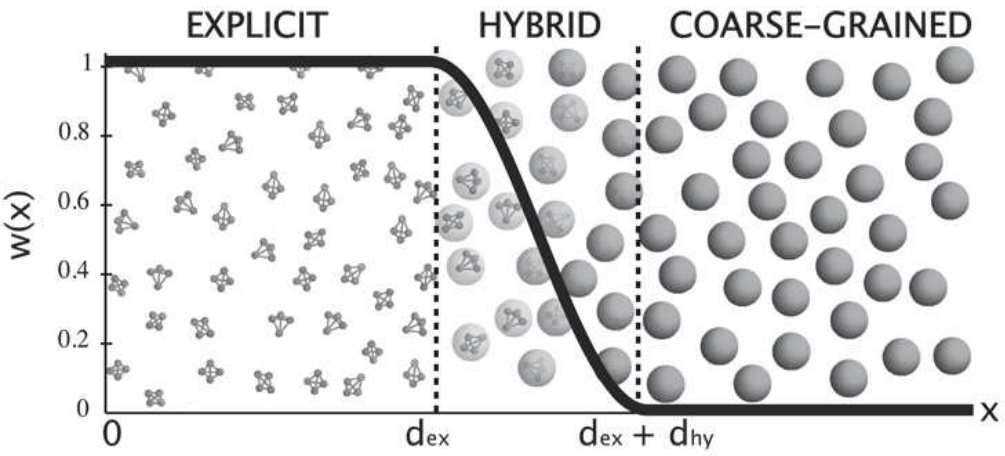

Figure 2.6: **Illustration of AdResS**: A schematic illustration of the AdResS method for a tetrahedral liquid. In this figure the $\cos^2$ weighting function is shown.

2.3.8 Summary

With the knowledge about the various coarse-graining methods and the implementation in the VOTCA package at hand, we can now focus on the AdResS scheme, which couples two resolutions based on the coarse-grained force field derived by these methods. However, we have to keep in mind, that the different methods of coarse-graining can depend on factors like chemical structure, system size, rare events due to small concentrations and possible correlations in time and space. Therefore, the derived potentials are only correct under certain assumptions and can depend on the sate point, where the development took place. Nevertheless, the savings in computer time and thus the extension of the time simulated can be enormous. This enhancement of *simulatable* time is also one of the major aims of the AdResS method, which is described in the next section.

2.4 Adaptive Resolution Scheme

The driving idea behind AdResS is to develop a scheme where the interchange between the atomistic and coarse level of description is achieved on the fly by changing the molecular degrees of freedom (DOFs). In order to develop this idea a test model for a molecule has been built.

In fig. 2.6 a pictorial representation of the tetrahedral molecule used and its corresponding spherical coarse-grained representation is given, derived in a way that it reproduces some chosen all-atom properties. The tetrahedral molecule consists of four atoms bond together by a spring-like potential

with a Lennard-Jones intermolecular potential (see explicit region of fig. 2.6); specific technical details of the model as well as of the coarse-grained procedure for the spherical representation are reported in the results section (sec. 4.1). The tetrahedral molecule can be seen as a simplified methane molecule, which has an additional center atom. As fig. 2.6 shows, the atomistic molecule, when passing to the coarse-grained region, slowly loses its vibrational and rotational DOFs. It passes through different stages of hybrid atomistic/coarse-grained representation and finally reduced its representation to a sphere whose DOFs are solely the translational ones of the center of mass with a proper excluded volume. A crucial point to keep in mind is that the different resolutions do not mean that the molecules are of different *physical* species, none of the resolutions should be preferred. The basic underlying physics should be the same in all regions and thus the process of exchange has to happen under the condition of thermodynamical and statistical equilibrium [3], which means thermal equilibrium $T^{\text{atom}} = T^{\text{cg}}$, and no net molecular flux $\rho^{\text{atom}} = \rho^{\text{cg}}$. Those conditions must be preserved by the numerical scheme and thus represent the conceptual basis of the method [92]. Next the effective dynamical coupling between the scales must be specified; this is reported in the following section.

2.4.1 Scale Coupling

Once the effective potential is derived on the basis of the reference all-atom system (see sec. 2.3) the atomistic and the coarse-grained scales are coupled via a position dependent interpolation formula on the atomistic and coarse-grained force [21, 93]:

$$\vec{F}_{\alpha\beta} = w(X_\alpha)w(X_\beta)\vec{F}^{\text{atom}}_{\alpha\beta} + [1 - w(X_\alpha)w(X_\beta)]\vec{F}^{\text{cg}}_{\alpha\beta} \ , \tag{2.112}$$

where α and β label two distinct molecules, $\vec{F}^{\text{atom}}_{\alpha\beta}$ is derived from the atomistic potential where each atom of molecule α interacts with each atom of molecule β, and $\vec{F}^{\text{cg}}_{\alpha\beta}$ is obtained from the effective (coarse-grained) pair potential between the centers of mass of the coarse-grained molecules. In the region where a smooth transition from one resolution to another takes place, a continuous monotonic *switching* function $w(x)$ is defined as can be seen in fig. 2.6 (where X_α, X_β are the x-coordinates of the centers of mass of the molecules α and β). A simple way to visualize the function $w(x)$ is the

[3] Pressure balance $P^{\text{atom}} = P^{\text{cg}}$ could be added as an additional condition, but is not necessary due to the fact that it is sufficient to have homogeneous pressure outside the transition region.

following: $w(x)$ is equal to one in the atomistic region and thus the switchable DOFs are fully counted, if $w(x)$ is zero in the coarse-grained region the switchable DOFs are turned off, between values of zero and one $w(x)$ provides a (continuous) hybrid representations of such DOFs (i. e. they count only in part).

In general, eq. 2.112, allows for a smooth transition from atomistic to coarse-grained trajectories without perturbing the evolution of the system in a significant way. More specific, the formula of eq. 2.112 works in such a way that when a molecule passes from the atomistic to the coarse-grained region, the molecular vibrations and rotations become less relevant until they vanish so that $w(x)$ smoothly *freezes* the dynamical evolution of those DOFs and their contributions to the interactions with the other molecules. Vice versa, when a molecule goes from the coarse-grained region to the atomistic one, $w(x)$ smoothly *reactivates* their dynamics and their contributions to the intermolecular interactions. This all happens in thermodynamical equilibrium (see next section) and the DOFs are in contact with a thermostat to satisfy this equilibrium and to avoid unphysical effects like artificial entropy production.

Basically one has the following cases: All molecules interacting with coarse-grained molecules interact as coarse-grained molecules independently of the region where they are. The coarse-grained molecule does not have any atomistic details, thus the other molecule can interact with this molecule only via the center of mass - center of mass interaction. Two atomistic molecules interact as atomistic ones, while for the other couplings, the interactions are governed by the $w(X_\alpha)w(X_\beta)$ combination. A very important point of eq. 2.112 is that, by construction, Newton's third Law is preserved. The diffusion of molecules between regions with different resolution must not be perturbed by the resolution change. Thus the conservation of the linear momentum dictated by Newton's third Law is crucial in adaptive resolution MD simulations.

2.4.2 Thermodynamical Equilibrium

From eq. 2.112 a potential can not be derived and thus a scheme based on eq. 2.112 would not have an energy to conserve. The natural subsequent question is, how to control the thermodynamic equilibrium. The conceptual problem for an adaptive scheme is that the free energy density can formally not be uniform since the number of DOFs varies in space. However, as the system should be uniform by construction and the underlying *physical* nature of the molecules should be equal everywhere, this is only an artifact of the interpolation formula used and not a real physical effect. This

non-uniformity leads to a non-physical preferential direction of the molecular flux. In fact, as numerical tests show, there is a preferential tendency of the atomistic molecules to migrate into the coarse-grained region and change the resolution in order to lower the free energy of the system. The free energy is an extensive quantity that is proportional to the number of DOFs. A simple qualitative way to picture this diode-like aspect is the following: When a molecule moves from an atomistic to a coarse-grained region it loses vibrational and rotational DOFs and thus in its interactions with the neighboring (coarse-grained) molecules it must accommodate only its excluded volume (i. e. find space). This becomes more complicated if a coarse-grained molecule moves into an atomistic region. In this case the molecule acquires rotational and vibrational DOFs and tries to enter into a region where other molecules are already locally in equilibrium. This means that in order to enter this region, the molecule should accommodate both rotational and vibrational DOFs according to the neighboring environment. Most likely the molecule would enter with vibrational and rotational motions, which do not fit the local environment and this would lead to a perturbation of the local equilibrium. Hence, for such a molecule the *way back* to the coarse-grained region is more convenient, and thus this free energy barrier works as a closed door (probabilistically) for the coarse-grained molecules and an open door for the atomistic ones so that a preferential molecular flux from the atomistic to the coarse-grained region is produced. In thermodynamic terms, as an artifact of the method, the different regions are characterized by a different chemical potential. However, since this aspect does not stem from the physics of the system but only from the formalism, we have to amend for this thermodynamical unbalance. This means that the use of eq. 2.112 alone can not assure thermodynamical equilibrium and further formal relations, linking the variables of the problem, should be determined in order to obtain equilibrium. This can be obtained, as shown in the next sections, by analyzing the meaning of the process of varying resolution in statistical and thermodynamical terms.

2.4.3 Principles behind the Thermodynamical Equilibrium

In this section we analyze the idea of describing thermodynamical equilibrium for a system where, formally, the number of DOFs is position dependent and yet the molecular properties are uniform in space.

2.4.3.1 Geometrically induced Phase Transitions

The space dependent change of resolution can be seen to have some similarities to a physical phase transition, as a fictitious geometrically induced phase transition. In simple words, the concept of latent heat is similar to that of a molecule, which, for example, goes from the liquid to the gas phase and in doing so needs a certain energy (latent heat) to activate those vibrational states that make the molecules free from the tight bonding of the liquid state. In the same way, a molecule in the adaptive scheme that passes from a coarse-grained to an atomistic resolution, needs a latent heat to formally (re)activate the vibrational and rotational DOFs and to reach equilibrium with the atomistic surrounding. Vice versa the heat is released when the molecule goes from gas to liquid phase and so the bond to the other molecules becomes tighter. In the same way in the adaptive scheme, the molecule passing from atomistic to coarse-grained, formally releases DOFs and thus automatically the associated heat. This concept can be formalized as:

$$\mu^{\text{atom}} = \mu^{\text{cg}} + \phi \ , \tag{2.113}$$

where μ^{cg} is the chemical potential calculated with the coarse-grained representation, μ^{atom} that of the atomistic one, and ϕ is the latent heat [94, 92]. Possible procedures for a formal derivation of an analytic or numerical form of ϕ and how to use it in the AdResS scheme is still a matter of discussion and subject of work in progress [95]. For the time being, a simpler and practical solution is to use a thermostat that is coupled to the system (see sec. 2.4.6), which automatically, as a function of the position in space, provides (or removes) the required latent heat assuring equilibrium of the system and stabilizing the algorithm itself. The coupling of the system to a thermostat leads to the natural question of how to define the temperature in the region of transition where the number of DOFs is space dependent.

2.4.3.2 Temperature in the Transition Region

In the atomistic and coarse-grained region the temperature can be defined employing the equipartition theorem:

$$T^{\text{atom/cg}} = 2\frac{< K^{\text{atom/cg}} >}{n^{\text{atom/cg}}} \ , \tag{2.114}$$

where $< K^{\text{atom/cg}} >$ is the average kinetic energy of the atomistic/coarse-grained region and $n^{\text{atom/cg}}$ is the total average number of DOFs. In the atomistic/coarse-grained region, such a quantity is a well defined number,

however it is not in the transition region where $n^{\text{trans}} = n(x)$. The question arising is how to define T^{trans} and above all what $< K^{\text{trans}} >$ means. To address this question we make the following observations: The switching procedure implies that a DOF, in calculating average statistical quantities, *fully counts* in the atomistic region, which formally means that an integral over its related phase space is performed ($\int ....\mathrm{d}q$; q being a generic switchable DOF). On the other hand in the coarse-grained region, q is not relevant to the properties of the system and thus it *does not count at all*, that is, no integration over its related phase space is required. In the transition region the situation is something in between and thus by switching the DOF q on or off, we effectively change the dimensionality (between zero and one) of its related phase space that is of its domain of integration. In simple words q in the transition region contributes to statistical averages with a weight. The mathematical tool, which allows to formalize this idea is provided by the technique of fractional calculus, where for a fixed resolution w the infinitesimal volume element is defined as [96]:

$$\mathrm{d}V_w = \mathrm{d}^w q\, \Gamma(w/2)/2\pi^{w/2}\Gamma(w) = |q|^{w-1}\mathrm{d}q/\Gamma(w) = d\mathrm{q}^w/w\Gamma(w)\ , \quad (2.115)$$

with $\Gamma(w)$ the well-known Γ function. Employing such a formalism to calculate the average energy for quadratic DOFs one obtains:

$$< K_q >_w = \frac{\int_0^\infty e^{-\beta q^2} q^{w+1}\mathrm{d}q}{\int_0^\infty e^{-\beta q^2} q^{w-1}\mathrm{d}q} \quad (2.116)$$

The solution of eq. 2.116 is found to be [96]:

$$< K_q >_w = \frac{w}{2}\beta^{-1} \quad (2.117)$$

This is nothing else than the analog of the equipartition theorem for non-integer DOFs. Here $< K_q >_w$ is the average kinetic energy of the switchable DOF q for the fixed resolution w. One can then think to use w as a continuous parameter and thus obtaining the definition of the kinetic energy for the switchable DOFs in the transition region. A further point needs to be explained, that is, we have implicitly used a Hamiltonian to perform the ensemble average and this would contradict the statement of the previous section about the non-existence of an energy within the coupling scheme used. To clarify this aspect we have to say that the coupling formula on the forces is not directly related to the derivation of the statistical average performed here. Here we have interpreted the process of changing resolution as the process of partially counting a DOF contribution into the statistical determination of an observable under the hypothesis that the underlying

Hamiltonian is the same all over the system. This is justified by the fact that the underlying physics is in principle the same all over the system but the formal representation and thus the analysis of the DOFs of interest and their contributions differs. In practice this term means that the derivation of the temperature and the principle of coupling of forces via spatial interpolation are two aspects of the same process but one can not formally derive both from a single general principle so that the connection between them, at this stage, must be intended as only being qualitative. However, we will use simulations where both eq. 2.112 and eq. 2.117 are employed in connection to each other to prove that they are numerically consistent. At this point the obvious question arises why to choose an approach based on the interpolation of the forces and not to choose the more natural one based on the smooth interpolation of the potential. This problem is treated in the next section.

2.4.4 Coupling via a Potential Approach

The coupling scheme analog to eq. 2.112 using potentials instead of forces would be:

$$U_{\alpha\beta} = w(x_\alpha)w(x_\beta)U^{\text{atom}}_{\alpha\beta} + [1 - w(x_\alpha)w(x_\beta)]U^{\text{cg}}_{\alpha\beta} \tag{2.118}$$

This approach leads to a series of problems whose solution is not trivial. In particular if one derives the forces from eq. 2.118 one obtains an extra term, which we will name *drift force*, of the following form:

$$\vec{F}^{\text{drift}} = U^{\text{atom}}\frac{\partial w}{\partial x} + U^{\text{cg}}\frac{\partial w}{\partial x} \tag{2.119}$$

There are two options at this point, one accepts this force as a result of a definition of a new force field in eq. 2.118, or one tries to remove it by a specific choice of $w(x)$ or by modifying $U_{\alpha\beta}$ in eq. 2.118. In the first case one has to be aware that, because the derivative of $w(x)$ enters into the equations of motion, the evolution of the system becomes highly sensitive to the choice of the form of $w(x)$. This means that different functions $w(x)$ may lead to completely different results, and as the choice of $w(x)$ is made on empirical basis, the dynamic becomes arbitrary and thus, most likely, unphysical. The limitation above applies in principle to the approach proposed in [23], where the scales are coupled by an interpolation of Lagrangians via a space dependent function. Moreover, the force obtained from eq. 2.118 does not preserve Newton's third law [94, 96].

Instead if one tries to follow the second possibility that is removing $\vec{F}_{\text{drift}}$, one encounters serious mathematical difficulties [97] since the condition

$\vec{F}_{\text{drift}} = 0$ leads to a system of partial differential equations of first order:

$$\begin{aligned} U^{\text{cg}} \frac{\partial f(X_\alpha, X_\beta)}{\partial X_\alpha} + U_{atom} \frac{\partial g(X_\alpha, X_\beta)}{\partial X_\alpha} = 0 \\ U^{\text{cg}} \frac{\partial f(X_\alpha, X_\beta)}{\partial X_\beta} + U_{atom} \frac{\partial g(X_\alpha, X_\beta)}{\partial X_\beta} = 0 \end{aligned} \tag{2.120}$$

Here $f(x)$ and $g(x)$ are the most general switching functions one can think of. For the system of eq. 2.120 each equation is characterized by two boundary conditions, thus the system is *overdetermined* and in general a solution *does not exist*. This is valid also if one tries to generalize eq. 2.118 as:

$$U^{\text{coupling}} = f(X_\alpha, X_\beta) U^{\text{cg}} + g(X_\alpha, X_\beta) U^{\text{atom}} + \Phi \tag{2.121}$$

The extra potential Φ does not improve the situation because in this case the overdetermination is shifted from f and g to Φ. This kind of problems, in principle, occurs for the conserving energy method proposed in [22], where the difference between the true (full atomistic) energy of the system and the one of the hybrid scheme is provided during the adaptive run via a book keeping approach while the forces are calculated with a scheme similar to that of AdResS. The problem of the overdetermination reported above would mean in this case that the conserved energy is not consistent with the dynamics of the system. In comparison, the AdResS method has the limitation of not even attempting to define an energy, but on the other hand the overall scheme is robust enough to keep the dynamics and the essential thermodynamics under control without the problem of energy conservation. The next step consists of combining the principles of the previous section with known algorithms to implement AdResS in an existing molecular code.

2.4.5 Algorithms of the Implementation

In this section we will discuss somewhat more technical details needed to understand how AdResS can be implemented. First, every molecule needs a well-defined mapping point in both representations, usually the center of mass, but any other linear combination of particle coordinates is also sufficient. Second, eq. 2.112 will be taken apart.

2.4.5.1 Force Fields and Interpolation

As mentioned above the force between two molecules is given by [21]:

$$\vec{F}_{\alpha\beta} = w_\alpha w_\beta \vec{F}^{\text{ex,mol}}_{\alpha\beta} + [1 - w_\alpha w_\beta] \, \vec{F}^{\text{cg,mol}}_{\alpha\beta} \ , \tag{2.122}$$

where α and β label the two molecules and w_α, w_β are the adaptive weights of the two molecules, which are a function of the position of their mapping points. In the next section we will show how to distribute these forces between molecules to their atoms using virtual sites. The first part, which results from the explicit interaction of the molecules, can be written as:

$$\vec{F}^{\text{ex,mol}}_{\alpha\beta} = \sum_{\substack{i \text{ in} \\ \text{mol. } \alpha}} \sum_{\substack{j \text{ in} \\ \text{mol. } \beta}} \vec{F}^{\text{ex}}_{ij} \ , \tag{2.123}$$

where $\vec{F}^{\text{ex}}_{ij}$ is the force between the ith atom in αth molecule and the jth atom in the βth molecule, which is given by a explicit force field. The second part of eq. 2.122 comes from the coarse-grained interaction of the molecules. The coarse-grained force field is usually derived from the atomistic system by structure-based and state point coarse-graining (see sec. 2.3).

The important element of this interpolation (see eq. 2.122) is the adaptive weight function (for illustration see fig. 2.6):

$$w(x) = \begin{cases} 1 & : \text{atomistic/explicit region} \\ 0 < w < 1 & : \text{hybrid region} \\ 0 & : \text{coarse-grained region} \end{cases} , \tag{2.124}$$

has a value between 0 and 1. Obviously this definition of w gives a purely explicit force in the explicit region and a purely coarse-grained force in the coarse-grained region, so essentially eq. 2.122 describes something new for the hybrid region only. Depending on the physical interest of the research, several functions can be implemented as long as the necessary boundary conditions are fulfilled [21]. These conditions are that the function is 1.) continuous, 2.) monotonic and has zero derivatives at the boundaries. This ensures a smooth transition and the absence of artificial barriers.

In most applications a $\cos^2$-like function is used as a weighting function:

$$w(x) = \begin{cases} 0 & : \quad x > d_{\text{ex}} + d_{\text{hy}} \\ \cos^2\left(\frac{\pi}{2d_{\text{hy}}}(x - d_{\text{ex}})\right) & : \quad d_{\text{ex}} + d_{\text{hy}} > x > d_{\text{ex}} \\ 1 & : \quad d_{\text{ex}} > x \end{cases} , \tag{2.125}$$

where d_{ex} and d_{hy} are the sizes of the explicit and the hybrid region, respectively. Other weighting functions have been tested in the early stage of AdResS [21], but except for different perturbations in the density profile no influences have been found.

In the literature both spherical [98] and one dimensional [93] splitting of the simulation box has been reported and depending on this, the distance x

to the center of the explicit region is calculated as follows:

$$x = \begin{cases} |(\vec{R}_\alpha - \vec{R}_{\rm ct}) \cdot \hat{e}| : \text{splitting in } \hat{e} \text{ direction} \\ |\vec{R}_\alpha - \vec{R}_{\rm ct}| \quad : \text{spherical splitting} \end{cases} , \tag{2.126}$$

where $\vec{R}_{\rm ct}$ is the center of the explicit zone, which can be fixed or moved in time (e. g. with the center of a molecule). $\vec{R}_\alpha$ is the mapping point of the αth molecule. For the center of mass mapping, it is given by:

$$R_\alpha = \frac{\sum_{i\in\alpha} m_i r_i}{\sum_{i\in\alpha} m_i} \tag{2.127}$$

Note that the value of the weighting function depends exclusively on the mapping of the molecule.

2.4.5.2 Virtual Sites

Virtual sites are needed in AdResS due to the fact that there are forces acting on the coarse-grained representation and on the explicit representation simultaneously, which need be to redistributed in the correct way to the explicit representation.

These sites are well-known constructions [87] that allow for interactions between arbitrary points in a molecule. One can imagine these interaction points as virtual atoms, whose movement is determined by a geometrical rule rather than by Newton's equation of motion. In formulas this means the potential between two real atoms is given by:

$$V_{ij} = V^{\rm ex}(\vec{r}_i, \vec{r}_j, \ldots) + V^{\rm vs}(\vec{r}^{\rm vs}(\vec{r}_i, \vec{r}_j, \ldots)) , \tag{2.128}$$

where the first part depends on the positions of the real atoms and the second part depends on the position of the virtual site $\vec{r}^{\rm vs}$, which is also a function of the positions of the real atoms. The force on the ith atom deriving from this interaction is simply given by:

$$\vec{F}_i = -\frac{\partial(V^{\rm ex} + V^{\rm vs})}{\partial \vec{r}_i} = \vec{F}_i^{\rm ex} + \vec{F}^{\rm vs}\frac{\partial \vec{r}^{\rm vs}}{\partial \vec{r}_i} , \tag{2.129}$$

where the last product is a matrix multiplication. In the case where a virtual site is located in the center of mass, which is the most important case in standard AdResS, eq. 2.129 simplifies to:

$$\vec{F}_i = \vec{F}_i^{\rm ex} + \frac{m_i}{\sum_{i\in\alpha} m_i}\vec{F}^{\rm vs} \tag{2.130}$$

2.4.5.3 Interface Pressure Correction

The interpolation of forces (see eq. 2.122) can produce inhomogeneities in the density and affect the structure and total pressure of the system in comparison to a purely explicit simulation. These artifacts are results of the inhomogeneous structure and pressure in the hybrid region, due to the fact that the interpolation formula does guarantee the correct structure in the transition region.

One way to reduce such effects in AdResS is to replace the coarse-grained potential $\vec{F}_{\alpha\beta}^{\mathrm{cg,mol}}$ of eq. 2.122 in the hybrid region by a corrected one of the form [98] [4]:

$$\begin{aligned} \vec{F}_{\alpha\beta}^{\mathrm{cg,mol}} \rightarrow \vec{F}_{\alpha\beta}^{\mathrm{cg,corrected,mol}} &= s(w_\alpha w_\beta)\vec{F}_{\alpha\beta}^{\mathrm{cg,mol}} \\ &\quad + [1 - s(w_\alpha w_\beta)]\vec{F}_{\alpha\beta}^{\mathrm{hy,mol}} \ , \end{aligned} \tag{2.131}$$

where $\vec{F}^{\mathrm{hy,mol}}$ is obtained from the structure and state point-based coarse-graining of a constant-w system with $w = 1/2$ (see eq. 2.122). A common choice for s is

$$s(x) = 4\left(\sqrt{x} - 1/2\right)^2 \ , \tag{2.132}$$

which ensures that the coarse-grained particles ($w = 0$) interact with the usual coarse-grained potential, while hybrid particles with $w = 1/2$ will feel the full, corrected coarse-grained force field $\vec{F}^{\mathrm{hy,mol}}$. The rest will interact with an interpolated force that varies smoothly with $s(x)$ (see fig. 2.7 for illustration). Depending on the system, other smoothing functions and more interpolation points could possibly lead to better results.

2.4.5.4 Thermodynamic Force

Another way of reducing the inhomogeneities due to the force interpolation is the so-called thermodynamic force [99]. It consists of a position-dependent external field applied in the region of interpolation in order to produce a flat density profile, based on thermodynamic consistency considerations. The force provides the amount of work necessary to compensate the difference in excess chemical potential responsible for the density artifacts. Its effect is exerted on the coarse-grained site of the molecules and can be used to correct local density inhomogeneities in the hybrid region. The force can also be used to couple explicit and coarse-grained regimes of different state points.

[4] Again, it is important to note that this interpolation still obeys Newton's third law.

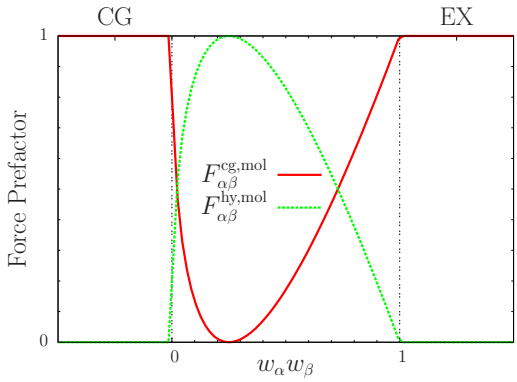

Figure 2.7: **AdResS interface pressure correction**: The force pre-factors of the interface pressure correction (see eq. 2.131) between the coarse-grained (CG) and explicit (EX) zone. Note that zone in the middle of the plot, $0 < w_\alpha w_\beta < 1$, is not necessarily the hybrid zone, but contains the hybrid zone.

2.4.6 Implementation in ESPResSo

In this section we explain how the AdResS method is integrated in the soft matter research package ESPResSo [59]. It is important to note that we are not aiming primarily for computational speed but rather for flexibility and modularity, which was the major aim of ESPResSo since the beginning. A fast implementation with a new adaptive integrator will be released in the successor of ESPResSo: ESPResSo ++ [88].

2.4.6.1 General Overview

We implemented the AdResS simulation as an explicit simulation with additional constraints, namely a virtual site in the center of mass of each molecule that represents the coarse-grained particle. Therefore, we have hybrid molecules throughout the simulation box and we integrate only the explicit representations of the particles using a velocity Verlet scheme, after we have distributed the forces from their coarse-grained representations. The idea of keeping the number of "virtual" particles constant over the whole box is well known from grand-canonical simulations [100]. On one hand, this makes our code slightly slower, but, on the other hand, one can omit the calculation of the explicit interaction in the coarse-grained region and the calculation of the coarse-grained interaction in the explicit region, which would lead to an increase in the total speed of an adaptive simulation compared to a pure explicit simulation. This *double representation* trick for every molecule saves having to destroy and create particles on the fly, which is a

very expensive operation in the sense of memory access and bookkeeping in the case of a parallel simulation. It is important to note that the force calculation is the most time consuming operation in most molecular dynamics simulations, which essentially leads us to this approach. Of course we do not calculate the forces on the particles where it is not needed. By not destroying the explicit particle this implementation favors molecules with only a few particles, so it must be rethought for molecules with more than 25 particles. However, this trick also allows us to perform constant-w simulations easily, which are sometimes needed to determine the interface pressure correction potential $F^{\mathrm{hy,mol}}$ (see sec. 2.4.5.3).

The AdResS feature of ESPResSo can be enabled by adding:

```
#define ADRESSO
```

to the ESPResSo configuration file, e. g. `myconfig.h` and the following line in the header of the steering `TCL` script:

```
adress set topo $kind width $width $hybrid_width \
           center x $R_x wf $wf
```

where `kind` determines the type of AdResS simulation:
0 disabled
1 constant weight function
2 one dimensional splitting
3 spherical splitting
`wf` the type of weighting function:
0 see eq. 2.125
1 user defined
`width` and `hybrid_width` are the widths of the explicit and hybrid regions (see eq. 2.125), respectively and `R_x` is the x position of the center of the explicit zone.

2.4.6.2 Integrator

The integration in ESPResSo is done by a standard velocity Verlet integrator with an addition to allow NVT ensemble simulations by a Langevin [44] or a DPD [46, 1] thermostat. It works in the following way, which is the numerical version of what we have described in fig. 2.1 on page 11.

1. $v(t + \Delta t/2) = v(t) + \Delta t/2 \cdot f(t)/m$
2. $p(t + \Delta t) = p(t) + \Delta t\, v(t + \Delta t/2)$
3. Calculate $f(t + \Delta t)$ from $p(t + \Delta t)$, $v(t + \Delta t/2)$

4. $v(t + \Delta t) = v(t + \Delta t/2) + \Delta t/2 \cdot f(t + \Delta t)/m$

The scheme must be extended in the case of an AdResS simulation:

1. $v(t + \Delta t/2) = v(t) + \Delta t/2\, f(t)/m$
2. $p(t + \Delta t) = p(t) + \Delta t\, v(t + \Delta t/2)$

2b. Update the positions, velocities and weighting functions $w(R)$ of the virtual sites (see eq. 2.124 and eq. 2.127)

3. Calculate $f(t + \Delta t)$ from $p(t + \Delta t)$, $v(t + \Delta t/2)$

3b. Distribute the force of the virtual sites to the real particles using eq. 2.130

4. $v(t + \Delta t) = v(t + \Delta t/2) + \Delta t/2\, f(t + \Delta t)$

The *double representation* of every molecule allows us to apply the actual integration (step 1,2,3,4) on the explicit particles only, regardless of their real representation given by the weighting function. This is obviously not the most effective method of implementation, but avoids the need to develop a new integrator that can switch the actual calculation between explicit and coarse-grained particles on the fly.

2.4.6.3 Parallelization Scheme

The ESPResSo package basically divides the simulation box into sub-boxes, which are distributed among different processors. To allow the calculation of interactions across the boundaries of these sub-boxes, every sub-box is surrounded by a so-called ghost layer. This layer contains copies of the particles, which are in a different sub-box (and so on a different processor), and are called ghost particles. These ghosts are updated after every propagation step of their real counterparts by communication between the neighboring processors. A detailed description can be found in the ESPResSo publication [59].

Only the communication part of the propagator needs to be modified in the presence of virtual sites. Essentially, the amount of communication *doubles* due to the fact that the positions of all the particles (including the ghost particles) have to be updated to calculate the positions of the virtual sites (first communication) and the position of the virtual sites has to be distributed again (second communication).

The same effect comes into play when the forces are collected from all the copies of one particle (ghosts and real particles). To distribute the force

from a virtual site one needs to sum up its forces first (first communication) and then distribute them, but as the forces are possibly distributed to ghost particles one will have to collect the forces again (second communication). It is clear that in the absence of virtual sites the second communication is not necessary.

2.4.6.4 Thermostat

ESPResSo provides several stochastic thermostats, which are of a local nature and can be extended to the AdResS scheme. We will discuss all the details for the Langevin thermostat [44], but they also apply for the DPD thermostat [46, 1]. We want to obtain a canonical distribution of the velocities of the atoms in the hybrid and explicit regions and of the centers of mass in the whole system. This requires the proper application of the thermostat to each region with a consistent initialization of the atoms' velocities whenever a molecule crosses from the coarse-grained to the hybrid regime. While we have chosen to apply the Langevin thermostat over the complete box on the explicit particles, independent of their nature, there are two tested ways of initializing the explicit velocities when the additional degrees of freedom are introduced:

- Copy the atoms' velocities relative to the center of mass of a molecule from a random explicit molecule when a molecule crosses the border to the hybrid region. This method was used in the original description of AdResS [21].

- The thermostat acts on the explicit representation of the particle everywhere, regardless if they are in the coarse-grained region or not. In this case the initialization does not matter, since the atoms are already thermalized. It is clear that the coarse-grained particles (technically virtual sites at the centers of mass) will automatically have the correct velocity distribution, if all the explicit particles have the correct distribution [5].

The main difference between these two approaches is that the second case is much easier to parallelize, since less communication is involved. Under both schemes the temperature and density profiles are unchanged and

[5]For the center of mass of a molecule, one just has to add up the individual Langevin thermostats of the atoms of the molecule and it follows that this "summed thermostat" obeys the dissipation-fluctuation theorem itself, if the friction of the explicit particles is scaled with their masses (for the detailed derivation see sec. 2.2.9)

the structural properties remain intact. Thus, the correct distribution of velocities is guaranteed.

The thermostat can be set up as follows:

```
thermostat langevin $temperature $friction
```

or

```
thermostat dpd $temperature $friction $cutoff
```

2.4.6.5 Intramolecular Initialization

For the initialization of the molecule's orientation and internal bond-lengths when crossing from the coarse-grained to the hybrid regime, we follow the same procedure as for the velocities in the previous subsection, that is, we assign relative positions by selecting a random explicit particle or, alternatively keep the explicit representations that come from the coarse-grained region, where the atoms feel no intermolecular interaction. In both cases, the behavior is similar, but in the second approach the results are much simpler to implement and faster during computation.

2.4.6.6 Bonded Interactions

The intra-molecular interactions are not described by the general interpolation equation (see eq. 2.122), but are fully included within the whole simulation box. For rigid molecules such as water there is no explanation for this approach needed, however, for flexible bonds, the bonded interactions are still calculated across the whole simulation box without any interpolation due to stability reasons. It is clear that this only works for coarse-grained models with one coarse-grained site per molecule. For more complicated structures where bonded interactions are also present in the low resolution regime, the force interpolation has to be rethought.

2.4.6.7 Cut-Offs

As the representation (explicit, hybrid or coarse-grained) of the molecule is determined on the basis of the position of the center of mass of the molecule it is natural to use a cut-off based on the distance between the centers of mass. This strategy is well known [101] in water simulations with Reaction-Field to avoid artificial dipole moments at the cut-off. In other words, a molecular cut-off ensures that either all particles or no particles of a molecule interact.

We have implemented the molecular cut-off as a virtual interaction:

```
inter $type1 $type2 molcut 1 $cut-off
```

where `type1` and `type2` are the particle types. Internally the distance between the centers of mass of the molecules to which these particle types belong is used. All other individual interaction cut-offs stay active and have to be increased to a value higher than the molecular cut-off in the case that one wants to simulate a system with molecular cut-offs only.

2.4.6.8 Molecules & Mappings

The basic support to join particles to molecules comes with the standard version of ESPResSo. The user is allowed to combine a set of particles into a molecule. We use this feature to equip every molecule with one virtual site to define the set of atoms from which the center of mass is calculated. The extension to center of geometry or multiple virtual sites in one molecule is in preparation.

A single tetrahedral molecule can be set up with:

```
set molecule_topology $molecule_type
part 1 pos $pos1x $pos1y $pos1z virtual 0
lappend molecule_topology 1
part 2 pos $pos2x $pos2y $pos2z virtual 0
lappend molecule_topology 2
part 3 pos $pos3x $pos3y $pos3z virtual 0
lappend molecule_topology 3
part 4 pos $pos4x $pos4y $pos4z virtual 0
lappend molecule_topology 4
part 5 pos $pos5x $pos5y $pos5z virtual 1
lappend molecule_topology 5
eval analyze set $molecule_topology
analyze set topo_part_sync
```

and the position of the virtual site (in this case particle number `5`) can be set to the right position by calling

```
integrate 0
```

From now on

```
integrate $steps
```

will integrate the whole system for `steps` steps and move the virtual sites accordingly.

2.4.6.9 Interface pressure correction

The interface correction is introduced as a new kind of non-bonded tabulated interaction.

```
inter $type1 $type2 adress_tab_ic $filename
```

where `type1` and `type2` are the types of particles subjected to the coarse-grained interaction that will be modified. They must be virtual particles. `filename` contains the tabulated fields corresponding to the interactions $\vec{F}_{\alpha\beta}^{\mathrm{cg,mol}}$ and $\vec{F}_{\alpha\beta}^{\mathrm{hy,mol}}$ according to eq. 2.131. The syntax is consistent with the usual tabulated interactions, the first four lines being the special character `#`, followed by the number of points N and the minimum and maximum separation distances r_{max} and r_{min}. After this information, the two potentials are introduced in five columns, as r, $\vec{F}_{\alpha\beta}^{\mathrm{cg,mol}}/r$, $\vec{U}_{\alpha\beta}^{\mathrm{cg,mol}}$, $\vec{F}_{\alpha\beta}^{\mathrm{hy,mol}}/r$, $\vec{U}_{\alpha\beta}^{\mathrm{hy,mol}}$. The number of points and the cut-off radius are assumed to be the same for both potentials, while the values of r are equally distributed between r_{min} and r_{max} with a fixed distance $(r_{max} - r_{min})/(N-1)$. The position and potential columns are ignored in the calculations and are included for readability only.

2.4.6.10 Thermodynamic Force

This new force is implemented as a tabulated external field that acts only in the switching region. Different kinds of particles can have different thermodynamic forces. The corresponding command is

```
thermodynamic_force $type $filename $prefactor
```

where `type` is the type of particle on which the tabulated force specified in `filename` will be exerted. `prefactor` is a coefficient that multiplies the force defined by the user.

The format of the tabulated external field follows the same definition as the tabulated pair potentials in standard ESPResSo. It starts with four special lines with the character `#`, the number of points and the minimum and maximum separation distances, r_{max} and r_{min}. Then, three columns should be entered, s, $-\frac{dU}{ds}$ and U. s corresponds to a dimensionless position $\frac{x}{d_{hy}}$ in the hybrid region, going from 0 (coarse-grained representation) to 1 (explicit representation). The derivative of the potential has to be expressed as a function of this variable. The third column is the potential energy associated with the field; it has no effect on the equations of motion and is included for readability only.

2.4.7 Summary

We have shown how to use AdResS in ESPResSo. In addition, we described several correction algorithms to obtain a flat density distribution over the entire box. We also discussed several choices in the implementation, which were made based on performance issues, namely the thermostat and the reinitialization of the molecule. A sample AdResS simulation was made available in the tutorial of ESPResSo version 2.2. However, ESPResSo ++ [88] is also on the way and will replace ESPResSo in the near future. A similar implementation has been done in GroMaCS [102] to study atomistic systems.

The implementation in ESPResSo will used to reproduce some results from previous studies of the tetrahedral liquid in sec. 4.1.While the implementation in GroMaCS will used to study water and the hydration of fullerenes in sec. 4.2 and sec. 4.3.

Chapter 3

Coarse-Graining

To remind the reader, coarse-graining methods aim at finding a way to map an all-atom model onto a less structured and simpler model, which is computationally much more efficient. A coarse-grained system should preserve as many properties of the underlying all-atom system as possible or as required by the physical question under consideration.

In this section we will focus on the description of water on the coarse-grained level, i. e. treating water molecules as ideal *spheres*. All non-spherical effects are neglected, which in some cases can be a rather crude approximation. We will see which atomistic properties can be reproduced using these simplified models.

After a detailed study of the common three-site water model using iterative Boltzmann inversion, different coarse-graining techniques will be compared. As the idealized water ball is often too simple to show certain effects, such as multicomponent effects or the effects of limited basis sets, one must also take a look at some slightly more complicated molecules like methanol, propane and hexane.

As the use of coarse-grained models aims at a simplified description of both static and dynamic properties, it will be shown how to tune some transport properties, namely the diffusion constant and the viscosity, of a coarse-grained water system to those obtained on the atomistic level. This shows that only parts of the dynamics, namely the diffusion part, can be recovered, but not the full dynamics of the atomistic system.

But let us first briefly discuss the atomistic water model to understand what we are actually coarse-graining.

3.1 Atomistic Water Models

The development of all-atom water models has a long history. The first ideas go back to Bernal and Fowler in 1933 [103]. The modern development of water modeling in computer simulation started in the early 1970s and several of the currently used models were developed in the 1980s. The SPC model was introduced by Berendsen and co-workers in 1981 [104] and TIP3P and TIP4P were published in 1983 [105]. The most widely studied extension of the SPC model, namely the SPC/E model (1987) [106], takes into account the averaged polarization effects. A more recent development is the TIP5P (2000) [107] model that is able to reproduce the density anomaly near 277 K as well as to maintain high-quality structural and thermal properties. Over the last 30 years a huge amount of work has been devoted to develop improved water models [108] and this will stay an active area in the future.

All-atom models are designed and parameterized to fit one or more physical properties, such as the radial distribution function, density anomaly, heat of vaporization, dipole moment, etc. Alternatively, they can be based on ab initio calculations of water dimers or higher clusters. None of the classical all-atom models is able to simultaneously reproduce all physical properties of water, as is pointed out and analyzed in literature [108, 109, 110, 111]. Moreover, all present models take only two-body interactions into account, while three-body interactions, which are shown to contribute less than 14.5% to the total internal energy [112], are neglected. This also gives an impression about the typical accuracy achieved by classical all-atom models used so far. Although the water molecule is small and one of the most basic molecules in nature, it still poses difficult problems. Nevertheless MD simulations employing classical atomistic water models play an increasingly important role in many areas of computational physics and chemistry.

In the present work, three simple rigid and non-polarizable three-site models (TIP3P [105], SPC[104] and SPC/E [106]) are studied. All more complex models are extensions thereof. The methods described below can be extended to any all-atom water model without substantial problems. Therefore, the simplest water models are chosen.

The oxygen atoms interact through a Lennard-Jones potential, which defines the overall size of the molecule, while the Coulomb interaction is assigned to charges on the hydrogen and oxygen nuclei to model hydrogen bonds. The intermolecular potential of these models can be expressed as

$$V_{\alpha\beta} = \sum_{i\in\alpha}\sum_{j\in\beta} \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} + \frac{C_{12}}{r_{\mathrm{OO}}^{12}} - \frac{C_6}{r_{\mathrm{OO}}^{6}} \ , \tag{3.1}$$

where α and β stand for two different molecules and r_{ij} denotes the distance

	TIP3P	SPC	SPC/E
r_{OH} [Å]	0.9572	1.0	1.0
∠HOH [°]	104.52	109.47	109.47
$10^3 C_6$ [kJ/mol nm^6]	2.4889	2.6171	2.6171
$10^6 C_{12}$ [kJ/mol nm^{12}]	2.4352	2.6331	2.6331
q_{O} [e]	−0.834	−0.820	−0.8476
q_{H} [e]	+0.417	+0.410	+0.4238

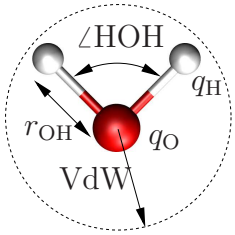

Table 3.1: **Parameters of all-atom water models**: The water models studied here are the rigid versions of TIP3P, SPC and SPC/E. All of them have Lennard-Jones interactions between the oxygens, whose C_6 and C_{12} is given above.

between atom i on molecule α and atom j on molecule β. r_{OO} denotes the distance between the oxygen atoms. The intramolecular interactions are not considered because the geometry of the molecule is kept rigid. The parameters of the interactions and geometrical constraints are listed in tab. 3.1. The intramolecular bonds are kept rigid by the LINCS algorithm [113].

Various water models with internal degrees of freedom have been studied as well. These internal degrees are bond length and angle vibrations [114, 115, 116, 117] or even polarization effects [118, 119, 120, 121, 122]. Such models are not considered here.

3.2 Coarse-Grained Water Models

Here, we map the entire water molecule onto one coarse-grained bead in the center of the oxygen. Recently, this procedure was adapted to study TIP3P water in an adaptive resolution simulation (AdResS method) [123]. In that study the parameterization of the coarse-grained model was based on TIP3P water with a bond angle of 112.19° instead of 104.52° [124]. This can led to a shift of the relative depths of the minima in the coarse-grained potential. This difference in coarse-grained models motivated, to some extent, this detailed study, in which different models are compared in a systematic way.

As already discussed in sec. 2.3, there are different options to parameterize the effective coarse-grained particle-particle interactions in order to match the overall water structure most closely. Coarse-grained models can not reproduce every property of the corresponding all-atom model [125]. Therefore one has to adjust the coarse-grained model based on which question it is supposed to answer. For now we are mostly concerned with the structural properties of liquid water at a given temperature and pressure. The construction of an effective interaction consists of two main steps:

1. Iterative Boltzmann inversion. This scheme constructs an effective potential that reproduces the center of mass or O-O radial distribution function of the all-atom model. So, if a property of interest is mostly relying on the RDF, then it is naturally reproduced by the resulting coarse-grained model. For the three water models considered in this chapter, the differences in the peaks and minima between the center of mass and O-O radial distribution functions are no more than 2%.

2. Adjust the obtained potential produced by the first step to reproduce further properties we are interested in. Of course the correction may sacrifice the accuracy of other properties that are not so important to the problem. This step can be iterated with every step of iterative Boltzmann inversion or after the whole inversion procedure.

3.2.1 Liquid Structure – Geometrical Consideration

A special property of liquid water is the tetrahedral packing due to hydrogen bonds. For the all-atom models this local structure is the result of the nearest neighbor Coulomb interaction of the partial charges on the oxygens and hydrogens. In contrast to the all-atom models, the effective potential is spatially isotropic. For that reason the structure is a result of two length scales in the effective potential. Here, we will refer to the molecule in the center of four nearest neighbors as the reference molecule and analyze whether such a cluster resembles the tetrahedral packing of the underlying all-atom model. The tetrahedral packing is measured by a parameter q_4 which is computed by a sum over the deviation between each "bond angle" (there are no actual bonds between the coarse-grained beads) and the perfect tetrahedral angle $\arccos(-1/3) \approx 109.471°$.

$$q_4 = 1 - C_4^{-1} \sqrt{\sum_{i<j} \left[\, \theta_{ij} - \arccos(-\frac{1}{3}) \, \right]^2} \, , \tag{3.2}$$

where θ_{ij} is the bond angle between particle i, the reference particle, and particle j, where i and j sum over the first four nearest neighbors of the reference particle. The bond angles are measured between the oxygen atoms in the all-atom description and between the beads in the coarse-grained model. $C_4 = 1.8165\,\text{rad}$ is a constant chosen such that $q = 0$ when "bond angles" are randomly distributed and $q = 1$ when the particles exhibit perfect tetrahedral packing. As the tetrahedral packing becomes weaker, the value of q becomes smaller. This definition is very similar to the parameter introduced in [126]. The difference is that our parameter measures the deviation from the "bond

	ϕ (1.155)	γ (1.633)	$\bar{q}_4$	$P(q_4 \geq 0.7)$
TIP3P AA			0.3512 (0.0009)	0.0251
TIP3P CG	1.120	1.617	0.2530 (0.0003)	0.0049
TIP3P PC CG	1.127	1.531	0.2525 (0.0003)	0.0048
SPC AA			0.3823 (0.0009)	0.0359
SPC CG	1.120	1.636	0.2592 (0.0003)	0.0059
SPC PC CG	1.128	1.575	0.2587 (0.0003)	0.0060
SPC/E AA			0.4109 (0.0009)	0.0479
SPC/E CG	1.121	1.643	0.2670 (0.0003)	0.0078
SPC/E PC CG	1.125	1.587	0.2662 (0.0003)	0.0071
POE	1.131	1.632	0.3142 (0.0004)	0.0251
POE PC	1.131	1.611	0.3238 (0.0004)	0.0297

Table 3.2: **Structural properties of different water models**: The results of the all-atom (AA), the coarse-grained (CG), the pressure-corrected coarse-grained (PC CG) models and the potential of extrapolation (POE) model, which optimizes the tetrahedral packaging (see eq. 3.6 for definition), are shown. The properties presented in the table are from left to right: ϕ (the ratio between peak position and first-well position), γ (the ratio between second-well position and first-well position), tetrahedral packing parameter and the probability that q_4 is a large value (not less than 0.7). The statistical error is given in parentheses.

angles" from the perfect tetrahedral angle rather than of the cosine values. We chose this term due to the fact that it *amplifies* the region around the perfect tetrahedral angle, which can easily be seen by calculating $\mathrm{d}q_4/\mathrm{d}\theta_{ij}$. The change of q_4 in our formula is much bigger than the one in [126]. Also, this definition can easily be extended to q_8 and q_{12} for particle arrangements on hcp and fcc lattices.

Back to the case of tetrahedral clusters, these can be characterized by two typical distances. The first is the distance between the reference molecule and the nearest neighbors (the center-corner distance), and the second is the distance between pairs of nearest neighbors (the corner-corner distance). The ratio between the second typical distance and the first distance is $2\sqrt{2} : \sqrt{3}$. Not surprisingly, the ratio between the distances of two minima in coarse-grained effective potential is roughly that number. This ratio is denoted by γ and its values for the different coarse-grained models are listed in tab. 3.2. It is clear that γ should not deviate from this perfect ratio too much, due to the fact that it is needed to reproduce the tetrahedral packing in liquid state. A minimum in the potential is always a favored distance between two molecules.

We denote the ratio of the distance of the peak between two first minima

	ρ $[\frac{\mathrm{g}}{\mathrm{cm}^3}]$	P [bar]	κ_T $[10^{-10}\frac{\mathrm{m}^2}{\mathrm{N}}]$	D $[10^{-9}\frac{\mathrm{m}^2}{\mathrm{s}}]$
TIP3P AA	0.9846	1.00 (0.65)	5.76 (0.02)	5.9319 (0.0737)
TIP3P CG		8536 (0.67)	4.79 (0.02)	19.3899 (0.0556)
TIP3P PC CG		0.70 (0.68)	27.12 (0.51)	19.4417 (0.2585)
SPC AA	0.9769	0.82 (0.67)	5.28 (0.02)	4.4374 (0.0643)
SPC CG		8994 (0.72)	4.66 (0.02)	17.9753 (0.0355)
SPC PC CG		1.67 (0.72)	28.62 (0.61)	17.8028 (0.2597)
SPC/E AA	0.9984	0.76 (0.73)	4.56 (0.02)	2.7866 (0.0310)
SPC/E CG		9886 (0.81)	4.38 (0.02)	15.7114 (0.0351)
SPC/E PC CG		0.45 (0.82)	29.71 (0.74)	15.6021 (0.2622)
POE	0.9984	8982 (1.10)	4.78 (0.03)	5.8013 (0.1331)
POE PC		3.61 (1.12)	28.15 (0.92)	4.7760 (0.0053)
exp. (298 K)	0.99705 [127]	-	4.599 [128]	2.272 [129]

Table 3.3: **Thermodynamic properties of different water models**: The results of the all-atom (AA), the coarse-grained (CG), the pressure-corrected coarse-grained (PC CG) models and the potential of extrapolation (POE) model, which optimizes the tetrahedral packaging (see eq. 3.6 for definition), are shown. The properties presented in the table are from left to right: density, virial pressure, isothermal compressibility and diffusion constant. In parentheses are the statistical errors. The last line shows data from experimental measurements from the literature.

and the first potential minimum by ϕ, its value is roughly $2:\sqrt{3}$. From the geometrical side one can see the peak as a barrier to prevent other molecules from entering the tetrahedral cluster. That is why it appears at the distance when one mirrors the reference particle at the connection line of two of its neighbors.

3.2.2 Effective Coarse-Grained Potentials

We performed NVT simulations for the all-atom water models TIP3P, SPC and SPC/E using GroMaCS 3.3 [130]. $24^3 = 13834$ molecules are studied inside a cubic box with periodic boundary conditions. The intramolecular bonds are kept rigid by the LINCS algorithm [113]. The temperature is adjusted to 300 K by a Berendsen thermostat [42], which was obviously not the best choice due to the fact that the Berendsen thermostat strongly influences the dynamics of the system. The densities (see tab. 3.3) of the systems are the average densities obtained from NPT simulations of the same systems, where the pressures were kept at one bar by the Berendsen barostat [42]. The long range electrostatic interactions are calculated by the particle mesh Ewald

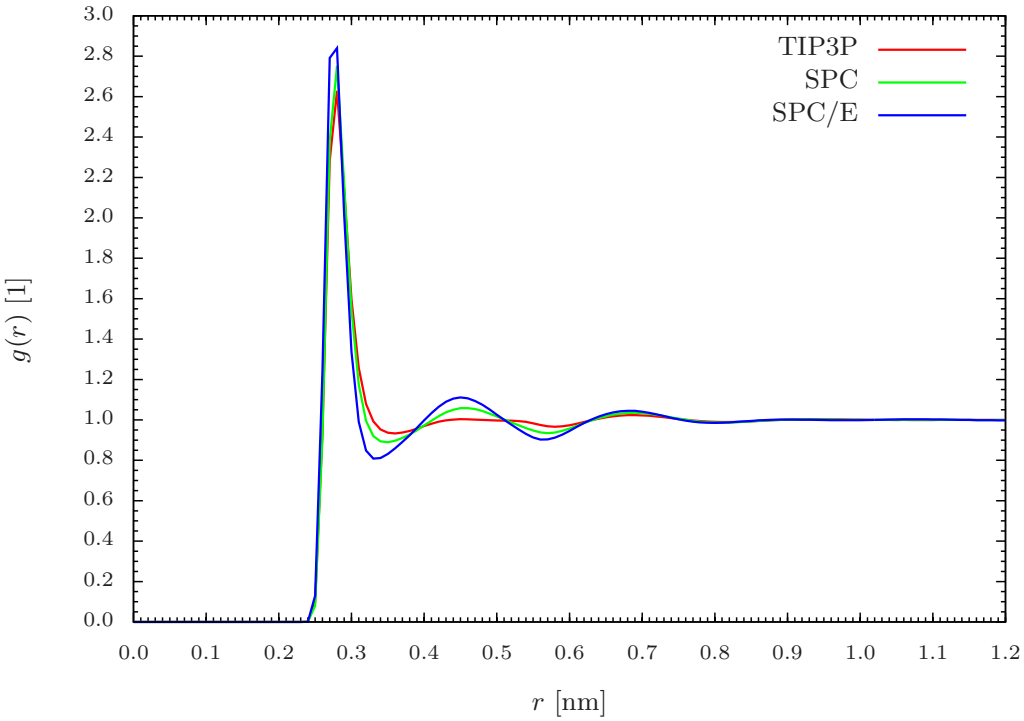

Figure 3.1: **Radial distribution function of all-atom models**: Radial distribution function between the oxygen atoms of the all-atom TIP3P, all-atom SPC and all-atom SPC/E model at ambient conditions.

(PME) method [131]. Parts of the coarse-grained and pressure-corrected coarse-grained simulations are carried out using the ESPResSo package [59] due to the fact that it is more flexible and interactive than GroMaCS. Later some of the routines used for this study were developed further with the iterative framework of the VOTCA package (see sec. 2.3.6).

Our coarse-grained simulations contain 10^4 particles and the size of the periodic simulation boxes is adjusted and fixed so that the densities are the same as for the corresponding all-atom simulations. The cut-off is set to 0.7 nm, which contains most of the structure (see fig. 3.1). Temperature is kept constant at $300\,K$ by a Langevin thermostat (see sec. 2.2.4) with a friction constant of $5\,\mathrm{ps}^{-1}$. The systems are integrated at a time step of 0.004 ps. Thus, the results are targeted towards static properties. As the thermostat can dramatically influence the dynamics, the diffusion constants of the coarse-grained models are measured by simulations employing GroMaCS 3.3 with the same system settings as for all-atom simulations (notably the Berendsen thermostat) to ensure comparable dynamics.

The RDFs of all the employed all-atom models are shown in fig. 3.1. One

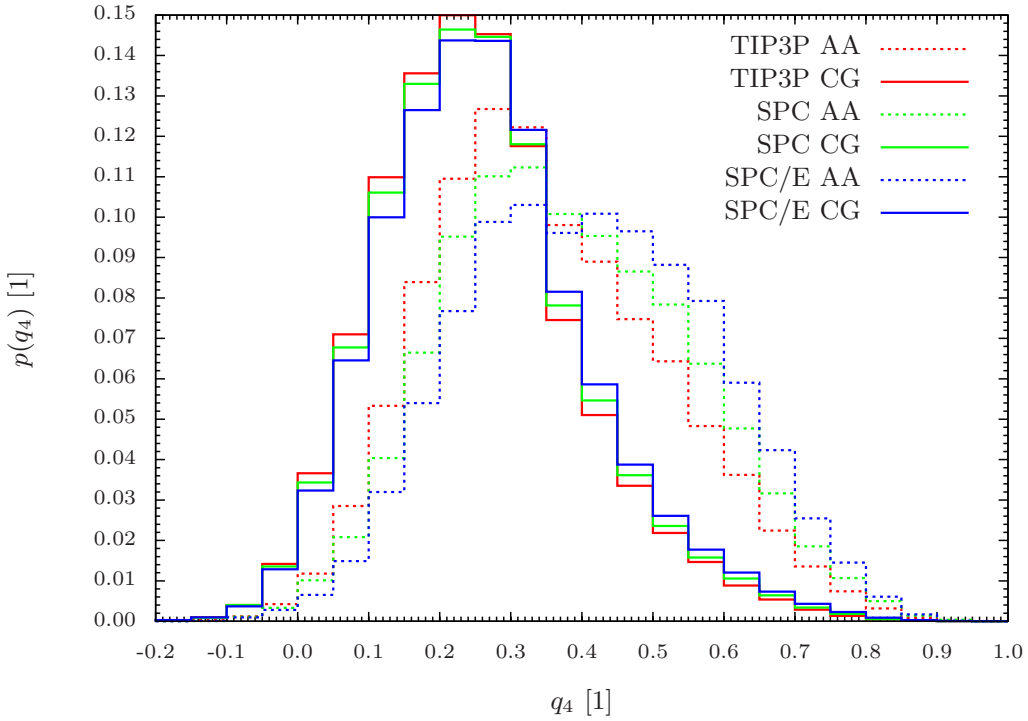

Figure 3.2: **Tetrahedral packing parameter of water**: The distribution function of the tetrahedral packing parameter q_4 of all-atom (AA) water models and coarse-grained (CG) water models. The results of pressure-corrected potentials are not plotted because they are almost indistinguishable from those without pressure-correction.

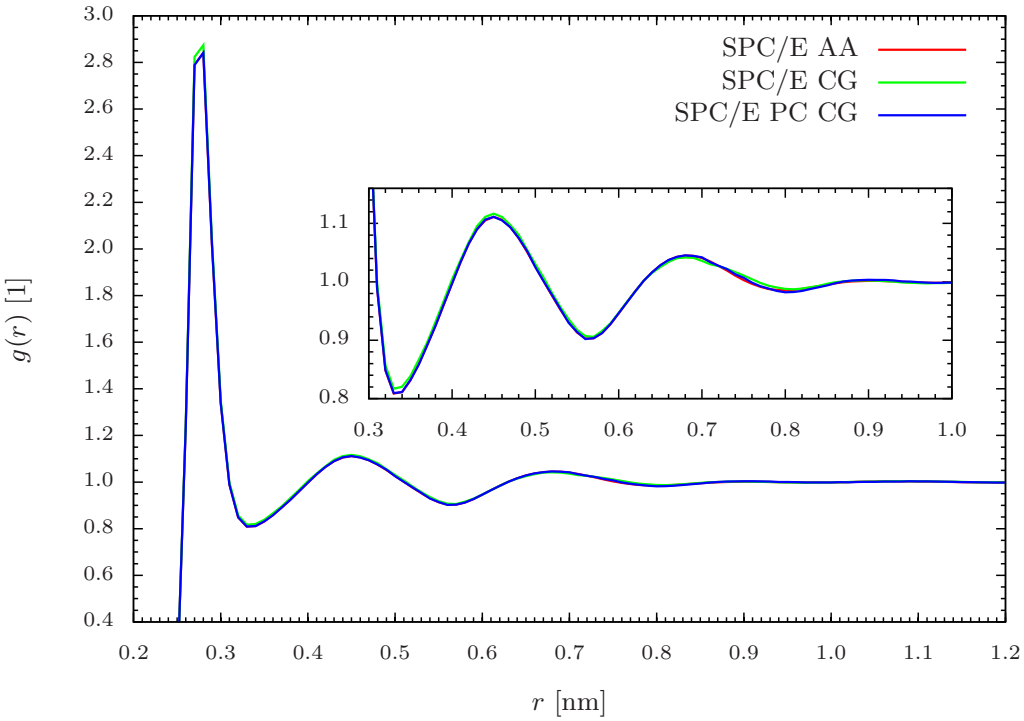

Figure 3.3: **Radial distribution function of coarse-grained SPC/E water**: Comparison between the RDF of all-atom SPC/E model (SPC/E AA), coarse-grained SPC/E model (SPC/E CG) and pressure-corrected coarse-grained SPC/E model (SPC/E PC CG). The all-atom and the coarse-grained RDFs coincide with each other quite well, while the pressure-corrected RDF deviates slightly at the minima and maxima as the insertion shows.

can see that the RDF of the SPC/E model has the strongest peaks and wells while the RDF of the TIP3P model has the weakest structure. A similar conclusion is reached for q_4 by looking at the plot of the distribution of tetrahedral parameter q_4, see the dashed lines in fig. 3.2.

The derived coarse-grained models match the all-atom RDFs (only the RDFs of SPC/E model are plotted in fig. 3.3) extremely well. The corresponding effective potentials are plotted as solid lines in fig. 3.4, from which it is observable that the larger magnitude of the peaks and wells in RDFs results in stronger peaks and wells in effective potentials. However, all coarse-grained models produce a significantly weaker tetrahedral order than the corresponding all-atom models as shown in fig. 3.2. SPC/E displays the most pronounced tetrahedral packing and TIP3P presents the least, again in agreement with the shape of the coarse-grained potential. The reason is that the hydrogen bonds originating from the Coulomb interactions of the partial

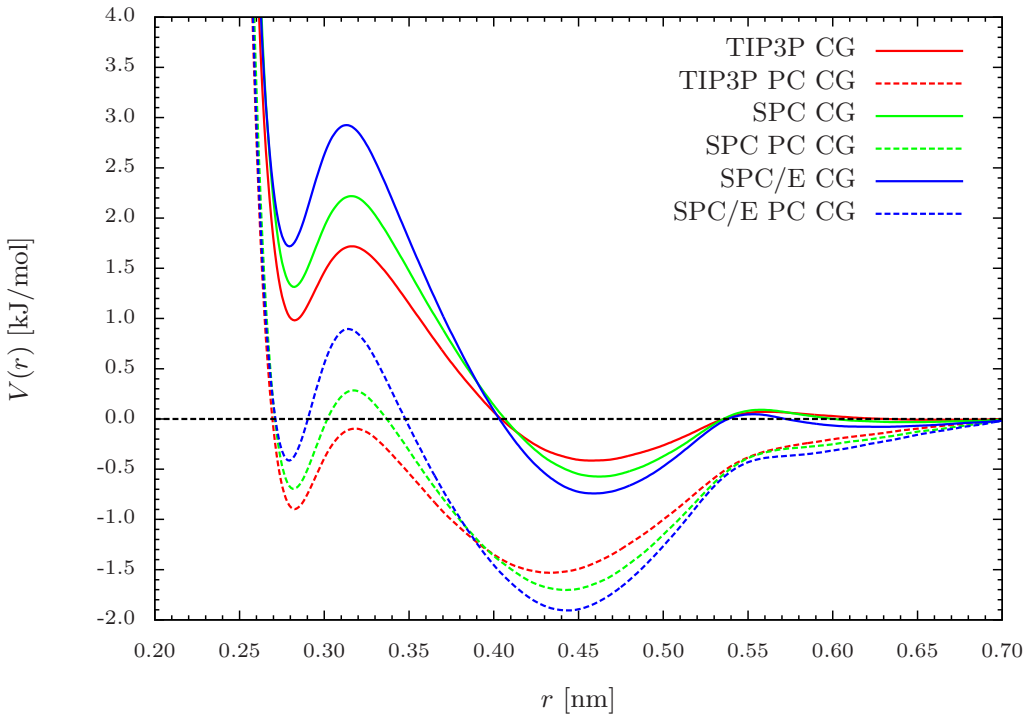

Figure 3.4: **Effective coarse-grained potentials**: Different effective coarse-grained (CG) potentials of several water models generated by iterative Boltzmann inversion. Solid lines correspond to those without pressure-correction (PC) and the dashed lines are those with pressure-correction.

charges of the molecule are replaced by isotropic potential wells interacting with a single site. Our data shows that the tetrahedral packing for electric interactions is stronger than for effective potentials. This effect is supported by the fact that the life times of tetrahedral clusters (see sec. 3.2.3) also decrease for the effective potentials.

We observe only a small discrepancy between the all-atom RDFs and the pressure-corrected coarse-grained RDFs (see fig. 3.3) even though the effective potentials look quite different from those of the coarse-grained models without pressure-correction, see fig. 3.4. Henderson [72] has shown that an effective potential designed to reproduce a given RDF is unique up to a constant. While this is a rigorous result, the coarse-graining procedure shows that rather small variations in the RDFs can result in significant changes of the potential. Despite the different effective potentials and a discrepancy in γ between coarse-grained and pressure-corrected coarse-grained models (see tab. 3.2), the distributions of q_4 are nearly the same. Thus, they will not be plotted in fig. 3.2 and we may refer the reader to the mean values and the probability of large q_4 in tab. 3.2.

As shown in tab. 3.3, the virial pressure of the coarse-grained models does not agree with the underlying all-atom model. This significant deviation reflects the fact that the coarse-grained models approximate a system with significantly more degrees of freedom, more spatially varying and effectively anisotropic interactions. This is well known for coarse-grained simulations [125, 32]. All pressure-corrected coarse-grained models reproduce the pressures of the all atom models within the error bars, supporting our pressure-correction strategy, which was to apply the pressure-correction at each step of the iterative Boltzmann inversion. When the RDF is approximated well enough, we apply an iteration of pressure-correction to obtain the correct pressure.

Nevertheless, in order to deal with a mixed all-atom and coarse-grained simulation (AdResS scheme, see sec. 2.4) it is more appropriate to adjust the compressibility rather than the pressure. One can determine the isothermal compressibility of the system by the following finite difference method,

$$\kappa_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T \approx -\frac{1}{V}\frac{V^+ - V^-}{p^+ - p^-} \ . \tag{3.3}$$

Here, we want to note that the resulting effective potential depends on the temperature T and density ρ [125] for which the coarse-graining procedure is carried out. This has been a well-known fact for over 10 years [26]. However, since the compressibility is quite small and the system is reasonably big, it is sufficient for this estimate not to account for any dependence of the coarse-grained potential on the states $\{N, V^+, T\}$ and $\{N, V^-, T\}$. As shown in

tab. 3.3, the side effect of the pressure-correction is a strong deviation in the isothermal compressibilities. The agreement between the isothermal compressibilities of the all-atom system and the coarse-grained system without pressure-correction is much better. The isothermal compressibility can also be determined by the following formula [132]

$$\rho k_B T \kappa_T = 1 + 4\pi\rho \int \mathrm{d}r \; r^2 \left[g(r) - 1 \right] . \tag{3.4}$$

The large deviations in the compressibility between the coarse-grained and the pressure-corrected coarse-grained results show that the compressibility is a very sensitive measure of the overall agreement between different RDFs.

And it also raises a valuable question: Is it possible to fit the pressure and compressibility at the same time? And if yes, how? We will come back to this point later, but let us first analyze the change of the involved time scales.

3.2.3 Time Scales

In tab. 3.2 it is also shown that the intrinsic time scale of the coarse-grained water models is approximately 4 times bigger than that of the all-atom simulations. One can calculate a so-called time scaling factor

$$s = \frac{D_{\mathrm{CG}}}{D_{\mathrm{AA}}} , \tag{3.5}$$

to quantify this speedup, which is due to the softer interaction potentials and the resulting reduced friction between molecules, as has already been observed in the context of polymer melts [32]. This speedup leads to the sampling of a bigger part of the phase space, but does not necessarily ensure that the dynamics of the coarse-grained system resemble the dynamics of the all-atom system.

Also the life time of the tetrahedral clusters has changed. We define the life time (see tab. 3.4) of tetrahedral clusters as the length of the time period, during which a cluster has a q_4 larger than 0.7. This value ranges from 0.04 ps up to 0.07 ps and the standard deviation is roughly the same as the mean value. The life time of individual hydrogen bonds is significantly longer, roughly 0.55 ps up to 1.36 ps. Of course the definition of a *hydrogen bond* is somewhat arbitrary within these models. Taking the RDFs we count a pair of molecules as connected by a hydrogen bond when the distance of two oxygens is less than 0.31 nm. This definition is not a problem for coarse-grained models, but obviously not truly adequate for all-atom models, nevertheless it is still well suited for a comparison of the different models.

Model	t_T [ps]	t_H [ps]
TIP3P	0.057 (0.059)	0.868 (0.928)
SPC	0.067 (0.072)	0.994 (1.120)
SPC/E	0.070 (0.073)	1.268 (1.544)
TIP3P CG	0.044 (0.042)	0.546 (0.456)
SPC CG	0.043 (0.040)	0.615 (0.527)
SPC/E CG	0.044 (0.045)	0.735 (0.654)
TIP3P PC CG	0.045 (0.046)	0.550 (0.459)
SPC PC CG	0.044 (0.043)	0.610 (0.530)
SPC/E PC CG	0.047 (0.043)	0.741 (0.664)
POE	0.061 (0.069)	1.306 (1.291)
POE PC	0.061 (0.072)	1.362 (1.380)

Table 3.4: **Life time analysis of different water models**: Life times of clusters of the all-atom (AA), the coarse-grained (CG), the pressure-corrected coarse-grained (PC CG) models and the potential of extrapolation (POE) model, which optimizes the tetrahedral packaging (see eq. 3.6 for definition), are compared. Mean values are measured and standard deviations (not the statistical error) are given in parenthesis. t_T is the life time of a tetrahedral cluster, while t_H is the life time of a hydrogen bond.

We conclude that the life times of the tetrahedral clusters and the life times of a hydrogen bond for the coarse-grained models are in general smaller than the value for the corresponding all-atom model. While the life times of a hydrogen bond are around 40% smaller for the coarse-grained models the life times of the tetrahedral clusters are only about 30% smaller. This effect also supports the weaker tetrahedral packing of the coarse-grained models.

3.2.4 Optimizing the Tetrahedral Packing

Let us now come back to the question if we can find a potential that reproduces the tetrahedral packing in the same way as the all-atom models. The coarse-grained models, which fit the RDFs extremely well, have a much weaker tetrahedral packing (see tab. 3.2).

In fig. 3.1 we find that the neighbor distribution of SPC/E displays a weaker structure than that of the SPC model, while the TIP3P model has a nearly flat distribution of neighbors beyond the first peak of the RDF. Along with this the strength of the tetrahedral packing also increases from TIP3P to SPC and even more to SPC/E (see fig. 3.2). The coarse-grained potentials (see fig. 3.4) suggest a connection on how the tetrahedral packing and the shape of the potentials are correlated.

From the analysis in sec. 3.2.1, one knows how to obtain a modified effective potential that can produce better tetrahedral packing, namely by keeping the position of the wells and the peaks, but varying their height/depth. We find that the distribution of q_4 is insensitive to the change of the depth of the first well as long as the barrier is lower than a critical value of roughly $kT \approx 2.5\,\mathrm{kJ/mol}$. When we increase the height of the peak or the depth of the second well, the tetrahedral packing also becomes more pronounced. This effect may be understood by the fact that the difference between the peak and the second well acts as a barrier, which prevents more particles from entering the first neighbor shell. This result suggests that in order to reach better tetrahedral packing, we have to increase the depth of the second well and the height of the peak of the potential while keeping them at the right position.

To check the ideas raised in the previous paragraph, we manipulate the effective coarse-grained potentials by linearly extrapolating two coarse-grained potentials under the constraint of keeping the region of equality nearly unchanged. Since it does not matter which potential is used as base potential in the extrapolation we have chosen a combination of the coarse-grained SPC and coarse-grained SPC/E potential. The extrapolation can be described by

$$V_{\mathrm{POE}}(r) = V_{\mathrm{SPC}}(r) + \lambda(r)(V_{\mathrm{SPC/E}}(r) - V_{\mathrm{SPC}}(r)) \;, \tag{3.6}$$

where $\lambda(r)$ is a first order continuously differentiable function, which was generated by cubic splines. The requirement of smoothness of $\lambda(r)$ is the necessary and sufficient condition for a continuous effective force resulting from the potential $V_{\mathrm{POE}}(r)$. A value $0 < \lambda < 1$, an interpolation, allows us to locally change the potential between the two original coarse-grained potentials considered. A value $\lambda > 1$, an extrapolation, corresponds to adding even more SPC/E structure. Outside this region we use the same potential as for SPC/E ($\lambda = 1$). Negative values of λ point more towards the coarse-grained version of the TIP3P model.

The weight function $\lambda(r)$ used in this study is plotted in fig. 3.5, where the values of λ at the position of the first well, the peak and the second well are 3, 3 and 8 respectively.

The RDF obtained by a simulation using the extrapolated potential is shown in fig. 3.6. Thermal and structural properties of this potential can be found in tab. 3.3 and tab. 3.2. From fig. 3.7 we deduce that the extrapolated potential can only fit the all-atom TIP3P result when $q_4 \geq 0.7$. There is still an obvious discrepancy for smaller q_4 values. We could improve the fit of the distribution of smaller q_4 by increasing the weight function λ, but then the agreement at large q_4 values would be destroyed. In practice, the region

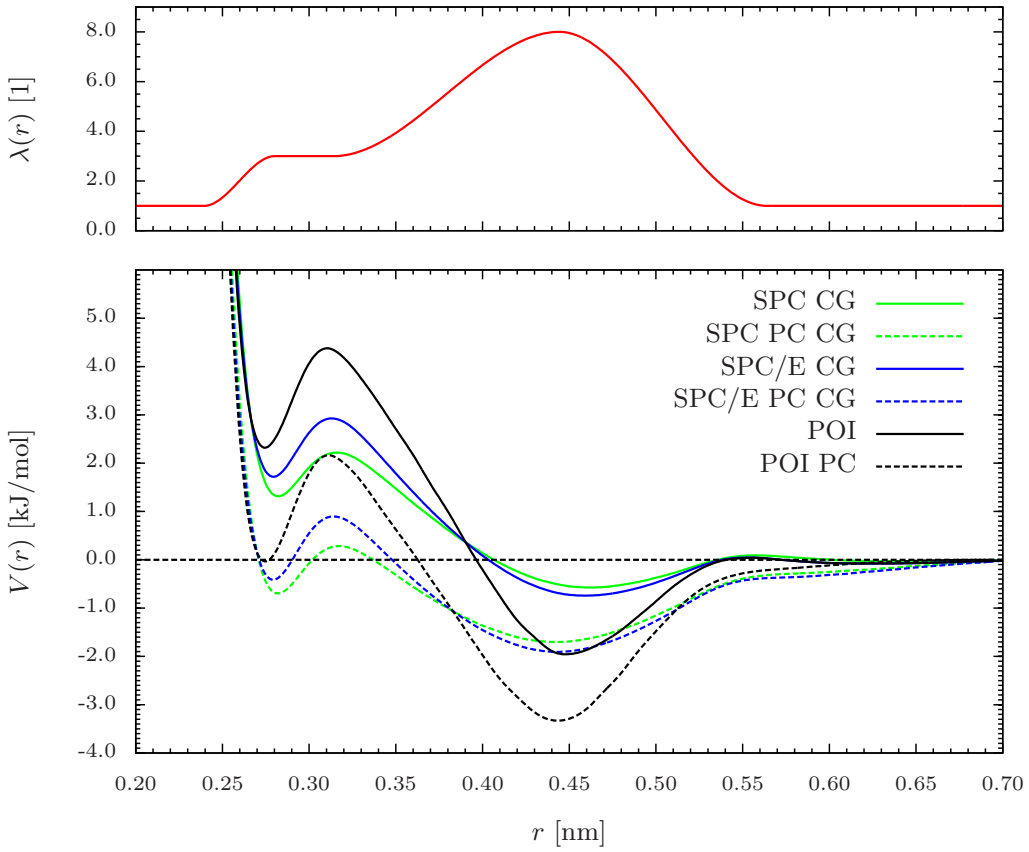

Figure 3.5: **Potential of extrapolation**: Weight function $\lambda(r)$, the potential of extrapolation (POE) and its pressure-corrected version (POE PC) are shown. These potentials are plotted in comparison with coarse-grained (CG) and pressure-corrected coarse-grained potentials (PC CG) of SPC and SPC/E.

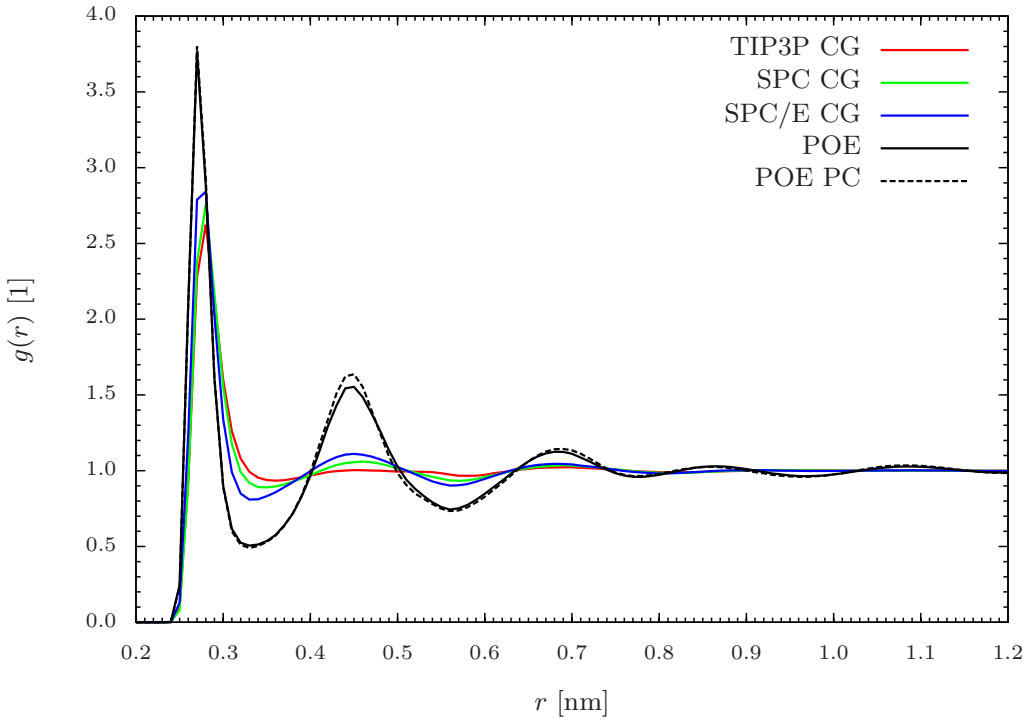

Figure 3.6: **Radial distribution function of the potential of extrapolation**: Comparison of the RDF of coarse-grained (CG) models and the extrapolated potential (POE). The extrapolated potential model is much more structured than all the other models.

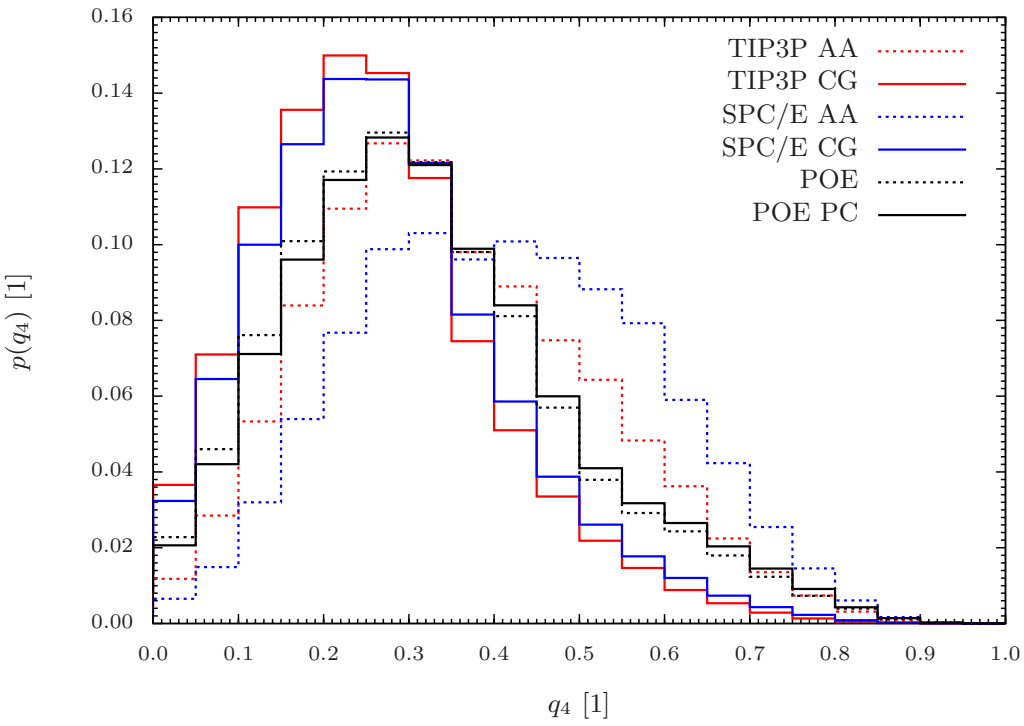

Figure 3.7: **Tetrahedral packing of the potential of extrapolation**: Distribution of tetrahedral packing parameter q_4 of the extrapolated potential (POE) model and its pressure-corrected version (POE PC), which optimizes the tetrahedral packaging (see eq. 3.6 for definition), compared to all-atom TIP3P model (TIP3P AA) and coarse-grained TIP3P model (TIP3P CG), which has the weakest tetrahedral packaging, and the all-atom SPC/E model (SPC/E AA) and coarse-grained SPC/E model (SPC/E CG), which has the strongest tetrahedral packaging.

of large q_4 (where clusters are tetrahedrally well packed) is more important than the region of small q_4. Thus we focus on changes of the coarse-grained potentials, which better fit q_4 in the $[0.7, 1]$ interval. For this new potential, fig. 3.6 shows that the RDFs deviate strongly. With deeper wells and a higher peak we fit the higher order structural aspects of the liquid better, but introduce more structure in the RDFs. It seems impossible to fit both the RDF and the q_4 distribution at the same time with isotropic central potentials.

Even considering deeper potential wells and higher peaks, we can not reach the q_4 distribution of the all-atom SPC and SPC/E models. Since the strength of the potential wells is then comparable to $kT \approx 2.5\,\mathrm{kJ/mol}$, the coarse-grained water is no longer a homogeneous liquid.

Surprisingly, the isothermal compressibility of the extrapolated potential is preserved very well though the fit of the RDF becomes poor. For the coarse-grained SPC and SPC/E model the virial pressure is too large, which has been shown above for all non-pressure-corrected coarse-grained potentials. To obtain a coarse-grained model that can reproduce the all-atom pressure, one can also extrapolate between pressure-corrected coarse-grained SPC and SPC/E models instead of between the original coarse-grained potentials. But then the resulting pressure is too low due to the stronger structure in the potential. To solve this problem, we apply a tail correction to the resulting potential and vary the size of the support of the correction to arrive at the correct pressure. In tab. 3.3 we show that the pressure is correct now, but the compressibility is wrong, which is the same observation for all pressure-corrected models. From the dashed lines in fig. 3.6 and fig. 3.7, we find no obvious difference in the structural properties between the pressure-corrected extrapolated potential and the non-pressure-corrected potential.

3.2.5 Summary

It was shown how various simple rigid water models can be coarse-grained in a straight forward way by replacing one molecule by a coarse-grained bead. An efficient coarse-grained interaction between the oxygen atoms was derived by the inverse Boltzmann method, which naturally gives a coarse-grained model a very similar structure and nearly the same compressibility as the all-atom model. Under the condition of losing the agreement of compressibility, we were able to adjust the pressure to the same value as in the all-atom case by a simple linear correction. It seems to be, however, impossible to simultaneously adjust the pressure and the compressibility by simple isotropic two-body potentials.

The lifetime analysis of the tetrahedral clusters showed that for all the models a cluster lives 50 fs on average, which makes our approach of coarse-

graining much more physical than the idea of replacing five water molecules (one tetrahedral cluster) by a coarse-grained bead. The lifetime gives a rough estimate of the maximal physical timescale of these kinds of configurations.

We also showed how to improve the tetrahedral packing of coarse-grained models by introducing a higher barrier in the coarse-grained potentials. This also leads to more structure in the RDF. This is why we think good tetrahedral packing and conserved structure is not possible at the same time with such a simple spherical interaction. In general for all these coarse-grained potentials there are two characteristic length ratios. First, the ratio of the first well to the second well, which is also the ratio between the nearest and the second nearest neighbor distances in the tetrahedral cluster. Second, the ratio of the first well to the first peak, which prevents molecules from entering the first neighbor shell.

Altogether such a coarse-graining procedure leads to an increase of computation speed by a factor of 10 (9 Coulomb + 1 Lennard Jones vs. 1 tabulated interaction) plus a gain due to the lack of electrostatic interaction and the four times larger intrinsic timescale. This accumulates to a speedup of the order of 50 in computer time. While for small systems this is not decisive, for huge simulations of e. g. biomolecular systems with surrounding water this will be crucial in many cases. Of course models with just isotropic interactions can not capture all properties compared to more complicated ones. But this is the same for all coarse-grained or simplified models and in order to overcome such shortcomings links to other approaches have to be employed. One possibility is given by the AdResS method (see sec. 4.2), where all atom simulations and coarse-grained simulations are coupled in a way that there is a free exchange of molecules and their representation. For the effect of long range electrostatic interactions the model can be linked to, e. g. a fast Poisson-Boltzmann solver [133]. However, all these extensions require the detailed understanding of the possibilities and limitations of both the underlying atomistic and coarse-grained models. That is why we will now discuss more advanced coarse-graining techniques, for water and other molecules.

3.3 Other Coarse-Graining Techniques

After we have studied different water models in the previous section, we will focus on different coarse-graining techniques. First, we will come back to water, as it is from the coarse-graining point of view one of the easiest models one can imagine. Nevertheless, it is much too simple to understand

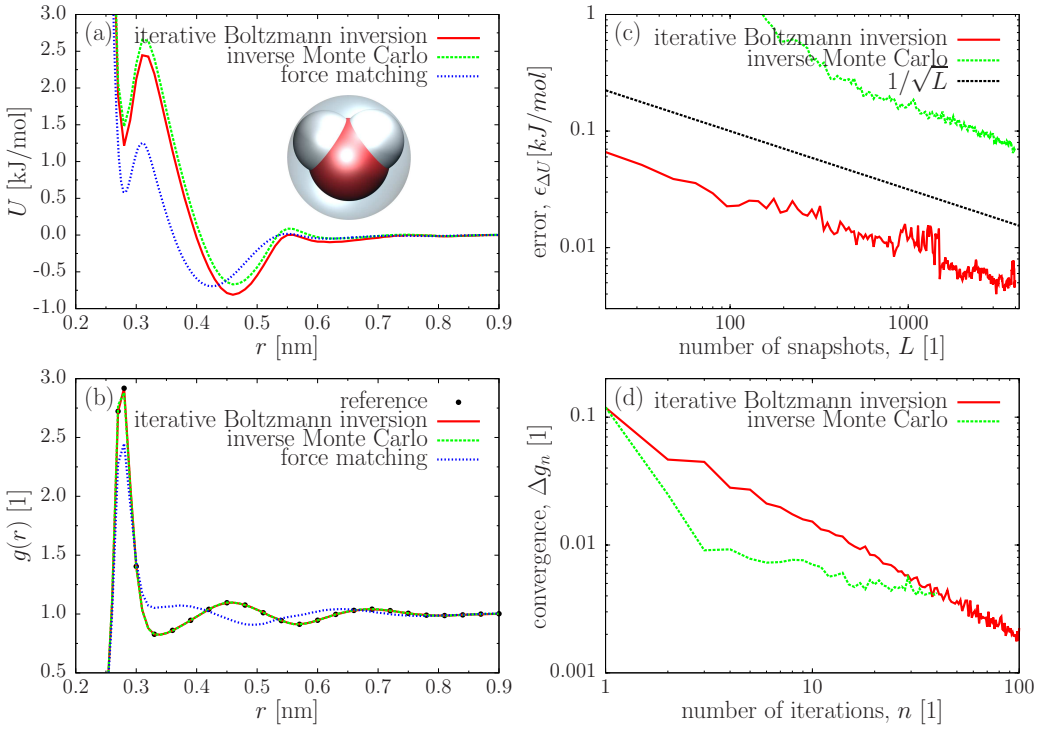

Figure 3.8: **Coarse-grained SPC/E Water**: (a) Coarse-grained potentials for SPC/E water obtained using different coarse-graining techniques. (b) Corresponding radial distribution functions. (c) Average error of the potential update function versus number of snapshots used for calculating the update function. (d) Root mean square deviation of reference and current radial distribution function versus iteration step. One can see that IMC converges faster than IBI. Inset of (a) shows the coarse-grained representations of a single water molecule, the parameters used can be found in tab. 3.1.

some effects going on in coarse-graining that is why we will also take a look at methanol, propane and hexane.

3.3.1 Comparative coarse-graining of SPC/E Water

In this section we coarse-grain the SPC/E water model again using different coarse-graining techniques, in principal every model from the previous section could be used. But we have chosen to study the SPC/E water model again due to the fact that it exhibits the most structured pair distribution function of all previous studied rigid water models (see fig. 3.1) and we will later use the SPC/E model in an adaptive resolution simulation to solvate fullerenes (see sec. 4.2). The corresponding parameters of this 3-site model are given in tab. 3.1. In contrast to the previous section, we used a one-site coarse-grained representation with a pair potential $U(R_{ij})$, where R_{ij} connects the

centers of mass of water molecules i and j and *not* the oxygens.

The all-atom system consisting of 2180 water molecules was first equilibrated in the NPT ensemble at 300 K and 1 bar for 100 ns using the Berendsen thermostat and barostat [42]. The last 80 ns were used to determine the equilibrium box size of 4.031 nm, which was then fixed during the 45 ns production run in the NVT ensemble using a stochastic dynamics algorithm [134]. Later it turned out that the change of the thermostat, from Berendsen to stochastic dynamics (a Langevin thermostat), will lead to a pressure in NVT simulation slightly higher than 1 bar. However, as the pressure of coarse-grained simulations is anyway to high this is a minor effect. For all further analysis, only the last 40 ns were used. The radial distribution function was calculated using a 0.01 nm grid spacing. The snapshots were output every 0.4 ps.

Force matching potentials were calculated using blocks of 6 snapshots each. Spline grid spacing of 0.02 nm was used in the interval from 0.24 to 1 nm. For the iterative procedures, the potential of mean force was taken as an initial guess for the interaction potential. The coarse-grained box had the same system size as in the atomistic simulations. Simulations of the coarse-grained liquid were done using a stochastic dynamics algorithm [134]. 300 iterations of 100 ps each were performed when using IBI. For IMC we used 10 iterations of 500 ps each. Additionally, two iterations of triangular smoothing were applied to the IMC potential update, ΔU. The cut-off was chosen at 0.9 nm with a grid spacing of 0.01 nm.

The reference radial distribution function, $g^{\mathrm{ref}}(r)$, coarse-grained potentials and corresponding radial distribution functions are shown in fig. 3.8a,b. IBI and IMC give practically the same interaction potential. Although the force-matched potential has a very similar structure with two minima, the corresponding radial distribution function is very different from the target one. Possible reasons for these discrepancies are discussed in the literature [79, 125, 73] and stem from the fact that FM aims to reproduce the many-body potential of mean force, which does not necessarily guarantee perfect pairwise distribution functions, considering the fact that the basis sets in the coarse-grained force field may be limited.

Note that all three methods lead to a different pressure of the coarse-grained system: 8000 bar (IBI), 9300 bar (IMC), and 6500 bar (FM). Different pressures for the iterative methods are due to different accuracy of the potential update. Indeed, small changes of pressure can significantly affect the potential, especially its long tail [64, 81]. However, they hardly change the radial distribution function due to the small compressibility of water. One can improve the agreement between the iterative methods by using pressure correction terms for the update as we have done in the previous section.

The performance of the iterative methods depends on two factors: (i) the

average (over all bins) error of the potential update $\epsilon_{\Delta U}$ and (ii) the number of iterations required for convergence. We define the average error as

$$\epsilon_{\Delta U} = \frac{1}{N} \sum_{i=0}^{N} \epsilon(\Delta U(r_i)) \ , \tag{3.7}$$

where N is the number of bins and $\epsilon(\Delta U(r_i))$ is the error of the update function at a separation r_i. $\epsilon(\Delta U(r_i))$ was calculated using a Jackknife analysis [135].

The average error of the potential update is shown in fig. 3.8c as a function of the run length. One can see that, for both methods, the error decreases as $1/\sqrt{L}$, where L is the number of snapshots used for averaging. However the prefactor for the IBI update error, which is based on the radial distribution function, is at least ten times smaller than for the IMC update error, which makes use of cross-correlations of S_α. This observation implies that, in order to have the same accuracy of the update function, IMC needs significantly longer sampling.

This disadvantage is of course compensated by the efficiency of the update function, which is assessed by computing the root mean square deviation, Δg_n, of the current and target radial distribution functions

$$\Delta g_n^2 = \int \left[g^{\mathrm{ref}}(r) - g^{(n)}(r) \right]^2 \mathrm{d}r \ . \tag{3.8}$$

Δg_n is plotted as a function of the number of iterations, n, in fig. 3.8d. It is clear that IMC converges much faster than IBI, though the root mean square deviation saturates after some number of iterations.

3.3.2 Comparative coarse-graining of Methanol

Liquid methanol (see the inset in fig. 3.9) is the second example of coarse-graining of *non-bonded* interactions that is presented here. In fact, FM has already been used to coarse-grain this system [81] and, contrary to water, the liquid structure (radial distribution function) is well reproduced by the FM coarse-grained potential. In addition, the excluded volume of methanol is larger than that of water and the undulations of the radial distribution function extend up to 1.5 nm. As we will see, this leads to pronounced finite size effects for IMC, since it has a non-local potential update. FM and IBI do not have this problem, since the IBI potential energy update is local and FM is based on pair forces. The range of the latter is much shorter than the correlation length of structural properties, which may propagate over the

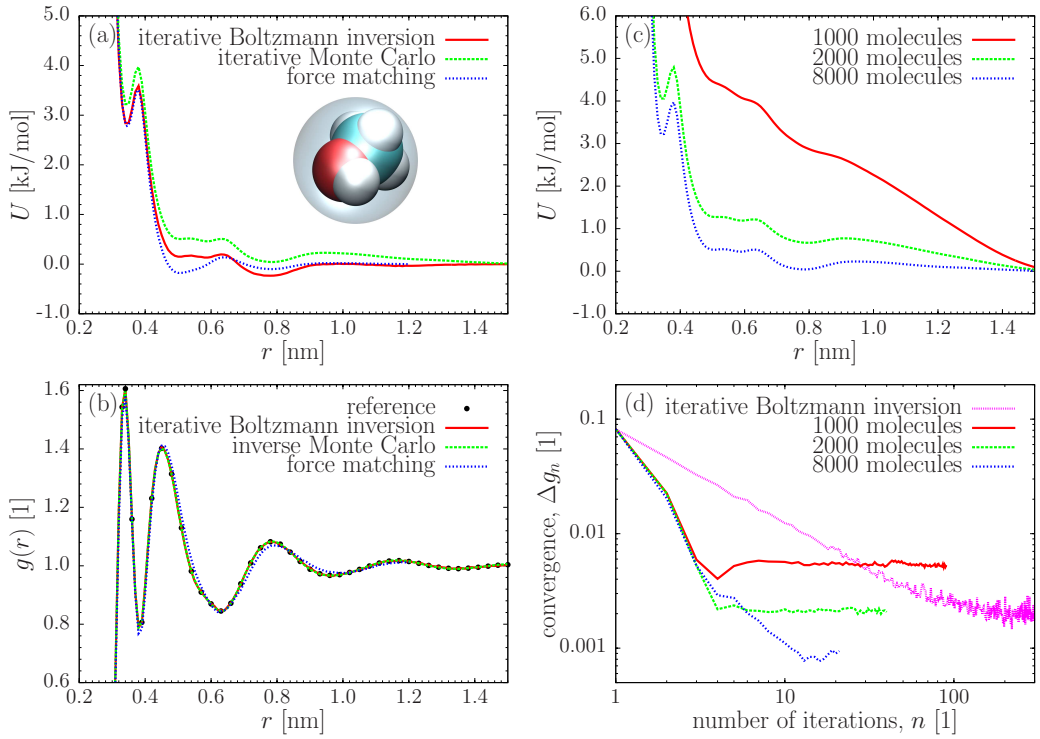

Figure 3.9: **Coarse-grained methanol**: (a) Coarse-grained potentials for methanol using different coarse-graining techniques. (b) Corresponding radial distribution functions. (c) Coarse-grained potentials using 10 IMC iterations for simulation boxes with 1000, 2000 and 8000 methanol molecules (cubic box with length 4.09 nm, 5.15308 nm, and 8.18 nm) equilibrated at the same density. (d) Root mean square deviation of reference and the current radial distribution function versus number of iterations. Similar to liquid water, IMC converges faster than IBI. The convergence saturates and the saturation error strongly depends on the system size. The inset of (a) shows the Van der Waals excluded volume and coarse-grained representations of a methanol molecule.

boundaries for small boxes. Such correlations could be undulations of the radial distribution function.

Simulation setup parameters were taken from the literature [81] and OPLS [136] all-atom force field was used. Atomistic simulations were performed with 1000 methanol molecules in a cubic box (4.09 nm box size) at 300 K using the Nose-Hoover thermostat [137, 138]. The system was equilibrated for 2 ns followed by a production run of 18 ns. The reference radial distribution function was calculated using snapshots every 0.5 ps and is shown in fig. 3.9b.

The FM potential was calculated using blocks of 6 frames each and a spline grid of 0.02 nm. With this potential, coarse-grained simulations were performed using a stochastic dynamics integrator, 1000 beads and with the same box size and the same temperature as in the atomistic simulations. The system was equilibrated for 40 ps followed by a production run of 160 ps. Snapshots were stored every 5 ps and used to calculate the radial distribution function.

For the iterative procedures, the potential of mean force was taken as an initial guess. The cut-off was chosen at 1.54 nm with a grid spacing of 0.01 nm. For IBI, 300 iterations were performed using stochastic dynamics with the same parameters used in the FM-based procedure. The IMC iterations were performed with 8000 molecules and a box size of 8.18 nm. The total length of the run was 1 ns and snapshots were stored every 0.2 ps. Two smoothing steps were used in each iteration step for the potential update, ΔU.

The coarse-grained potentials for all methods are shown in fig. 3.9a. In spite of small differences between the coarse-grained potentials, the agreement between the reference and coarse-grained radial distribution functions is excellent, as can be seen in fig. 3.9b.

It is important to mention that the IMC method, which has a non-local update, is prone to systematic errors due to finite size effects and hence requires much larger simulation boxes in order to calculate the potential update. This is due to artificial cross-correlations of S_α at large distances, which lead to a small difference of tails between the coarse-grained and the reference radial distribution functions and as a consequence, to a much higher pressure of the coarse-grained system and a significantly different coarse-grained potential. In contrast, IBI and FM work well with system sizes in the order of two radial distribution function cut-off lengths.

To illustrate this point, we prepared simulation boxes of three different sizes, with 1000, 2000 and 8000 methanol molecules The simulations were performed with cubic boxes with box lengths of 4.09 nm, 5.15308 nm and 8.18 nm and simulation times of 3 ns, 2 ns and 1 ns, respectively. The IMC

iterative procedure was repeated until the potentials converged, and these are shown in fig. 3.9c. One can see that the potentials significantly differ from each other. These differences lead to small deviations in the tail of the radial distribution function, which, however, vanish, in a systematic way, for bigger boxes, as illustrated in fig. 3.9d, where we plot the integral of the difference of the reference and current distribution functions.

A more detailed analysis has shown that, for small boxes, an additional linear term in the potential update at large separations appears. To determine the origin of this term, ΔU was calculated using the full matrix $A_{\alpha\beta}$ as well as only its diagonal elements. The potential after 50 IBI iterations was taken as an initial guess. Without the off-diagonal elements ΔU was small once the reference and coarse-grained radial distribution functions were matching each other. Inclusion of the off-diagonals elements always resulted in an additional, practically linear, term in the potential update, which became smaller for large boxes. Based on this observation we concluded that the off-diagonal elements of the matrix $A_{\alpha\beta}$ systematically change with the box size.

To summarize, IMC should be used with care for small systems. The potential update (or the coarse-grained potential) must be converged with respect to the simulation box size. In the case of methanol coarse-graining, a box of size three times the radial distribution function cut-off was not enough to achieve the converged potential for IMC, even though this is sufficient for IBI and FM methods.

3.3.3 Comparative coarse-graining of Liquid Propane

So far we have illustrated coarse-graining of *non-bonded* degrees of freedom using liquid water and methanol as examples. Here we show how *bonded* interactions can be coarse-grained by deriving a united atom model, i. e. hydrogens embedded into heavier atoms, from an all-atom model of liquid propane. The united atom model we use here shall not be confused with the united atom models commonly used in the atomistic force-field community, for example OPLS-UA force-field [136]. The latter models map the potentials, which are analytical functions of bonds, angles, and dihedral angles, onto thermodynamic properties of the corresponding substances. In our case coarse-grained potentials are tabulated functions of coarse-grained variables and only the mapping, embedding hydrogens into heavier atoms, is similar to that of the united atom force-fields. The mapping scheme, as well as the bonded coarse-grained variables, in this case two bonds, b, and one angle, θ, are shown in the inset of fig. 3.10. Note that this coarse-graining scheme has two different bead types: an inner bead, of type B, with two hydrogens, and

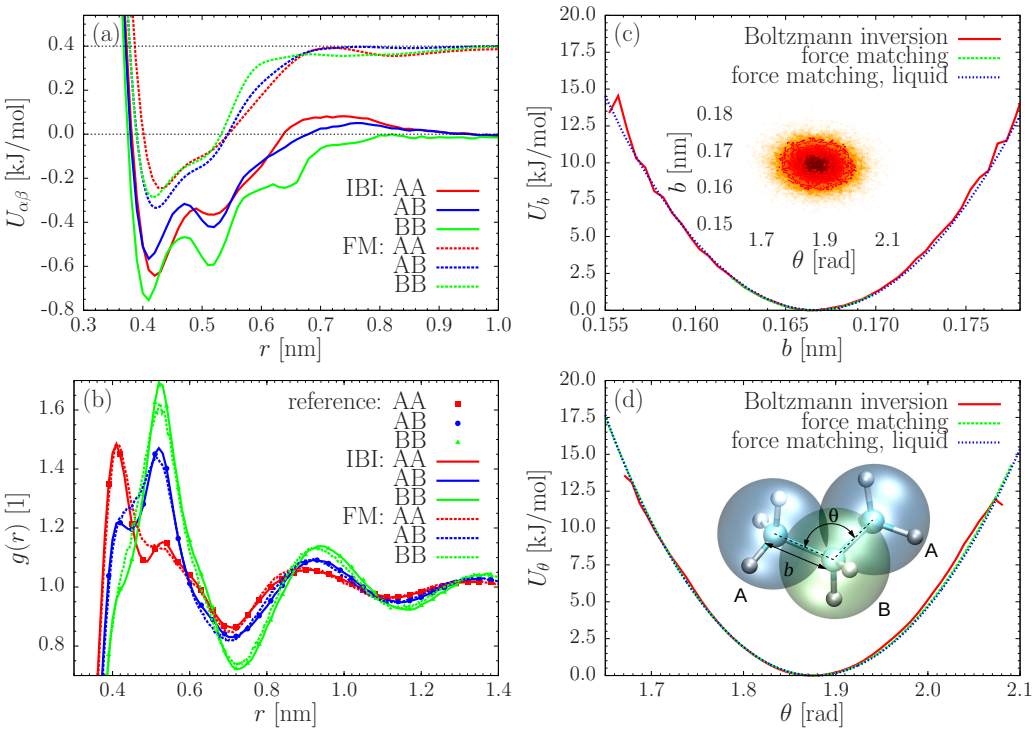

Figure 3.10: **Coarse-grained propane**: (a) Non-bonded interaction potentials U_{AA}, U_{BB}, and U_{AB} for propane obtained with IBI and FM methods. For clarity, FM potentials are offset along the y axis. (b) Corresponding radial distribution functions, plotted together with the atomistic radial distribution function. (c) Bond potential obtained for a single molecule in vacuum by Boltzmann-inverting the corresponding distribution function, using FM for a single propane molecule in vacuum and force matching for liquid propane. (d) Angular coarse-grained potentials. FM-based distributions for a single molecule and the liquid are on top of each other. The inset of (c) shows the correlations of b and θ. The inset of (d) shows all-atom and coarse-grained representations of a propane molecule, bead types, and coarse-grained bonded degrees of freedom (bond b and angle θ).

two outer beads, of type A, with three hydrogens. As a result, three types of non-bonded interactions, U_{AA}, U_{BB}, and U_{AB} must be determined.

The atomistic simulations were performed using the OPLS all atom force field [136]. A box of liquid propane was first equilibrated at 200 K and 1 bar in the NPT ensemble for 10 ns using the Berendsen thermostat and barostat [42]. The equilibrated box of the size $4.96337 \times 5.13917 \times 4.52386\,\mathrm{nm}^3$ was then simulated for 10 ns in the NVT ensemble at 200 K using velocity rescaling with a stochastic term [43]. No bond constraints were used during the simulations and hence the integration time step was 1 fs. Snapshots were written every 1 ps.

In the case of iterative methods, the bonded potentials (bond and angle) were calculated by Boltzmann-inverting the corresponding distribution functions of a single molecule in vacuum. The potentials have to be normalized with the right volume factor (see eq. 2.90). The propane molecule in vacuum was simulated in an stochastic dynamics run [134] for 100 ns with snapshots stored every 2 ps. Non-bonded potentials were iteratively refined by using IBI with a grid spacing of 0.01 nm and a cut-off of 1.36 nm for A-A, A-B and 1.38 nm for B-B interaction types, respectively. The run length for each iteration was 50 ps with snapshots written every 0.5 ps. At every iteration step only one interaction type was corrected. When using the FM method, both bonded and non-bonded potentials were obtained at the same time, since FM does not require the explicit separation of bonded and non-bonded interactions.

The obtained potentials are shown in fig. 3.10a,c,d. FM and Boltzmann inversion-derived bond and angle potentials (fig. 3.10c,d) perfectly agree with each other. The non-bonded potentials, shown in fig. 3.10a, are not completely identical, but have similar shapes and barrier heights. This of course results in a good reproducibility of the propane liquid structure by the FM-based coarse-grained potentials, as can bee seen from the radial distribution functions shown in fig. 3.10b. Again, as expected, IBI reproduces the reference radial distribution functions exactly.

To summarize, the united atom model of liquid propane is an ideal example of coarse-graining where the structure- and force-based methods result in similar bonded and non-bonded interaction potentials. As we will see later, this is due to (i) the completeness of the basis set used to construct the coarse-grained force-field; and (ii) independence of bond and angular degrees of freedom. The latter can be understood with the help of a histogram showing the correlation of b and θ, depicted in the inset of fig. 3.10c. In the next section we will look at coarse-graining of a single molecule of hexane, for which this completeness of the basis set is not the case.

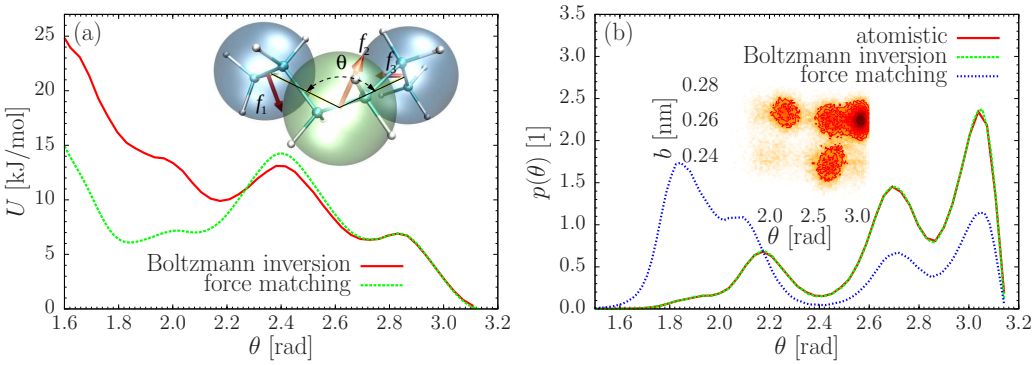

Figure 3.11: **Coarse-grained hexane**: (a) Coarse-grained angular potentials obtained using Boltzmann inversion (no iterations) and FM for a single hexane molecule in vacuum. The inset of (a) shows a hexane molecule and its coarse-grained representation. Arrows indicate the directions of the forces on three beads for a specific snapshot. (b) Probability density (probability distribution normalized by the interval) obtained from the atomistic run, as well as from the runs using coarse-grained angular potentials. The inset of (b) shows the correlation of b and θ.

3.3.4 Angular Potential of a Hexane Molecule

Another example we would like to discuss shortly is the angular potential of a hexane coarse-grained into a three-bead chain, with two carbon atoms per bead (see the inset in fig. 3.11a). A detailed analysis of the coarse-grained potentials of hexane can be found in the thesis of V. Rühle [86]. Atomistic simulations of a single hexane molecule in vacuum were performed using the all-atom OPLS force field and a stochastic dynamics integrator [134]. The run length was 1000 ns and snapshots were stored every 2 ps.

The coarse-grained angular potential was again obtained by Boltzmann-inverting the angular distribution function or by using the FM method. In the case of FM, we used blocks of 50000 frames each, a spline grid of 0.05 nm and sampling in the $\theta \in [1.6, 3.14]$ interval. Both coarse-grained potentials are shown in fig. 3.11a. The corresponding distribution functions, together with the reference function obtained from the atomistic simulations, are shown in fig. 3.11b.

It is obvious that the distribution, which corresponds to simple Boltzmann inversion, is practically identical to the reference distribution, while the FM-based distribution samples small angles much more often, which is a direct consequence of a very deep local minimum in the angular potential at these angles. It is easy to understand why FM fails to predict the relative height of this minimum. On a coarse-grained level the change of the angle from large to small values corresponds to the reorientation of the dihedral angles at the atomistic level. This reorientation results in instantaneous forces, $\vec{f}_1$,

$\vec{f}_2$, $\vec{f}_3$, on the beads, which have an *out of plane* component, where the plane is defined by the centers of the beads (see also the inset of fig. 3.11a). The coarse-grained potential, however, has only an angular term, U_θ, and hence can only capture forces, which lie *in the plane* in which the angle θ is defined. Hence, only the *projections* of the forces on this plane are used in FM, and this clearly leads to underestimation of the position of the second minimum, since the work conducted by the out-of-plane forces is completely ignored. For condensed phase systems, the error introduced by the off-plane component of the force might be compensated by some other pair interactions. In this particular case, however, coarse-graining of liquid hexane with both bonded and non-bonded degrees of freedom treated simultaneously results in a very similar angular distribution to that of a single molecule in vacuum [86].

Additionally, this mapping scheme does not have independent variables, e. g. bond and angle degrees of freedom are coupled, as can be seen from the Ramachandran plot shown in the inset of fig. 3.11(b). This means that, even though Boltzmann inversion reproduces correct distributions, sampling of the configurational space is incorrect because of the lack of cross-correlation terms in the coarse-grained potential.

This example clearly shows that coarse-graining shall be used with understanding and caution, the methods should be cross-checked with respect to each other, as well as with respect to the reference system.

3.3.5 Summary

The fact that different coarse-graining techniques turn out to be best suitable for specific systems is due to their different theoretical concepts, as well as the underlying assumptions. Generally it appears that the simplest method, iterative Boltzmann inversion, is the most reliable one. This has several reasons, first of all because correlations between different distances and particle types are neglected, much less statistics is needed and artificial correlations due to finite size effects in system size and sampling time can not occur. Second, the simplicity of the method allows for a better tuning of the iterative process. Nevertheless, for coarse-grained simulations, in which not only the structural properties are important, force matching can be more appropriate. Also due to the fact that for more complicated systems the iterative nature of some of the techniques can lead to a much higher consumption in computer-time. Nonetheless, none of the techniques allows to have a real influence on the change of the time scale of the coarse-grained system, that is why we will discuss in the next section how to do this again using water as an example.

3.4 Dynamics of Coarse-Grained Water

Influencing the dynamics of a coarse-grained system can be important when modeling diffusive processes in AdResS simulations [139] and to obtain consistent fluxes in triple-scale simulations [140]. In this section we show how this is done for coarse-grained TIP3P water, but this mechanism can be applied to all one-site coarse-grained models.

3.4.1 Atomistic Model

All-atom water NVT simulations at ambient conditions are performed using a rigid TIP3P water model with a bond angle of 112.19° [124]. The electrostatics is described by the reaction field method [141, 142], in which all molecules outside a spherical cavity of a molecular based cut-off radius $R_c = 9\,\text{Å}$ are treated as a dielectric continuum with a dielectric constant $\epsilon_{RF} = 80$. Here, in order to reproduce the hydrodynamics correctly we employed the DPD thermostat (see sec. 2.2.5) with the friction constant $\zeta^{\parallel} = 0.038\,\text{ps}^{-1}$ and cut-off radius as for the reaction field. The constant $\zeta^{\parallel}$ is small compared to the intrinsic friction coefficient ξ of the TIP3P water system, i. e. $\xi = 288.6\,\text{ps}^{-1}$, so that this stochastic approach generates a dynamics comparable to NVE [139]. Basically, the setup of the water simulation was exactly the same as in [98] that is why the TIP3P model used here had an angle of 112.19° [124], which makes it, as mentioned in sec. 3.2, a *modified* TIP3P model. We did not correct the bond angle due to the fact that the same setup was later used in [143].

3.4.2 Coarse-grained Model

For the coarse-grained water simulations with the Transverse DPD thermostat we have employed the single-site water model from [123], which reproduces essential thermodynamics and structural properties, e. g. the pressure, density, and radial distribution functions, of the all-atom rigid TIP3P water model at standard conditions. Other simulation details are the same as given in [123].

3.4.3 Results

Having shown that the new thermostat enables us to tune the diffusion constant and viscosity of a simple liquid (see sec. 2.2.8) we apply it to a slightly more important biophysical example, i. e. liquid water at ambient conditions.

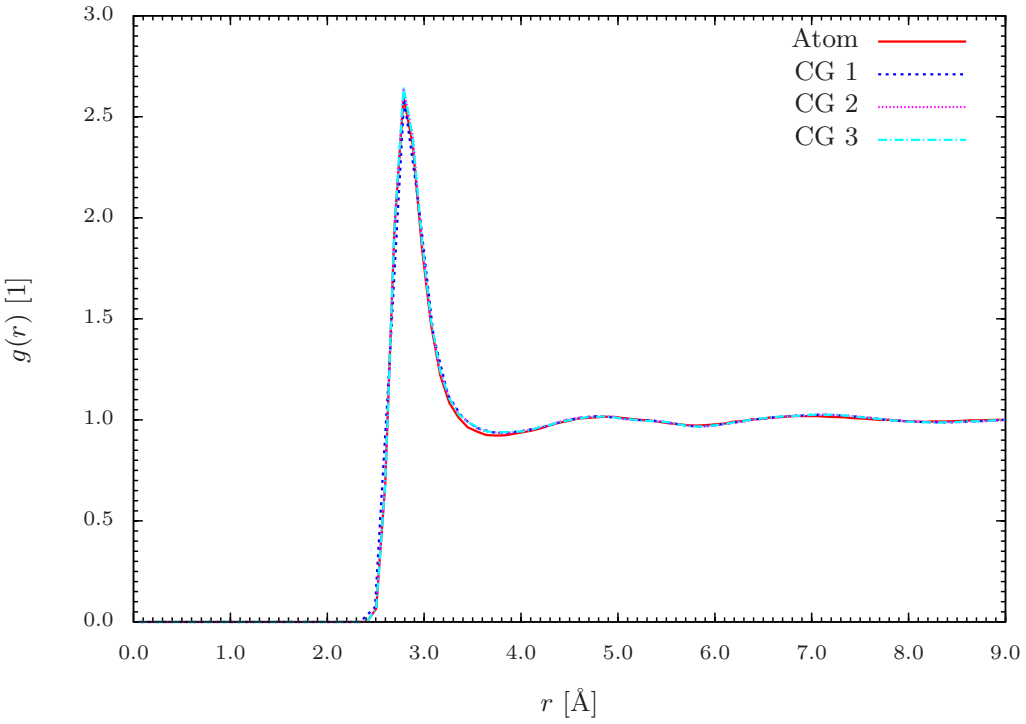

Figure 3.12: **Radial distribution functions of the modified TIP3P water**: The center of mass radial distribution functions of the all-atom ($\zeta^{\parallel} = 0.5\,\mathrm{ps}^{-1}$, $\zeta^{\perp} = 0$) and several coarse-grained simulations ($\zeta^{\parallel} = 0$ and $\zeta_1^{\perp} = 0.5\,\mathrm{ps}^{-1}$, $\zeta_2^{\perp} = 0.75\,\mathrm{ps}^{-1}$ and $\zeta_3^{\perp} = 1.0\,\mathrm{ps}^{-1}$).

We first check that the structural properties do not depend on the thermostat and also that we obtain consistency between the coarse-grained and atomistic simulations. The center of mass radial distribution functions of the all-atom and coarse-grained system using different values of $\zeta^{\perp}$ match within the line thickness (see fig. 3.12).

There is an intrinsic time scale difference in the diffusive dynamics of the coarse-grained water system, because of the reduced number of DOFs and the much smoother interactions. For this reason the self-diffusion constant for the coarse-grained water model is larger than the corresponding all-atom one. It is important to note that the diffusion dynamics depends on the used thermostat and its friction constant that is why a speedup between 2, for a Langevin thermostat [123], and 4, for a Berendsen thermostat (see sec. 3.2), is possible.

Due to correct hydrodynamics the standard DPD thermostat makes the diffusive dynamics less dependent on the friction constant. However, in this case we want to retain the diffusive dynamics of the all-atom model and just

change the one of the coarse-grained model. We used $\zeta^{\perp} = 0.75\,\mathrm{ps}^{-1}$ for the Transverse DPD thermostat ($\zeta^{\parallel} = 0$) to match the diffusion constant of the coarse-grained water model to the corresponding value $D = 3.2 \cdot 10^{-5}$ cm^2/s (see fig. 3.13) obtained from the all-atom simulation with the standard DPD thermostat ($\zeta^{\parallel} = 0.038\,\mathrm{ps}^{-1}$). One may raise the question, why the friction constant in the atomistic simulation is an order of magnitude smaller than the one in the coarse-grained simulation? But this is just due to the fact that the diffusive dynamics is nearly independent of the friction constant $\zeta^{\parallel}$ of the standard DPD thermostat, and so this value of $\zeta^{\parallel}$ in atomistic simulation is somewhat arbitrary and the diffusion constant will not decrease dramatically, if we increase the friction constant by a factor of 10. The value of $0.038\,\mathrm{ps}^{-1}$ was just chosen due to the reference simulation related to [139].

With a friction strength of $\zeta^{\perp} = 0.6\,\mathrm{ps}^{-1}$, which is, as desired, very close to the above value for matching the diffusion constants, we were also able to match the viscosity to the desired value $\eta = 0.5 \pm 0.1 \cdot 10^{-3}\,\mathrm{Pa} \cdot \mathrm{s}^{-1}$ for the TIP3P water model from our atomistic simulation with standard DPD thermostat ($\zeta^{\parallel} = 0.038\,\mathrm{ps}^{-1}$). In this case the characteristic times for the atomistic model are much longer than the time scale of the shearing ($1/\dot{\gamma}$). Therefore, even with a shear rate of $\dot{\gamma} = 0.01\tau^{-1}$, where τ is 1.579 ps, we are in the no shear limit and hence no extrapolation is required. The obtained diffusion constants and viscosities are in good agreement with the published data [144]. The difference in transverse friction constant $\zeta^{\perp}$ for fitting the diffusion constants and viscosity can be explained by the fact that the viscosity was measured in a NEMD simulation while the diffusion constant in equilibrium simulations. The shearing can lead to extra rotation of the particles and takes over some work of the thermostat, which corresponds to a smaller friction constant. Nevertheless, the better approach would be to study the viscosity in the Green-Kubo picture [53, 54] by looking at the correlation of the stress tensor.

3.4.4 Summary

By employing the new thermostat one can reproduce both the structure and the diffusive dynamics of the all-atom liquid water with the single-site coarse-grained water model. This is essential if one wants to synchronize the time scales of the all-atom and coarse-grained regimes in the adaptive resolution scheme (see sec. 2.4).

We have shown that this Galilean invariant local thermostat can be used in coarse-grained simulations to tune the diffusion constant and viscosity of the system to the desired values. This can be done in an iterative fashion using the iterative framework of VOTCA (see sec. 2.3.6.2). With this ex-

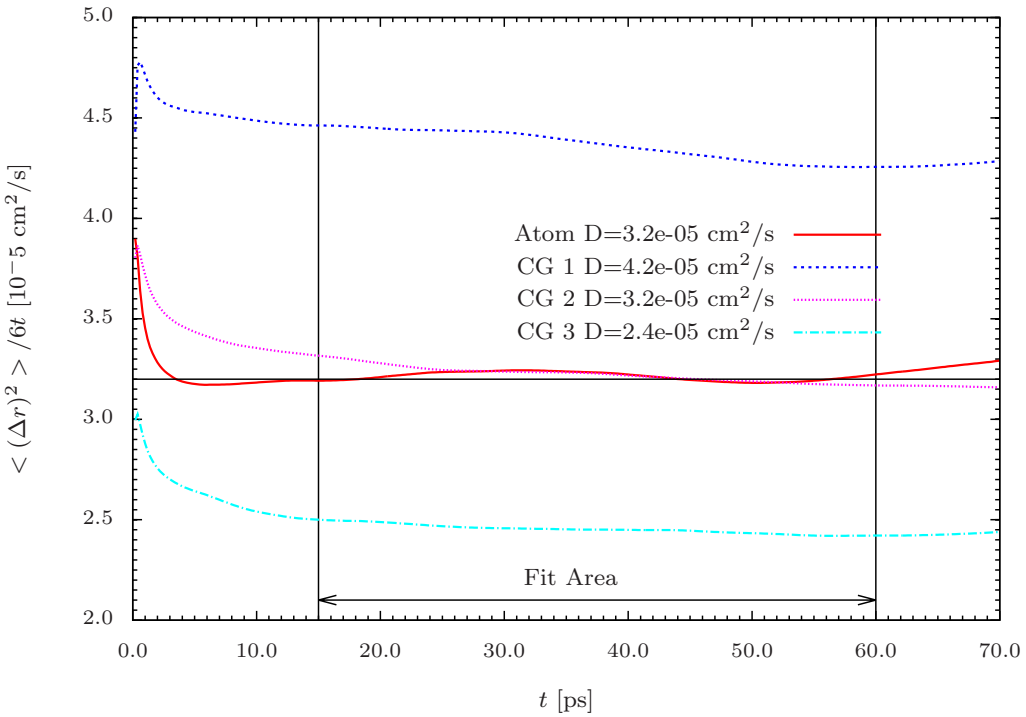

Figure 3.13: **Diffusion process for different friction constants**: The mean square displacements over time plot of the all-atom ($\zeta^{\parallel} = 0.038\,\mathrm{ps}^{-1}$, $\zeta^{\perp} = 0$) and several coarse-grained simulations ($\zeta^{\parallel} = 0$ and $\zeta_1^{\perp} = 0.5\,\mathrm{ps}^{-1}$, $\zeta_2^{\perp} = 0.75\,\mathrm{ps}^{-1}$ and $\zeta_3^{\perp} = 1.0\,\mathrm{ps}^{-1}$).

tended DPD thermostat one has the possibility of reproducing the diffusive dynamics of the atomistic system in a coarse-grained simulation.

Chapter 4

Adaptive Resolution Simulations

After we have discussed properties of several coarse-grained water models we will put them to use in the adaptive resolution scheme. Since AdResS is a rather complicated method and has a relatively short history of development we will discuss some technical aspects of the tetrahedral liquid, which was devised as a first test system for AdResS.

The AdResS method [21] is a scale-bridging method used in molecular simulations and soft matter studies. So far it has been applied to dense systems of the tetrahedral liquid [93] and to a rather small system of liquid TIP3P water [123, 139]. It has been used to couple atomistic to continuum simulations via an intermediate coarse-grained level, thus combining three scales [140]. The way it is implemented in ESPResSo is described in sec. 2.4.6.

4.1 Tetrahedral Liquid

A fluid made of tetrahedral molecules [21] is the most suitable case to demonstrate AdResS itself, the interface correction, and the action of thermodynamic forces.

The system is composed of 2520 molecules, each consisting of four atoms. All atoms interact according to a purely repulsive, shifted 12-6 Lennard-Jones potential:

$$U_{LJ}^{ex}(r_{i\alpha j\beta}) = \begin{cases} 4\epsilon \left[(\sigma/r_{i\alpha j\beta})^{12} - (\sigma/r_{i\alpha j\beta})^{6} + \frac{1}{4} \right] & : r_{i\alpha j\beta} \leq 2^{1/6}\sigma \\ 0 & : r_{i\alpha j\beta} > 2^{1/6}\sigma \end{cases} , \quad (4.1)$$

while the atoms within a molecule are bonded by an additional attractive

finitely extensible nonlinear elastic (FENE) potential:

$$U^{ex}_{FENE}(r_{i\alpha j\alpha}) = \begin{cases} -\frac{1}{2}kR_0^2 \ln\left[1-(r_{i\alpha j\alpha}/R_0)^2\right] : r_{i\alpha j\beta} \leq R_0 \\ \infty \qquad\qquad\qquad\qquad\qquad\quad : r_{i\alpha j\beta} > R_0 \end{cases} , \tag{4.2}$$

where i and j are the atoms belonging to molecules α and β, respectively. For this case we chose $\sigma = \epsilon = 1$, $k = 30\epsilon/\sigma^2$ and $R_0 = 1.5\sigma$. The molecule in the high resolution we call "explicit". In the coarse-grained representation the molecule is replaced by a spherical bead, which interacts via a tabulated interaction.

The box dimensions, in units of the fundamental length σ, are $36\times20\times20$, split in the x direction. The explicit and the coarse-grained region have the same size of 6 reduced units. To show the influence of the force interpolation on the density, the size of the hybrid region was chosen twice the size of the explicit one, $d_{hy} = 12$.

On the contrary, the coarse-grained interaction consists of a one-site tabulated potential able to reproduce satisfactorily the radial distribution function and the state point of the liquid [21].

For a comparison of the basic structure of the simulations, fig. 4.1 shows the radial distribution functions for the centers of mass of the molecules in the explicit and AdResS simulations (normal, interface corrected and under the effect of the thermodynamic force).

The small deviations at the peaks of the distribution can be removed by improving the coarse-grained potential of the tetrahedral liquid. However, as the deviations are less than 5% one basically sees that the structure is not influenced by the different correction methods. This is clear due to the fact that the coarse-grained and explicit systems already have the right radial distribution and only the distribution in the hybrid zone is violated. For that reason it is more suitable to look at the density profile.

The density profile is defined as the number of particles per volume, where we split the volume in the same direction as the resolution with AdResS (x). In fig. 4.2 we show the density profiles of the simulations for three cases: (a) simple AdResS interpolation, (b) AdResS with interface-pressure correction and (c) the same system under the effect of the thermodynamic force. The thermodynamic force was determined by calculating the excess chemical potential for several systems at constant weighting function by means of the test particle insertion method [145] in the GroMaCS package [87]. The force was obtained as the derivative of a continuous interpolation of this excess chemical potential profile. A detailed discussion about the density inhomogeneities and their elimination can be found in the literature [99].

In conclusion, the application of the thermodynamic force works best for this system that is why we are confident that it will work for water as well.

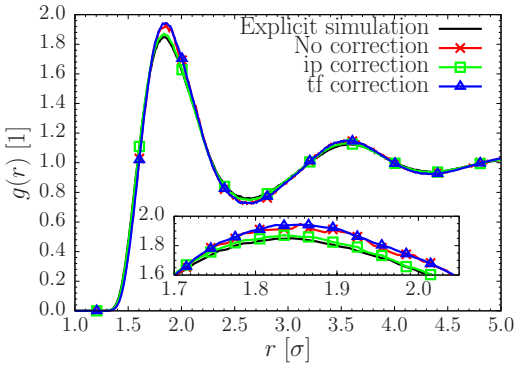

Figure 4.1: **Structure of the tetrahedral liquid**: Radial distribution functions between centers of mass for explicit and AdResS simulations of the tetrahedral liquid with several corrections. The inset shows a close-up of the first peak.

Also from the theoretical background it is clear that the thermodynamic force works better due to the fact that the interface-pressure correction corrects the structure in the middle of the hybrid zone, which is not directly related to the density profile, while the thermodynamic force acts on every molecule dependent on its position and hence is much closer related to the density profile.

The tetrahedral liquid has been chosen as a model system for AdResS and this section has shown that with our implementation in ESPResSo one can easily reproduce results of previous studies [21, 93, 94] and build the foundation of future developments in AdResS [95]. After we have seen how AdResS performed for this simple liquid, we will now apply it to liquid water.

4.2 Water

In this section we will couple one of the coarse-grained water models discussed in sec. 3.2 to the atomistic resolution using AdResS. The analysis in this section is the foundation for the study of the influence of the hydrogen bonding network on the solvation of molecules, which is presented in the next section. The water model shows several complications compared to the tetrahedral liquid, which appear already at the stage of coarse-graining:

- Water in the atomistic resolution has partial charges (long-ranged interactions)

- Bonds between Oxygen and Hydrogen in one molecule are rigid

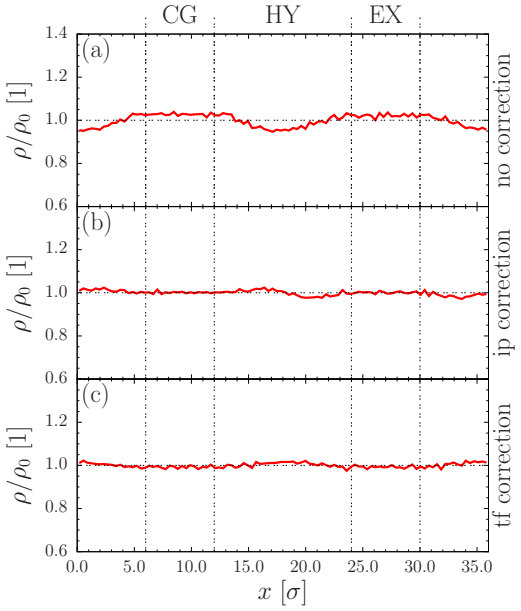

Figure 4.2: **Density profiles of the tetrahedral liquid**: AdResS simulation of tetrahedral molecule system with a hybrid region width of 12σ with (a) no correction in the hybrid region, (b) interface-pressure (ip) correction and (c) external thermodynamic force (tf) correction.

- Pressure fluctuations are much higher (in the order of 100 bar)

Let us discuss the ideas to overcome these problems first, before we come to the actual simulation setup and the conclusion we can draw from that.

The problem of charged molecules is delicate due to the fact that on the one hand electrostatic interactions are long-ranged. On the other hand in AdResS two molecules far away from each other interact as coarse-grained molecules, which are uncharged. Therefore, the question is, how to make two explicit molecules interact in a local, short-ranged, way? Due to the fact that long-range Coulomb interactions are also computationally very expensive, this problem was already considered by others [141, 142]. Hence, we can adopt the idea of the reaction field, which replaces the long-range Coulomb interaction using a mean-field approximation. In this approximation a charge experiences all charges beyond a cut-off as a dielectric background resulting in an effective short-ranged electrostatic interaction. The drawback is that the coarse-grained model (from sec. 3.2, which was obtained using the particle-mesh Ewald method) has to be reparameterized for the reaction field method, which, however, can be done easily using VOTCA (see sec. 2.3.6).

The rigid bonds within the water molecule pose a more technical difficulty, since first of all, a system with constraints (in this case rigid bonds) has to be integrated with a different class of integrators [146, 147, 148, 113]. These account for the constraint forces and make the system obeying the constraints up to a certain accuracy.

However, as these integrators are implemented in ESPResSo and GroMaCS we will not discuss their technical details here, but rather consider how we can smoothly switch the rigid interactions on and off. One approach could be to make the bonds flexible in the first step and removing them in the second step, when a molecule goes from the explicit to the coarse-grained region in AdResS and vice versa in the other direction. But from a statistical point of view this is problematic due to the fact that making a bond flexible means increasing the number of degrees of freedom. Hence, the hybrid region would have more degrees of freedom per molecule than the explicit region, which would make the transition region more complicated in addition to the fact that the switch from flexible to rigid is not trivial and involves a metric force [149, 150, 151]. Due to all these considerations and the way how the bonded interactions are implemented in the simulation package (see sec. 2.4.6.6), we have chosen the simplest transition, i. e. keeping rigid bonds in the hybrid region.

We note in passing that the pressure fluctuations at ambient conditions are about 300 bar, which makes it technically difficult to determine a pressure corrected model. It also shows how different the pressure in the same system

can be.

4.2.1 Simulations Setup

The atomistic, explicit model for water used in this study is the SPC/E model [106] at ambient conditions, whose parameterization can be found in tab. 3.1. The NVT ensemble was generated by a Langevin thermostat with $\xi = 5\,\mathrm{ps}^{-1}$ and a time step of 2 fs. The electrostatic interactions were calculated using the reaction field approximation. All cut-offs were set to 0.9 nm. Different sizes of the explicit region were studied using spherical splitting (see eq. 2.126) of the periodical simulation box of size $(4.04538\,\mathrm{nm})^3$ filled with 2180 water molecules. The spherical atomistic cavity was connected to the coarse-grained region by a hybrid zone of 1.2 nm.

In the coarse-grained region an atomistic water molecule is replaced by a sphere, which has its center on the oxygen atom. The reason for not mapping it to the center of mass was basically that we planned to take a look at the influence of flexibility on the atomistic model, however, later it turned out that this influence of the mapping point is nearly negligible, because the difference in the radial distribution functions of center of mass and oxygen mapping is less than 2%.

The coarse-grained potential was obtained with an iterative version of the inverse Monte Carlo method (see sec. 2.3.4.3) using VOTCA (see sec. 2.3.6). The potential was cross-checked with a potential obtained by iterative Boltzmann inversion (see sec. 2.3.4.1). The potentials were essentially the same, which we already found in the comparison of different coarse-graining methods (see fig. 3.8).

No correction was applied in the hybrid zone, due to the fact that we later want to study the influence of the size of the coarse-grained region on the hydrogen bonding. Nevertheless, a correction can be required in other cases and is in principle possible. The thermodynamic force $\vec{F}_{\mathrm{th}}$ (see sec. 2.4.5.4) for a box split in x-direction with a explicit region of 2 nm and a hybrid region of 3 nm can be found in fig. 4.3. This force was obtained in an iterative procedure within 5 steps using the iterative framework of VOTCA (see sec. 2.3.6.2) by S. Fritsch [152].

All simulations were performed using a modified, but freely available version of GroMaCS [102]. The algorithms are implemented in the same fashion as described in sec. 2.4.6.

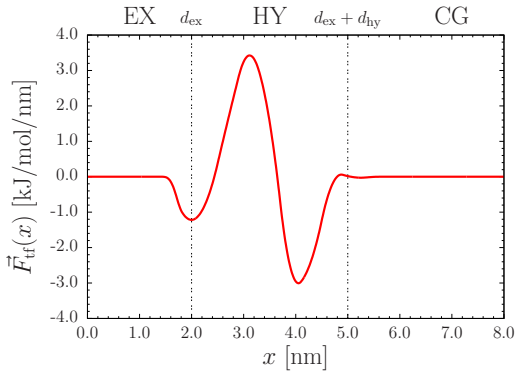

Figure 4.3: **Thermodynamic force of SPC/E coarse-grained water**: The box was split in x-direction, the explicit region and the hybrid region had a size of 2 nm and 3 nm, respectively. The force was obtained in an iterative procedure within 5 steps using the iterative framework of VOTCA (see sec. 2.3.6.2) by S. Fritsch [152].

4.2.2 Dependence on the Size of the explicit Region

We already know that the force interpolation of AdResS can lead to small density inhomogeneities (see fig. 4.2). Immediately the question arises how far inside the explicit region the structure is still perturbed by the presence of the coarse-grained region. For that we calculate the radial distribution function of a water molecule fixed in the middle of the explicit region with the other molecules. This radial distribution function is compared for several AdResS simulations, where the spherical explicit region either contains no, only the first, or all hydration shells of the atomistic structure. The distribution function is also compared to a common atomistic simulation, which is equivalent to having only the explicit region of the AdResS simulation.

From the purely atomistic simulation (see fig. 3.1) we know that the structure begins, $g(r) > 0$, at $d_{ex} = 0.2\,\text{nm}$ as well as that the first hydration shell ends at $d_{ex} = 0.35\,\text{nm}$ and that most of the structure is contained in the sphere of radius $d_{ex} = 1\,\text{nm}$ around a water molecule. These values are confirmed by the distribution function of a purely explicit simulation (see fig. 4.4), despite the fact that the statistics is in this case of course much smaller.

In fig. 4.4 the distribution functions for different sizes of the explicit region are shown. First of all one can see that with bigger explicit regions the structure tends more towards the atomistic structure, which is not surprising due to the fact that AdResS is designed to keep the structure of the atomistic model. Nevertheless because of the force interpolation this structure will be perturbed near to the hybrid region, but with a bigger explicit region

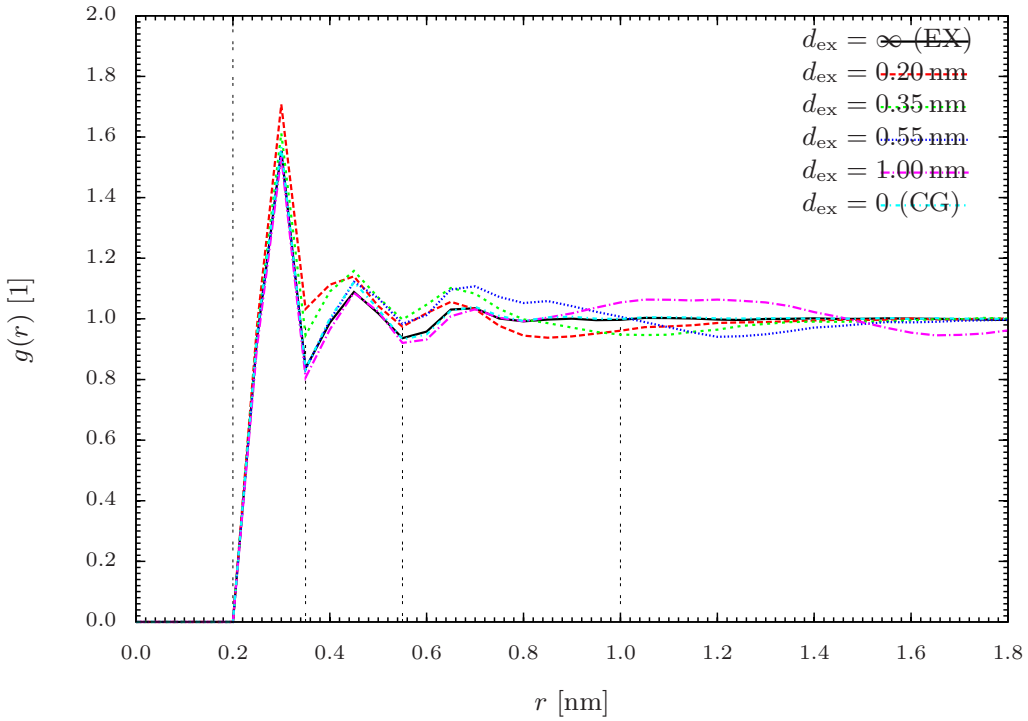

Figure 4.4: **Radial distribution functions for SPC/E water**: The size of the spherical explicit region was changed from 0.2 nm to 1.0 nm, while the hybrid region was kept at a constant size of $d_{\mathrm{hy}} = 1.2\,\mathrm{nm}$. Additionally distribution functions from the purely atomistic and the purely coarse-grained simulation are shown. The perturbation depth into the explicit region can be determined from this picture and is approximately 0.3 nm.

this perturbation will move more and more away from the center of the atomistic region. It is also interesting that the perturbation of the pair distribution functions appears also inside the atomistic region and not only in the hybrid region. This effect can be understood from the point of view that molecules near the hybrid zone still interact partly coarse-grained. This partly coarse-grained interaction follows from eq. 2.112 with $w(X_\alpha) = 1$ (molecule α is atomistic) and $w(X_\beta) < 1$ (molecule β is hybrid), which then leads to a contribution of coarse-grained interaction $\vec{F}^{\mathrm{cg}}_{\alpha\beta}$. The perturbation depth, which can be determined from fig. 4.4, is approximately 0.3 nm (a third of the interaction cut-off), which additionally depends on the splitting of box. For the one-dimensional splitting we found a perturbation depth of about 0.1 nm, which is smaller due to the fact that an atomistic molecule near the boundary to the hybrid region is encased by more atomistic molecules than in the case of spherical splitting.

4.2.3 Summary

In the last section we have shown how to apply AdResS to a water system. This setup can be now be used to study the solvation of bigger molecules in the next section or to couple the coarse-grained system to the continuum as done in the literature [143]. However, one has to keep in mind that a lot of simplifying assumptions were made. For the calculation of the electrostatics the reaction field approach was used, which is a crude approximation. Furthermore a coarse-grained water molecule is represented by a sphere, which can not represent the dipolar nature of water. And last, the compressibilities in the both resolutions are not the same. Nevertheless, with all these points in mind we will use this setup as a foundation to study the solvation of fullerenes in the next section.

4.3 Solvated Fullerenes

In this section we will extend the study from the previous section by replacing the water molecule in the center of the explicit region by a bigger molecule from the $60n^2$ series of icosahedral fullerenes. A simulation snapshot of the biggest fullerenes, C_{2160}, is shown in fig. 4.5.

Due to the fact that coarse-grained water is not capable of forming real hydrogen bonds we can study the influence of these bonds on the structure forming at the surface of the solute molecule using AdResS, which allows us to switch the coarse-grained nature on or off in certain regions.

4.3.1 Introduction

Liquid water, has a rather simple chemical structure, but is capable of forming highly complex hydrogen bond networks, which directly affect the way in which biological molecules move and function [153, 154]. In this context understanding the solvation of hydrophobic molecules is a key to understanding crucial processes occurring in biomolecular systems [155, 156].

Some limits have already been studied extensively. In general, the structure of liquid water around a solute molecule results from a delicate competition between the disruption of the local tetrahedral order of bulk water and the formation of a two dimensional surface-like order at the solute interface. For very small hydrophobic solutes, such as methane [157], water molecules can encapsulate the guest molecule, and form so called clathrate hydrates [158]. An adaptive resolution investigation comparable to this one has been done in Ref. [159].

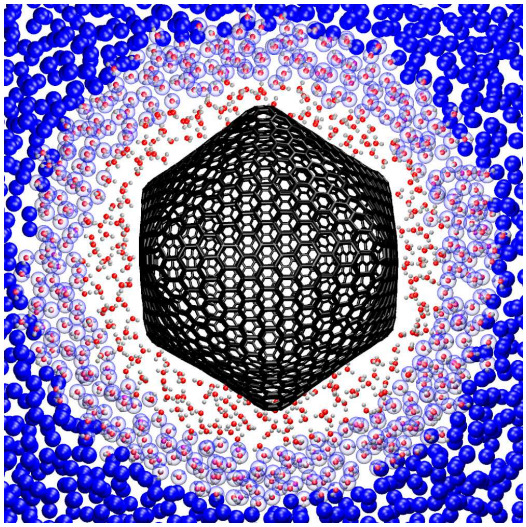

Figure 4.5: **Snapshot of a fullerene in an AdResS simulation**: Adaptive resolution simulation scheme for hydrophobic solutes illustrated for the case of C_{2160} icosahedral fullerene.

For very large solutes, the structure of water close to the solute approaches the limiting behavior of an interface with an infinite hydrophobic surface [160]. Empirical theories have been proposed, which suggest that there is a solute size-dependent critical length scale. On this scale the structure changes from a local (surface dominated) to non-local (bulk dominated) effect with a crossover length of about 1 nm. [161, 154] However, quantifying the concept of local and non-local effects and the related crossover length in experimental or computational studies remains an open challenge [162, 163].

In order to computationally address the question of the locality of the hydrophobic effect, a tool is needed that can slowly switch on and off the hydrogen bonds in a well defined region around the solute, so that their relevance for the rest of the network can be determined in an unequivocal way. The switching of the hydrogen bond degrees of freedom must occur without affecting the thermodynamic equilibrium of the whole system.

For the problem treated here, AdResS can be used to interface an atomistic model of water, which explicitly forms hydrogen bonds, with a spherical coarse-grained (CG) representation of water developed in sec. 3.2 that does not form hydrogen bonds. This allows to systematically determine the role of hydrogen bonds of the bulk onto the structure of water around hydrophobic solutes of different sizes, spanning the proposed 1 nm crossover length. It will turn out that the locality of hydrophobic hydration does not depend on the solute size, but rather on the nature of the hydrophobic solute-water in-

teraction, which exhibits local behavior for Lennard-Jones (LJ) solutes, but mixed (local/non-local) behavior for purely repulsive solutes.

If adaptive resolution simulations with a minimal explicit region can reproduce the results of fully explicit all-atom simulations for a hydrophobic solute, then the effect of the solute on water structure is deemed local. That means, it is not significantly supported by the bulk hydrogen bond network, as the bulk is modeled by CG water that is unable to form real directional hydrogen bonds (see sec. 3.2.3 and Ref [123, 139]).

Hydrophobes are typically non-polar molecules that can not form hydrogen bonds with water. The specific character of the interaction of hydrophobic molecules with water has been debated in the literature [164, 165, 166, 167]. Electronic structure calculations for benzene in water [168] suggest the use of a Lennard-Jones type potential of the C-O non-bonded interaction. Alternatively, the C-O interaction is sometimes treated as a purely repulsive interaction, and modeled with either a Weeks-Chandler-Andersen potential [169] or with a modified LJ potential with the $\left(\frac{\sigma}{r}\right)^6$ term removed [170]. It is beyond the scope of this study to assess the physical relevance of one C-O potential function over the others. However, the results we present here suggest that the choice for the C-O interaction has determining consequences on the locality of the disturbance on the hydrogen bond network produced by the solute.

4.3.2 Simulation Setup

To test the locality of the hydrophobic hydration, AdResS simulations of the $60n^2$ series of icosahedral fullerenes were performed with varying thickness d_{ex} for the layer with atomistic molecular representation. The fullerenes have effective radii of $0.35 - 2.1\,\text{nm}$, going from C_{60} to C_{2160}.

All the results presented were obtained by running NVT simulations using the same modified version of GroMaCS [102] as above, whose implementation details can be found in sec. 2.4.6.

The simulation setup was quite similar to the one of the previous section. A Langevin thermostat with a friction constant of $\xi = 5\,\text{ps}^{-1}$ and a time step of 2.0 fs was used. The volume of the system was obtained from all-atom NPT simulations with $P_{ref} = 1\,\text{atm}$ using the reaction-field method for the electrostatics, and Berendsen Barostat for pressure coupling.

In this case the reaction-field method was a rather poor approximation due to the fact that there was no water inside the fullerenes and hence the mean field assumption of the the method did not hold anymore. However no better approximation was found so far.

Atom	σ (nm)	ϵ (kJ/mol)	q (e)	m (u)
C	0.355	0.293	0.0	12.011
O	0.316557	0.650194	-0.8476	15.9994
H	0.0	0.0	0.4238	1.008
Interaction	σ (nm)	ϵ (kJ/mol)	(Lorentz-Berthelot)	
C-O	0.335779	0.436471		
C-H	0.0	0.0		
O-H	0.0	0.0		

Table 4.1: **Forcefield parameters for fullerenes and SPC/E water**: These force field parameter for the fullerenes [170] and SPC/E water model were used. Other interaction parameters were obtained using the Lorentz-Berthelot mixing rules.

Periodic boundary conditions and minimum image convention were applied in all directions, and a cut-off of 0.9 nm was used for the non-bonded interactions. After warm-up and equilibration of 1 ns, several trajectories of 2 ns were collected for each interaction and AdResS setup. The coarse-grained potential was obtained using inverse Monte Carlo simulations with the VOTCA package (see sec. 2.3.6) after this iterative procedure a pressure correction (see sec. 2.3.4.2) was applied. The coarse-grained water potential interaction is again centered in the oxygen atom, which leads to a smoother transition between the different resolutions and ensures an unique mapping point even in the case of flexible water models.

The parameters for the SPC/E water model can be found in tab. 3.1. A previously described OPLS model for buck-minster fullerene (C_{60}) was adapted for the $60n^2$ fullerenes [170], the exact force field parameters in combination with SPC/E water can be found in tab. 4.1. It is important to note that the typical energy of a hydrogen bond (on average 23.5 kJ/mol for SPC/E), is roughly 50 times larger than the optimal C-O Lennard-Jones interaction. [171] The functional form of the non-bonded potentials was:

$$U = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] & : \text{Lennard-Jones} \\ 4\epsilon \left(\frac{\sigma}{r}\right)^{12} & : \text{Purely Repulsive} \end{cases} \tag{4.3}$$

A snapshot of the computational setup is illustrated in fig. 4.5. As the fullerenes do not possess spherical symmetry, the distribution of the distances of the C atoms from the solute center of mass is spread over an interval Δr, with Δr ranging from less than 0.05 nm for C_{60} to about 0.5 nm for C_{2160}. Hence, the sizes of the explicit region were chosen according to the sizes of the hydration shells, which are the positions of these minima in the distribution

function, determined in a purely atomistic simulation and can be found in tab. 4.2. The width of the hybrid region d_{hy} was 1.2 nm in order to smoothly couple the layer of explicit water around the solute to the bulk CG water region. Small inhomogeneities in the density profile are observed, but no correction was applied due to the fact that they do not appear to affect the results discussed here. As the explicit region is still a sphere there is an edge thinning effect around the solute, which is not considered here.

The sizes of the hydration shells lead to a nearly constant relative thickness $d_{\mathrm{ex}}^{\mathrm{rel}}$ of the explicit zone, when measuring from the surface of the solute (see last line of tab. 4.2). But as the absolute size of the hydration shells of the Lennard-Jones and purely repulsive solutes differ, the corresponding sizes of the explicit regions used in the two models have slightly different sizes.

When speaking of a surface and a distance to the surface of a solute we mean the closest carbon of the solute. Obviously this is not a unique definition of the surface, one could think about the distance from the plane through the three closest carbons, but this will lead to other definition problems at the edges of the solute. However, to understand our results even this very rough estimate of the surface is sufficient.

With this surface definition one identifies the distribution function $S(r^{\mathrm{rel}})$, which is defined as the average number of water molecules observed at a distance r^{rel} from the solute surface. Because of the non-spherical shape of the fullerenes (see fig. 4.5), the function $S(r^{\mathrm{rel}})$ is reported in addition to the traditional radial distribution function $g(r)$. Approximations for $g(r)$ based on a fit to the volume of each fullerene as a function of their effective radii could be used as well [167], but would unnecessarily complicate the understanding of the results.

To ensure that the results are not influenced by the choice of the water model, simulations were tested for rigid and flexible versions of the TIP3P and SPC/E point charge models (see tab. 3.1). Although only the results for rigid SPC/E are shown here, the general conclusions drawn hold the same for each three-site model, as the detailed simulation results were qualitatively the same for each of the models, even for flexible TIP3P, which can include Lennard-Jones interactions with the hydrogen atoms. Introducing explicit interactions for the hydrogens does not affect the results, because the geometry of the molecules prevents the hydrogen atoms from penetrating the fullerene even in the absence of explicit hydrogen interactions.

Solute	*Hydration shell width* (nm)			
	Lennard-Jones		Purely Repulsive	
	1st	2nd	1st	2nd
C_{60}	0.85	1.05	0.95	1.25
C_{240}	1.20	1.40	1.30	1.60
C_{540}	1.53	1.73	1.63	1.93
C_{960}	1.90	2.10	2.00	2.30
C_{1500}	2.25	2.45	2.35	2.65
C_{2160}	2.55	2.75	2.65	2.95
$d_{\mathrm{ex}}^{\mathrm{rel}}$	0.5	0.7	0.6	0.9

Table 4.2: **Hydration shell sizes of fullerenes**: The sizes of the first and second hydration shells measured from the center of mass of the solute. The radii of the shells were then used as the size of the explicit region d_{ex}, the relative size $d_{\mathrm{ex}}^{\mathrm{rel}}$ of the explicit region measured from surface of the solute is given in the last line. As these radii are different for the Lennard-Jones and purely repulsive solutes, the corresponding size of the explicit region used in the two models is different.

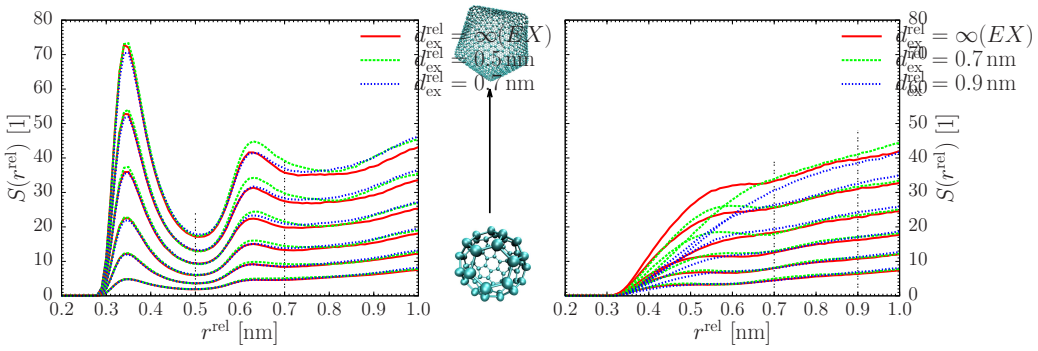

Figure 4.6: **Water distribution near the surface of a fullerene**: The distribution $S(r^{\mathrm{rel}})$ of the number of SPC/E water molecules is plotted as a function of the distance from the surface of the solute, r^{rel}. Results for the LJ solutes are shown on the left and for the purely repulsive solutes on the right. Different sets of curves correspond to different solutes from C_{60} to C_{2160} as the cartoon in the middle shows. Vertical lines indicate the average size of the explicit region in each AdResS simulation. Colored curves correspond to different sizes of the explicit region: Up to the first hydration shell (green) or to the second hydration shell (blue). The red line represents the reference all-atom simulation. The different curves are shown up to distances corresponding to the size of the explicit region used.

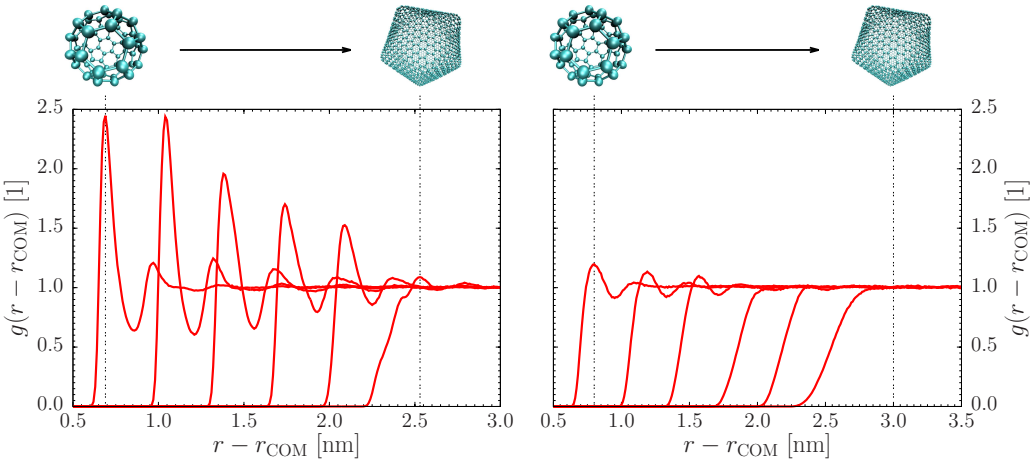

Figure 4.7: **Radial distribution function of water near the fullerene surface**: The radial distribution function $g(r)$ of SPC/E water molecules around the solutes obtained from the fully explicit simulations is plotted as a function of the distance from the center of mass r_{COM} of the solute. Results for the LJ solutes are shown on the left and for the purely repulsive solutes of the right. On top cartoons from C_{60} to C_{2160} are displayed to identify the curves. It is important to note that the purely repulsive solutes have nearly no first peak.

4.3.3 Results

The relative ability of each adaptive resolution simulation to reproduce the local density and structure of water around the solute is shown in fig. 4.6 and fig. 4.8. A comparison of the right and left panels of fig. 4.6 shows that limiting the size of the explicit region to the first hydration shell is sufficient to reproduce the density of water around the LJ solute (left panel). The situation is very different for the purely repulsive solutes, where the radial distribution of the water is greatly disturbed by the coarse-graining of the bulk (right panel). For much larger explicit regions (i. e. $d_{ex} = 2.25\,nm$ for C_{60}, not shown) we notice a consistent radial distribution. However, explicit region sizes comparable to the first and the second hydration shells are insufficient to reproduce accurately the results obtained in fully explicit simulations. Thus for purely repulsive hydrophobes, changes in the bulk directly affect the local hydration structure.

This difference between LJ and purely repulsive solutes is also evident from the distribution of the tetrahedral order parameter $q_4(r)$ around the solutes shown in fig. 4.8. In contrast to sec. 3.2.1 this time we defined $q_4(r)$

as in [126]:

$$q_4 = 1 - \frac{3}{8}\sum_{i<j}\left[\cos(\theta_{ij}) - \frac{1}{3}\right]^2 , \tag{4.4}$$

where θ_{ij} is the bond angle between particle i, the reference particle, and particle j, where i and j sum over the first four nearest neighbors of the reference particle. In this study we used this definition due the fact that we are more interested in the full range of θ_{ij} and not in the narrow range around the perfect tetrahedral packing like in sec. 3.2.1.

However, in fig. 4.8 we see that the repulsive solutes (right panel) exhibit almost negligible tetrahedral order at the solute interface. In particular, the water orientation very close to the repulsive solute surface appears to be increasingly more random for increasing solute size that means q_4 tends towards 0 (no tetrahedral packing) for the larger repulsive solutes. On the contrary, the LJ solutes (left panel) present a smaller decrease, even though significant. q_4 decreases from the bulk value of about 0.6 to 0.3-0.4 closer to the solute surface. However, the decrease is nearly independent of the thickness of the all-atom regime. These results suggest that LJ solutes induce a locally ordered yet very flexible hydrogen bond network, completely consistent with recent vibrational sum frequency spectroscopy results for water [162].

In addition, the height of the first peak of the radial distribution function $g(r)$ for the LJ solutes shown in the left panel of fig. 4.7, does not appear to decrease for increasing solute size, consistent with a previous study of spherical solutes [172].

For the repulsive solutes on the contrary, a complete *non-wetting* is observed, as the $g(r)$ function for the two largest repulsive solutes has essentially no first peak, shown in the right panel of fig. 4.7. The peak of the first hydration shell in the radial distribution function can be used to identify the formation of an interface layer similar to that of a gas-liquid interface. The phenomenon associated with the depletion of this interface layer is usually referred to as *non-wetting* [154, 165]. As the surfaces of the largest solutes approach the limit of a graphene sheet [173] the comparison to previous studies of attractive and repulsive hydrophobic sheets [174] is reasonable, where a consistent difference in non-wetting behavior observed with the two potentials was found.

This interpretation explains why the AdResS simulations are capable of reproducing local structures for the LJ solutes, but are not capable to do so for the repulsive solutes. The latter do not support a specific water structure in any significant way and the influence of the bulk water propagates towards the interface. As a consequence, adaptive resolution simulations that approximate the bulk with a CG model can not match the local density and

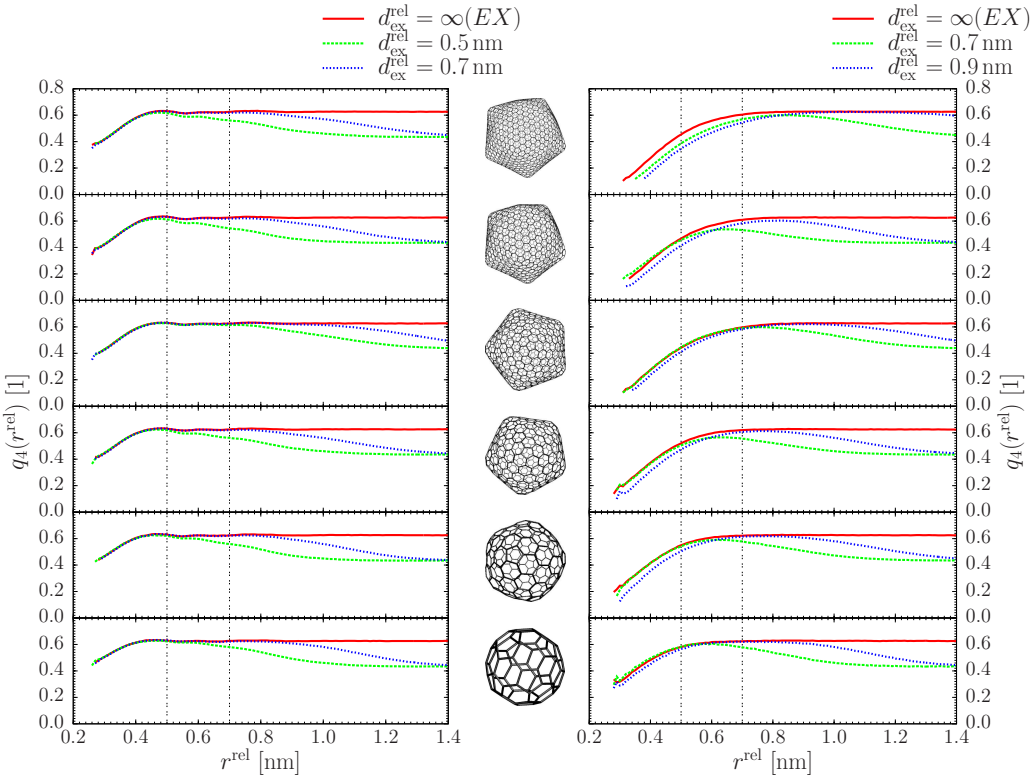

Figure 4.8: **Tetrahedral packing of water near the fullerene surface**: The average tetrahedral order parameter $q(r)$ of SPC/E water as a function of the distance r (in nm) from the surface of the solute is plotted for the LJ on the left and for purely repulsive solutes on the right. Vertical lines and colors are the same as in fig. 4.6. Results are reported for the solutes (see also the corresponding cartoon) from C_{60} (bottom plots) to C_{2160} (top plots).

structure of water around the repulsive solutes.

4.4 Summary

The locality of hydrophobic hydration is primarily controlled by the nature of the interaction between the hydrophobe and water. For purely repulsive hydrophobes, the hydration structure extends its effect to the bulk. On the same footing, perturbations in the bulk, such as the CG approximation, affect the structure and density of water at the solute interface. The size of the solute does not appear to affect the locality of the solvation significantly.

Chapter 5

Conclusion & Outlook

In this thesis we have discussed water models on two different resolutions: On the atomistic level and on the coarse-grained level. These two scales have been coupled using the adaptive resolution scheme, and the coupling was then used to describe the solvation of a series of icosahedral fullerenes. We will first summarize the results of applying different coarse-graining techniques to atomistic water models and other molecules. Then, the outcomes of the AdResS study using these coarse-grained force fields will be discussed. Finally we will conclude with an outlook on future research plans.

5.1 Coarse-Graining

The coarse-grained water models developed in this thesis treat water as a sphere. The atomistic and the coarse-grained resolutions are related by mapping the atomistic coordinates to the coarse-grained ones. Two different mapping schemes were applied: Either the center of mass or the oxygen of a single molecule was mapped onto the coarse-grained coordinates.

From the detailed study of structure-based coarse-graining of three different rigid water molecules [3], one can conclude that the coarse-grained interaction is governed by two length ratios, which are necessary to allow for tetrahedral packing. This special feature of water can be characterized by the tetrahedral order parameter, which measures the extend to which the 4 neighboring molecules of a water molecule form a tetrahedron. The two length ratios can be seen in the shape of the coarse-grained potential, which exhibits two minima with the one at larger distances being deeper.

For all coarse-grained models, the compressibility fits its corresponding atomistic counterpart with at most 10% deviation. However, the pressure in the coarse-grained model is approximately by a factor of 8000 higher than in

the atomistic case with the same density. When the pressure is constrained to match the atomistic pressure, by adding a pressure correction term, the compressibility increases by a factor 5. This leads to the conclusion that, at least for these simple coarse-grained water models, compressibility and pressure can not be fitted simultaneously.

Furthermore, for the coarse-grained models, be they pressure-corrected or not, the tetrahedral packing decreases significantly with the mean value of the tetrahedral order parameter, dropping by 30%. Nevertheless, we have developed a coarse-grained model with a packing behavior comparable to the atomistic models. This model exhibits an over-structured radial distribution function, and hence does not match the one of the atomistic model anymore, which should actually be the main outcome of a structure-based coarse-graining approach. We relate this weaker packing to the fact that coarse-grained models interact isotropically and therefore can not form a hydrogen bond network. This is supported by the observation that the average life time of the tetrahedral cluster decreases. Recently, it was shown that the tetrahedral packing and the radial distribution function can be matched simultaneously using a coarse-grained water model with a 3-body interaction [175].

In this study, the coarse-graining was done by iterative Boltzmann inversion, although the results do not depend on the method employed. To prove this, we compared different coarse-graining methods, namely iterative Boltzmann inversion, iterative inverse Monte Carlo and force matching [5]. While iterative Boltzmann inversion and iterative inverse Monte Carlo basically led to the same results for water, force matching generated a potential similar in shape, but different in the pair distribution function. This deviation can be understood from the point of view that on the one hand force matching is neither a structure-based nor an iterative method, while on the other hand, the basis set of the coarse-grained force field, an isotropic interaction, is too limited to fit the many-body potential of mean force. With an increased basis set, namely a 3-body interaction, a consistent pair distribution function can be obtained using an extension of the force matching technique called multi-scale coarse-graining [175].

To judge the suitability of the different coarse-graining methods from an objective point of view, the open-source package VOTCA [5] was developed and published on the web [84]. Aside from water the package was used to study liquid methanol, liquid propane, and a single chain of hexane.

For methanol, force matching and iterative Boltzmann inversion worked beautifully, whereas iterative inverse Monte Carlo caused problems when calculating the correct correlation for small systems due to the finite size of the simulation box. After solving this issue by studying bigger systems, all

three methods agreed in terms of structure and shape of the coarse-grained potential.

Liquid propane was taken as an example for a multi-component system. The coarse-grained mapping incorporates two kinds of beads. Iterative Boltzmann inversion and force matching resulted in similar bonded and non-bonded interaction potentials, while iterative inverse Monte Carlo exhibited stability problems and thus no potentials could be determined.

For the example of a single chain of hexane, the influence of coupled degrees of freedom and the choice of an incomplete basis set for a coarse-grained force field were examined and arising problems in the force matching algorithm were discussed.

In summary, the VOTCA package enables an automated and systematic comparison of various coarse-graining techniques and the coarse-graining community is already benefiting from this package.

Nevertheless, pair distributions are only one aspect of coarse-graining. In general, the dynamics of a coarse-grained system is much faster than that of the atomistic reference system, which is due to the coarse-grained systems having a much smoother energy landscape.

The faster diffusion has the advantage of an intrinsically faster system, which samples a bigger part of the phase space. The disadvantage is, of course, that the dynamics of the coarse-grained system does not necessarily resemble the one of the all-atom system. In a multicomponent system every component will have a different speedup, which makes it hard to study processes involving multiple components and to determine the speedup of the total system.

To slow down the diffusive dynamics, we proposed an extension of the DPD thermostat [1]. This so-called transversal DPD thermostat acts on the degrees of freedom perpendicular to the interatomic axis of two particles. It has been shown that with this approach, we can influence the diffusive part of the dynamics in a way that is compatible with hydrodynamics due to the fact that our thermostat inherits momentum conservation from the DPD thermostat.

5.2 Adaptive Resolution Scheme

With the knowledge about the development of coarse-grained models we then focused on AdResS [4], which is a simulation method that couples two resolutions of the same physical system. The coupling is done via force interpolation, which ensures the validity of Newton's third law and hence the free exchange of particles between the two regions. The basic ansatz of the

method is easy to understand, but exhibits some issues in its implementation. For this reason we fully explain the implementation of AdResS into the ESPResSo [6] and GroMaCS packages [102]. Due to the fact that this is the first freely available implementation we call it a reference implementation and validate it with former results namely the tetrahedral liquid.

With the confidence of having a working and efficient implementation we investigated an atomistic water system coupled to a coarse-grained system using one on the models mentioned above. The coupling introduces a small perturbation in the atomistic region, which is determined to be a third of the interaction length in the case of a spherical atomistic region. Despite this small perturbation, the radial distribution function is conserved and molecules are freely exchanged between the regions.

As the coarse-grained water is not able to form hydrogen bonds, we were able to study the influence of the absence of the hydrogen bond network on the hydration of small molecules, in this case a series of icosahedral fullerenes. In AdResS one can increase the size of the atomistic region and allow a bigger hydrogen bond network to form. We found that the locality of hydrophobic hydration is primarily controlled by the nature of the interaction between the molecule and water. For purely repulsive interactions between solute and water the hydration structure is influenced by the bulk, while for Lennard-Jones interaction an influence was not measurable.

5.3 Outlook

In the coarse-grained approaches discussed in this thesis the Coulomb interaction was always reduced to an effective interaction, which was short ranged. However, this only reproduces the correct properties if the coarse-grained beads can be chosen to be uncharged. A fitting procedure for effective coarse-grained charges is possible, but technically very demanding. In AdResS the problem is comparable, the interaction between two molecules has to be short ranged, otherwise the force interpolation is not possible. That means AdResS has to be extended for systems with charged coarse-grained beads.

VOTCA has provided significant insight in the process of systematic coarse-graining, but the development process of the package is not yet complete. First, an implementation of the simplex algorithm, which allows the parameterization of the coarse-grained model for thermodynamic properties, for example pressure or free energies, is in preparation. Second, an iterative method, which combines the iterative inverse Monte Carlo update at the beginning of the coarse-graining procedure with the iterative Boltzmann in-

verse update in later stage, could be beneficial for the development of coarse-grained models. Third, coarse-graining of long range interactions should be implemented. Fourth, back-mapping to atomistic details is another planned feature of the VOTCA package. Since a coarse-grained model only reproduces certain properties of the higher resolution model, atomistic details can be reintroduced after equilibration at a coarse-grained level to measure the atomistic properties [26]. With this approach only a short equilibration run at atomistic level is required to locally equilibrate the system.

Due to the development of automated coarse-graining an increasing number of coarse-grained force fields will be developed. To avoid the constant reinventing of force fields these should be stored and categorized in an electronic library. Due to growing availability of coarse-grained force fields it will be also much more convenient to use AdResS to access much bigger time and length scales. With our implementations of the scheme we have started the transition from the method development stage to the application stage, which will allow the method to be used by the biochemical community.

Bibliography

[1] **C. Junghans**, M. Praprotnik, and K. Kremer. Transport properties controlled by a thermostat: An extended dissipative particle dynamics thermostat. *Soft Matter*, 4(1):156–161, 2008.

[2] M. Praprotnik, **C. Junghans**, L. Delle Site, and K. Kremer. Simulation approaches to soft matter: Generic statistical properties vs. chemical details. *Computer Physics Communications*, 179(1-3):51–60, 2008.

[3] H. Wang, **C. Junghans**, and K. Kremer. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *The European Physical Journal E: Soft Matter and Biological Physics*, 28(2):221–229, 2009.

[4] **C. Junghans**, M. Praprotnik, and L. Delle Site. Adaptive resolution schemes. In J. Grotendorst, N. Attig, S. Blügel, and D. Marx, editors, *Multiscale Simulation Methods in Molecular Sciences: Lecture Notes*, page 359. Forschungszentrum Jülich, 2009.

[5] V. Rühle, **C. Junghans**, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile Object-Oriented toolkit for Coarse-Graining applications. *Journal of Chemical Theory and Computation*, 5(12):3211–3223, 2009.

[6] **C. Junghans** and S. Poblete. A reference implementation of the adaptive resolution scheme in ESPResSo. *Computer Physics Communications*, 181(8):1449–1454, 2010.

[7] B. P. Lambeth, **C. Junghans**, K. Kremer, C. Clementi, and L. Delle Site. Communication: On the locality of hydrogen bond networks at hydrophobic interfaces. *The Journal of Chemical Physics*, 133(22):221101, 2010.

[8] S. Koneshan, J. C. Rasaiah, R. M. Lynden-Bell, and S. H. Lee. Solvent structure, dynamics, and ion mobility in aqueous solutions at 25 °C. *The Journal of Physical Chemistry B*, 102(21):4193–4204, 1998.

[9] M. R. Shirts and V. S. Pande. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *The Journal of Chemical Physics*, 122(13):134508, 2005.

[10] D. Sebastiani and L. Delle Site. Adsorption of water molecules on flat and stepped nickel surfaces from first principles. *Journal of Chemical Theory and Computation*, 1(1):78–82, 2005.

[11] Y. M. Rhee, E. J. Sorin, G. Jayachandran, E. Lindahl, and V. S. Pande. Simulations of the role of water in the protein-folding mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, 101(17):6456–6461, 2004.

[12] B. Dünweg and A. Ladd. Lattice boltzmann simulations of soft matter systems. In *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, pages 89–166. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 2009.

[13] P. J. Hoogerbrugge and J. M. V. A. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhysics Letters*, 19(3):155–160, 1992.

[14] D. Andrienko and K. Kremer. Simulations. In K. Matyjaszewski, Y. Gnanou, and L. Leibler, editors, *Macromolecular Engineering: Precise Synthesis, Materials Properties, Applications*. Wiley-VCH Verlag GmbH & Co. KGaA, 1st edition, 2007.

[15] G. Lu, E. B. Tadmor, and E. Kaxiras. From electrons to finite elements: A concurrent multiscale approach for metals. *Physical Review B*, 73(2):024108, 2006.

[16] J. Rottler, S. Barsky, and M. O. Robbins. Cracks and crazes: On calculating the macroscopic fracture energy of glassy polymers from molecular simulations. *Physical Review Letters*, 89(14):148304, 2002.

[17] G. Csanyi, T. Albaret, M. C. Payne, and A. D. Vita. "Learn on the fly": A hybrid classical and Quantum-Mechanical molecular dynamics simulation. *Physical Review Letters*, 93(17):175503, 2004.

[18] A. Laio, J. van de Vondele, and U. Rothlisberger. A hamiltonian electrostatic coupling scheme for hybrid CarParrinello molecular dynamics simulations. *The Journal of Chemical Physics*, 116(16):6941, 2002.

[19] D. Jiang and E. A. Carter. First principles assessment of ideal fracture energies of materials with mobile impurities: implications for hydrogen embrittlement of metals. *Acta Materialia*, 52(16):4801–4807, 2004.

[20] G. Lu and E. Kaxiras. Hydrogen embrittlement of aluminum: The crucial role of vacancies. *Physical Review Letters*, 94(15):155501, 2005.

[21] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *The Journal of Chemical Physics*, 123(22):224106–14, 2005.

[22] B. Ensing, S. O. Nielsen, P. B. Moore, M. L. Klein, and M. Parrinello. Energy conservation in adaptive hybrid Atomistic/Coarse-Grain molecular dynamics. *Journal of Chemical Theory and Computation*, 3(3):1100–1105, 2007.

[23] A. Heyden and D. G. Truhlar. Conservative algorithm for an adaptive change of resolution in mixed Atomistic/Coarse-Grained multiscale simulations. *Journal of Chemical Theory and Computation*, 4(2):217–221, 2008.

[24] C. Peter and K. Kremer. Multiscale simulation of soft matter systems - from the atomistic to the coarse-grained level and back. *Soft Matter*, 5(22):4357–4366, 2009.

[25] A. P. Lyubartsev, M. Karttunen, I. Vattulainen, and A. Laaksonen. On Coarse-Graining by the inverse Monte Carlo method: Dissipative particle dynamics simulations made to a precise tool in soft matter modeling. *Soft Materials*, 1(1):121, 2002.

[26] W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, and T. Bürger. Simulation of polymer melts. II. From coarse-grained models back to atomistic description. *Acta Polymerica*, 49(2-3):75–79, 1998.

[27] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein. A coarse grain model for phospholipid simulations. *The Journal of Physical Chemistry B*, 105(19):4464–4470, 2001.

[28] C. F. Abrams and K. Kremer. Combined Coarse-Grained and atomistic simulation of liquid bisphenol a – polycarbonate: Liquid packing and intramolecular structure. *Macromolecules*, 36(1):260–267, 2003.

[29] T. Murtola, E. Falck, M. Patra, M. Karttunen, and I. Vattulainen. Coarse-grained model for phospholipid/cholesterol bilayer. *The Journal of Chemical Physics*, 121(18):9156, 2004.

[30] S. Izvekov and G. A. Voth. Multiscale coarse graining of liquid-state systems. *The Journal of Chemical Physics*, 123(13):134105, 2005.

[31] Q. Sun and R. Faller. Systematic Coarse-Graining of a polymer blend: polyisoprene and polystyrene. *Journal of Chemical Theory and Computation*, 2(3):607–615, 2006.

[32] V. A. Harmandaris, N. P. Adhikari, N. F. A. van der Vegt, and K. Kremer. Hierarchical modeling of polystyrene: From atomistic to Coarse-Grained simulations. *Macromolecules*, 39(19):6708–6719, 2006.

[33] L. Yelash, M. Müller, W. Paul, and K. Binder. How well can Coarse-Grained models of real polymers describe their structure? the case of polybutadiene. *Journal of Chemical Theory and Computation*, 2(3):588–597, 2006.

[34] A. Y. Shih, A. Arkhipov, P. L. Freddolino, and K. Schulten. Coarse grained Protein-Lipid model with application to lipoprotein particles. *The Journal of Physical Chemistry B*, 110(8):3674–3684, 2006.

[35] A. Lyubartsev. Multiscale modeling of lipids and lipid bilayers. *European Biophysics Journal*, 35(1):53–61, 2005.

[36] J. Zhou, I. F. Thorpe, S. Izvekov, and G. A. Voth. Coarse-Grained peptide modeling using a systematic multiscale approach. *Biophysical Journal*, 92(12):4289–4303, 2007.

[37] A. Villa, N. F. A. van der Vegt, and C. Peter. Self-assembling dipeptides: including solvent degrees of freedom in a coarse-grained model. *Physical Chemistry Chemical Physics*, 11(12):2068–2076, 2009.

[38] Google Books. "computer simulation". http://books.google.de/books?q=%22computer+simulation%22, 2010.

[39] Google Books. "molecular dynamics". http://books.google.de/books?q=%22molecular+dynamics%22, 2010.

[40] H. Goldstein, C. P. Poole, and J. L. Safko. *Classical Mechanics*. Pearson, 3rd edition, 2001.

[41] H. Risken. *The Fokker-Planck Equation. Methods of Solution and Applications*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K, 1989. 2nd printing edition, 1989.

[42] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[43] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101–7, 2007.

[44] B. Dünweg. Langevin methods. In B. Dünweg, A. I. Milchev, and D. P. Landau, editors, *Computer Simulations of Surfaces and Interfaces: Proceedings of the NATO Advanced Study Institute, Albena, Bulgaria, from 9 to 20 September 2002*, pages 77–92. Springer Netherlands, 2004.

[45] B. Dünweg and W. Paul. Brownian dynamics simulations without gaussian random numbers. *International Journal of Modern Physics C*, 2(3):817–827, 1991.

[46] T. Soddemann, B. Dünweg, and K. Kremer. Dissipative particle dynamics: A useful thermostat for equilibrium and nonequilibrium molecular dynamics simulations. *Physical Review E*, 68(4), 2003.

[47] P. Español. Hydrodynamics from dissipative particle dynamics. *Physical Review E*, 52(2):1734, 1995.

[48] P. Español and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhysics Letters*, 30(4):191–196, 1995.

[49] C. Pastorino, T. Kreer, M. Müller, and K. Binder. Comparison of dissipative particle dynamics and Langevin thermostats for out-of-equilibrium simulations of polymeric systems. *Physical Review E*, 76(2):026706, 2007.

[50] C. Marsh, G. Backx, and M. Ernst. Fokker-Planck-Boltzmann equation for dissipative particle dynamics. *Europhysics Letters*, 38(6):411, 1997.

[51] P. Español and M. Revenga. Smoothed dissipative particle dynamics. *Physical Review E*, 67(2):026705, 2003.

[52] H. Noguchi and G. Gompper. Transport coefficients of off-lattice mesoscale-hydrodynamics simulation techniques. *Physical Review E*, 78(1):016706, 2008.

[53] M. S. Green. Markoff random processes and the statistical mechanics of Time-Dependent phenomena. II. Irreversible processes in fluids. *The Journal of Chemical Physics*, 22(3):398, 1954.

[54] R. Kubo. Statistical-Mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[55] R. Zwanzig. Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics*, 33(5):1338–1341, 1960.

[56] R. Zwanzig. Memory effects in irreversible thermodynamics. *Physical Review*, 124(4):983, 1961.

[57] D. Forster. *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions*. Perseus Books, new edition, 1995.

[58] **C. Junghans** and B. Dünweg. Mori-Zwanzig projector formalism for dissipative particle dynamics thermostat. (unpublished), 2008.

[59] H. Limbach, A. Arnold, B. Mann, and C. Holm. ESPResSo–an extensible simulation package for research on soft matter systems. *Computer Physics Communications*, 174(9):704–727, 2006.

[60] B. Dünweg and K. Kremer. Molecular dynamics simulation of a polymer chain in solution. *The Journal of Chemical Physics*, 99(9):6983, 1993.

[61] B. Efron. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial Mathematics, 1982.

[62] F. Ercolessi and J. B. Adams. Interatomic potentials from First-Principles calculations: The Force-Matching method. *Europhysics Letters*, 26(8):583–588, 1994.

[63] B. Hess, C. Holm, and N. F. A. van der Vegt. Modeling multibody effects in ionic solutions with a concentration dependent dielectric permittivity. *Physical Review Letters*, 96(14):147801, 2006.

[64] D. Reith, M. Pütz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry*, 24(13):1624–1636, 2003.

[65] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Physical Review E*, 52(4):3730, 1995.

[66] A. K. Soper. Empirical potential Monte Carlo simulation of fluid structure. *Chemical Physics*, 202(2-3):295–306, 1996.

[67] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: a new method for force-matching. *The Journal of Chemical Physics*, 120(23):10896–10913, 2004.

[68] G. Toth. Effective potentials from complex simulations: a potential-matching algorithm and remarks on coarse-grained potentials. *Journal of Physics: Condensed Matter*, 19(33):335222, 2007.

[69] G. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, 7(1):95–99, 1963.

[70] V. A. Harmandaris, D. Reith, N. F. A. van der Vegt, and K. Kremer. Comparison between Coarse-Graining models for polymer systems: Two mapping schemes for polystyrene. *Macromolecular Chemistry and Physics*, 208(19-20):2109–2120, 2007.

[71] A. Villa, C. Peter, and N. F. A. van der Vegt. Self-assembling dipeptides: conformational sampling in solvent-free coarse-grained simulation. *Physical Chemistry Chemical Physics*, 11(12):2077, 2009.

[72] R. L. Henderson. A uniqueness theorem for fluid pair correlation functions. *Physics Letters A*, 49(3):197–198, 1974.

[73] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen. Multiscale modeling of emergent materials: biological and soft matter. *Physical Chemistry Chemical Physics*, 11(12):1869–1892, 2009.

[74] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864, 1964.

[75] J. T. Chayes and L. Chayes. On the validity of the inverse conjecture in classical density functional theory. *Journal of Statistical Physics*, 36(3):471–488, 1984.

[76] J. T. Chayes, L. Chayes, and E. H. Lieb. The inverse problem in classical statistical mechanics. *Communications in Mathematical Physics*, 93(1):57–121, 1984.

[77] S. Leon, N. van der Vegt, L. Delle Site, and K. Kremer. Bisphenol a polycarbonate: Entanglement analysis from Coarse-Grained MD simulations. *Macromolecules*, 38(19):8078–8092, 2005.

[78] T. Murtola, E. Falck, M. Karttunen, and I. Vattulainen. Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints. *The Journal of Chemical Physics*, 126(7):075101, 2007.

[79] W. G. Noid, J. Chu, G. S. Ayton, and G. A. Voth. Multiscale Coarse-Graining and structural correlations: connections to Liquid-State theory. *The Journal of Physical Chemistry B*, 111(16):4116–4127, 2007.

[80] W. G. Noid, J. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen. The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *The Journal of Chemical Physics*, 128(24):244114, 2008.

[81] W. G. Noid, P. Liu, Y. Wang, J. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth. The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models. *The Journal of Chemical Physics*, 128(24):244115, 2008.

[82] A. Das and H. C. Andersen. The multiscale coarse-graining method. III. A test of pairwise additivity of the coarse-grained potential and of new basis functions for the variational calculation. *The Journal of Chemical Physics*, 131(3):034102, 2009.

[83] V. Krishna, W. G. Noid, and G. A. Voth. The multiscale coarse-graining method. IV. Transferring coarse-grained potentials between temperatures. *The Journal of Chemical Physics*, 131(2):024103, 2009.

[84] Versatile object-oriented toolkit for coarse-graining applications (VOTCA). http://www.votca.org, 2010.

[85] Apache license, 2004. (Version 2.0).

[86] V. Rühle. *Morphology and charge transport in conjugated polymers.* PhD thesis, Johannes -Gutenberg-Universität, Mainz, 2010.

[87] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, Load-Balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[88] ESPResSo++. http://www.espresso-pp.de/, 2010.

[89] R. Faller, H. Schmitz, O. Biermann, and F. Müller-Plathe. Automatic parameterization of force fields for liquids by simplex optimization. *Journal of Computational Chemistry*, 20(10):1009–1017, 1999.

[90] B. Hess, C. Holm, and N. F. A. van der Vegt. Osmotic coefficients of atomistic NaCl (aq) force fields. *The Journal of Chemical Physics*, 124(16):164509, 2006.

[91] E. Darve. Thermodynamic integration using constrained and unconstrained dynamics. In A. Pohorille and C. Chipot, editors, *Free Energy Calculations: Theory and Applications in Chemistry and Biology*, page 119. Springer, Berlin, 1st edition, 2007.

[92] M. Praprotnik, L. Delle Site, and K. Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Annual Review of Physical Chemistry*, 59(1):545–571, 2008.

[93] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. *Physical Review E*, 73(6), 2006.

[94] M. Praprotnik, K. Kremer, and L. Delle Site. Adaptive molecular resolution via a continuous change of the phase space dimensionality. *Physical Review E*, 75(1), 2007.

[95] S. Poblete. *Thermodynamic Consistency in Adaptive Resolution Simulations.* PhD thesis, Johannes -Gutenberg-Universität, Mainz, 2010.

[96] M. Praprotnik, K. Kremer, and L. Delle Site. Fractional dimensions of phase space variables: a tool for varying the degrees of freedom of a system in a multiscale treatment. *Journal of Physics A: Mathematical and Theoretical*, 40(15):F281–F288, 2007.

[97] L. Delle Site. Some fundamental problems for an energy-conserving adaptive-resolution molecular dynamics scheme. *Physical Review E*, 76(4), 2007.

[98] M. Praprotnik, L. Delle Site, and K. Kremer. A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation. *The Journal of Chemical Physics*, 126(13):134902–8, 2007.

[99] S. Poblete, M. Praprotnik, K. Kremer, and L. Delle Site. Coupling different levels of resolution in molecular simulations. *The Journal of Chemical Physics*, 132(11):114101, 2010.

[100] L. A. Rowley, D. Nicholson, and N. G. Parsonage. Monte Carlo grand canonical ensemble calculation in a gas-liquid transition region for 12-6 argon. *Journal of Computational Physics*, 17(4):401–414, 1975.

[101] T. N. Heinz, W. F. van Gunsteren, and P. H. Hunenberger. Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations. *The Journal of Chemical Physics*, 115(3):1125–1136, 2001.

[102] B. Lambeth, S. Fritsch, and **C. Junghans**. AdressMacs. http://repo.or.cz/w/gromacs/adressmacs.git, 2010.

[103] J. D. Bernal and R. H. Fowler. A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions. *The Journal of Chemical Physics*, 1(8):515, 1933.

[104] H. J. C. Berendsen, J. P. M. Postma, and W. F. V. Gunsteren. Interaction models for water in relation to protein hydration. In A. Pullman, editor, *Intermolecular Forces: Symposium Proceedings*, pages 331–342. Springer Netherlands, 1981.

[105] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926, 1983.

[106] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[107] M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *The Journal of Chemical Physics*, 112(20):8910, 2000.

[108] B. Guillot. A reappraisal of what we have learnt during three decades of computer simulations on water. *Journal of Molecular Liquids*, 101(1-3):219–260, 2002.

[109] A. Baranyai, A. Bartok, and A. A. Chialvo. Limitations of the rigid planar nonpolarizable models of water. *The Journal of Chemical Physics*, 124(7):074507, 2006.

[110] A. Brodsky. Is there predictive value in water computer simulations? *Chemical Physics Letters*, 261(4-5):563–568, 1996.

[111] J. L. Finney. The water molecule and its interactions: the interaction between theory, modelling, and experiment. *Journal of Molecular Liquids*, 90(1-3):303–312, 2001.

[112] V. E. Petrenko, M. L. Dubova, Y. M. Kessler, and M. Y. Perova. Water in computer experiment: Contradiction in parameterization of potentials. *Russian Journal of Physical Chemistry*, 74:1777–1781, 2001.

[113] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[114] S. Amira, D. Spångberg, and K. Hermansson. Derivation and evaluation of a flexible SPC model for liquid water. *Chemical Physics*, 303(3):327–334, 2004.

[115] L. X. Dang and B. M. Pettitt. Simple intramolecular model potentials for water. *The Journal of Physical Chemistry*, 91(12):3349–3354, 1987.

[116] K. Toukan and A. Rahman. Molecular-dynamics study of atomic motions in water. *Physical Review B*, 31(5):2643, 1985.

[117] S. Zhu and C. F. Wong. Sensitivity analysis of water thermodynamics. *The Journal of Chemical Physics*, 98(11):8892, 1993.

[118] L. X. Dang. The nonadditive intermolecular potential for water revised. *The Journal of Chemical Physics*, 97(4):2659, 1992.

[119] G. Lamoureux, A. D. MacKerell, and B. Roux. A simple polarizable model of water based on classical drude oscillators. *The Journal of Chemical Physics*, 119(10):5185, 2003.

[120] S. W. Rick, S. J. Stuart, and B. J. Berne. Dynamical fluctuating charge force fields: Application to liquid water. *The Journal of Chemical Physics*, 101(7):6141, 1994.

[121] H. A. Stern, F. Rittner, B. J. Berne, and R. A. Friesner. Combined fluctuating charge and polarizable dipole models: Application to a five-site water potential function. *The Journal of Chemical Physics*, 115(5):2237, 2001.

[122] H. Yu, T. Hansson, and W. F. van Gunsteren. Development of a simple, self-consistent polarizable model for liquid water. *The Journal of Chemical Physics*, 118(1):221, 2003.

[123] M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer, and C. Clementi. Adaptive resolution simulation of liquid water. *Journal of Physics: Condensed Matter*, 19(29):292201, 2007.

[124] M. Praprotnik, S. Matysiak, L. Delle Site, K. Kremer, and C. Clementi. Corrigendum: Adaptive resolution simulation of liquid water. *Journal of Physics: Condensed Matter*, 21(49):499801, 2009.

[125] M. E. Johnson, T. Head-Gordon, and A. A. Louis. Representability problems for coarse-grained water potentials. *The Journal of Chemical Physics*, 126(14):144509, 2007.

[126] J. R. Errington and P. G. Debenedetti. Relationship between structural order and the anomalies of liquid water. *Nature*, 409(6818):318–321, 2001.

[127] F. Franks. *Water: A Matrix of Life*. Amer Chemical Soc, 2nd edition, 1994.

[128] M. Rodnikova. A new approach to the mechanism of solvophobic interactions. *Journal of Molecular Liquids*, 136(3):211–213, 2007.

[129] D. Eisenberg and W. Kauzmann. *Structure and Properties of Water*. Oxford University Press, 1969.

[130] D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005.

[131] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An $N \log N$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089, 1993.

[132] J. Hansen and I. R. Mc Donald. *Theory of Simple Liquids*. Academic Pr Inc, 3rd edition, 2006.

[133] M. Baptista, R. Schmitz, and B. Dünweg. Simple and robust solver for the Poisson-Boltzmann equation. *Physical Review E*, 80(1):016705, 2009.

[134] W. F. V. Gunsteren and H. J. C. Berendsen. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1(3):173, 1988.

[135] W. Janke. Statistical analysis of simulations: Data correlations and error estimation. In J. Grotendorst, D. Marx, and A. Muramatsu, editors, *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms: Lecture Notes. NIC-Serie, Band 10*, pages 423–445. Forschungszentrum Jülich, 2002.

[136] W. L. Jorgensen and J. Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6665–6670, 2005.

[137] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, 52(2):255, 1984.

[138] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695, 1985.

[139] S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer, and L. Delle Site. Modeling diffusive dynamics in adaptive resolution simulation of liquid water. *The Journal of Chemical Physics*, 128(2):024503–12, 2008.

[140] R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik. Concurrent triple-scale simulation of molecular liquids. *The Journal of Chemical Physics*, 128(11):114110–9, 2008.

[141] M. Neumann. Dipole moment fluctuation formulas in computer simulations of polar systems. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, 50(4):841, 1983.

[142] M. Neumann. The dielectric constant of water. computer simulations with the MCY potential. *The Journal of Chemical Physics*, 82(12):5663, 1985.

[143] R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik. Coupling atomistic and continuum hydrodynamics through a mesoscopic model: Application to liquid water. *The Journal of Chemical Physics*, 131(24):244107, 2009.

[144] Y. Wu, H. L. Tepper, and G. A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *The Journal of Chemical Physics*, 124(2):024503, 2006.

[145] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2nd edition, 2001.

[146] J. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3):327–341, 1977.

[147] H. C. Andersen. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34, 1983.

[148] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.

[149] R. Schmitz, P. Ranganathan, and B. Dünweg. Brownian dynamics with constraints. (unpublished), 2010.

[150] M. Fixman. Simulation of polymer dynamics. I. General theory. *The Journal of Chemical Physics*, 69(4):1527, 1978.

[151] D. C. Morse. Theory of constrained brownian motion. In S. A. Rice and I. Prigogine, editors, *Advances in Chemical Physics: Volume 128*, page 65. John Wiley & Sons, 1st edition, 2004.

[152] S. Poblete, S. Fritsch, G. Ciccotti, and L. Delle Site. Adaptive resolution simulation: A grand canonical interpretation. (unpublished), 2010.

[153] D. M. Huang and D. Chandler. Temperature and length scale dependence of hydrophobic effects and their possible implications for protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 97(15):8324–8327, 2000.

[154] D. Chandler. Interfaces and the driving force of hydrophobic assembly. *Nature*, 437(7059):640–647, 2005.

[155] G. A. Papoian, J. Ulander, M. P. Eastwood, Z. Luthey-Schulten, and P. G. Wolynes. Water in protein structure prediction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(10):3352–3357, 2004.

[156] P. Ball. Water as an active constituent in cell biology. *Chemical Reviews*, 108(1):74–108, 2008.

[157] J. S. Tse, M. L. Klein, and I. R. Mc Donald. Molecular dynamics studies of ice Ic and the structure I clathrate hydrate of methane. *The Journal of Physical Chemistry*, 87(21):4198–4203, 1983.

[158] E. D. Sloan and C. A. Koh. *Clathrate hydrates of natural gases*. CRC Press, 2008.

[159] B. Lambeth, M. Praprotnik, C. Clementi, K. Kremer, and L. Delle Site. Adaptive resolution simulation: Methane in water. (in preparation), 2010.

[160] K. Lum, D. Chandler, and J. D. Weeks. Hydrophobicity at small and large length scales. *The Journal of Physical Chemistry B*, 103(22):4570–4577, 1999.

[161] J. D. Weeks. Connecting local structure to interface formation: A molecular scale van der Waals theory of nonuniform liquids. *Annual Review of Physical Chemistry*, 53(1):533–562, 2002.

[162] L. F. Scatena, M. G. Brown, and G. L. Richmond. Water at hydrophobic surfaces: Weak hydrogen bonding and strong orientation effects. *Science*, 292(5518):908–912, 2001.

[163] F. Pizzitutti, M. Marchi, F. Sterpone, and P. J. Rossky. How protein surfaces induce anomalous dynamics of hydration water. *The Journal of Physical Chemistry B*, 111(26):7584–7590, 2007.

[164] S. Sarupria and S. Garde. Quantifying water density fluctuations and compressibility of hydration shells of hydrophobic solutes and proteins. *Physical Review Letters*, 103(3):037803, 2009.

[165] B. J. Berne, J. D. Weeks, and R. Zhou. Dewetting and hydrophobic interaction in physical and biological systems. *Annual Review of Physical Chemistry*, 60(1):85–103, 2009.

[166] N. Choudhury and B. M. Pettitt. The dewetting transition and the hydrophobic effect. *Journal of the American Chemical Society*, 129(15):4847–4852, 2007.

[167] H. S. Ashbaugh and M. E. Paulaitis. Effect of solute size and solute – water attractive interactions on hydration water structure around hydrophobic solutes. *Journal of the American Chemical Society*, 123(43):10721–10728, 2001.

[168] J. Ma, D. Alfe, A. Michaelides, and E. Wang. The water-benzene interaction: Insight from electronic structure theories. *The Journal of Chemical Physics*, 130(15):154303, 2009.

[169] D. Chandler, J. D. Weeks, and H. C. Andersen. Van der Waals picture of liquids, solids, and phase transformations. *Science*, 220(4599):787–794, 1983.

[170] D. R. Weiss, T. M. Raschke, and M. Levitt. How hydrophobic buckminsterfullerene affects surrounding water structure. *The Journal of Physical Chemistry B*, 112(10):2981–2990, 2008.

[171] S. J. Suresh and V. M. Naik. Hydrogen bond thermodynamic properties of water from dielectric constant data. *The Journal of Chemical Physics*, 113(21):9727, 2000.

[172] J. Mittal and G. Hummer. Static and dynamic correlations in water at hydrophobic interfaces. *Proceedings of the National Academy of Sciences*, 105(51):20130–20135, 2008.

[173] C. H. Xu and G. E. Scuseria. An O(N) tight-binding study of carbon clusters up to c8640: the geometrical shape of the giant icosahedral fullerenes. *Chemical Physics Letters*, 262(3-4):219–226, 1996.

[174] N. Choudhury and B. M. Pettitt. Local density profiles are coupled to solute size and attractive potential for nanoscopic hydrophobic solutes. *Molecular Simulation*, 31(6):457, 2005.

[175] L. Larini, L. Lu, and G. A. Voth. The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *The Journal of Chemical Physics*, 132(16):164107, 2010.

Acknowledgment

Resume

RESUME

RESUME