Scale bridging concepts in molecular simulation: coarse-graining and thermodynamic coupling

Dissertation

zur Erlangung des Grades

"Doktor der Naturwissenschaften"

am Fachbereich Physik
der Johannes Gutenberg-Universität Mainz

Sebastian Fritsch

geb. in Speyer

Abstract

This thesis deals with simulations of liquids on a molecular level, making use of different multiscale techniques. These methods allow efficient, computationally less demanding simulations of the liquid, which enables phenomena on longer time- and length-scale to be investigated. One crucial component is a simplified (coarse-grained) model, which is derived in a systematic way from a more detailed atomistic reference model, such that selected properties of the detailed model (such as pair-correlation function, pressure, etc.) are reproduced.

In addition, algorithms that allow a concurrent coupling of the detailed and simplified model (Adaptive Resolution Scheme, AdResS) are investigated. Thereby, the detailed model is used in a well-defined sub-volume, whereas the remainder is treated using the coarse-grained model. A method was developed (thermodynamic force) to enable the coupling also in the case that the two coupled models are at different state points. Furthermore, a novel algorithm (H-AdResS) is presented, which formulates the coupling with the aid of a Hamiltonian. In this algorithm, a compensation analogous to the Thermodynamic Force is available at lesser computational cost.

As an application of these basic techniques, path integral MD simulations of liquid water were studied. Using this method it is possible to include nuclear quantum effects (delocalization, zero-point energy) in the simulation. In the first step, a multiscale technique (force-matching) was used to parametrize an effective interaction based on a detailed simulation employing density functional theory (DFT). The path integral technique improves the description of the intramolecular structure in comparison with the experiment. The model derived in this way is also suitable for concurrent coupling, in which a water molecule (represented by 48 point particles in the path integral model) is coupled to a coarse-grained model (single point particle). In this fashion, a water-vacuum interface was simulated, where the interface is treated in the detailed path integral model, while the bulk is described by the coarse-grained model.

Zusammenfassung

In dieser Arbeit wurden Simulation von Flüssigkeiten auf molekularer Ebene durchgeführt, wobei unterschiedliche Multi-Skalen Techniken verwendet wurden. Diese erlauben eine effektive Beschreibung der Flüssigkeit, die weniger Rechenzeit im Computer benötigt und somit Phänomene auf längeren Zeit- und Längenskalen beschreiben kann. Ein wesentlicher Aspekt ist dabei ein vereinfachtes (coarse-grained) Modell, welches in einem systematischen Verfahren aus Simulationen des detaillierten Modells gewonnen wird. Dabei werden ausgewählte Eigenschaften des detaillierten Modells (z.B. Paar-Korrelationsfunktion, Druck, etc.) reproduziert.

Es wurden Algorithmen untersucht, die eine gleichzeitige Kopplung von detaillierten und vereinfachten Modell erlauben (Adaptive Resolution Scheme, AdResS). Dabei wird das detaillierte Modell in einem vordefinierten Teilvolumen der Flüssigkeit (z.B. nahe einer Oberfläche) verwendet, während der Rest mithilfe des vereinfachten Modells beschrieben wird. Hierzu wurde eine Methode (Thermodynamische Kraft) entwickelt um die Kopplung auch dann zu ermöglichen, wenn die Modelle in verschiedenen thermodynamischen Zuständen befinden. Zudem wurde ein neuartiger Algorithmus der Kopplung beschrieben (H-AdResS) der die Kopplung mittels einer Hamilton-Funktion beschreibt. In diesem Algorithmus ist eine zur Thermodynamischen Kraft analoge Korrektur mit weniger Rechenaufwand möglich.

Als Anwendung dieser grundlegenden Techniken wurden Pfadintegral Molekulardynamik (MD) Simulationen von Wasser untersucht. Mithilfe dieser Methode ist es möglich, quantenmechanische Effekte der Kerne (Delokalisation, Nullpunktsenergie) in die Simulation einzubeziehen. Hierbei wurde zuerst eine Multi-Skalen Technik (Force-matching) verwendet um eine effektive Wechselwirkung aus einer detaillierten Simulation auf Basis der Dichtefunktionaltheorie zu extrahieren. Die Pfadintegral MD Simulation verbessert die Beschreibung der intra-molekularen Struktur im Vergleich mit experimentellen Daten. Das Modell eignet sich auch zur gleichzeitigen Kopplung in einer Simulation, wobei ein Wassermolekül (beschrieben durch 48 Punktteilchen im Pfadintegral-MD Modell) mit einem vereinfachten Modell (ein Punktteilchen) gekoppelt wird. Auf diese Weise konnte eine Wasser-Vakuum Grenzfläche simuliert werden, wobei nur die Oberfläche im Pfadintegral Modell und der Rest im vereinfachten Modell beschrieben wird.

Contents

VIII

List of Figures

List of Tables

Related Publications

[1] S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer. Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir. *Physical Review Letters*, 108(17):170602, 2012.

[2] S. Fritsch, C. Junghans, and K. Kremer. Structure formation of toluene around c60: Implementation of the adaptive resolution scheme (AdResS) into GROMACS. *Journal of Chemical Theory and Computation*, 8(2):398–403, 2012.

[3] R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio. Hamiltonian adaptive resolution simulation for molecular liquids. *(manuscript under review)*.

Chapter 1

Introduction

The physical properties of complex fluids often arise from an interplay between phenomena occurring at different scales, ranging from a quantum mechanical description, classical mechanics, up to (mesoscale) hydrodynamics and thermodynamics.

In recent decades, it has become routine to investigate such systems with the aid of computer simulations [4]. Such "computer experiments" can be seen as a third way to study nature, besides experiments and analytical theory. The advantages are especially apparent for complex systems, where a large number of interconnected degrees of freedom must be described simultaneously.

In many cases, however, limited computational resources force the researcher to compromise on the level of detail in order to simulate long enough time and length scales. For polymeric fluids, for example, the timescales at the atomistic level are characterized by the bond vibrations $\tau \approx 10^{-13}$ s, while typical relaxation times of the overall chain are of the order of $\tau \approx 10^{-7} - 10^{-4}$ s [5]. If the bond vibrations need to be resolved in the simulation, the relaxation time can hardly be reached. For inferring generic properties, such as scaling behavior, it is often advisable to neglect chemical details and build a simple model which contains the most important physical aspects [6].

In other cases, the chemical details need to be incorporated. The properties of water, for example, depend crucially on its ability to form hydrogen bonds, which establishes a near-ordering in the liquid. Despite being the most studied of all liquids [7], there is still debate about many issues in liquid water, one of them being the exact hydrogen bonding arrangement. This fact was highlighted by Science magazine, which listed the determination of the exact structure of water as one of the remaining unanswered questions with high impact for research [8].

To arrive at a model which includes all the "a priori" knowledge we have about a system and which still is simple enough to be simulated efficiently, a way of transferring properties from one scale to another is needed. For example, in water, we know that electronic effects play an important role, which are described by

3

Shrödinger's equation. It is, however, computationally very demanding to treat larger systems with full electronic interactions everywhere (even though efficient approximate methods, such as the density functional theory (DFT) [9] are available). Therefore, often classical effective models are constructed by fitting a simple model potential to either quantum simulations at the level of Shrödinger's equation or to experimentally observed properties.

For some applications even an classical atomistic model is not efficient enough and a simpler model at the level of "super-atoms" (or "coarse-grained" beads) is necessary. This approach has especially been explored for biomolecules and polymer melts [10, 11, 12, 13, 14, 15]. Again, a method is needed to transfer properties from the detailed to the coarse level of description. Different strategies are adopted, which focus on reproducing the structure, forces or thermodynamic properties of the underlying atomistic system. Such techniques are often called *hierarchical* multi-scale techniques: First, the atomistic model is simulated as to to construct a "coarse-grained" model. In a second step, the "coarse-grained" model is then used to build larger systems which can be simulated with less computer time.

In many cases the "coarse-grained" model indeed enables larger time and length scales to be reached, but some physical properties are lost. One possibility is then to re-introduce atomistic degrees of freedom in a "back-mapping" procedure [16]. There are, however, systems where the process studied needs a high level of detail locally during the whole simulation. A classic example for this is crack propagation in a solid, where the breaking of the bonds locally needs quantum mechanical details. For these situations *simultaneous* multi-scale methods have been developed, which describe one region in space according to a more detailed model and the remainder using a simplified "coarse-grained" model. The simulation scheme then has to provide a way of coupling the different sub-regions, whereby boundary conditions inferred from physical arguments need to be defined. Such approaches have been established for solids [17, 18, 19], quantum-classical coupling of electronic systems (QM/MM) [20, 21] and recently also for liquids [22, 23, 24]. Due to the importance of fluctuations in liquids, the coupling needs to be performed such that molecules can leave and enter either region, which poses special challenges to the simulation scheme. The commonly used QM/MM schemes do not allow for such fluctuations as molecules have to be assigned to be represented by either a quantum or classical model. Furthermore, the fact that the "coarse-grained" models can often not reproduce all relevant thermodynamic properties requires a coupling which ensures that the simultaneous coupling is performed at the state point of interest.

1.1 Outline

In chapter 1, the fundamental aspects of molecular dynamics simulations and re-
lated coarse-graining techniques are presented. The adaptive resolution scheme
(AdResS) is introduced, which allows for the simultaneous treatment of atomistic
and coarse-grained models based on force interpolation.

Chapter 2 introduces a method for ensuring the correct thermodynamic bound-
ary conditions in AdResS simulation through a "thermodynamic force". The
method is tested for an atomistic model of liquid water coupled to a coarse-grained
water model.

Chapter 3 describes adaptive simulations where the ability to preserve the near-
ordering in a C60-fullerene embedded in a toluene solvent is tested. The most
important technical aspects of the implementation of the AdResS scheme developed
within this work are outlined.

In chapter 4 a combination of hierarchical methods is used to obtain a water
model based on quantum-level simulations (DFT) and its representations on the
all-atom and coarse-grained level. It is then demonstrated that this can be used
to simulate a water-vacuum interface where the interface region is treated within
the path-integral MD picture (and hence including nuclear quantum effects) and
the bulk liquid as a coarse-grained system.

Chapter 5 describes the design and testing of an improved adaptive resolution
method, which is based on a Hamiltonian interpolation scheme. This new scheme
(H-AdResS) can be used to run energy-conserving adaptive simulations and allows
for a straightforward application of standard concepts from statistical mechanics
to derive the appropriate thermodynamic coupling conditions.

Chapter 2

Theory

Typical condensed and soft matter applications deal with inferring properties for a system consisting of a large number of particles, which altogether form the thermodynamic state of the system. The connection between the microscopic state (i.e. a set of atoms described by their positions and momenta) and the macroscopic state (i.e. pressure, density, heat capacity, etc.) is given by the laws of statistical mechanics. In molecular dynamics simulations, the aim is to use computers to calculate the trajectory for a sufficiently large number of particles in order to infer information on the macroscopic properties. Even with the power of present day computers, which are able to simulate a few ten-thousand of particles for up to microseconds [25], the time- and length scales available are still small compared to macroscopic sizes (where the number of particles is typically $\approx 10^{23}$).

In the following section, the fundamental physical principles which are used in molecular dynamics simulations are briefly outlined. Subsequently, techniques for systematic coarse-graining are discussed, which aim to extend the time and length scales available in simulations. Finally, the adaptive resolution scheme is introduced, which is able to concurrently couple different scales.

2.1   Molecular Dynamics Simulations

The purpose of Molecular Dynamics (MD) simulations is to generate a trajectory, given a set of particles and their interaction potentials. From this trajectory, dynamical properties or averages can be extracted. These results can be related to experimental measurements or be used for example to predict properties of new materials.

A system consisting of N particles is described by a set of point particles with positions $\mathbf{r}_i$ and momenta $\mathbf{p}_i$. The set of particle positions can be formally denoted as a 3N dimensional vector

$$\mathbf{r}^N \equiv \{\mathbf{r_1}, \ldots, \mathbf{r_N}\} \tag{2.1}$$

for the particle positions and

$$\mathbf{p}^N \equiv \{\mathbf{p_1}, \ldots, \mathbf{p_N}\} \tag{2.2}$$

for the momenta, spanning a 6N dimensional phase space. What would be the observable in an experiment, can be expressed as an ensemble average in classical statistical mechanics

$$\langle A \rangle = \int f(\mathbf{r}^N, \mathbf{p}^N) A(\mathbf{r}^N, \mathbf{p}^N) \mathrm{d}\mathbf{r}_1, \ldots \mathrm{d}\mathbf{r}_N, \mathrm{d}\mathbf{p}_1, \ldots \mathrm{d}\mathbf{p}_N \ , \tag{2.3}$$

where A is an observable and $f(\mathbf{r}^N, \mathbf{p}^N)$ a weighting function which is different for each ensemble. Given that the system is ergodic, i.e. all points in the phase space are visited during the time evolution, it is possible to write this integral as a time average[26]

$$< A > = < A >_t = lim_{\tau \to \infty} \int_0^\tau A(\mathbf{r}_1(t), \ldots, \mathbf{r}_N(t), \mathbf{p}_1(t), \ldots, \mathbf{p}_N(t)) \mathrm{d}t \ . \tag{2.4}$$

The assumption of ergodicity is very strong. In practice it has to be carefully checked whether the simulated time τ was sufficient to represent the ensemble average

In classical mechanics, the dynamics is completely specified by the 6N initial conditions (3N positions $\mathbf{r}_i^0$ and 3N momenta $\mathbf{p}_i^0$) and the Hamiltonian which has the form [27]

$$H = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}^N) \ . \tag{2.5}$$

The microcanonical ensemble describes the situation of an isolated system with a fixed number of particles N in a Volume V at a total energy E. The corresponding weighting function appearing in the average eq. 2.3 is

$$f(\mathbf{r}^N, \mathbf{p}^N) = \delta(H(\mathbf{r}^N, \mathbf{p}^N) - E) \ . \tag{2.6}$$

The Hamiltonian gives rise to the microscopic dynamics through the canonical equations of motion, which for the Hamiltonian in eq. 2.5 simplifies to

$$\frac{d\mathbf{r}_i}{dt} = \frac{\partial H}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m} \ , \tag{2.7}$$

for the positions and

$$\frac{d\mathbf{p}_i}{dt} = -\frac{\partial H}{\partial \mathbf{r}_i} = -\nabla_{\mathbf{r}_i} U(\mathbf{r}^N) \ , \tag{2.8}$$

for the momenta. In order to numerically determine the time average eq. 2.4 in the microcanonical ensemble, the system is evolved according to the canonical equations.

2.1.1 Integration

The set of first order differential equations in eqs. 2.7 and 2.8 can be solved approximately using the velocity-verlet algorithm [28]. A Taylor expansion (discarding terms of the order Δt^3) with respect to the coordinates gives

$$\mathbf{r}_i(t + \Delta t) \approx \mathbf{r}_i(t) + \Delta t \mathbf{v}_i(t) + \frac{\Delta t^2}{2}\mathbf{a}_i(t) \tag{2.9}$$

and for the momenta

$$\mathbf{p}_i(t + \Delta t) = \mathbf{p}_i(t) + \Delta t \dot{\mathbf{p}}_i(t) + \frac{\Delta t^2}{2}\ddot{\mathbf{p}}(t) + \mathcal{O}(\Delta t^3) \tag{2.10}$$

$$\approx \mathbf{p}_i(t) + \frac{\Delta t}{2}(\dot{\mathbf{p}}_i(t + \Delta t) + \dot{\mathbf{p}}_i(t)) \,, \tag{2.11}$$

where an additional Taylor expansion of $\dot{\mathbf{p}}(t + \Delta t)$ neglecting terms of the order Δt^2 has been used to eliminate the unknown $\ddot{\mathbf{p}}$. This algorithm has the advantage of evaluating the positions and momenta at the same time, whereas for example the leap frog algorithm [29] needs to evaluate the velocities at time $t + \frac{1}{2}\Delta t$ [30]. Practically on a computer, the algorithm can be performed as follows:

1. Calculate positions $\mathbf{r}_i(t+\Delta t)$ using velocities and momenta at time t (eq. 2.9).

2. Calculate accelerations $\mathbf{a}_i(t + \Delta t)$ from the interaction forces which depend on positions $\mathbf{r}_i(t + \Delta t)$.

3. Calculate momenta $\mathbf{p}_i(t + \Delta t)$ (eq. 2.11).

Note that the simulation is started with positions and momenta as input. It can be easily shown that the velocity-verlet algorithm is time reversible and thus preserves an important property of the Hamiltonian [30].

The dynamics generated on a computer are however very sensitive on the initial conditions. This can be seen from the fact that two trajectories, started from very close points in phase space, diverge exponentially fast, which can be characterized by so called Lypanov exponents [4]. Hence to predict the exact trajectory of a starting point, the initial conditions have to be known with infinite precision. However, in most cases one is not interested in knowing the exact time propagation of a given initial condition, but instead in a statistical average which is well represented on long time scales.

2.1.2 Thermostats

The microcanonical ensemble, which describes an isolated system at constant energy E, is usually not the most appropriate ensemble to work in, given that experiments usually consider a subsystem which exchanges energy with its surroundings. If the subsystem is small compared to the surrounding reservoir, the energy

exchange imposes a temperature (as described in the canonical ensemble). In principle one could just explicitly include the reservoir in the simulation, but this would consume a lot of computational resources as the reservoir needs to be large compared to the sub-system of interest. Instead a thermostating algorithm is employed which models the heat transferred between the system of interest and the heat bath.

2.1.2.1 Langevin thermostat

The Langevin thermostat is inspired by the Brownian motion of a heavy particle immersed in a bath. For heavy particles, the frequent collisions with the solvent particles can be represented by a random force and a friction term which is proportional to the velocity. The equations of motion for the particle in a bath are [31]

$$\frac{dv(t)}{dt} = -\frac{\gamma}{m}v(t) + \frac{1}{m}\xi(t) \ , \tag{2.12}$$

$$\frac{dr(t)}{dt} = v(t) \ , \tag{2.13}$$

where γ is a friction constant and $\xi(t)$ a stochastic random variable such that $< \xi(t) > = 0$. The equations are written here for simplicity as one-dimensional equations, but the extension to higher dimensions is straightforward. The noise term is delta correlated in time:

$$\langle \xi(t_1)\xi(t_2) \rangle = g\delta(t_2 - t_1) \ . \tag{2.14}$$

where the strength g is a yet to be determined factor. In order to find g one can explicitly integrate eq. 2.12, which yields [31]

$$v(t) = v_0 e^{-\frac{\gamma}{m}t} + \frac{1}{m}\int_0^t ds \ e^{-\frac{\gamma}{m}(t-s)}\xi(s) \ . \tag{2.15}$$

From this solution, it is then possible to compute the autocorrelation function, using the fact that $\langle v_0\xi(t) \rangle = 0$ and eq. 2.14

$$\langle v(t_2)v(t_1) \rangle = v_0^2 e^{-\frac{\gamma}{m}(t_2+t_1)} \tag{2.16}$$

$$+ g\int_0^{t_2} ds_2 \int_0^{t_1} ds_1 \delta(s_2 - s_1)e^{\frac{\gamma}{m}(s_1-t_1)}e^{\frac{\gamma}{m}(s_2-t_2)} \tag{2.17}$$

$$= \left(v_0^2 - \frac{g}{2m\gamma}\right)e^{-\frac{\gamma}{m}(t_2+t_1)} + \frac{g}{2m\gamma}e^{-\gamma(t_2-t_1)} \ . \tag{2.18}$$

In equilibrium, this velocity autocorrelation needs to be stationary, i.e. it cannot have an explicit dependence on time. Hence the first term in eq. 2.18 (containing

$t_2 + t_1$) must vanish, which is only true if

$$v_0^2 = \frac{g}{2m\gamma} \ . \tag{2.19}$$

In the canonical ensemble, in equilibrium, each degree of freedom has a kinetic energy according to the equipartition theorem

$$\frac{m}{2} \left\langle v^2 \right\rangle = \frac{1}{2} k_B T, \tag{2.20}$$

where k_B is the Boltzmann constant and T the target temperature. This has to be valid at all times, and hence, the v_0 can be taken to be the average velocity. Inserting the average velocity into eq. 2.19 yields:

$$g = 2\gamma k_B T \ . \tag{2.21}$$

Thus, the temperature T fixes the ratio between the friction γ and the noise strength g (Fluctuation-Dissipation theorem). A more formal way to arrive at this result is given in the Fokker-Planck picture [32].

For some applications (e.g. analytical models of large Brownian particles), one can use the Stokes friction $\gamma = 6\pi\eta R$ and model for example diffusion processes (Brownian motion) without explicitly considering the solvent. When used as a thermostat in MD simulations, the noise and friction terms are added to the forces acting on each degree of freedom. The only free parameter is then the friction constant γ, which needs to be carefully chosen(in many publications, the characteristic time $\tau_{\text{Langevin}} = 1/\gamma$ is specified instead of γ). This is also applied to systems where the solvent is modeled explicitly. The "random kicks" on each atom can be imagined to originate from the reservoir controlling the temperature.

The strength of the friction constant determines the time the system takes to arrive at the target temperature T. If the initial temperature deviates from the equilibrium temperature, a strong friction constant quickly relaxes the velocities. However, dynamical properties (such as the self-diffusion constant) measured from such a simulation, are significantly altered from the realistic values. Furthermore, hydrodynamic correlations between particles, which occur on timescales much larger than the average time between collisions, are lost [33]. As a consequence, when interpreting dynamical quantities, it should be ensured that the coupling was sufficiently weak. The friction coefficient is commonly chosen proportional to the mass of a particle $\gamma_i' = m_i\gamma$ in order to have the same acceleration from the random forces for each atom [34].

One appealing feature of the method is that it always generates a canonical ensemble (in contrast to other thermostating methods). Moreover, the Langevin thermostat is local, meaning that the force on each atom does not depend on the other degrees of freedom.

2.1.2.2   Other thermostats

The problem of generating a canonical trajectory without perturbing the dynamics in the simulation has a long history in literature [34]. The thermostats can be categorized to belong either to the class of global or local thermostats. In global thermostats, the kinetic energy of a molecule is coupled to all other degrees of freedom while for local thermostats no such coupling is present.

The most simple approach is based on a rescaling of velocities at each time step according to the target temperature, which is done for randomly selected atoms in the Anderson thermostat [35]. In the Berendsen [36] thermostat, a weak-coupling relaxation scheme is employed to reach the target temperature. In its original form, this does not give the correct canonical fluctuations, but a stochastic term can be added to eliminate this deficiency [37]. A different class of thermostats starts from constructing an "extended system". The Hamiltonian is extended by one or more degrees of freedom which represent the reservoir, which controls the temperature by exchanging energy with the other degrees of freedom ("Nosé-Hoover thermostat") [38, 39]. A local stochastic thermostat which also preserves the linear momentum and thus leads to the correct hydrodynamics is the DPD thermostat [40, 33, 41].

All those algorithms allow to sample a canonical ensemble, but perturb the dynamics in some way. Therefore, the results in terms of dynamics have to be analyzed carefully as to exclude a strong influence of the thermostat.

2.1.3   Model potentials

In order to solve the equations of motion, knowledge of the interaction potential $U(\mathbf{r}^n)$ is required. There are several strategies for finding an appropriate potential depending on the required accuracy and context.

In general, $U(\mathbf{r}^n)$ is a multi-body potential, where the potential energy of each atom depends on all degrees of freedom. To simplify the the intermolecular interactions, in many liquids the assumption of pairwise, radial symmetric interactions is often made, i.e. $U(\mathbf{r}^n) = \sum_{ij} U(|\mathbf{r}_i - \mathbf{r_j}|)$. A commonly used form of the interaction potential is the Lennard-Jones type:

$$U_{LJ}(r_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] , \qquad (2.22)$$

where r_{ij} is the distance between atoms i and j and σ_{ij}, ϵ_{ij} are coefficients.

Electrostatic effects are included using "partial" charges which refer to using non-integer charge coefficients q_i in the Coulomb potential $U_C(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon r_{ij}}$ (see next section 2.1.4).

The chemical bonds in molecules are either treated as rigid using a constrained integration [42, 43, 44] or modeled as a combination of linear (two-body), angular

(three-body), and dihedral (four-body) potentials. The bond potential is often simply modeled as a harmonic potential

$$U_b(r_{ij}) = \frac{1}{2}k_{ij}^b(r_{ij} - b_{ij})^2 \ , \qquad (2.23)$$

which takes the bond length b_{ij} and the force constant $k_{r_{ij}}^b$ as parameters. For polymers, it is often useful to limit the bond distance to a maximum value as to avoid crossing of chains, which can be achieved by the FENE potential [6]. For the three-body interaction, the harmonic angular term is frequently used

$$U_\theta(\theta_{ijk}) = \frac{1}{2}k_{ijk}^\theta(\theta_{ijk} - \theta_{ijk}^0)^2 \ . \qquad (2.24)$$

In general the unknown parameters can be obtained by fitting to known experimental data (e.g. density, heat capacity, etc.) or to calculations based on more accurate physical models (for example solving the Schrödinger equation)[45]. For bio-molecular contexts several groups have developed sets of parameters bundled in so called "force-fields" (e.g. OPLS[46], CHARMM[47], AMBER[48]).

In most cases, periodic boundary conditions are used[29]. This means that the simulation volume represents a unit cell which is thought to be replicated an infinite number of times. Particles which are out of the unit cell after the Verlet integration step are placed back on the opposite side of the unit cell. The interactions are then calculated using the "minimum image convention": Each particle interacts only with the closest replica, which can be computed easily from geometrical considerations.

The evaluation of the pairwise terms can become computationally untraceable due to the scaling $\propto N^2$. Therefore, the sum is often performed such that only terms with $r_{ij} < r_{\rm cut}$ are included. The cutoff $r_{\rm cut}$ then defines a sphere within which a particle interacts. This approximation however can cause serious artifacts, among which is a drift in the total energy due to the discontinuous forces at $r_{\rm cut}$. Hence, a modified potential which is shifted to be zero at the cutoff is often used[29].

2.1.4 Electrostatic interactions

Electrostatic interactions are difficult to handle in MD simulations because of the slow decay with $\frac{1}{r}$. If the interaction of charges which are further then the cut-off distance r_{cut} away from an atom i are simply neglected, the Coulomb energy reads

$$U_{\rm Coulomb}^{\rm direct} = \sum_{i=1}^{N_{atom}} \sum_{\substack{j<i \\ r_{ij}<r_{cut}}}^{N_{atom}} \frac{q_i q_j}{r_{ij}} \ , \qquad (2.25)$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ denotes the distance between the atoms i and j. The truncation at r_{cut} can cause artifacts to occur due to the remaining net charges on the

13

surface of the sphere defined by the cutoff radius. To overcome this, many different techniques have been used in the past. The most frequently used technique is the Ewald-summation and variants thereof [29]. In the Ewald-summation, a charge interacts with all its periodic images. This interaction can be calculated efficiently by splitting the interaction in short- and long-ranged parts and solving the latter in Fourier space. The Ewald-method has rarely been used in this thesis due to unsolved difficulties with long range electrostatics in the AdResS scheme, which will be explained in more detail later. Instead, the reaction-field technique [49, 50] was mostly used.

In the reaction-field technique, the cutoff radius r_{cut} defines as spherical cavity with volume $V_{r_{cut}}$ centered on molecule α within which charges interact explicitly. Outside the cavity, a dielectric medium with a dielectric constant ϵ_{RF} is assumed. The charges within the cavity polarize the surrounding medium, which in turn produces a field. This field can be shown to be of the form [51]

$$\mathbf{E}_\alpha = \frac{(2\epsilon_{RF} - 1)}{2\epsilon_{RF} + 1} \frac{1}{r_c^3} \sum_{\beta \in V_{r_\alpha}^{cut}} \boldsymbol{\mu}_\beta \qquad (2.26)$$

where $\boldsymbol{\mu}_\beta = \sum_i q_{\beta_i} \mathbf{r}_{\beta_i}$ is the dipole moment of a molecule β and $V_{r_\alpha}^{cut}$ denotes a sphere with radius r_{cut} centered on molecule α. The contribution to the potential energy from the interaction with the reaction field for a molecule α is $-\frac{1}{2}\boldsymbol{\mu}_\alpha \mathbf{E}_\alpha$ and hence the total electrostatic energy is

$$U_{RF} = -\frac{1}{2} \sum_{\alpha=1}^{N_{molecules}} \boldsymbol{\mu}_\alpha \cdot \mathbf{E}_\alpha + U_{\text{Coulomb}}^{\text{direct}} , \qquad (2.27)$$

where the first sum runs over all molecules. The only unknown is the dielectric constant ϵ_{RF} which has to be determined e.g. from preliminary simulations employing long-range electrostatics. The force due to charges used in the simulation then simply follow from $\mathbf{F}_i = -\nabla_{r_i} U_{RF}$.

It has to be noted that the reaction-field technique can lead to problems with energy conversion as jumps in the energy occur whenever a molecule enters or leaves the cavity [29]. Furthermore, the assumption of a constant dielectric surrounding is only valid for a homogeneous system. An advantage over the Ewald-technique is however the reduced computational cost.

2.2 Properties of liquids

The presence of near-order and the importance of inter-molecular interactions are characteristic features of liquids. In contrast to solids, where the potential energy is much larger than the kinetic energy and gases, where the potential energy is much smaller than the kinetic energy, both terms are comparable in liquids [26].

To characterize the liquid structure, pair correlation functions are frequently used. The pair correlation function, which is also called radial distribution function (RDF), is defined in the canonical ensemble for homogeneous systems as [29]

$$g(\mathbf{r}_1, \mathbf{r}_2) = \frac{\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2)}{\rho^2} \ , \tag{2.28}$$

where ρ is the density and $\rho(\mathbf{r}_1, \mathbf{r}_2)$ the 2-particle density defined as

$$\rho^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{N(N-1)}{Z} \int \exp[-\beta U(\mathbf{r}^n)] d\mathbf{r}_3, \ldots, d\mathbf{r}_N \ . \tag{2.29}$$

The factor $N(N-1)$ reflects that the particles r_1 and r_2 can be chosen arbitrarily. Equation 2.29 thus describes the probability density of finding a particle at $\mathbf{r}_2$ given a particle at $\mathbf{r}_1$. For isotropic systems, the RDF depends only on the distance of particles and is thus often written as $g(r)$, which can be computed from simulation as

$$g(r) = \frac{V}{N^2} \left\langle \sum_i \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle \ , \tag{2.30}$$

where the the brackets denote the ensemble average.

For most liquids this function shows a peak at low r values, which represents the next neighbors (on average). In the limit of $r \to \infty$ it approaches the ideal gas value $g(r) \to 1$. In case of pairwise additive interactions and a homogeneous, isotropic system, some thermodynamic properties are completely specified by the RDF. For example the excess energy U^{ex}, which is defined as the difference in internal energy $U^{ex} = U - U^{id} =$ with respect to the ideal gas value U^{id} is: [26]

$$U^{ex}/N = 2\pi\rho \int_0^\infty U(r)g(r)r^2 dr \ . \tag{2.31}$$

The $g(r)$ is also related to the pressure and the compressibility, which will be made use of in section 2.3.4.2.

2.3 Coarse-graining

MD simulations usually treat the system of interest with atomistic details, meaning that each atom is represented by a point particle. The cost for evaluating the interactions on a computer increases with the number of atoms and limits the length and time scales available. Coarse-graining (CG) techniques aim at a simplified, effective description with less degrees of freedom. These simpler models can then be simulated using significantly less computational resources and thus bring larger length and timescales within reach. This usually means to replace a subset of

atoms by a single "superatom" or CG bead. The interaction potentials of these CG beads are unknown a priori, and have to be determined within a numerical scheme. Other CG exist which aim for a continuum [17], or lattice [52] based description.

The central question is how to preserve the physical properties of interest in the coarse-grained system. These properties of interest can be structural properties, such as pair correlation functions, free energies or other thermodynamic properties. [53]

This requirement dictates the choice of mapping from the all-atom to the CG representation, but also the method used to obtain the CG integration potential. Formally, the question can be posed as follows: Given an atomistic Hamiltonian

$$H^{\text{AA}} = \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \ldots, \mathbf{r}_n) \;, \tag{2.32}$$

which spans a $6n$ dimensional phase space, we are looking for and effective Hamiltonian

$$H^{\text{CG}} = \sum_{\alpha=1}^{N} \frac{\mathbf{p}_\alpha^2}{2M_\alpha} + U^{\text{CG}}(\mathbf{R}_1, \ldots, \mathbf{R}_N) \;, \tag{2.33}$$

where the $\mathbf{R}_1, \ldots, \mathbf{R}_N$ denote the CG particles with masses M_α and momenta $\mathbf{p}_\alpha$. The CG Hamiltonian is only $6N$ dimensional ($N < n$) and preserves certain features of the system of interest. How the CG positions are connected to the underlying all-atom Hamiltonian is discussed next.

2.3.1 Mapping

The mapping scheme is a 'geometric rule' defining how the CG beads are constructed from the underlying all-atom simulation. The most intuitive rule is to use the center of mass of a subset of atoms to represent a CG bead. This kind of mapping was used throughout this thesis, however, other mapping schemes (e.g. representing an entire water molecule by its oxygen position [54]) are generally possible. Note that for big molecules (e.g. polymers) it is desirable to represent one molecule by multiple CG beads, depending on the local chemical details which need to be preserved. In this case, the CG model also has to include an effective potential for the internal interactions of the CG beads [11] . In the following paragraphs, only simple mappings of one molecule to one CG bead will be discussed. It is thus convenient to denote the distances between CG particles by $R \equiv R_{\alpha\beta} = |\mathbf{R}_\alpha - \mathbf{R}_\beta|$ and the CG configuration by

$$\mathbf{R}^N \equiv \{\mathbf{R_1}, \ldots, \mathbf{R_N}\} \;. \tag{2.34}$$

The connection between all-atom and CG systems is formally given by application of the mapping operator [55]

$$\mathbf{M}(r^n) = \{\mathbf{M}_1(\mathbf{r}^n), \dots, \mathbf{M}_N(\mathbf{r}^n)\} \ . \tag{2.35}$$

For the case of a center of mass mapping, each component is given by

$$\mathbf{M}_\alpha(r^n) = \frac{\sum_{i=1}^{n_\alpha} m_{\alpha_i} \mathbf{r}_{\alpha_i}}{\sum_{i=1}^{n_\alpha} m_{\alpha_i}} \ , \tag{2.36}$$

where the index i denotes n_α different atoms within a molecule α. The atomic masses are denoted as m_{α_i}.

2.3.2 Potential of mean force

Any CG model should fulfill certain consistency requirements with regard to the all-atom Hamiltonian it is constructed from. In the optimal case, properties measured in the CG simulation give identical results to the atomistic simulation. This can be enforced by matching the phase-space probabilities of the all-atom and CG system. The consistency condition in coordinate space is thus written as matching the probabilities of finding a phase space point in the atomistic (p_r) and CG system (P_R) [55]:

$$P_R(\mathbf{R}_1, \dots, \mathbf{R}_N) = p_r(\mathbf{R}_1, \dots \mathbf{R}_N) \ . \tag{2.37}$$

This directly implies

$$\exp(-\beta U(\mathbf{R}_1, \dots, \mathbf{R}_N)) \propto \int \exp(-\beta U(\mathbf{r}_1, \dots, \mathbf{r}_n)) \delta(\mathbf{M}(\mathbf{r}_1, \dots, \mathbf{r}_n) - \mathbf{R}^N) d\mathbf{r}^n \ , \tag{2.38}$$

where $\mathbf{R}^N$ denotes a $3N$ dimensional vector of the CG coordinates. An equivalent condition can be written for the momentum space. Taking the logarithm on both sides of eq. 2.38, the potential on the left hand side is a multi-dimensional potential of mean force (PMF), which has the form of a free energy. The PMF $U(R_1, \dots, R_N)$ depends on all CG degrees of freedom, and hence is hard to evaluate. Even if it was available, in the CG simulation, the calculation of the force at each time step would involve all the CG degrees of freedom and thus be too inefficient.

Therefore, the multi-body PMF represents only a formal solution of the problem. All the methods introduced in the following paragraphs try to approximate the multi-body PMF by finding a two-body effective potential. In the literature, there exist also methods considering three-body CG potentials [56], which where not used in this thesis.

2.3.3 Boltzmann inversion

A very simple guess of the effective potential can be made by inverting the corresponding distribution function [11] measured in the all-atom simulation. In the canonical ensemble, an independent degree of freedom q is distributed as $p(q) \propto e^{-\beta U(q)}$. An effective potential can thus simply be calculated as

$$U(q) = -\mathrm{k_B}T \ln p(q) \ , \tag{2.39}$$

where the distribution function $p(q)$ is as a normalized histogram in the reference all-atom simulation. Hence the potential $U(q)$ is the inverted Boltzmann distribution, which gives the scheme its name. In practice this technique is mostly used for internal interactions (e.g. bonds, angles, etc). When using these, the appropriate normalization factors have to be included, i.e.

$$U_r = -\mathrm{k_B}T \ln \left[H_r(r)/(4\pi r^2) \right] \ , \tag{2.40}$$

for bonds and

$$U_\theta = -\mathrm{k_B}T \ln \left[H_\theta(\theta)/\sin\theta \right] \ , \tag{2.41}$$

for angular interactions [57]. However, in many cases the degrees of freedom cannot be taken to be independent, which is why more sophisticated approaches have to be taken.

2.3.4 Iterative Boltzmann inversion

The iterative Boltzmann inversion (IBI) method [12] is based on the relation between the pair distribution function $g(r)$ and the 2-body potential of mean force $w(r)$. The latter corresponds to the reversible work of pulling two arbitrary particles in the system from infinite separation to a distance r [58]:

$$w(r) = -k_B T \ln g(r) \ . \tag{2.42}$$

As a first step, the initial guess for the CG is thus simply taken to be the inversion of the all-atom distribution function:

$$U_{CG}^0(R) = -k_B T \ln g_{\mathrm{AA}}(R) \ , \tag{2.43}$$

where $g_{\mathrm{AA}}(R)$ is the RDF of the CG beads calculated in the all-atom reference simulation. The resulting potential $U_{CG}^0(R)$ is then used in a CG simulation from which the distribution function $g^1(R)$ is calculated. This new distribution function will however usually deviate from the reference.

To correct for this deviation, the approach taken in iterative Boltzmann inversion is to iteratively correct the CG potential based on the difference of distributions [12]. Thus an update is constructed which reads:

$$\Delta U^{(i)}(R) = w^{(\mathrm{AA})}(R) - w^{(i)}(R) = k_B T \ln[g^{(i)}(R)/g_{\mathrm{AA}}(R)] \ . \tag{2.44}$$

This is then used to correct the CG potentials in step $i + 1$:

$$U_{\mathrm{CG}}^{(i+1)}(R) = U_{\mathrm{CG}}^{(i)}(R) + \Delta U^{(i)}(R) \ . \tag{2.45}$$

The iteration is considered to be converged if the difference between the 2-body PMFs small. A suitable criterion can be defined as

$$\Xi^{(i)} \equiv \int (g^{(i)}(R) - g^{(\mathrm{AA})}(R))^2 dR \ , \tag{2.46}$$

where the iteration is converged once $\Xi^{(i)}$ has dropped below some chosen value. The iteration scheme is outlined in fig. 2.1. The CG potential from a converged iterative Boltzmann inversion thus closely reproduces the 2-body-PMF of an atomistic simulation. It can be seen from the explicit appearance of the temperature in the IBI equations, that the CG potentials will depend on a specific state point. Thus it is necessary to perform the IBI procedure for each state point of interest. To understand the need for determining iteratively the CG potential, we can set coarse-graining aside and consider a simple mono-atomic liquid interacting through a pairwise potential. Given the corresponding structure $g(r)$, is it possible to reconstruct the interaction potential $U(r)$? This problem is also known as the inverse problem of statistical mechanics [53].
The full relationship is given by a density expansion[26]

$$g(r) = exp[-\beta U(r)](1 + \sum_{n=1}^{\infty} \rho^n g_n(r)) \ , \tag{2.47}$$

where the g_n are coefficients of the power series from a density expansion of $g(r)$ and ρ is the density. The fact that eq. 2.47 and eq. 2.43 are only equivalent for the limit $\rho \to 0$ already indicates that the 2-body PMF eq. 2.42 used as potential gives the correct distribution only when 3-body and higher order correlations are unimportant. This is the case in the limit of zero density. In the case of coarse-graining the situation can be even more complex because of the additional reduction in degrees of freedom. No exact relation as in eq. 2.47 can be derived and hence the CG potential is approximated numerically.

The IBI procedure iteratively converges to a solution which reproduces, within specified accuracy, the target structure $g(R)$. Hence, the two-body PMF between two CG particles matches the reference all-atom two-body PMF. The effective potential $U^{CG}(R)$ which generates this structure is unique up to a constant. This is known as the Henderson theorem, which is discussed next.

2.3.4.1 The Henderson theorem

The IBI technique relies entirely on the knowledge of the $g(r)$ to calculate a CG potential. Weather a unique potential exists, which generates a given $g(r)$ was
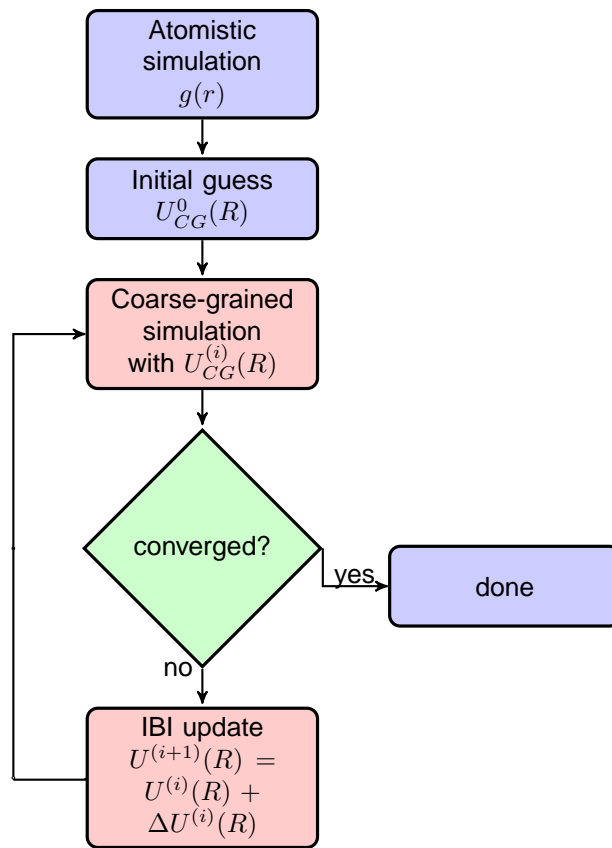
Figure 2.1: Flow chart of the IBI update procedure

answered by Henderson in 1974 [59, 53]: The Henderson theorem states that given a $g(r)$, the pairwise potential generating this g(r) is unique up to a constant. The proof is based on the Gibbs-Bogoliubov inequality, which states for the free energies F_1 and F_2 of two Hamiltonians H_1 and H_2:

$$F_1 \leq F_2 + \langle H_2 - H_1 \rangle_1 \; , \tag{2.48}$$

where

$$\langle H_2 - H_1 \rangle_1 = \int \mathrm{d}r^n \mathrm{d}p^n e^{-H_1(r^n, p^n)/k_\mathrm{B}T} (H_2(r^n, p^n) - H_1(r^n, p^n)) \; , \tag{2.49}$$

is the canonical average at the configurations compatible with H_1. The equality in eq. 2.48 only holds if $H_2 - H_1$ is independent of all coordinates. Henderson's theorem can be proven by constructing a contradiction: Assume two identical systems with different pair potentials u_1 and u_2, but identical RDFs $g_1(r) \equiv g_2(r)$. Now if $u_1 - u_2$ is not constant, equality in eq. 2.48 does not hold and thus

$$f_1 < f_2 + \frac{1}{2}\rho \int \mathrm{d}r[u_2(r) - u_1(r)]g_1(r) \; , \tag{2.50}$$

where f_i is the free energy per particle in each system and ρ the particle density. The inequality must also hold if the labels of the two systems are exchanged

$$f_2 < f_1 + \frac{1}{2}\rho \int \mathrm{d}r[u_1(r) - u_2(r)]g_2(r) \; . \tag{2.51}$$

However, inserting the assumption $g_1(r) \equiv g_2(r)$ and adding eq. 2.50 and eq. 2.51 yields $0 < 0$. Hence the assumption that $u_1 - u_2$ is not constant must be wrong. For a system of identical particles and assuming a pairwise interaction potential, it is thus proven that if a solution is found that reproduces a given $g(r)$, the solution is unique up to a constant. The theorem, however, makes no statement about the existence of such a potential. This means that all structure based CG methods, including IBI, must generate the same CG potential (if it exists) once the structure $g(r)$ is exactly reproduced. As will be shown numerically in section 3.1.2, the statement has to be treated with care in the context of numerical simulations. The reason is that the Henderson theorem does not give an upper boundary as to how much a deviation in the $g(r)$ affects the corresponding potential.

2.3.4.2 Pressure correction

The pressure of the CG model often differs significantly from the reference all-atom pressure (at the same density and temperature). This has been found by many studies using Iterative Boltzmann Inversion [12, 54], but is also true for other

coarse-graining methods [57]. In a pairwise interacting, mono-atomic liquid, the pressure can be related exactly to the $g(r)$ by the formula [26]

$$p = \rho k_B T - 2/3\pi N \rho \int_0^\infty r^3 \frac{dU(r)}{dr} g(r) dr \; , \tag{2.52}$$

where ρ is the density and N the number of particles. IBI preserves the $g(r)$ by construction, the derivative of the potential $\frac{dU(r)}{dr}$ however is not necessarily the one that gives the reference all-atom pressure in eq. 2.52. The reason is that no knowledge about the pressure has been used for the update eq. 2.44.

In order to correct the IBI potential to give the pressure of the reference all-atom system, a linear correction term was proposed in [12]. This correction is defined as

$$\Delta U_{pressure}(r) = A \left(1 - \frac{r}{r_{cut}} \right) \; , \tag{2.53}$$

where r_{cut} is the cutoff distance and A a constant prefactor. The correction yields a constant force and is defined to be zero at the cutoff. The prefactor A can be estimated based on the difference between the current pressure at step i and the target by exploiting equation eq. 2.52 [54].

$$A \approx (P - P_{target})V / \left[-\frac{2\pi N \rho}{3 r_{cut} \int_0^\infty r^3 \frac{dU(r)}{dr} g(r) dr} \right] \; . \tag{2.54}$$

As the target pressure is usually not reached within one correction, iterations of the correction need to be performed. Note that the pressure correction can possibly alter the potential such that the structure is not exactly preserved. This can lead to a significant change in the compressibility, the implications of which will be discussed in chapter 3.

2.3.5 Force matching

The force-matching (FM) [60] (or Multi-Scale Coarse-Graining [61, 55, 62]) approach takes a different strategy to find the effective potential. The idea is to use a minimization technique to find the CG force which matches the total force on a CG bead as closely as possible. Although this approach seems to be very different from IBI, a careful analysis reveals some similarities which will be discussed in the next section.
FM starts with a representation of the CG force

$$\mathbf{F}^{CG} = \mathbf{F}^{CG}(f_1, \ldots, f_N) \tag{2.55}$$

in terms of unknown parameters $f_1, \ldots, f_N$. If no functional form is known, the f_i can be defined as discrete values of the CG force $\mathbf{F}^{CG}$. Additionally, the smoothness of the CG force can be enforced by using a cubic spline representation [63].

A minimization is carried out to find the parameters of the CG force, which closely reproduce the forces in the all-atom simulation:

$$\chi^2 = \frac{1}{3N_I N_{\mathrm{CG}}} \sum_I^{N_I} \sum_i^{N_{\mathrm{CG}}} \left| \mathbf{F}_i^{I,\mathrm{AA}} - \mathbf{F}_i^{I,\mathrm{CG}}(f_1, \ldots, f_N) \right|^2 . \tag{2.56}$$

Here I labels the N_I instantaneous force vectors $\mathbf{F}_i^{I,\mathrm{AA}}$ at different times taken from a reference all-atom simulation, i labels the N_{CG} CG beads. The minimization $\partial \chi^2 / \partial f_j = 0$ leads to an overdetermined system of linear equations, which can be solved using standard techniques. In contrast to IBI, FM is thus based on the knowledge of positions and all-atom forces.

2.3.5.1 Relation to structure based coarse-graining

Quite remarkably, FM can be shown to be equivalent to solving an equation relating the 2- and 3-body structural correlations to the potential [64], known as the YBG equation. To make the connection, it is first assumed that the CG force is pairwise and radially symmetric

$$\mathbf{F}_i^{I,\mathrm{CG}} = \sum_{j \neq i} \mathbf{u}_{ij} f(r_{ij}) , \tag{2.57}$$

where $\mathbf{u}_{ij}$ denotes the unit distance vector $\mathbf{u}_{ij} = \mathbf{r}_{ij}/|r_{ij}|$. The interaction force can be calculated using discrete bins of size Δr. For the notation of this discretized force, it is convenient to define a discrete delta-functions δ_D, which is 1 if $-\Delta r/2 \leq r < \Delta r/2$ and 0 otherwise. Using this definition, the force is written as

$$f(r) = \sum_d^{N_D} f_d \delta_D(r - r_d) , \tag{2.58}$$

where the sum runs over the N_D discrete points representing the force. Inserting these definitions into eq. 2.56 leads to

$$\chi^2 = \frac{1}{3N_I N_{\mathrm{CG}}} \sum_I^{N_I} \sum_i^{N_{\mathrm{CG}}} \left| \mathbf{F}_i^{I,\mathrm{AA}} - \sum_{j \neq i} \mathbf{u}_{ij} \sum_d f_d \delta_D(r - r_d) \right|^2 . \tag{2.59}$$

The least-square minimization $\partial \chi^2 / \partial f_d = 0$ then yields

$$0 = \sum_I^{N_I} \sum_i^{N_{\mathrm{CG}}} \left(\mathbf{F}_i^{I,\mathrm{AA}} - \sum_{j \neq i} \mathbf{u}_{ij} \sum_{d'} f_{d'} \delta_D(r - r'_d) \right) \left(\sum_{j \neq i} -\mathbf{u}_{ij} \delta_D(r - r_d) \right)$$

$$= \left\langle -\sum_{i, j \neq i} \mathbf{F}^{I,\mathrm{AA}} \mathbf{u}_{ij} \delta_D(r_{ij} - r_d) \right\rangle + \sum_{d'} f_{d'} \left\langle \sum_i \sum_{j \neq i} \sum_{k \neq i} \mathbf{u}_{ij} \mathbf{u}_{ik} \delta_D(r_{ij} - r_d) \delta_D(r_{ik} - r'_d) \right\rangle ,$$

$$\tag{2.60}$$

where the brackets denote the average over the N_I frames. The minimization thus yields an equation connecting the force coefficients f_d with structural averages. The forces on the CG molecule from the reference simulation only appear in the first term. The sum in the second term contains contributions coming from configurations of three distinct particles $i \neq j \neq k$, as well as two particles ($j = k$). The three-particle contributions are weighted by the projection of the direction vector of the i, k pair onto the i, j pair. Because of this, it is clear that FM explicitly incorporates three-body correlations.

What happens if the three-body terms are neglected? In order to do this, all terms in the sum with $j \neq k$ can be set to zero, which then because of $\mathbf{u}_{ij}\mathbf{u}_{ij} = 1$ simplifies to

$$\left\langle \sum_{i,j\neq i} \mathbf{F}_i^{I,AA} \mathbf{u}_{ij}\delta_D(r_{ij} - r_d) \right\rangle = f_d \left\langle \sum_{i,j\neq i} \delta_D(r_{ij} - r_d) \right\rangle \tag{2.61}$$

The expression can be recast in a familiar form noting that the definition of the radial distribution function (see also section 2.2) is

$$g(r) = \frac{1}{4\pi\rho^2 V r^2} \left\langle \sum_{i,j\neq i} \delta_D(r_{ij} - r) \right\rangle , \tag{2.62}$$

where the brackets denote the average in the canonical ensemble. In [64] it was proven (making use of the properties of the δ_D function in spherical coordinates) that the derivative $\mathrm{d}g(r)/\mathrm{d}r$ is related to the atomistic forces on a CG molecule as

$$\frac{\mathrm{d}g(r)}{\mathrm{d}r} = \frac{1}{k_B T} \frac{1}{4\pi\rho^2 V r^2} \left\langle \sum_{i,j} \mathbf{F}_i^{I,AA} \mathbf{u}_{ij}\delta_D(r_{ij} - r) \right\rangle \tag{2.63}$$

Using eq. 2.62 and eq. 2.63, the FM equation (without 3-body terms) can then be rewritten in the continuum limit (i.e. the limit of infinitesimally small discretization steps r_d) as

$$k_B T \frac{\mathrm{d}g(r)}{\mathrm{d}r} = f(r)g(r) \tag{2.64}$$

Integration of both sides then leads to

$$- k_B T \ln(g(r)) = U(r) + \mathrm{const.} \tag{2.65}$$

This means that neglecting the three-body terms only the 2-body potential of mean force is left, which is exactly the starting point of IBI in eq. 2.43. When the three-body terms are kept, FM can be shown to be equivalent to the Yvon-Born-Green-equation[64]

$$(k_B T \nabla_{\mathbf{r}_i} - \mathbf{F}_{i,j})g^{(2)}(\mathbf{r}_i, \mathbf{r}_j) = \rho \int \mathrm{d}\mathbf{r}_k \mathbf{F}_{i,k} g^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) \tag{2.66}$$

where $\mathbf{F}_{i,j}$ is the force between CG molecules, under the assumption that the force is pairwise. This well-known result from liquid-state theory [26] connects the two-and three-body correlation functions in a pairwise interacting liquid. [15]. Thus, "starting" with the same first order term, IBI and FM incorporate different corrections to the two-body PMF. In IBI the correction is based on the difference in structure (eq. 2.44) using the 2-body potential of mean force. FM on the other hand, tries to find a "compromise" in the least-square sense in matching both 2- and 3-body correlations.

In practice, FM is quite efficient as no iterations are needed.

Overall, there are two main sources of error: First, the two-body effective potential might not have enough degrees of freedom to represent two-body as well as three-body PMFs. The FM potential found is a "compromise" (in the least-square sense) between matching two- and three-body correlations. One solution is then to extend the CG model by another degree of freedom which can represent the "lost" correlations. This, on the other hand, raises the computational costs again which might not be desired. Secondly, higher order correlations (4-body, etc..) might as also be important reference all-atom structure. Both sources of error will lead to deviations in the CG structure with respect to the all-atom structure.

The connection to structural averages allows to solve a set of equations equivalent to the FM equations making use only of structural (2- and 3-body) information [15, 65, 66]. An iterative extension, which tries match the structure by construction (similar to the iteration in IBI) has also been discussed [67].

2.3.6 Inverse Monte Carlo

For completeness, the Inverse Monte Carlo (IMC) technique [10] is also mentioned here although it was not used within this thesis. The technique can be used to generate CG potentials from either MD or Monte Carlo simulations [57].

The approach starts with decomposing the potential part of the Hamiltonian into a sum of discrete contributions

$$H = \sum_d U_d S_d \{\mathbf{r}_i\} \tag{2.67}$$

where the U_d are constants representing the interaction potential and the S_d functions of the coordinates. The goal is to express known measurables (e.g. the $g(r)$) as averages $\langle S_d \rangle$ and use these to refine an initial guess of the unknown CG potential. For the case where the $g(r)$ is the known measurable in the canonical ensemble it is suitable to choose

$$\langle S_d \rangle = \frac{N(N-1)}{2} \frac{4\pi r_d^2 \delta r}{V} g(r_d) \tag{2.68}$$

where δr is the bin width and V the volume. Because of eq. 2.67 the $\langle S_d \rangle$ can also be seen to be a function of U_d. In order to arrive at an update scheme one can

Taylor expand with respect to small changes ΔU_d in the potential

$$\langle S_d \rangle = \langle S_d \rangle \,|_{U_d^0} + \sum_{d'} \frac{\partial \langle S_d \rangle}{\partial U_{d'}} \Delta U_{d'} + \mathcal{O}(\Delta U^2) \tag{2.69}$$

the derivative can be carried out explicitly by writing out the expectation value in the canonical ensemble. The final result is

$$A_{dd'} \equiv \frac{\partial \langle S_d \rangle}{\partial U_{d'}} = \beta(\langle S_d \rangle \langle S_{d'} \rangle - \langle S_d S_{d'} \rangle) \tag{2.70}$$

Using this definition, eq. 2.69 is easily written as a system of linear equations

$$\langle S_d \rangle - \langle S_d \rangle \,|_{U_d^0} = A_{dd'} \Delta U_{d'} \tag{2.71}$$

These expressions are used to construct an iterative scheme: Using an initial guess for $U_d^{(0)}$ a first simulation is performed which yields a correlation $\langle S_d \rangle \,|_{U_d^0}$ possibly deviating from the target. The first guess U_d^0 can be taken to be the 2-body potential of mean force as in IBI (eq. 2.43). By solving the linear eq. 2.69, a new guess

$$U_d^{(1)} = U_d^{(0)} + \Delta U_d^{(0)} \tag{2.72}$$

is then calculated. The procedure is repeated until the update term is sufficiently small i.e. the CG potential generates a distribution close to the target. It is important to restrict the range of interest to d values which are reached within the simulated time since the matrix $A_{dd'}$ has to be well-defined for the inversion.

IMC thus presents a coarse-graining approach which uses the linearization in eq. 2.69 to refine the initial guess of the RDF based on the structure. Similar to FM, 3-body correlations enter explicitly through the cross-correlation term $\langle S_d S_{d'} \rangle$. The detailed differences between this and the other methods are beyond the scope of this thesis, a numerical comparison can be found in [57].

2.3.7 Summary

In this section different methods for systematic coarse-graining were introduced. All methods approximate in some way the multi-body PMF, which represents the exact solution of the problem. The CG potentials obtained in this way are likely to be system- and state- dependent and should thus be carefully analyzed in the context of the desired application. Additionally, due to the simplification of matching only selected properties, physical properties calculated from the CG simulations can differ from the ones calculated in the reference simulation. Hence the coarse-graining technique has to be chosen such that the desired properties are reproduced.

However, CG techniques offer possibilities to speed up MD simulations by several orders of magnitude. This is due to their simplicity, which allows to calculate

the interactions much faster on a computer. Furthermore, in the CG system also dynamical quantities, such as the diffusion constant, are found increased due to a smoother energy landscape [54, 68, 69]. One attempt to exploit the advantages and at the same time overcome the loss of the atomistic details, is presented in the Adaptive Resolution Scheme introduced in the next section.

2.4 Adaptive resolution scheme (AdResS)

With the knowledge of the coarse-grained interaction a simplified, computational efficient model is available. It is desirable to use this simplified model to reach longer length and time scales, using the same amount of computer time. However, the CG model is constructed based on selected properties (such as the radial distribution function), while other properties may be lost.

The idea of adaptive resolution schemes is to subdivide the simulation volume into regions of different resolution. The full, computationally expensive interactions are calculated where it is absolutely necessary, while using the efficient simplified CG interactions are used for the rest. The molecules are allowed to leave either region and adapt their interactions based on the local level of resolution. As an example, one may think of aggregation phenomena where full detail is important close to a surface. The "bulk" supply of aggregating molecules could then be described in an efficient CG fashion. Another example is the solvation of larger molecules where the first solvation shell is treated in full resolution.

To this end, the ADaptive RESolution Scheme (AdResS) [70, 71, 72] was designed. The simulation volume is divided into an all-atom (AA, high-resolution) and coarse-grained (CG, low-resolution) region and the exchange between them is mediated by a transition (hybrid) region. A pictorial representation is given in fig. 2.2. Note that, for simplicity of the notation, the "low-resolution" region model is assumed to stem from systematic coarse-graining techniques (as discussed in the previous section 2.3). In principle, however, arbitrary models can be coupled in the AdResS scheme. Such a coupling scheme has to fulfill certain minimal criteria:

- It should allow for the free exchange of molecules between the regions

- There should be no free energy barrier between the regions

- The resolution switching in the transition region should be smooth as to avoid discontinuous jumps in the forces

- In the high-resolution region the structure and dynamics should be unperturbed

How can such a scheme be realized?

First, in the hybrid region, there has to be a smooth transition between the molecular interactions in the all-atom and CG region. For this, the interactions of the all-atom and CG models need to be interpolated. Secondly it has to be well defined what happens to the internal degrees of freedom which are not present in the CG model. The following part is divided into the discussion of the non-bonded forces and the treatment of the internal degrees of freedom.

28
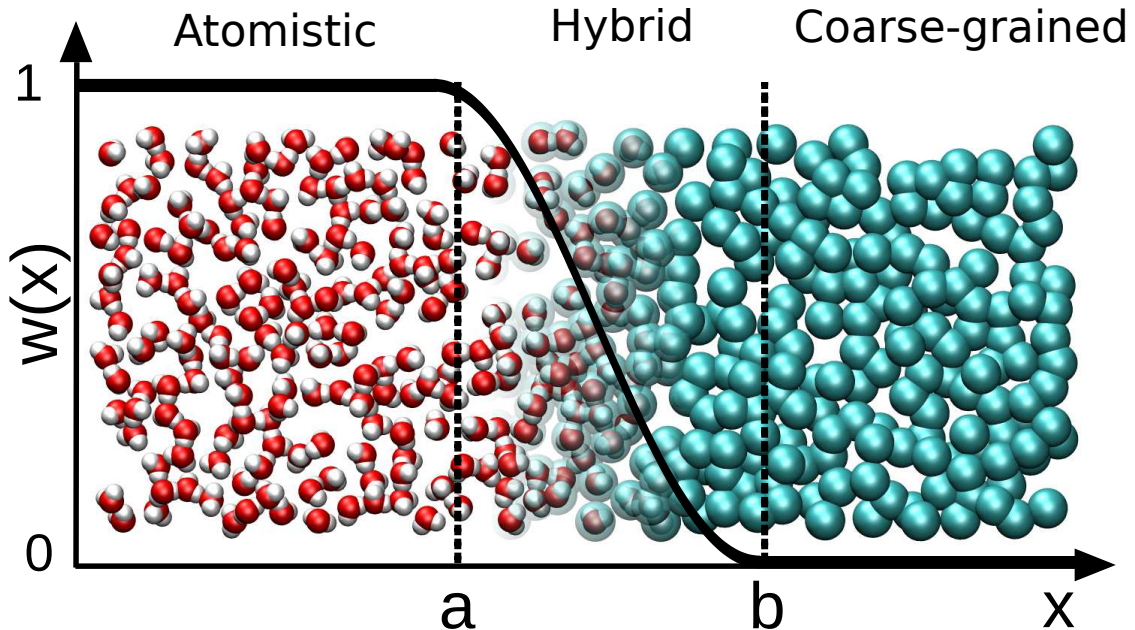
Figure 2.2: Pictorial representation of the adaptive simulation box. On the right is the coarse-grained (CG) region and on the left is the all-atom (AA) region. In between is the transition (hybrid) region with the switching function $w(x)$ (curve in grey).

2.4.1 Force interpolation

For the interpolation, which is necessary to adapt the interactions, a position-dependent weighting function is introduced. The weighting function $w(\mathbf{X}_\alpha)$ (depicted in fig. 2.2), assigns to molecule α a "level of resolution", where $w = 1$ represents the all-atom and $w = 0$ the CG resolution. To accomplish a smooth transition, a continuous, differentiable function is chosen to interpolate between the two representations:

$$w(r) = \begin{cases} 0 & : r > b \\ \cos^2\left[\frac{\pi}{2(b-a)}(r-a)\right] & : b > r > a \\ 1 & : a > r \end{cases} , \qquad (2.73)$$

where a is the width of the all-atom region and $b-a$ the width of the hybrid region. The geometry of the regions and the origin can be suitably chosen depending on the problem studied. For example, if the resolution change is required to be along the x-axis, $r = \hat{\mathbf{e}}_x(\mathbf{x} - \mathbf{x}_0)$ can be chosen, where $\mathbf{x}_0$ denotes the center of the all-atom region.

The next step is to derive interactions for the hybrid scheme. For simplicity the following notations are defined: Molecules are labeled with Greek indices α, β which consists of atoms (Latin indices) α_i and β_j. For the derivation it is assumed that there are N molecules of the same type, consisting of n atoms each, but the

treatment of different kinds of molecules is straightforward.

The vector $\mathbf{X}_{\alpha_i}$ denotes the coordinate of atom i which is part of molecule α and if no Latin index is specified $\mathbf{X}_\alpha$ refers to the center of mass of the molecule:

$$\mathbf{X}_\alpha = M_\alpha^{-1} \sum_{i=1}^n m_{\alpha_i} \mathbf{X}_{\alpha_i} \ , \tag{2.74}$$

where M_α is the total mass of the molecule.

The aim of the coarse-graining procedure (as discussed in section 2.3) was to eliminate the internal degrees of freedom and represent their intermolecular interaction through the CG potential. Thus, the interactions of the internal degrees of freedom has to be "switched off" in the CG region. How exactly this can be treated will be explained in more detail in the following section 2.4.2. For the non-bonded interactions, requiring that there is no influence of internal degrees of freedom which are in the CG region implies that

$$\tilde{U}^{AA}(\mathbf{X}_{\alpha_i}, \mathbf{X}_{\beta_j}) = 0 \ \ \text{if} \ \ w(\mathbf{X}_\alpha) = 0 \vee w(\mathbf{X}_\beta) = 0 \ , \tag{2.75}$$

i.e. the all-atom interaction potential in the adaptive simulation $\tilde{U}^{AA}$ between any atom α_i and β_j is zero once any of the two molecules α or β is in the CG region.

In all cases studied so far, atoms are assumed to interact with a pairwise additive potential as described in section 2.1.3. One of the fundamental design principles in AdResS was to ensure that the forces between molecules are anti-symmetric with respect to the exchange of molecule labels, i.e. the force exerted by molecule α on molecule β is the negative of the force exerted by molecule β on molecule α. This statement is simply Newton's third law, stated for the interaction of molecules. For the interpolation scheme to fulfill this, the weight assigned to the pairwise force must hence depend on both molecules.

The condition eq. 2.75 suggest that the pairwise interaction should be weighted by the product of weighting functions, i.e. for the non-bonded interaction

$$\tilde{U}^{AA}_{\alpha_i\beta_j} = w(\mathbf{X}_\alpha)w(\mathbf{X}_\beta)U^{AA}_{\alpha_i\beta_j}(|\mathbf{X}_{\alpha_i} - \mathbf{X}_{\beta_j}|) \qquad \alpha \neq \beta \ , \tag{2.76}$$

where U^{AA} denotes the all-atom pairwise interaction. This interaction is thus zero if one of the molecules α, β is in the CG region because of the definition of the weighting function eq. 2.73. The CG interaction could then be defined as

$$\tilde{U}^{CG}_{\alpha\beta} = (1 - w(\mathbf{X}_\alpha)w(\mathbf{X}_\beta))\, U^{CG}_{\alpha\beta}(|\mathbf{X}_{\alpha_i} - \mathbf{X}_{\beta_j}|) \ , \tag{2.77}$$

where U^{CG} is the CG effective potential. The total Hamiltonian would thus be

$$H = \sum_{\alpha<\beta} \sum_{i,j} \tilde{U}^{AA}_{\alpha_i\beta_j} + \sum_{\alpha<\beta} \tilde{U}^{CG}_{\alpha\beta} + H_{intra} \ , \tag{2.78}$$

which contains the full internal Hamiltonian H_{intra}, the contribution of which will be discussed later. There is however a fundamental problem with this approach as is visible when taking the derivative in order to calculate the forces. For simplicity a system containing only two molecules with no internal interactions will be considered here. Then the force on atom α_i due to atom β_j ($\alpha \neq \beta$) is:

$$\mathbf{F}_{\alpha_i \beta_j} = \quad -\nabla_{\alpha_i}[w(\mathbf{X}_\alpha)w(\mathbf{X}_\beta)U_{\alpha_i\beta_j}^{AA} + (1 - w(\mathbf{X}_\alpha)w(\mathbf{X}_\beta))U_{\alpha\beta}^{CG}] \quad (2.79)$$

$$= \quad -w(\mathbf{X}_\alpha)w(\mathbf{X}_\beta)\nabla_{\alpha_i}U_{\alpha_i\beta_j}^{AA} - (1 - w(\mathbf{X}_\alpha)w(\mathbf{X}_\beta))\nabla_{\alpha_i}U_{\alpha\beta}^{CG} \quad (2.80)$$

$$-[\nabla_{\alpha_i}w(\mathbf{X}_\alpha)]w(\mathbf{X}_\beta)(U_{\alpha_i\beta_j}^{AA} - U_{\alpha\beta}^{CG})$$

$$\equiv \quad \mathbf{F}_{\alpha_i\beta_j}^{\text{Newton}} - \nabla_{\alpha_i}w(\mathbf{X}_\alpha)w(\mathbf{X}_\beta)(U_{\alpha_i\beta_j}^{AA} - U_{\alpha\beta}^{CG}) \,, \quad (2.81)$$

where terms involving the gradient of the potentials have been included in $\mathbf{F}_{\alpha_i\beta_j}^{\text{Newton}}$. The total force consists of these "Newtonian" forces and a term involving the gradient of the weighting function. This contribution is nonzero only in the hybrid region since the gradient of the weighting function vanishes in the all-atom and CG region. The strength of this "drift" force, which would not arise in a simulation employing a single resolution ($w = \text{constant}$), is proportional to the difference in potential energies between the AA and CG representation.

Moreover, the pairwise force defined in this way violate Newton's third law which states that $\mathbf{F}_{\alpha_i\beta_j} = -\mathbf{F}_{\beta_j\alpha_i}$. This can be seen by taking the gradient for atom β_j

$$\mathbf{F}_{\beta_i\alpha_j} = \mathbf{F}_{\beta_i\alpha_j}^{\text{Newton}} - w(\mathbf{X}_\alpha)[\nabla_{\beta_j}w(\mathbf{X}_\beta)](U_{\beta_i\alpha_j}^{AA} - U_{\beta\alpha}^{CG}) \,. \quad (2.82)$$

The forces eq. 2.81 and eq. 2.82 can only fulfill Newton's third law for a constant weighting function, which would correspond to no change in the resolution [73]. In order to avoid the drift forces in the in the AdResS method, only the Newtonian part $\mathbf{F}_{\alpha_i\beta_j}^{Newton}$ of eq. 2.81 is used for the forces in the simulation. The drawback is however that the potential energy according to $\mathbf{F}_{\alpha_i\beta_j}^{Newton}$ is then path-dependent and thus no Hamiltonian exists which can generate such a force. Therefore, the integration using the forces of eq. 2.81 leads to a constant increase in the total energy of the system [74]. As will be shown later numerically, the perturbation due to the elimination of the non-Newtonian terms, which only appear in the hybrid region, is however rather weak. If used in an NVT ensemble where the temperature is controlled by an appropriate local thermostat (see section 2.1.2), the excess heat coming from the non-conservative nature of the forces is removed, generating a well defined temperature everywhere.

In this work, the CG interaction was always chosen to depend on the center of mass of the molecule, which means that the CG force acting between molecules gets redistributed onto the atoms according to:

$$\mathcal{F}_{\alpha_i,\beta} = \frac{m_i}{M_\alpha}\mathbf{F}_{\alpha\beta}^{\text{CG}} \,. \quad (2.83)$$

The total AdResS non-bonded forces acting on atom α_i is then given by summing up all pairwise forces $\mathbf{F}^{\text{Newton}}_{\alpha_i \beta_j}$ in eq. 2.82:

$$\mathbf{F}^{\text{AdResS}}_{\alpha_i} = \sum_{\beta=1}^{N} \sum_{j=1}^{n} w(X_\alpha) w(X_\beta) \mathbf{F}^{AA}_{\alpha_i \beta_j} + \sum_{\beta=1}^{N} (1 - w(\mathbf{X}_\alpha) w(\mathbf{X}_\beta)) \frac{m_{\alpha_i}}{M_\alpha} \mathbf{F}^{\text{CG}}_{\alpha\beta} \qquad \alpha \neq \beta \;.$$

$$(2.84)$$

In summary, the interactions due to these forces can be categorized in 4 cases:

(i) $w(X_\alpha) = w(X_\beta) = 0$: Only the centers of mass of CG molecules interact via the CG potential U_{CG}.

(ii) $w(X_\alpha) = w(X_\beta) = 1$: The full atomistic interaction is used, which is exactly the same as in a conventional all-atom simulation.

(iii) $0 < w(X_\alpha) < 1$ and $w(X_\beta) \neq 0$: At least one molecule is in the hybrid regions. The molecules interact in a hybrid way, which is the weighted sum of all-atom and CG interaction as specified in eq. 2.84.

(iv) $w(X_\alpha) \neq 0$ and $w(X_\beta) = 0$: One molecule is in the CG region. The interaction between the two molecules is only the CG potential.

Using this scheme the forces are well defined in the whole simulation. Atomistic degrees of freedom in the CG region do not contribute to the force and Newton's third law is fulfilled. In the cases (i) and (ii) the full interactions of the all-atom and CG systems are recovered, and thus bulk AA and CG interactions are described. The hybrid zone smoothly fades out atomistic details which contribute less to the forces if a molecule diffuses towards the CG region. How the internal degrees of freedom can be treated in this process is discussed in the next section.

2.4.2 Internal degrees of freedom

In the CG description, information about the internal degrees of freedom (e.g. the intra-molecular energy) is lost. What does that mean for a scheme, like AdResS, where molecules are exchanged between CG an all atom subsystems?

Once a molecule has left the hybrid region to enter the CG region, the information about position and momenta of the internal degrees of freedom would have to be "deleted" from the memory of the computer. If the trajectory is reversed (i.e. the velocity of all molecules is inverted) the molecule could not enter in the same configuration as before, as there would be no unique way of reintroducing the full internal degrees of freedom in the instance the molecule reenters into the hybrid region [1]. Thus, a scheme, where molecules can diffuse back and forth between the two regions would be non-reversible by construction. This line of arguing is mostly

---

[1] This is true even if no stochastic thermostat is used. A conventional integration scheme, like Verlet, is time-reversible in principle (see section 2.1.1)

of theoretical interest as in practice, the reversibility (for longer times) does not play a major role (see section 2.1.1).

However, in practice the removal of internal degrees of freedom also introduces the the problem of reintroducing velocities to the internal degrees of freedom once a molecule enters from the CG region. There is no unique way of assigning a velocity, since only the average kinetic energy is known (in the canonical ensemble). One approach could hence be to draw momenta randomly from a Maxwell-Boltzmann distribution at the target temperature.

In the implementation developed within this work, a simpler approach is taken, described in [75]. The internal degrees of freedom are integrated in the entire simulation volume in the CG region, as defined in eq. 2.84. The "resolution-change" is in practice only performed in terms of the non-bonded forces and can thus be seen as a decoupling of the internal degrees of freedom from the center of mass of the molecule. When calculating structural properties of the CG region, the internal degrees of freedom are discarded. From a practical perspective this is still advantageous since most of the computational time is typically spent on the calculation of the non-bonded forces, and therefore the decoupling can be used to reach longer simulation times. Because a Hamiltonian that generates the forces in eq. 2.84 cannot be derived, the thermodynamical equilibrium properties need to be carefully analyzed. As discussed above, the forces are non-conservative only in the (ideally small) hybrid region. Using a local thermostat with a coupling strong enough to remove the excess heat generated by this non-equilibrium process, the equilibrium temperature distribution can be ensured everywhere.

In previous publications, the transition of a molecule from the all-atom to CG region was described as an continuous change of phase space dimensionality [73, 72]. In this picture, the process is formally a projection from higher to lower-dimensional phase space. The phase space which an atom in the hybrid region occupies is of fractional dimension, which allows for the generalization of the equipartition theorem such that the kinetic energy for a degree of freedom

$$\langle K \rangle_\alpha = \frac{w(\mathbf{X}_\alpha) \mathrm{k_B} T}{2} \; , \tag{2.85}$$

where T is the temperature and $w(\mathbf{X}_\alpha)$ the weighting function introduced in the previous section. This means that each degree of freedom contributes to the kinetic energy according to its weight which depends on the position. This is thought to reflect the change in dimensionality of the phase space. The approach is theoretically appealing, but was not further considered in this thesis.

As described above, the "resolution change" is more easily realized using a decoupling of the internal from the external degrees of freedom. The internal energy term H_{intra} was treated simply as the atomistic kinetic and internal energy

terms in all the regions:

$$H_{intra} = \sum_{\alpha_i} \frac{\mathbf{p}_{\alpha_i}^2}{2m_{\alpha_i}} + U_{intra}(\mathbf{X}_{\alpha_1}, \ldots, \mathbf{X}_{\alpha_n}) \qquad (2.86)$$

where U_{intra} denotes internal interactions (e.g. bonds, angles, etc., as introduced in section 2.1.3), which are left unaffected by the force interpolation. For calculating the temperature thus all the degrees of freedom contribute. In equilibrium, by the equipartition theorem each degree of freedom has a kinetic energy $\langle K \rangle = \frac{k_B T}{2}$. When calculating the temperature from the simulation in equilibrium the fractional and conventional formula thus lead to the same results.

2.4.3 Thermostat

Due to the non-conservative nature of the AdResS forces in the hybrid region, it is necessary to apply a local thermostatting algorithm which takes care of the excess heat. For simplicity of the implementation, the thermostat acts on all degrees of freedom including the decoupled atomistic ones in the CG region [75]. This approach has the practical advantage that the internal degrees of freedom are thermalized according to the target temperature when they enter the hybrid region. However, it has to be shown that the CG sub-system follows a canonical trajectory which is equivalent to the the canonical trajectory a system composed of only CG particles would follow.

According to eq. 2.13 when thermostatting using a Langevin thermostat (as discussed in section 2.1.2.1), the force on each atom in molecule α is

$$\mathbf{F}_{\alpha_i} = -\gamma m_i \mathbf{v}_{\alpha_i} + \boldsymbol{\eta}_{\alpha_i}(t) \ , \qquad (2.87)$$

where the noise term fulfills the property

$$\left\langle \boldsymbol{\eta}_{\alpha_i}(t) \otimes \boldsymbol{\eta}_{\beta_j}(t) \right\rangle = 2m_i \gamma k_B T \delta(t - t') \delta_{\alpha\beta} \delta_{ij} \boldsymbol{I} \ , \qquad (2.88)$$

where $\boldsymbol{I}$ denotes the unit matrix. The total force, including the Langevin terms, on a CG molecule can be calculated as follows: The intra-molecular atomistic interactions do not contribute to the total force on a molecule. In the CG region, the only force acting on the molecule's center of mass is due to the CG potential

and the thermostat acting on all degrees of freedom:

$$\mathbf{F}_\alpha = \sum_{i \in \alpha} \mathbf{F}_{\alpha_i} \tag{2.89}$$

$$= \mathbf{F}_\alpha^{\text{CG}} - \sum_{i=1}^{n} m_i \gamma \mathbf{v}_{\alpha_i} + \sum_{i=1}^{n} \boldsymbol{\eta}_{\alpha_i}(t) \tag{2.90}$$

$$= \mathbf{F}_\alpha^{\text{CG}} - M_\alpha \gamma \mathbf{v}_\alpha + \sum_{i=1}^{n} \boldsymbol{\eta}_{\alpha_i}(t) \ . \tag{2.91}$$

$$\tag{2.92}$$

where M_α is the total mass and $\mathbf{v}_\alpha$ the velocity of the center of mass of the molecule. The noise part has the property

$$\left\langle \sum_{i=1}^{n} \boldsymbol{\eta}_{\alpha_i}(t) \sum_{j=1}^{n} \boldsymbol{\eta}_{\alpha_j}(t') \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \left\langle \boldsymbol{\eta}_{\alpha_i}(t) \boldsymbol{\eta}_{\alpha_j}(t') \right\rangle \tag{2.93}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} 2 k_B T \gamma m_{\alpha_i} \delta_{\alpha\beta} \delta_{ij} \delta(t - t') \boldsymbol{I} \tag{2.94}$$

$$= \sum_{i=1}^{n} 2 k_B T \gamma m_{\alpha_i} \delta_{\alpha\beta} \delta(t - t') \boldsymbol{I} \tag{2.95}$$

$$= 2 k_B T \gamma M_\alpha \delta_{\alpha\beta} \delta(t - t') \boldsymbol{I} \ . \tag{2.96}$$

$$\tag{2.97}$$

By comparison to the Langevin equation for a single atom (eq. 2.18), it is apparent that thermostatting the atomistic degrees of freedom can be seen as an effective Langevin thermostat acting on CG particles with mass M_α. The relationship between friction coefficient and noise strength given through the fluctuation-dissipation theorem is preserved, which means that a canonical ensemble for the CG particles is generated.

2.4.4 Equilibrium conditions

It is natural to ask what are the appropriate thermodynamical variables to describe a situation where the CG and atomistic regions are in equilibrium. One might think of this in analogy to the phase-coexistence condition [58] between two phases A and B

$$\mu_A(p, T) = \mu_B(p, T) \ , \tag{2.98}$$

i.e. the chemical potentials μ are equal at given pressure p and temperature T
In the AdResS simulation the aim should be to reach a state where the thermo-dynamical properties of the all-atom reference system are preserved at least in

the all-atom region [72, 76]. As will be shown later, in most cases the CG model does not reproduce the atomistic state point. The system will then still reach a "phase-coexistence" but at different thermodynamic conditions μ', p', T' eq. 2.98.

In order to overcome this discrepancy, several strategies have been developed, which are able to preserve the equilibrium thermodynamic conditions of the all-atom system in the adaptive simulation. An approach based on measuring the chemical potential in simulations with constant weighting function w was presented in [76]. In this thesis, new methods to ensure the thermodynamic coupling are presented in chapter 3 and chapter 6.

2.4.5 Summary

The AdResS scheme, which enables to treat all-atom and CG models concurrently in one simulation, was studied throughout this thesis. The resolution coupling poses many challenges, among which are a good physical understanding of the CG potentials, consistent description in terms thermodynamics and technical challenges of the MD simulation code. Within this work an efficient implementation in the GROMACS software package was achieved which is discussed in chapter 4 and appendix B.

In chapter 3, an extension to the method is proposed which allows to couple systems at arbitrary state points. The ability to preserve the local order in a toluene solvent is analyzed in chapter 4. The method can also be used to couple a quantum simulation (through use of path-integral MD) to an effective, classical model as discussed in 5. Finally, in chapter 6, a variant is presented which allows energy conserving adaptive simulations by relaxing the requirement of fulfilling Newton's third law instantaneously.

2.5 Path integral molecular dynamics

2.5.1 Introduction

So far, only systems interacting in an effective classical way were discussed. It is assumed that atoms are point particles which interact through pairwise effective potentials. On the molecular scale, however, the laws of quantum mechanics can be of importance depending on the temperature and system studied. In the case of heavy atoms at high temperatures, the nuclei can be safely treated as classical point particles. However, for hydrogen at room temperature, the thermal de Brogli wavelength is [77]

$$\Lambda = \frac{h}{\sqrt{2\pi m k_B T}} \approx 0.1 \, \text{nm} \qquad (2.99)$$

which is roughly the same as the average "bond-length" (i.e. the hydrogen-oxygen distance) of a hydrogen in a water molecule. From this it can be seen that there is some interest in including quantum effects in Monte-Carlo or MD simulations even at room temperature [78, 79, 80].

The technique studied here to include nuclear quantum effects is derived from the path integral picture of quantum mechanics, originally developed by Richard Feynman [81, 82]. For calculating equilibrium properties (no time dependence) a relatively simple derivation exists making use of the "Trotter expansion". In the following, the path integral approach will be discussed where the focus is not on presenting a complete derivation, but on understanding the technique used later in the thesis for (adaptive) path integral MD (PIMD). A more detailed derivation can be found in [30].

In quantum statistical mechanics, the properties of a system are contained in the density matrix [30]

$$\rho = \sum_n p_n |\Psi^{(n)}\rangle \langle \Psi^{(n)}| \, , \qquad (2.100)$$

where $|\Psi^{(n)}\rangle$ denotes a unique microscopic quantum state and the sum runs over all such states, weighted by p_n. In the canonical ensemble, the weight is $p_n = e^{-\beta\hat{\mathcal{H}}}/\text{Tr}[e^{-\beta\hat{\mathcal{H}}}]$. Once the density operator is known the expectation value of an operator is simply given as $< \hat{A} > = \text{Tr}(\rho A)$, which has to be evaluated in an appropriate basis.

The basis can also be continuous, as it is the case for the position representation, where the density matrix becomes

$$\rho(x, x') \equiv \langle x'|e^{-\beta\hat{\mathcal{H}}}|x\rangle \, , \qquad (2.101)$$

where $\hat{\mathcal{H}} = \hat{U} + \hat{K}$ are the non-commuting kinetic and potential energy operators. The goal is now to arrive at an approximation for calculating the quantum expectation value of an operator A, $< \hat{A} > = \text{Tr}(\rho A)$ in the position basis. Thereby, it

is first assumed that the system consists of a single particle in one dimension, the extension to a many-particle system is discussed in section 2.5.2.

The trotter theorem [83] states that for non-commuting operators A, B

$$e^{A+B} = \lim_{P\to\infty} \left[e^{B/2P} e^{A/P} e^{B/2P} \right]^P , \qquad (2.102)$$

for integer P. The proof of this theorem is rather involved and thus does not fit this introduction. Using the Trotter theorem, the density operator can be written as

$$\rho(x, x') = \langle x' | \left[e^{\hat{U}/2P} e^{\hat{K}/P} e^{\hat{U}/2P} \right]^P | x \rangle . \qquad (2.103)$$

Defining $\hat{\Omega} = e^{\hat{U}/2P} e^{\hat{K}/P} e^{\hat{U}/2P}$ the expression can more simply be written as

$$\rho(x, x') = \langle x' | \hat{\Omega}^P | x \rangle \qquad (2.104)$$

By inserting the identity operator $\hat{I} = \int \mathrm{d}x |x\rangle\langle x|$ P-1 times one arrives at

$$\rho(x, x') = \lim_{P\to\infty} \int \mathrm{d}x_2 \cdots \mathrm{d}x_P \langle x' | \hat{\Omega} | x_P \rangle \langle x_P | \hat{\Omega} | x_{P-1} \rangle \langle x_{P-1} | \cdots | x_2 \rangle \langle x_2 | \hat{\Omega} | x \rangle \quad (2.105)$$

which expresses the density matrix as a P-1 dimensional integral. Now the aim is to explicitly calculate the elements $\langle x_P | \hat{\Omega} | x_{P-1} \rangle$ in order to evaluate the multi-dimensional integral eq. 2.105. Since the potential energy operator depends only on coordinates, the eigenvalues of the potential operator are simply given by $\exp(-\beta U(x_k)/2P)$. The eigenvalues of the kinetic energy operator $\hat{K}$ are only know in the momentum space, which can be made use of by inserting the identity operator for the momentum space $\hat{I} = \int \mathrm{d}p |p\rangle\langle p|$. The integration can then be carried out explicitly (and using $\langle x | p \rangle = \frac{1}{\sqrt{(2\pi\hbar}} e^{ipx/\hbar}$) to yield

$$\langle x_{k+1} | e^{-\beta\hat{K}/P} | x_k \rangle = \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{1/2} \exp\left[-\frac{mP}{2\beta\hbar^2} (x_{k+1} - x_k)^2 \right] . \qquad (2.106)$$

Substituting eq. 2.106 back into the density matrix, it can be written:

$$\rho(x, x') = \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \int \mathrm{d}x_2 \cdots \mathrm{d}x_P$$

$$\times exp\left\{ -\frac{1}{\hbar} \sum_{k=1}^{P} \left[-\frac{mP}{2\beta\hbar^2} (x_{k+1} - x_k)^2 + \frac{\beta\hbar}{2P} (U(x_{k+1}) - U(x_k)) \right] \right\} \Bigg|_{x_1=x}^{x_{P+1}=x'} . \qquad (2.107)$$

Now, in order to arrive at the partition function $Z = \mathrm{Tr}(\rho)$ the diagonal elements of the matrix are integrated:

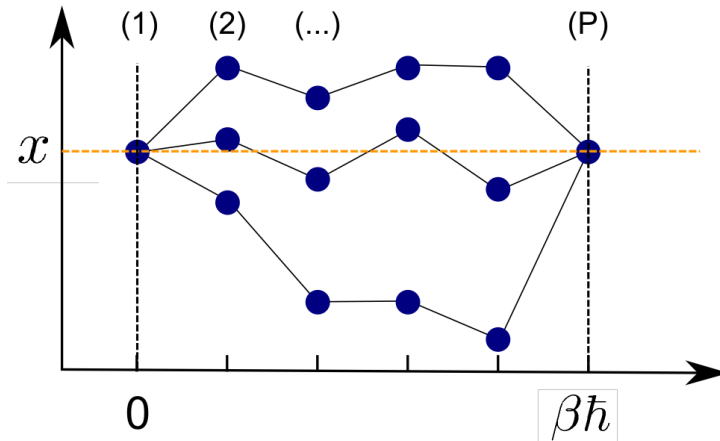$$Z = \int_0^L \mathrm{d}x \rho(x, x) , \qquad (2.108)$$

Figure 2.3: Illustration of the sum over path in the discretized path integral (adapted from [30]). Three representative configurations are shown. The integral runs over all possible (discrete) path which lead from 0 to $\beta\hbar$.

where the integration has to be restricted to an arbitrary spatial domain $[0, L]$, in order to have a finite integral. Because now only diagonal elements of the density matrix are calculated, the endpoints of the integral in eq. 2.107 are $x_1 = x_{P+1} = x$. Hence the integration over x can be equilvalently written as integration over x_1, and the expression becomes

$$Z = \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} \int_{D(L)} \mathrm{d}x_1 \cdots \mathrm{d}x_P e^{-\beta\Phi(x_1,\ldots,x_P)} , \qquad (2.109)$$

where the symbol $D(L)$ denotes the restriction to the spatial domain of length L. The "effective Hamiltonian" $\Phi(x_1, \ldots, x_P)$ is given by

$$\Phi(x_1, \ldots, x_P) = \sum_{k=1}^{P} \left[\frac{1}{2}m\omega_P^2(x_k - x_{k+1})^2 + \frac{1}{P}U(x_k)\right] , \qquad (2.110)$$

where the frequency $\omega_P = \sqrt{P}/\beta\hbar$ was defined. The partition function in eq. 2.109 is called the discretized path integral representation of the partition function. The partition function in eq. 2.109 on the other hand also resembles a completely classical partition function which describes a polymer necklace with P beads each connected by harmonic springs of strength ω_P (the polymer beads do not have momenta, which will be introduced later). The external potential acting on each of the beads is reduced by a factor $\frac{1}{P}$. In the limit of P going to infinity it provides an exact solution of Schrödingers equation for one particle (in one dimension). The sum over all paths of the discretized path integral is illustrated in fig. 2.3 [2]. The

_____

[2]For many applications the path integral is formulated as an integral over continuous paths, which however cannot be calculated by MD simulation.

integration is over all possible path in the domain $D(L)$ which lead from x to x in imaginary time $\beta\hbar$.

The meaning of "imaginary time" becomes apparent from the connection between density matrix and time propagator. The amplitude A of finding a particle prepared in position x at a position x' after time t is given by the time propagator $\hat{U}(t) = e^{-i\hat{\mathcal{H}}t/\hbar}$, such that $A = \langle x'|\hat{U}(t)|x\rangle$. The time propagator is connected to the density matrix by setting $t = -i\beta\hbar$, i.e. $\hat{\rho}(\beta) = \hat{U}(-i\beta\hbar)$. This way, knowing the propagator in principle the density matrix can be computed and vice-versa. The path in fig. 2.3 can thus be seen to be generated by a propagator evaluated at imaginary time. In principle, eq. 2.107 can also be used to obtain quantum dynamical information. However, transforming back to real time makes the exponential a rapidly oscillating function which is extremely difficult to converge in practice.

For this work only the equilibrium statistical mechanics results are relevant which are computed from the diagonal elements of the density matrix. How the multi-dimensional integral in eq. 2.109 can be computed in practice will be explained in the following section.

2.5.2 Path integral MD

So far only the single particle (1-dimensional) case has been considered. The extension to 3 dimension is straightforward, however describing multiple particles is more difficult and in general needs to incorporate important quantum physical phenomena. The reason is that the symmetry of the wavefunctions needs to be taken account for. For bosons the wavefunction is symmetric and for fermions antisymmetric upon exchange of particle labels. This means that for two particles labeled 1 and 2, the boundary conditions $x_1(0) = x_1$, $x_2(0) = x_2$ and $x_1(\beta\hbar) = x_2$, $x_2(\beta\hbar) = x_1$ have to be included in the partition sum. This represents a particle exchange where particle 1 has switched roles with particle 2 and vice-versa. For fermions, including this exchange term leads to two terms in the exponential in eq. 2.109 which have opposite sign and are on the same order of magnitude. This, so called "Fermi sign problem" makes it very hard to calculate properties of fermionic systems. The bosonic exchange terms however do not have this problem and the path integral formalism can be used to investigate for example superfluid He^4 [84].

Within this thesis, the path integral Molecular Dynamics (PIMD) is used to include nuclear quantum effects in the (adaptive) MD simulation [85, 86, 87]. In this approach, the exchange term is neglected, which is still a reasonable approximation for most liquids at room temperature. The path integral is then used to model the nuclear quantum effects, i.e. zero-point energy and delocalization. For light nuclei, like hydrogen, the nuclear quantum effect has significant consequences even at room temperature.

For the PIMD algorithms, the resemblance with a classical ring polymer system of eq. 2.109 is exploited further. The prefactor is simply written as an integral over

P momenta $p_1, \ldots, p_P$

$$\left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} = \int dp_1 \cdots dp_P \left(-\beta \sum_{i=1}^{P} \frac{p_i^2}{2m'}\right) \qquad (2.111)$$

where use of the Gaussian integral has been made and the "mass" $m' = \frac{mP}{(2\pi\hbar)^2}$ has been assigned. This mass does not correspond to the physical mass of the particles m. However, if only equilibrium properties are of interest, setting $m' = m$ yields the same results because a constant term in the partition function has no influence on equilibrium averages. These momenta thus are taken to be the momenta conjugate to the imaginary time beads x_k. The partition function is then the one of a ring polymer at temperature T with an effective bead mass m'. The corresponding Hamiltonian (now in 3-dimensions), reads

$$\mathcal{H} = \sum_{k=1}^{P} \left[\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i'} + \sum_{i=1}^{N} \frac{1}{2} m_i \omega_P^2 (\mathbf{x}_i^{(k+1)} - \mathbf{x}_i^{(k)})^2 + \frac{1}{P} U(\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_N^{(k)}) \right] \ . \quad (2.112)$$

This Hamiltonian can be evaluated in a fairly simple fashion using MD or Monte Carlo [84]. The potential acts only between beads at the same imaginary time. For PIMD, some care is needed for converging the simulation. The reason is that the harmonic springs connecting the imaginary time beads often constitute the fastest motion in the system and thus limit the convergency [88]. Thus, the so called "staging" or "normal-mode" path integral techniques [89] are often used. These algorithms use analytical transformations to evolve the different frequencies at different timescales.

The problem of non-ergodicty can also be overcome with the help of local thermostatting [90, 91]. When simulating the canonical ensemble the thermostat (in this thesis the Langevin thermostat, section 2.1.2.1) couples to all 3NP degrees of freedom. Hence also the slow modes receive energy from the random kicks of the noise term. The computational effort of a PIMD simulation is roughly P times bigger compared to a conventional simulation since the cost of evaluating the non-bonded forces increases roughly by P. The trade-off between speed and accuracy also determines the value of P used in practice; the higher P the more accurate and expensive is the simulation.

Thus, in summary the following effects are *included* in the PIMD

- nuclear zero point energy

- nuclear delocalization

- nuclear tunneling

and the following effects are *neglected*

- exchange (symmetry of wavefunction)

- electronic quantum effects

One important aspect to be discussed in chapter 5 is how to obtain the interaction potentials which are valid in the path integral MD simulation. This is of importance as most interaction potentials are derived from fitting to experimental quantities. These quantities already include the nuclear quantum effect and thus lead to a double counting when used with the PIMD technique [92].

2.5.2.1 Measuring quantities from PIMD simulations

Measuring observables from the PIMD follows from applying the usual concepts of quantum statistical mechanics. In the following, the most important aspects will be discussed. First the simple one-dimensional case is considered again. The expectation value of an observable A in quantum statistical mechanics is [30]

$$\left\langle \hat{A} \right\rangle = \frac{1}{Z} Tr[\hat{A} e^{-\beta \hat{H}}] \ . \tag{2.113}$$

In case the observable can be written as a function of the position operator with $\hat{A}(\hat{x})|x\rangle = a(x)|x\rangle$, the expectation value in the discrete PI approximation becomes

$$\left\langle \hat{A} \right\rangle = \frac{1}{Z} \int \mathrm{d}x \ a(x)\langle x|e^{-\beta \hat{H}}|x\rangle \tag{2.114}$$

$$= \frac{1}{Z} \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} \int_{D(l)} \mathrm{d}x_1 \cdots \mathrm{d}x_P a(x_1) \ e^{-\beta\Phi(x_1,\dots,x_P)} \ . \tag{2.115}$$

Where the previous result for the diagonal elements of the density matrix from eq. 2.109 has been used. This expression can be used to evaluate the expectation values from simulation, it is however not very efficient. The reason is that one imaginary time bead x_1 is preferred over the others. This shortcoming can be overcome by noting that the integral is invariant under relabeling of the particles due to the cyclic path. The average could hence be performed on any of the P imaginary time beads. Thus an completely equivalent expression is

$$\left\langle \hat{A} \right\rangle = \frac{1}{Z} \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2}\right)^{P/2} \int_{D(l)} \mathrm{d}x_1 \cdots \mathrm{d}x_P \left(\frac{1}{P}\sum_{k=1}^{P} a(x_k)\right) e^{-\beta\Phi(x_1,\dots,x_P)} \ . \tag{2.116}$$

The observable can be measured by taking the average over imaginary time slices and the ensemble average. In case the operator cannot be expressed in terms of coordinate eigenfunctions, the calculation becomes much more tedious as also off-diagonal elements need to be included. In the canonical ensemble however,

thermodynamic relations can be used to avoid this and evaluate the total energy using only diagonal elements of the density matrix. This results in expressions for pressure and total energy which are easily available from the simulation. In general, evaluated at finite trotter number P, the expectation value is called an estimate and the estimator $a_P(x_1, \ldots, x_P)$ is a function which allows to write the estimate $< a_P >$ as

$$\langle a_P(x_1, \ldots, x_P) \rangle = \frac{1}{Z} \left(\frac{mP}{2\pi\beta\hbar^2} \right)^{P/2} \int_{D(l)} \mathrm{d}x_1 \cdots \mathrm{d}x_P \; a_P(x_1, \ldots, x_P) e^{-\beta\Phi(x_1, \ldots, x_P)} \; . \tag{2.117}$$

The estimator for the total energy can be derived from thermodynamics by using elementary thermodynamic relations, for example for the total energy

$$E = -\frac{\partial}{\partial\beta} lnZ(N, V, T) \tag{2.118}$$

which can be used to define the estimator for the total energy:

$$\epsilon_P(\mathbf{x}_1, \ldots \mathbf{x}_P) = \frac{dNP}{2\beta} - \sum_{k=1}^{P} \sum_{i=1}^{N} \frac{1}{2} m_i \omega_P^2 \left(\mathbf{x}_i^{(k)} - \mathbf{x}_i^{(k+1)} \right)^2 + \frac{1}{P} \sum_{k=1}^{P} U(\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_i^{(k)}) \; , \tag{2.119}$$

and similarly for the pressure which can be derived from the thermodynamic relation

$$P = \mathrm{k_B}T \frac{\partial}{\partial V} lnZ(N, V, T) \; . \tag{2.120}$$

The estimator for the pressure is then given by inserting the path integral representation (eq. 2.109) for the partition sum Z:

$$\mathcal{P}_P(\mathbf{x}_1, \ldots \mathbf{x}_P) = \frac{NP}{\beta V} - \frac{1}{dV} \sum_{k=1}^{P} \sum_{i=1}^{N} \left[\frac{1}{2} m_i \omega_P^2 \left(\mathbf{x}_i^{(k)} - \mathbf{x}_i^{(k+1)} \right)^2 + \frac{1}{P} \mathbf{x}_i^{(k)} \nabla_{r_i^{(k)}} U \right] \; . \tag{2.121}$$

Using these estimators, the thermodynamically relevant measurables can be easily calculated from the PIMD simulations. In some cases, numerical problems are observed due to large fluctuations of the estimators in the simulation. For these measurables, special estimators have been developed which provide a more efficient convergency [93].

2.5.2.2 Summary

The PIMD technique presents a way of including nuclear quantum effects in MD simulations. There is a considerable interest in including this effect, even for liquid water at room temperature [80]. This will be explored in chapter 5, where the

determination of interaction potentials suitable for use in PIMD simulation of liquid water are discussed.

As has been demonstrated in [85, 86, 87], this technique can also be used in conjunction with the Adaptive Resolution Scheme. In some sense, one might view the path integral picture as another scale or level of detail, just as the CG representation introduced before was a change in resolution (although in general finding a classically equivalent description of quantum system is not straightforward or not possible at all). It is therefore interesting to explore the possibilities of coupling this "quantum-scale" to the classical scale. The application of the adaptive scheme to PIMD simulations will be discussed in chapter 5 at the example of a hybrid simulation of a slab of liquid water slab in vacuum.

Chapter 3

Adaptive resolution simulation of water

This part of the thesis lead to the following publication:
"Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir"
S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site and K. Kremer
Physical Review Letters, **108**, 17, 170602, 2012

In this part of the thesis, the application of the adaptive resolution scheme was probed as a method for simulating an explicit particle reservoir coupled to the all-atom region. Such a method is interesting, as there is a number of problems where the full detail is required only in a small sub-region, while a supply of particles from outside the subregion is needed. As an example one might think of aggregation on a surface, where only the surface region needs to be resolved with full details.

In Monte Carlo simulations, related techniques based on direct insertion of molecules have been used to provide a grand canonical ensemble [94, 95, 96]. These techniques are similar in the sense that the number of particles in the simulation box is allowed to fluctuate. Extensions were also made to perform hybrid molecular dynamics/monte carlo simulations where the insertion attempts are restricted to special regions [97].

Here, the AdresS concept was generalized to allow for the coupling of systems with different chemical potential and different pressure. The type of models that are coupled is rather arbitrary in this approach. For practical applications, one way to make use of this would be the case where the CG region is much larger and thus acts like a grand-canonical reservoir. Properties measured in the high-resolution region are thus effectively in the μVT ensemble, the accuracy of the approximation can be tuned with varying the size of the reservoir simulated. The main advantage

is that the CG "reservoir" can be simulated more efficiently because of the reduced number of intermolecular interactions.

The test system was liquid water at room temperature. This choice is natural considering the immense importance of liquid water in simulation of biological and soft matter systems [7], but also the enormous complexity of its physical properties. Water thus represents a benchmark which is important to master for any new simulation scheme.

The following section starts with introducing the atomistic model of liquid water followed by the development of a CG potential. Subsequently the coupling between the all-atom and CG model in the adaptive resolution scheme is discussed. In the last part, the concept of the thermodynamic force is developed, which takes care of the differences of the all-atom and CG model in the adaptive simulation.

3.1 Coarse-graining of liquid water

The properties of water arise as a complex interplay of hydrogen bonding and polarization [98]. The structure is commonly thought to follow on average a tetrahedral arrangement, formed by the four next-neighbor molecules [99]. In a coarse-grained description of a complex liquid like water, many of the properties of the liquid are lost. Hence, a choice has to be made beforehand as to which properties are needed in the CG model. Since this also plays a role for the adaptive simulations using the AdResS scheme, introduced in section 3.2, the results from CG simulations of water are first discussed here. The CG interaction was determined from the IBI scheme discussed in section 2.3.

The all-atom model used was the simple point charge SPC/E water model[1] [101, 102] which is widely used in computer simulations [7]. In this model, three point charges represent the two hydrogens and oxygen and a Lennard-Jones interaction (see section 2.1.3) accounts for the non-covalent interaction between molecules. The length of the bond (OH) and angle (HOH) are fixed. In the simulation a constrained integration (using the SETTLE [44] algorithm) was used to maintain the rigidity. The parameters are given in table 3.1 and a schematic illustration is given in fig. 3.1.

All simulations were carried out using the GROMACS [103] package. The IBI method is implemented in the "Versatile Object-oriented Toolkit for Coarse-graining Applications" (VOTCA) [57], which provides an interface to GROMACS.

For the atomistic reference simulation 2180 molecules were simulated in a cubic box. In order to equilibrate the simulation of the SPC/E water at ambient pressure ($p = 1\,\text{bar}$), initial simulations were run for of 10^7 steps, employing a Berendsen barostat [104] with a timestep of $\Delta t = 2\,\text{fs}$. The simulations were run with the

---

[1]Note that this model was originally parametrized for use with Ewald-summation long-range interactions [100], which were not considered here.
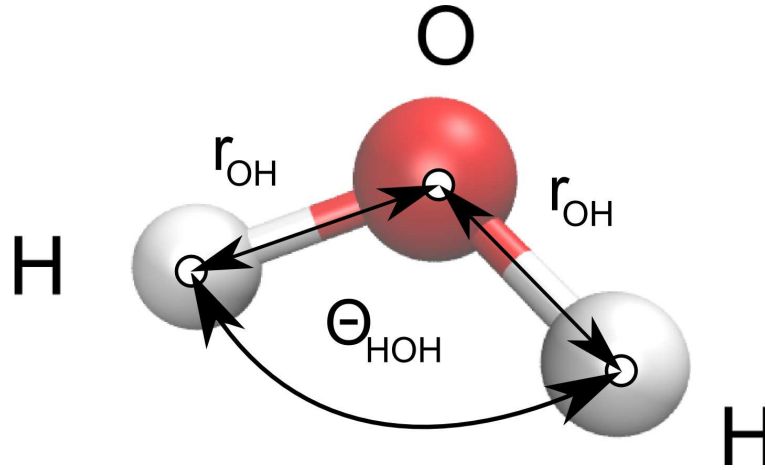
Figure 3.1: Schematical representation of the SPC/E water model

Hydrogen	Oxygen
$q_H = -0.8476e$	$q_O = 0.4238e$
$\sigma = 0$	$\sigma = 0.316557\,\mathrm{nm}$
$\epsilon = 0$	$\epsilon = 0.650194\,\mathrm{kJ/\,mol}$
Bond	Angle
$r_{OH} = 0.1\,\mathrm{nm}$	$\theta_{HOH} = 109.47°$

Table 3.1: Parameters of the SPC/E water model

Langevin thermostat at a temperature $T = 300\,\text{K}$ and a coupling constant of $\tau_{\text{Langevin}} = 0.5\,\text{ps}$. Electrostatic interactions were treated using the reaction-field (see section 2.1.4) approach using a cut-off length of $1.2\,\text{nm}$. The equilibrium density was calculated from NPT simulations using the Berendsen barostat[104] and found to be $\rho = 994.19 \pm 0.03\,\frac{\text{kg}}{\text{m}^3}$.

In a subsequent simulation (10^7 steps) at fixed density in the NVT ensemble, the structure in terms of the $g(r)$ of the molecules center of mass was calculated (shown in fig. 3.2). In the CG model, one water molecule is represented by its center of mass. The corresponding effective potential was derived using IBI (as described in section 2.3.4) using the atomistic center of mass $g(r)$ as target distribution. 100 steps of IBI were carried out, where for each iteration step 50×10^3 steps of MD using a timestep of $\Delta t = 2\,\text{fs}$ were run.

The evolution of the CG potential during the IBI is shown in fig. 3.2. The potential in the first step, which is simply $U_0 = w(r) = k_B T \ln g(r)$ does give liquid structure significantly different from the target. During the iteration, the initial is corrected based on the difference of current and reference $g(r)$ (see section 2.3.4). The iteration converges once the structure produced by the CG model is sufficiently close to the target structure. To quantify this, the error defined as mean squared deviation with respect to the target $g(r)$ (eq. 2.46) is calculated in each step (show in fig. 3.3). After approximately 100 steps the accuracy was considered sufficient ($\Xi < 0.001$). This error is mostly determined by the numerical precision (the binning of the numerical representation of the potential) and can thus be further improved by reducing the bin width and increasing the simulation length as to decrease the statistical fluctuations in each bin.

The RDF calculated from the CG simulation very closely matches the reference RDF. It is interesting to note the change of the first two minima in the CG potential during the IBI. These minima correspond to the first and second neighbors of a water molecule. Initially, the first minimum is deeper and the corresponding $g(r)$ over-structured with respect to the atomistic reference. As the iteration proceeds, the minima change their relative depth and converge to a state where the second-neighbor minimum is deeper than the first-neighbor minimum. Hence, the initial guess for the two-body potential of mean force (PMF) (see section 2.3.4 for definition) overestimates the interaction between next neighbors. This overestimation gets successively correct during the IBI.

3.1.1 Properties of the coarse-grained model

The previous section showed that using IBI it is possible converge a CG potential in terms of the matching the radial distribution function (and hence the two-body PMF). Other structural properties are however not guaranteed to be reproduced. One example is the tetrahedral order parameter defined as [105]

Coarse-grained potential
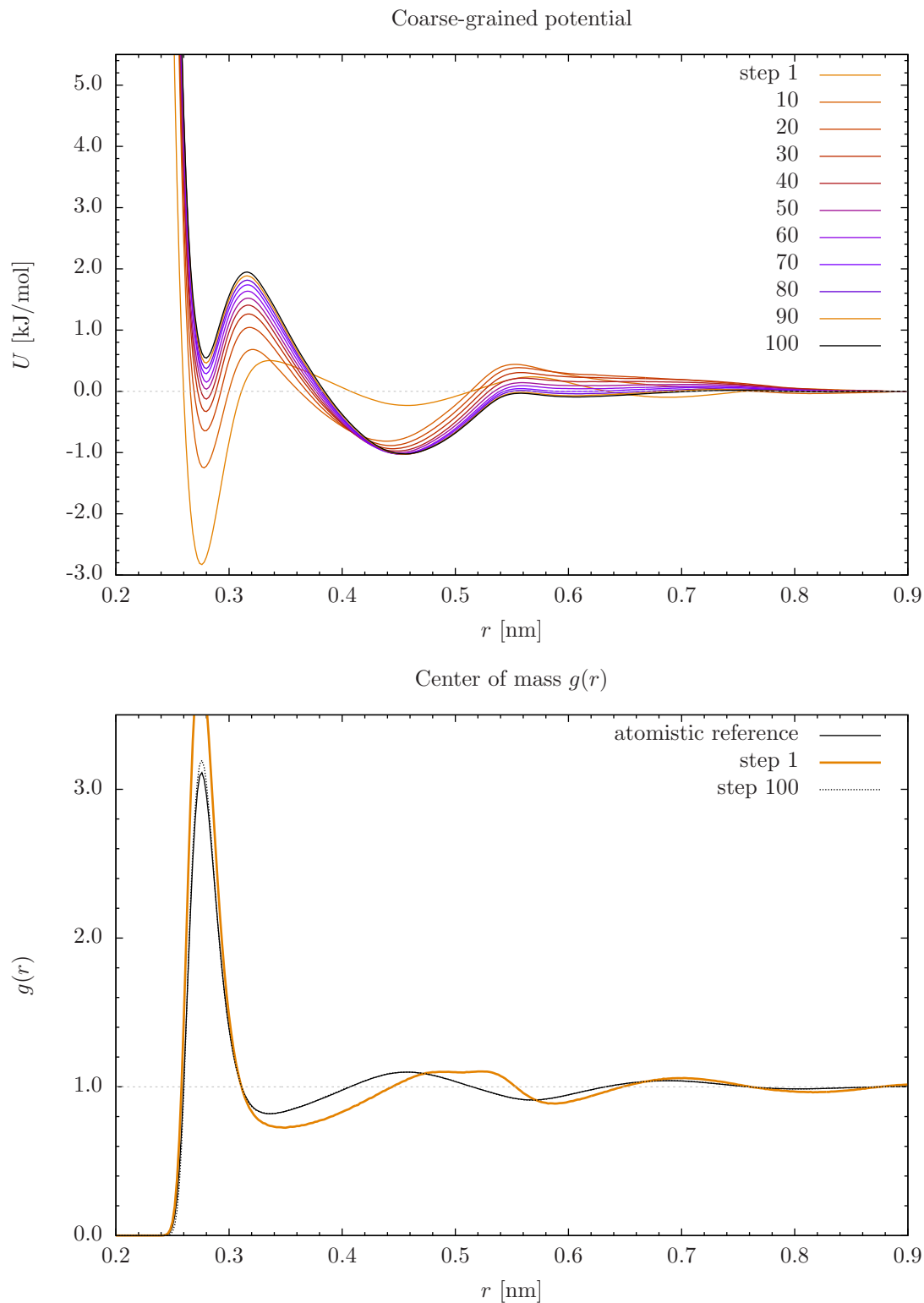

Center of mass $g(r)$

Figure 3.2: IBI of SPC/E water. Top: The evolution of the coarse-grained potential during IBI. Bottom: The radial distribution function for the final iteration compared to the all-atom reference (center of mass).
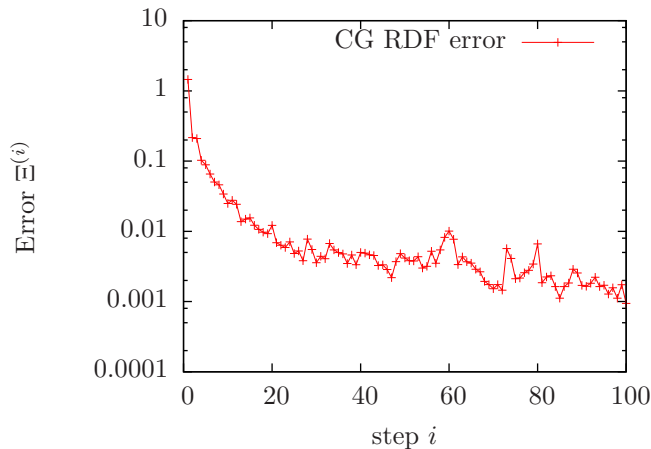
Figure 3.3: Error in IBI for finding a CG interaction potential that matches the $g^{(ref)}(R))$ of SPC/E water. The error is defined as $\Xi^{(i)} \equiv \sqrt{\int_0^{r_{cut}} (g^{(i)}(R) - g^{(ref)}(R))^2 dR}$

$$q_4 = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos(\Psi_{jk}) + \frac{1}{3} \right)^2 , \tag{3.1}$$

where $\Psi_{jk} = \arccos(\hat{\mathbf{r}}_{ij}\hat{\mathbf{r}}_{ik})$ is the angle between a selected oxygen atom i and the vectors connecting the oxygen position $\mathbf{r}_i$ to its four next neighbors $\mathbf{r}_j$ and $\mathbf{r}_k$ (with $j, k \leq 4$). The q_4 order parameter characterizes the orientational structure of next neighbor water molecules by measuring the deviation from the ideal tetrahedral-angle $\arccos(-1/3) \approx 109.47°$. In an ideal gas, where no near ordering exists the order parameter is $q_4 = 0$, while in a perfect tetrahedral network $q_4 = 1$. The value for liquid water at room temperature is expected to lie between these two limits. From the atomistic water model simulations the tetrahedral order parameter was found to be $< q_4 >_{All-atom} = 0.6282 \pm 0.0002$ where the average is taken over time and molecules. Measuring the order parameter in the CG simulation however gives $< q_4 >_{CG} = 0.4393 \pm 0.0002$. A qualitatively similar result was found by Wang et al. [54]. This result illustrates that the orientational properties are not carried over to the coarse-grained model. The reason is the hydrogen bonding which is very crucial property of water. In the atomistic model this is included as oxygens and hydrogens attract each other due to the partial charges with opposite sign (this is in fact plays a major role for the design of atomistic water models [7]). The IBI based CG model has no degree of freedom to describe the orientation as each molecule is represented by a point particle with isotropic interactions. The measured $< q_4 >_{CG}$ thus simply corresponds to the near-ordering in a liquid of spherical particles.

So far, no knowledge about thermodynamic properties has been used in the con-

struction of $U^{CG}(R)$ (with the exception of the compressibility which is completely specified by the $g(r)$ [26]). An analysis of the CG simulation (at the density of the reference atomistic simulation) gives an average pressure of $p_{CG} = 6234 \pm 0.5$ bar, which deviates almost five orders of magnitude from the pressure $p_{at} = 1$ bar the atomistic reference simulation was prepared at. This finding might seem surprising at first but is in fact consistent with other findings in the literature [57, 54]. In order to overcome this mismatch, different techniques will be discussed which include either a correction to the potential (described in the following section 3.1.2) or use of a "thermodynamic force" (discussed in section 3.3).

Dynamical properties like diffusion rates or relaxation times measured from the CG simulation, show strong deviations from the reference all-atom values as well. It has been found in [54], for example, that the self-diffusion constant is increased by approximately a factor 5 in the CG model. This can be explained by the fact that the friction depends on the roughness of the local energy landscape, which is smoother in the CG model due to the reduced number of degrees of freedom. Thus, if time dependent properties are calculated from the CG simulation it is not possible to directly compare them to the timescales in the referce all-atom simulation (also in complex systems, such as polymers, the dynamics of different parts of the system might change non-uniformly [69]). This very complex matter was not in focus of this thesis, however, whether this has an affect on the dynamics of the adaptive resolution simulations will be discussed in chapter 4.

3.1.2 Pressure corrected coarse-grained model

The most simple technique for correcting the pressure is to add a linear attractive potential to the coarse-grained interaction as explained in detail in section 2.3.4.2. For the pressure correction, 20 additional steps of IBI were carried out, 10 of which are performed with an update

$$\Delta U^{(i+1)}(r) = \Delta U_{IBI}^{(i)}(r) + A_i \left(1 - \frac{r}{r_{cut}} \right) , \tag{3.2}$$

where the first term is the standard IBI update (eq. 2.45) and the second is the pressure correction (eq. 2.53). This correction is able to significantly reduce the pressure to $\langle p \rangle_{CG} \approx 300$ bar within 5 iteration steps. After these 5 steps the pressure does not change further. The reason is that the two terms in eq. 3.2 represent competing contributions to the potential. In order to overcome this the final 10 steps of the iteration were carried out using only

$$\Delta U^{(i+1)} = A_i \left(1 - \frac{r}{r_{cut}} \right) , \tag{3.3}$$

which leads to a pressure of $p_{CG} = 0.7 \pm 2$ bar in the final step.
The pressure corrected potential is shown together with the unaltered potential
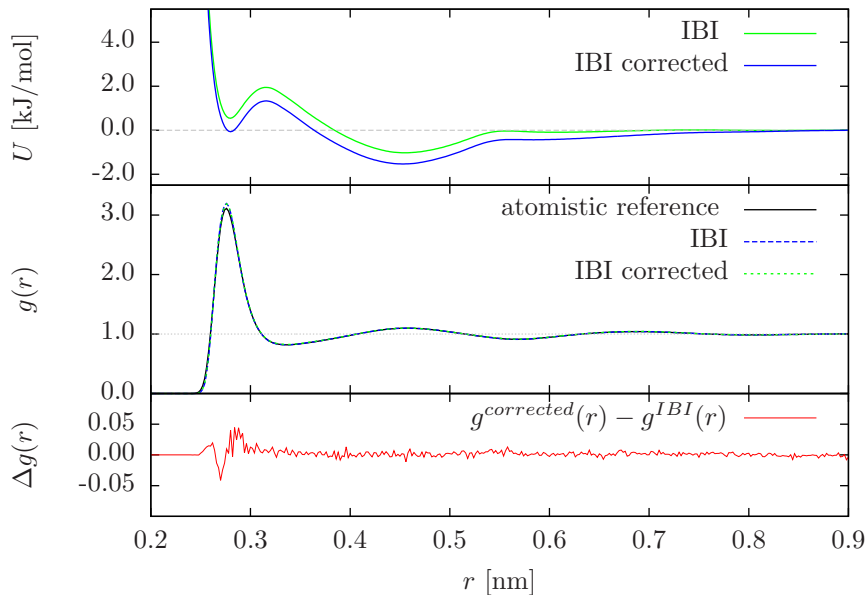
Figure 3.4: Pressure correction for IBI of water: The top plot shows the difference in the CG potential introduced by the pressure correction (as explained in the text). The bottom plot shows the difference introduced by this correction in terms of the $g(r)$.

in fig. 3.4. The pressure correction shifts the potential down, rendering it overall more attractive. While this change is sufficient to reach the desired pressure, the difference between corrected $g^{\text{corrected}}(r)$ and the reference $g(r)^{\text{IBI}}$ is very small ($\Delta g(r) < 0.1$, as shown in fig. 3.4). This outcome might seem a bit surprising at first considering the uniqueness statement in the Henderson theorem. A discussed in section 2.3.4.1, the Henderson theorem states that there exists a unique correspondence between $g(r)$ and $U^{CG}(r)$. As a consequence, the IBI potential which gives a perfect agreement with the atomistic reference RDF, is unique. Thus, any modification, including the pressure correction must introduce an error in the structure.

The statement of uniqueness however makes no sense in the context of numerical simulations as the numerical error was not considered. Due to finite precision of the numerical calculation and numerical approximation formulas (e.g. for performing integration) the properties always include an numerical error ($g(r)+\delta g(r)$, $U^{CG}(r)+\delta U^{CG}(r)$). The Henderson theorem shows the uniqueness of the potential, but it does not make a statement on how an error $\delta U^{CG}(r)$ affects $g(r) + \delta g(r)$. Thus, a priori, large changes $\delta U^{CG}(r)$ could have little effect on $g(r) + \delta g(r)$. This is indeed what has been found here (and by other studies [57]) and is shown in figure 3.4.

The isothermal compressibility on the other hand is very sensitive to changes
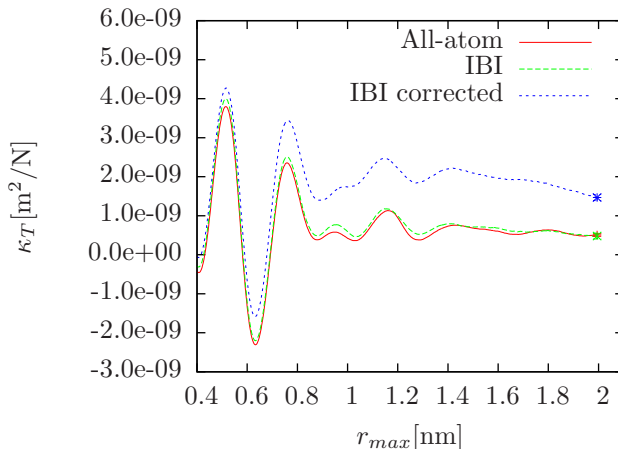
Figure 3.5: Isothermal compressibility estimate as obtained from eq. 3.4 as a function of the upper integration boundary r_{max}. The estimate converges in the limit of $r_{max} \rightarrow \infty$. At large r_{max} the compressibility of the pressure corrected CG model deviates significantly from the atomistic reference due to the sensitivity on the $g(r)$.

in structure. Here, the isothermal compressibility was calculated from the virial expression [26]:

$$\rho k_B T \kappa_T = \lim_{r_{max} \rightarrow \infty} 1 + 4\pi\rho \int_0^{r_{max}} r^2(g(r) - 1)\mathrm{d}r \; , \qquad (3.4)$$

where κ_T is the isothermal compressibility and ρ is the density. For IBI without pressure correction, the compressibility is matched by construction (up to numerical errors), if the $g(r)$ of the reference simulation is reproduced. Using eq. 3.4, the compressibility was estimated integrating the radial distribution function up to r_{max} for both all-atom and CG simulations. In fig. 3.5 the estimate of the compressibility is shown as function of r_{max}. The pressure corrected model gives an compressibility which is almost 5 times higher as shown in table 3.2. Although the CG model now still represents the structure well, this is only possible at the expense of changing the compressibility. This fact is quite relevant for the usage of the CG potential in the multi-scale simulation. The isothermal compressibility is, in the grand canonical ensemble [26], connected to the number density fluctuation by the formula $\rho k_B T \kappa_T = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle}$. Therefore, an increase in compressibility will promote density fluctuations, which is not desirable in applications where finite size effects are relevant. One example is the calculation of Kirkwood-Buff integrals used to determine solvation free energies [106, 107].

	All-atom	IBI	IBI PC
Pressure p	$p_{target} = 1\,\text{bar}$	$6243\,\text{bar} \pm 1\,\text{bar}$	$0.7 \pm 2\,\text{bar}$
Compressibility κ_T $[10^{-10}\,\text{Pa}^{-1}]$	5.1 ± 0.3	4.8 ± 0.7	14.7 ± 0.7

Table 3.2: Summary of properties measured in NVT simulations using the CG potential compared to the all-atom reference simulation. Errors for the compressibility are calculated by considering the statistical error of the respective $g(r)$, which enters in the integral in eq. 3.4. The systematic error of truncating at finite r_{max} is not included.

3.1.3 Summary coarse-graining

Using the IBI technique it is possible to obtain a CG potential, which accurately reproduces the fluid structure in terms of pair correlation function and the compressibility. Other properties are however not necessarily reproduced. Hence these properties should be carefully checked before interpreting results which are obtained in CG simulations. For the example of the pressure it was shown that compromises have to be made if it is necessary to fit a thermodynamic property in addition to structure.

It has been frequently reported in the literature that in addition to this problem, which can be called a representability problem [108] there exist also transferabilty problems, meaning that the coarse-grained model fails at state points different from the one it was parametrized in [109]. This is a consequence of explicit appearance of the temperature in the derivation of the IBI formula eq. 2.43. Overall, there exists no unique way to derive a CG model, only different strategies (discussed in section 2.3) to ensure the representation of a set of selected properties.

In the next section results from coupling CG and all-atom representation of water in the adaptive resolution scheme (see section 2.4) are presented. For this, the differences in the two models play a major role.

3.2 Adaptive resolution simulations

In this part, the AdResS method is extended to couple sub-systems at arbitrary state points. This is done by introducing the concept of a "thermodynamic force", which takes care of the pressure differences occurring due to the mismatch of all-atom and CG pressures. Using the thermodynamic force it is possible to use of the uncorrected CG model in the adaptive simulation.

First the adaptive simulations using the pressure corrected CG model is presented. The thermodynamic force is introduced in section 3.3 and applied to simulation of liquid water in section 3.4.
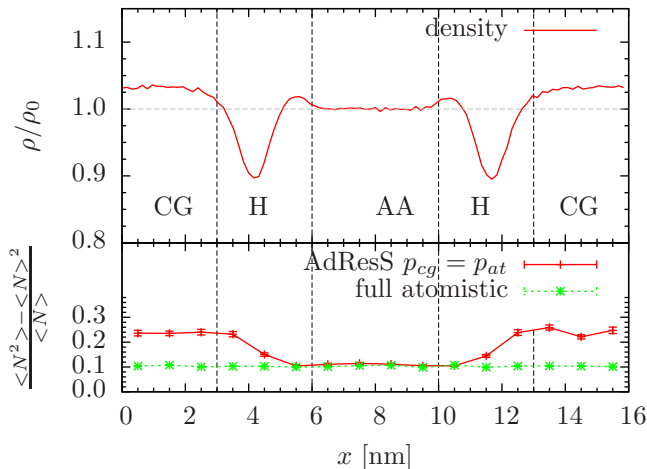
Figure 3.6: Density profile (top) and local particle number fluctuation (bottom) in the AdResS simulation employing the pressure corrected CG model. The particle fluctuation is calculated in sub-volumes of size $1\,\text{nm} \times 4\,\text{nm} \times 4\,\text{nm}$ along the x-axis.

3.2.1 Pressure corrected CG model

This section briefly presents the numerical results obtained with the pressure corrected CG model described in previous section 3.1.2. A similar study of water (using the TIP3P water model) was already reported in [110].

The system consisted of 8507 SPC/E water molecules in a box of dimensions $16\,\text{nm} \times 4\,\text{nm} \times 4\,\text{nm}$, employing periodic boundary conditions. All simulations were carried out in a modified version of GROMACS[103] [2]. The implementation is discussed in section 4.1 and appendix B.

Electrostatic interactions were treated with the reaction-field approach using a cut-off length of 1.2 nm (see section 2.1.4). The temperature was controlled using a Langevin thermostat (see sec. 2.1.2.1) with a coupling constant of $\tau_{\text{Langevin}} = 0.5ps$. The change of resolution is performed along the x-axis of the simulation box, with the center of the all-atom region at $x = 8\,\text{nm}$. The width of all-atom region was chosen to be $a = 2\,\text{nm}$ while, $b - a = 3\,\text{nm}$ were admitted for the hybrid region width (compare eq. 2.73 for the definitions). Note that this system represents a somewhat idealized case, in practice, it is more useful to have a thinner hybrid region and a larger CG "reservoir".

The system was then evolved for 10^6 MD steps employing a time step of $\Delta t = 2\,\text{fs}$. The resulting density profile along the x-axis is shown in fig. 3.6. The profile shows deviations from the reference flat density profile ρ_0 in agreement with previous studies [71, 110]. In the all-atom region the density is preserved,

---

[2]This implementation of the AdResS method will be included in the GROMACS release version 4.6 (www.gromacs.org)

while the hybrid region and also the CG region show deviations on the order of 10%. The artifact of the density depletion is well know from previous AdResS studies and several rationalizations have been discussed. In the first article [70] it was noted that even when the pressures of the "pure" all-atom and CG systems (i.e. a constant $w(x) = w = 1$ and $w = 0$, respectively) are the same, a system simulated with a constant $w(x) = w = \frac{1}{2}$ has a different pressure. The pressure in the simulation with $w = \frac{1}{2}$ is higher in this case and thus particles are pushed out of the hybrid region. In a subsequent paper [71] an "interface-pressure" correction was introduced which can, in first approximation, correct for this by calculating the pressure from a single hybrid simulation with $w(x) = w = \frac{1}{2}$. A more detailed study [76], the density depletion was rationalized to stem from a dependency of the chemical potential $\mu(w)$ on the weighting function. A correction was derived from considering several independent simulations at constant w and calculating the chemical potential difference using Widoms insertion method [111].

Although the density in the all-atom region remains unchanged, for some applications it might be necessary to restore a full density profile everwhere. With the concept of the thermodynamic force, introduced in section 3.3, another way of compensating the pressure artifact is available.

Since the pressure corrected model has a five times higher compressibility, it is interesting to know the compressibility of the adaptive simulation. In a conventional simulation, the compressibility would be accessible for example from the fluctuations of the volume at constant pressure [30]. Here however, since the compressibilities in all-atom and CG model is different, it is more appropriate to define a local compressibility by making use of the expression [26]

$$\rho k_B T \kappa_T = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} \ , \tag{3.5}$$

where N is the number of molecules. This expression is valid only in the thermodynamic limit, but can be used to analyze the qualitative behavior of the compressibility by comparing the molecule number fluctuations in different sub-regions. To this end, the molecular number fluctuations are evaluated in slabs of width $\Delta x = 1\,\mathrm{nm}$ along the x-axis. The bottom part of figure fig. 3.6 shows that for all-atom simulation, the molecule number fluctuations are constant along the x-axis as is expected for a homogeneous system. In the adaptive resolution system however, the fluctuations are different for the all-atom and CG region. The fluctuations in the CG region are higher, which corresponds well with the already discussed 5-fold increase in compressibility in the CG model. In the hybrid region, a smooth transition in fluctuations is observed.

Although this setup has been used in various studies in the past, for certain applications it is of importance to keep the molecule fluctuations everywhere (e.g. the calculation of Kirkwood-Buff integrals in mixtures [106, 107]). For matching the pressure of the all-atom and CG model, a sacrifice in accuracy of the pair structure

had to be made which causes the difference in molecule number fluctuations. A more consistent scheme should certainly follow the premise of structural coarse-graining and preserve the structure as accurately as possible. Thus in the following, a scheme is presented which allows to couple the IBI CG model directly by taking care of the pressure difference through the "thermodynamic force".

3.3 Thermodynamic force

As was explained in the previous sections, the structural coarse-graining does not necessarily generate a system at the reference state-point. By construction, the compressibilities match ($\kappa_{T,CG} = \kappa_{T,AA}$), but pressure, free energies and other thermodynamic properties are not preserved. If one attempts to couple this CG model to the all-atom model in a concurrent scheme, the system evolves into a new state driven by the free energy and/or pressure differences. Usually however, the aim is to reproduce the all-atom state point at least in the region of interest (which is usually the all-atom region). One possible solution would be to alter the CG two-body potential, but this leads to drastic changes in the compressibility (see section 3.1.2). Since it is desirable to preserve compressibility (and thus particle number fluctuations), a mechanism for coupling the two sub-systems at different state points is needed.

The approach taken here was to introduce an external force, $\mathbf{F}_{th}(x)$ which removes the effect of the mismatching state points. $\mathbf{F}_{th}(x)$ can be restricted to be non-zero only in the hybrid region, as to minimize direct influence on properties studied in the all-atom and CG regions. The force defined in this way provides the thermodynamic work necessary to adjust the pressure felt by a molecule crossing the hybrid region.

In the following, an iterative scheme will be presented which makes it possible to evaluate the thermodynamic force. As suggested by the requirement that the two sub-systems A and B exchange particles, the ansatz is made in the grand-canonical ensemble with μ, V, T as natural variables (provided that A and B contain a sufficient amount of particles). The corresponding thermodynamic potential is the grand potential $\Phi(\mu, V, T) = pV$. To apply this to the present problem, one can first consider the sub-regions as two independent systems A and B at the same volume and temperature. The grand potential of the system are unequal:

$$\Phi_A = p_A(\mu_A, T)V \neq p_B(\mu_B, T)V = \Phi_B \;, \tag{3.6}$$

with the chemical potentials μ_A and μ_B being different for each system. The difference in Φ can be compensated by adding a constant term, which is the integral of the thermodynamic force $\mathbf{F}_{\text{th}}(x)$:

$$\left(p_A(\mu_A, T) + \hat{\rho}_0 \int_a^b \mathbf{F}_{\text{th}}(x)\mathrm{d}x \right) V = p_B(\mu_B, T)V \;, \tag{3.7}$$

where $\hat{\rho}_0$ is the molecular number density. The integral corresponds to the amount of thermodynamic work equivalent to the difference in grand potential (per molecule) between systems A and B. This thermodynamic work does not have a meaning yet for the two isolated systems, but can be associated with the work that has to be provided when molecules cross the hybrid region in the adaptive simulation.

The connection to the adaptive simulation is now made by considering a situation, where the systems has been enforced to be homogeneous in terms of pressure p, temperature T and chemical potential μ. Note that these are the conditions of two-phase equilibrium (see section 2.4.4). Since in the adaptive simulation, different representation of the same liquid (in the same phase) are coupled to each other, it is in addition useful to require the density to be uniform and match the reference density ρ_0 of the all-atom system.

The ansatz is now made considering the adaptive simulation to be constrained at the reference density of the all-atom system ρ_0 everywhere. Under this constraint, the pressure becomes a function of the spatial coordinates $p(x, \rho_0)$. This is due to the AdResS forces which interpolate between models with different pressure and chemical potential. For a nonuniform pressure profile, an analogous to expression eq. 3.7 can be written as

$$\hat{\rho}_0 \int_a^b \mathbf{F}_{\text{th}}(x)\mathrm{d}x = p(x, \rho_0) \; . \tag{3.8}$$

The thermodynamic force $\mathbf{F}_{\text{th}}$ can hence be taken to be the force, which balances the pressure gradient in the adaptive simulation where $\rho(x) = \rho_0$ is enforced. This is written as

$$\mathbf{F}_{\text{th}}(x) = \frac{1}{\hat{\rho}_0}\nabla p(x, \rho_0) \; . \tag{3.9}$$

Note that the local pressure profile $p(x, \rho_0)$ is not measurable directly since there is a priori no way to enforce the constraint $\rho(x) = \rho_0 = \text{const}$. Instead, an approximation is made by considering the pressure which is reached once the adaptive simulation reaches a mechanical equilibrium. The pressure without a constrained on the density, denoted as $p(x, \rho(x))$, is expanded as a function of the local density $\rho(x)$

$$p(x, \rho(x)) = p(x, \rho_0) + (\rho(x) - \rho_0)\left.\frac{\partial p}{\partial \rho}\right|_{\rho_0} + \mathcal{O}(\rho(x)^2) \; . \tag{3.10}$$

In mechanical equilibrium however, the pressure gradient must disappear, i.e. $\nabla p(x, \rho(x)) = 0$. By taking the gradient on both sides, eq. 3.10 becomes

$$\nabla p(x, \rho_0) = -\nabla \rho(x)\left.\frac{\partial p}{\partial \rho}\right|_{\rho_0} \; , \tag{3.11}$$

where terms of order $\rho(x)^2$ are neglected. Because of the elementary thermodynamic relation $\frac{\partial p}{\partial \rho}|_{T,V} = \frac{1}{\rho \kappa_T}$ [26], where κ_T is the isothermal compressibility, this

can be simplified to

$$\nabla p(x, \rho_0) = -\frac{1}{\rho_0 \kappa_T} \nabla \rho(x) \ . \tag{3.12}$$

The compressibility κ_T can be assumed to be identical for A and B, given that B is obtained by structural coarse-graining of A. By substituting eq. 3.10 into eq. 3.9, the thermodynamic force becomes

$$\mathbf{F}_{\text{th}}^0(x) = -\frac{M_\alpha}{\rho_0^2 \kappa_T} \nabla \rho^{(0)}(x), \tag{3.13}$$

where the molecular number density $\hat{\rho}_0 = \rho_0/M_\alpha$ has been written as a mass density for a system composed of molecules with mass M_α. This approach thus relates (to the first order) the density profile $\rho(x)$ visible in an unperturbed adaptive simulation to the force $\mathbf{F}_{th}$ which compensates the pressure differences in the two models.

As will be clear from the numerical results, this first guess is not fully sufficient to reach a constant density in the simulation, although it is represents a good approximation. Because of this an iterative refinement is attempted as

$$\mathbf{F}_{\text{th}}^{i+1}(x) = \mathbf{F}_{\text{th}}^i(x) - \frac{M_\alpha}{\rho_0^2 \kappa_T} \nabla \rho^{(i)}(x), \tag{3.14}$$

which can be carried out until the gradient in the density $\nabla \rho^{(i)}(x)$ is sufficiently small. The work-flow of the iteration is depicted in fig. 3.7.

Once system is simulated with the converged thermodynamic force, the effect of the pressure difference is removed and a flat density profile is restored. The condition of matching the chemical potential μ was included implicitly by assuming that the local pressure is a function of the chemical potential[3]. The thermodynamic force therefore couples the two representation by providing the work necessary to overcome the difference in pressure and chemical potential. As the homogeneous temperature distribution is ensured by the thermostat, the conditions of matching μ, p and T for the two sub-regions are hence fulfilled. At the same time, the CG model retains its original properties as no modification is necessary to enforce the correct pressure. This will be put to a practical test in the next section 3.4.

3.4 Numerical results

For evaluating the concept of the thermodynamic force, simulations with the same geometry and parameters as in section (3.2) were prepared, now employing the CG model without pressure correction, as described in section 3.1.2. The pressure difference, which is approximately 6×10^3 bar, drives the adaptive system into a stationary state with non-uniform density profile, the density in the all-atom
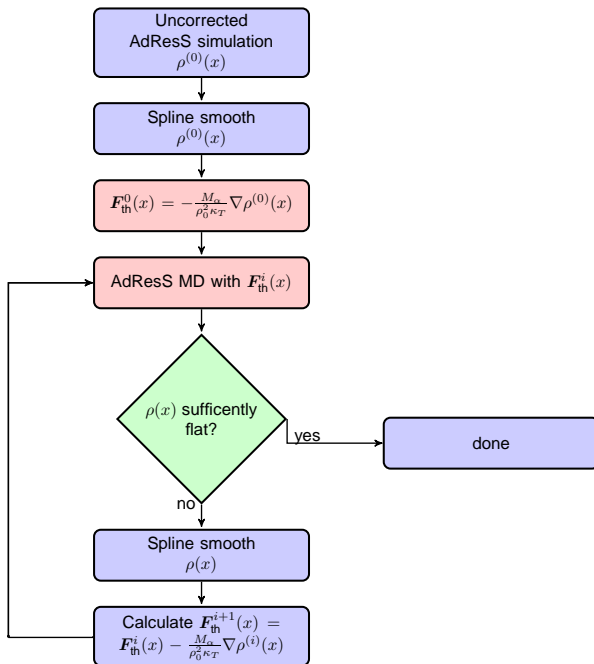
Figure 3.7: Work flow of the thermodynamic force iteration. The criterion for "sufficiently flat density profile" can be realized e.g. as checking that $max((\rho(x) - \rho(0))/\rho_0) < \epsilon$.

region being higher (top graph in fig. 3.8). The thermodynamic force iteration was implemented into the VOTCA package [57]. The work-flow is depicted in figure fig. 3.7 .

5 steps of the iteration were performed, each using an MD simulation of 10^6 steps. The first 10^5 steps were discarded and the density profile was averaged over the remaining frames taken at an interval of 500 MD steps. Due to the sensitivity of the thermodynamic force on the local density fluctuations, a spline fit [112] using a grid 0.4 nm was used in order to smoothen the force. The thermodynamic force was then calculated in the range $1.4\,\text{nm} < l < 4.2\,\text{nm}$ where $l = |x - x_0|$ and x_0 denotes the center of the all-atom region (see section 2.4, for definition). The setup thus extends the thermodynamic force slightly into the all-atom region. This is necessary because molecules can interact in a hybrid way (i.e. $w(X_\alpha)w(X_\beta) \neq 1$) up to one cut-off distance away from the hybrid zone. As is visible from fig. 3.9 this perturbation of the all-atom region is however very small. For the compressibility appearing in eq. 3.14, the all-atom value of $\kappa_T = 0.3 \times 10^{-10}$ was used. The change of the thermodynamic force during the iteration is depicted in fig. 3.9.

The iteration converges to an accuracy of $max((\rho(x) - \rho(0))/\rho_0) = 0.4\%$ after 4 steps of the procedure. It is important to stress here that a sufficient time for

---

[3]The contribution of the chemical potential is further clarified in chapter 6

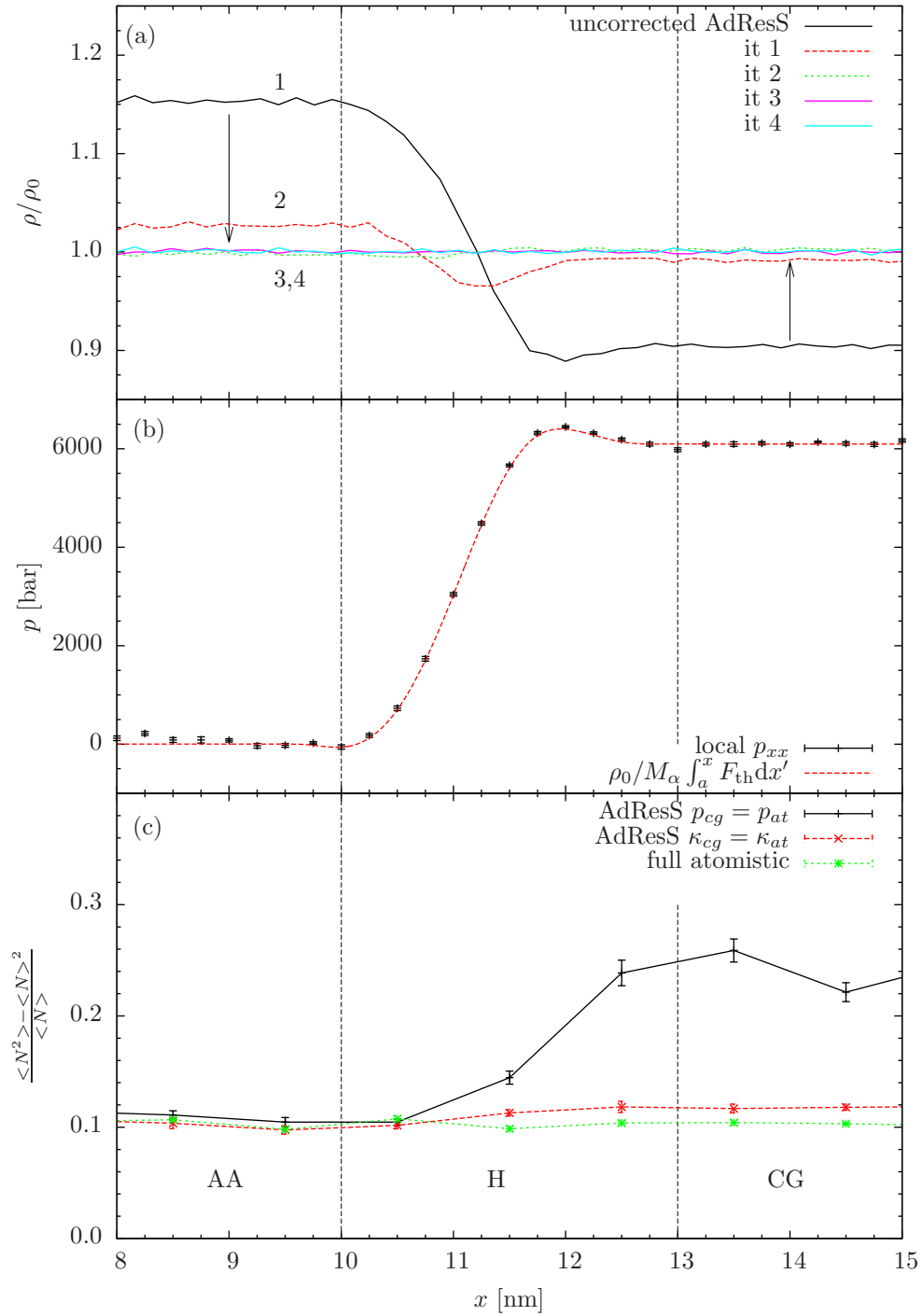Figure 3.8: (a) Density profile $\rho(x)$ in the adaptive simulation without and during the iteration of the thermodynamic force. (b) Local pressure as measured through the method of the planes in the simulation employing the converged thermodynamic force in iteration 4. Also shown is the integral of the thermodynamic force. (c) Molecule number fluctuations calculated in bins of width $\Delta x = 1\,\text{nm}$.
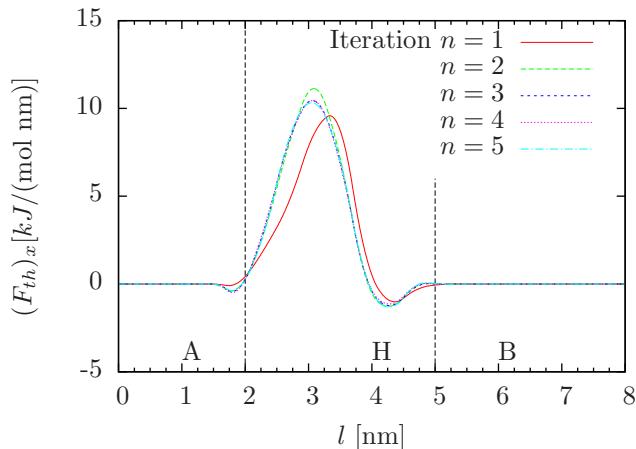
Figure 3.9: Evaluation of the thermodynamic force during the iteration. Since the change of resolution is performed in x-direction $(F_{th})_y = (F_{th})_z = 0$.

each simulation has to be admitted for the relaxation of the system from the initial condition to the new density profile.

The density profile is shown in fig. 3.8a) for the first 4 steps of the iteration. Due to the effect of the thermodynamic force, the density imbalance is removed and the original density restored.

The derivation for the thermodynamic force started with the ansatz of balancing the pressure gradient, given that the system is constraint to be at ρ_0. It is thus of interest to check numerically that the pressure is exactly balanced in the simulation with the thermodynamic force. To that end, the p_{xx} component of the pressure tensor was computed locally a long the x-axis using the formula derived by Todd, Evans and Daivis [113] (for a description, see appendix A.1). In figure fig. 3.8b) the resulting pressure profile is compared to the integral of the thermodynamic force (i.e. the work related with pulling a particle across the hybrid region). The two curves overlap within the error bars. Thus the contribution of the thermodynamic force and pressure in x-direction on a molecule are equal and of opposite sign and cancel each other. The local virial pressures (not counting the contribution of the external force) are however different. The thermodynamic force therefore fulfills the initial requirement of canceling the drift and coupling the systems at different state points.

It is natural to ask how this coupling affects physical properties. In fig. 3.8c) the particle number fluctuations in slabs (calculated as in the previous section) along the x-axis are shown. In the simulation with the thermodynamic force the fluctuations are now constant over the full simulation box. This indicates that also the compressibility is now preserved everywhere. The remaining deviation between all-atom and CG fluctuations can be attributed to the small difference in
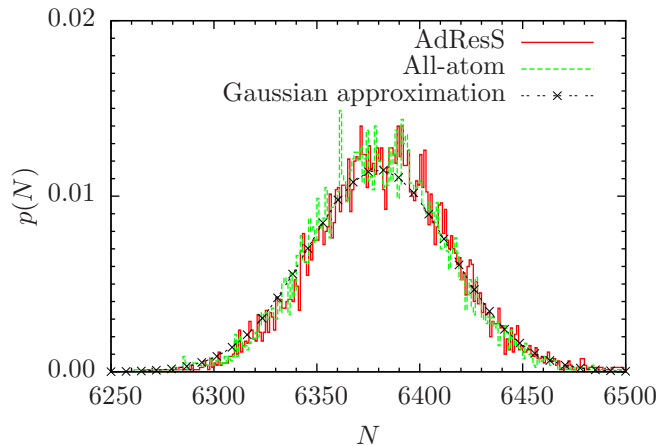
Figure 3.10: Particle number distribution in a slab of width $l = 4$ nm and for an atomistic zone in adaptive resolution with the same width. Both distributions match a Gaussian behavior with mean $\mu = \langle N \rangle$ and variance $\sigma^2 = \rho k_B T \kappa_T \langle N \rangle$

compressibility, which is due to the residual error in the CG RDF observed for the CG model without pressure correction (see section 3.1.2). The accuracy could be further approved by tuning the discretization of the numerical representation of the potential and performing more iterations.

For a more detailed analysis, the distribution of the number of molecules $p(N)$ in the all-atom region was calculated and compared to a sub-volume of the same size in an all-atom simulation. This represents an important check using the CG sub-system as a particle reservoir, the particle fluctuations provided should match the all-atom ones. The all-atom distribution was found to match a Gaussian distribution with mean $< N >= 6380$ and variance $\sigma^2 = \rho k_B T \kappa_T < N >$. The $p(N)$ in the adaptive simulation matches very closely the reference all-atom distribution (shown in fig. 3.10) which suggests that the hybrid setup is able to provide exactly the same fluctuations as the all-atom simulation.

The results so far still do not prove that there is free molecule exchange overall in the system, i.e. there a no (free-) energy barriers for crossing the hybrid region. Numerically, this can be checked by "monitoring" a selected set of molecules which are initially in specified region $x_0 < x < x_1$ at $t = t_0$. The density profile as a function of time is averaged over different initial times t_0, which can be denoted as $< \rho(x, \Delta t) >_{t_0}$, where $\Delta t = t - t_0$. The result is shown in fig. 3.11. This analysis is repeated for molecules which are initially all in the all-atom and all in the CG region. The diffusion profiles spread in time and cross the hybrid region, which proves that the sub-systems exchange particles. Furthermore, the profiles are asymmetric, which can be explained with the fact that the diffusion constant are different in all-atom and CG simulation (see section 3.1). The fact that the
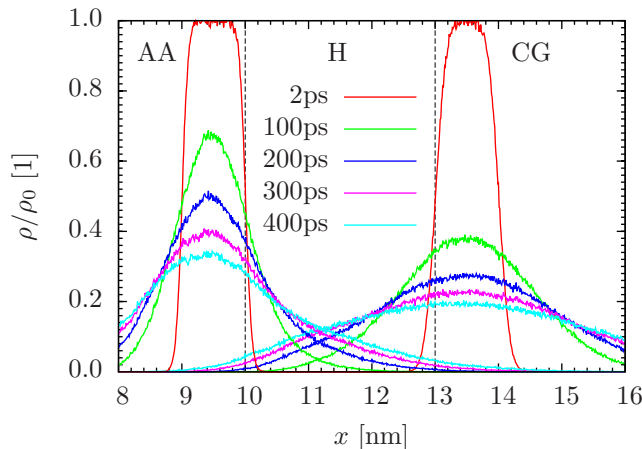
Figure 3.11: Inter-diffusion between the two regions. This shows the density profiles over time for particles originally in a cubic sub-volume with width $\Delta x = 1$nm at $t = t_0$. By this it is clear that both domains constantly exchange particles. The profiles are asymmetric as diffusion constants are different in each representation.

CG system can provide and take out particles in a manner consistent with the all-atom simulation suggest a new way of performing effectively grand-canonical simulations. Of course, for the all-atom sub-system to be in truly grand-canonical equilibrium an infinite reservoir is required. However, the present method can be used for approximating such a reservoir explicitly in a computer simulation since the CG model is cheap to compute.

3.5 Transferability

The thermodynamic force approach enables the coupling of sub-systems which are at different state points. One issue with the present technique is the transferability of the thermodynamic force. Certainly, the shape and magnitude of the force depends on the both atomistic and coarse-grained interactions and thus has to be re-parametrized for each new model. For the use of thermodynamic force in other geometries at least some transferability is expected. Figure 3.12 shows the result from a simulation using a smaller hybrid region width ($b - a = 1.8$), for which the thermodynamic force iteration was re-done. The converged thermodynamic force is shown together with the thermodynamic force for the simulation with $b - a = 3.0$ and rescaled force with $F_{th}^{scaled}(x') = \frac{1}{s}F(sx)$ where s is a scaling parameter defined as the relative width of the hybrid zones $s = 3.0/1.8$. The scaled thermodynamic force is similar to the originally calculated one, with some small deviations.

This suggest that the thermodynamic force shows some transferability between different geometries. Therefore, the iteration can be performed using smaller sys-
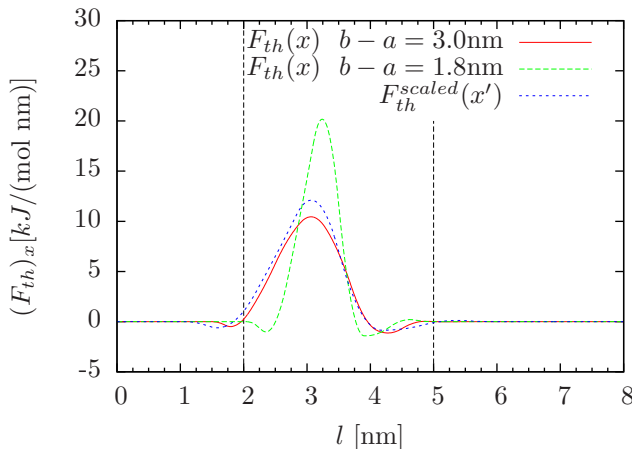
Figure 3.12: Transferability of the thermodynamic force for different geometries. F_{th} was calculated for a hybrid zone width of $b - a = 3\,\text{nm}$ and $1.8\,\text{nm}$. The thermodynamic force rescaled with $s = 3/1.8$ matches quite closely the reference thermodynamic force

tems, which saves computer time and then be rescaled to fit the larger system. If this approximation is not accurate enough, it can be at least used as an initial guess in the larger system. One possible explanation for the differences is the presences of correlations in the liquid, which lead to different effective pressures along the direction of the switching function $w(x)$.

3.6   Summary

In this part of the thesis it was shown how sub-systems at very different state points can be coupled in a concurrent adaptive simulation. The concept of the thermo-dynamic force was used to derive an external field, which removes efficiently the drift coming from the mismatch based on the density profile. With this approach it is possible to preserve the structure everywhere in the adaptive simulation. This was not possible before as the fitting of the pressure into the CG model severely alters the compressibility. The CG system can be viewed as a computationally efficient way to consider a particle reservoir, which yields a grand-canonical MD simulation in the limit of the CG system being much larger then the all-atom systems. In addition, to provide an even better approximation to the grand-canonical system, coupling to continuum models has also been explored [114, 115]. In these approaches, the CG model facilitates the insertion of molecules into the liquid.

The results for liquid water indicate that the method works for complex, realistic system where local ordering, mediated by the hydrogen bonds, is of importance. Potential practical applications include structure formation phenomena (crystal growth, etc) where a supply of particles is needed. The method was already used

for studying co-solvent mixtures [106, 107].

Despite the robustness of the thermodynamic force iteration, there is also room for improvement. One drawback certainly is that the relaxation time of the density profile is quite long and thus long simulations are required to reach the mechanical equilibrium. Furthermore, the fluctuations in density control the accuracy and thus the simulation time has be increased. This can be partially overcome by simulating smaller systems and scaling the results as described in the previous sections. However, it would be preferable to have an analytical first guess of the zeroth order thermodynamic force. An alternative approach, which does not rely the density approximation, is discussed in chapter 6.

More generally, it has to be evaluated carefully for each new application if the adaptive resolution method is able to preserve the properties of interest. The coarse-graining procedure itself is connected with some loss of thermodynamic properties which might influence the physical properties of interest. Further testing of the adaptive resolution scheme was done in the following chapter, where a C60 fullerene was simulated with the solvent described in adaptive resolution.

Chapter 4

Adaptive resolution simulation of toluene

In the following chapter, the adaptive resolution technique is put to the test in a study of a C60 fullerene [116] in a toluene solvent. This system was used as a reference for the implementation of the AdResS method into the GROMACS [103] molecular dynamics software package.

The C60 fullerene was selected as a test molecule considering that there are many potential applications in organic electronics [117], biopharmaceutics and drug delivery [118]. For these applications, a microscopic understanding available from a computer simulation can be of aid with the interpretation of experiments. An efficient simulation scheme could extend the length and time scales available for example for studies of the aggregation behavior. The C60 is surrounded by a spherical all-atom region, where the full all-atom interactions for the toluene solvent are used. Outside this region, the toluene solvent is represented by a CG interaction potential obtained by IBI as described in the previous chapters (a schematic representation of this setup is given in fig. 4.1). Toluene was chosen as solvent as this molecule is very planar and thus near-ordering is expected.

The adaptive resolution technique then allows to answer questions, which cannot be answered by conventional simulations: Since the size of the all-atom and hybrid layers can be varied, it is possible to ask how much of the full detailed structure in the surrounding is required to support the all-atom ordering. Besides

this being an interesting study by itself, for practical AdResS simulations it will be required to maintain the all-atom ordering in the high-resolution region. The test system studied thus provides a benchmark as to what size of all-atom and hybrid regions has to be chosen in order to maintaint the order close to the solute. In a similar study, this approach was used to investigate the influence of the bulk water hydrogen bonding network on the solvation structure around fullerenes [119].

This chapter is organized as follows: First the development of a CG model for toluene is discussed. An order parameter is introduced for characterizing the local solvent structure. Then, results from adaptive simulations of the CG model with a C60 molecule in the center of the all-atom region are presented. The main focus lies on how the local ordering is affected by contact with the CG reservoir and what effects on the systems dynamics can be observed. At the end of the chapter the challenges of an efficient implementation of the AdResS scheme are outlined.

4.0.1 Coarse-grained model of liquid toluene

For the all-atom reference system, 1000 molecules of pure toluene were simulated in a box of size $5.6\,\mathrm{nm} \times 5.6\,\mathrm{nm} \times 5.6\,\mathrm{nm}$, employing periodic boundary conditions. The parameters used to model the all-atom toluene interactions were taken from the OPLS forcefield [120, 121] where the CH_3 group is represented by a single interaction site (united atom model). Electrostatic interactions needed due to the partially charged atoms in toluene were evaluated using the reaction field method (as described in section 2.1.4). In order to determine the unknown effective dielectric constant for this model, initial simulations employing long range electrostatics (PME) [122] were performed. The dielectric constant was calculated using the fluctuation of the total dipole moment [4], which yielded $\epsilon_{\mathrm{PME}} = 1.1619$. Using ϵ_{PME} as reaction-field constant however yields a system with a slightly deviating dielectric constant $\epsilon_{\mathrm{dipole}}^{(i)}$. A series of simulations was then performed, where $\epsilon_{\mathrm{RF}}^{(i+1)} = \epsilon_{\mathrm{dipole}}^{(i)}$ until both values were close. The final iteration yielded $\epsilon_{\mathrm{RF}} = 1.16735$ for the reaction-field constant. In order to relax the system to the target pressure $p = 1\,\mathrm{bar}$, simulations with a Berendsen barostat [101] were performed, which resulted in an equilibrium density of $\rho = 5.636/\,\mathrm{nm}^3$. For thermostatting, a Langevin thermostat (as explained in section 2.1.2.1) with a time constant of $\tau_{\mathrm{Langevin}} = 3\,\mathrm{ps}$ was used.

The CG potential of toluene was then determined using the iterative Boltzmann technique (as described in 2.3.4). The interaction potential was chosen to be a function of the toluene center of mass. Using the VOTCA [57] package, 150 steps of IBI were carried out, each of a $2\,\mathrm{ns}$ length. The cutoff of the CG potential was set to be $r_{cut} = 1.32\,\mathrm{nm}$, which corresponds to the second minimum in the RDF of the all-atom system, depicted in fig. 4.2.

As described for the case of water (section 3.1.2), the pressure in the CG model was found to be severely altered. A pressure correction protocol was used as
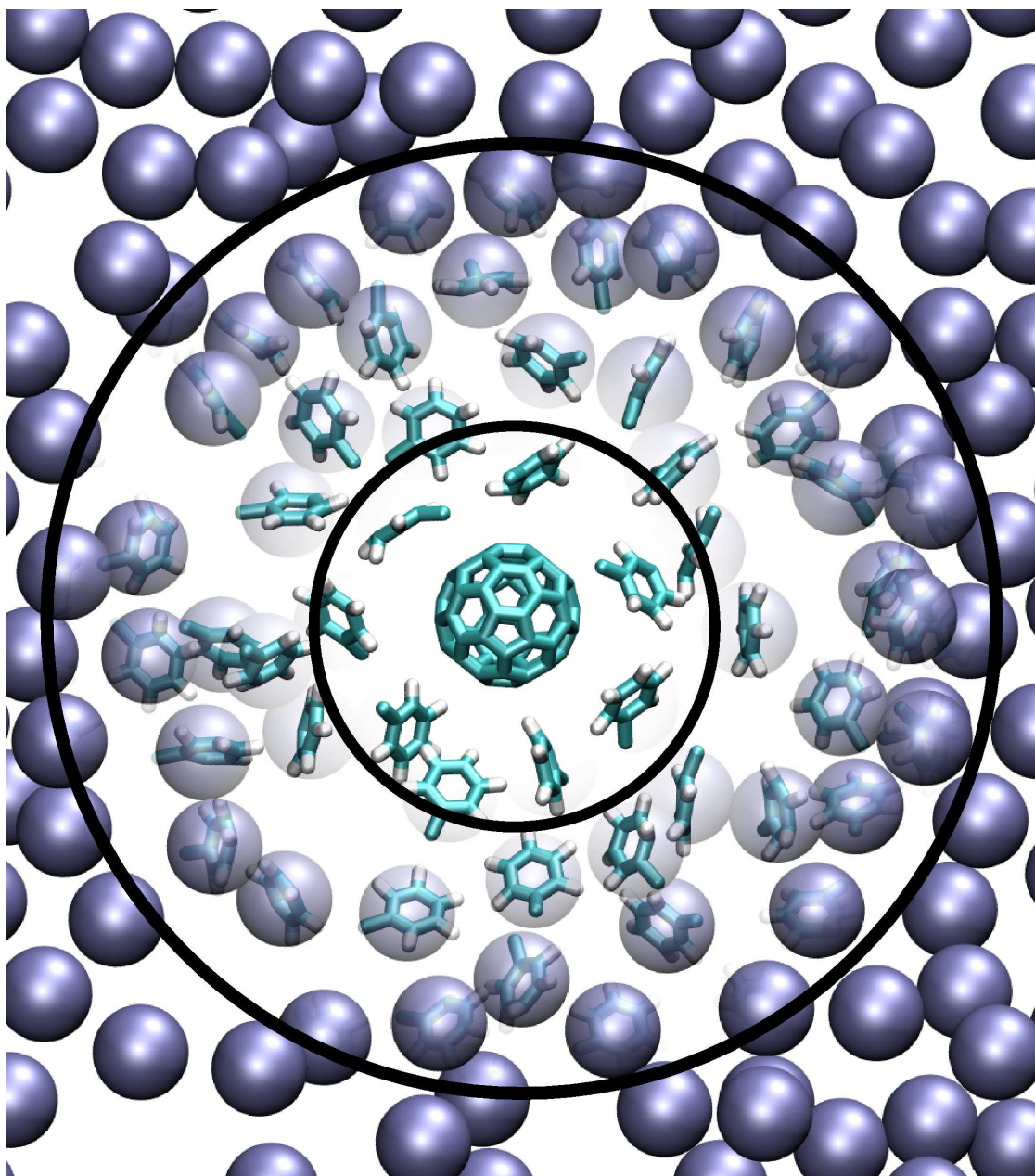
Figure 4.1: Illustration of the adaptive simulation setup with spherical all-atom region. The solvent molecules close to the solute are treated with full detail. In the CG region a toluene molecule only interacts according to its center of mass.
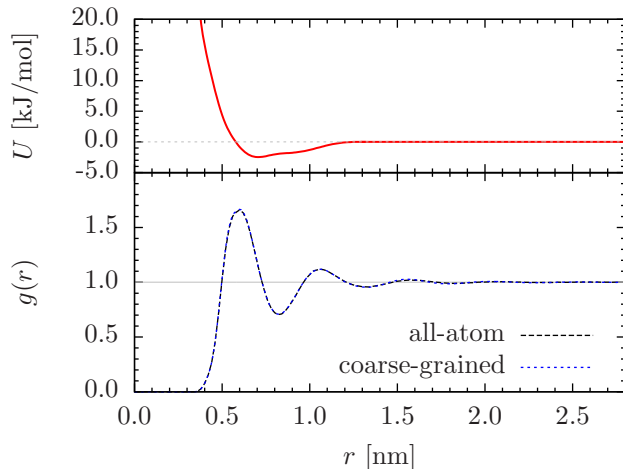
Figure 4.2: Toluene center of mass radial distribution function in all-atom and coarse-grained representation. The RDFs match within line thickness.

described in section 3.1.2, here using 17 steps with pressure and potential update terms (eq. 3.2) and the last 3 with only the pressure term (eq. 3.3). The resulting CG potential is shown together with the RDF in fig. 4.2. The CG RDF matches perfectly the reference distribution. The pressure correction is however known to alter the compressibility, which was measured from the volume fluctuations during an NPT simulation, using the Parrinello-Rahman barostat [123]. For the all-atom compressibility it was found that $\kappa_T^{\text{all-atom}} \approx 1.18 \times 10^{-4}\,\text{bar}^{-1}$, while in the CG model $\kappa_T^{\text{CG}} \approx 7.7 \times 10^{-4}\,\text{bar}^{-1}$. The CG model hence reproduces well the overal pair-wise structure and the pressure, while the compressibility is not preserved.

4.0.2 All-atom simulation

First all-atom reference simulations consisting of 3987 molecules of pure toluene and the C60 fullerene were performed in a box of size $8.9\,\text{nm} \times 8.9\,\text{nm} \times 8.9\,\text{nm}$, employing periodic boundary conditions. The C60 interactions were chosen as in Weiss et al. [124], where the same non-bonded interaction parameters for the carbon-carbon interaction as the OPLS force-field ($\sigma = 0.355\,\text{nm}$ and $\epsilon = 0.293\,\text{kJ/mol}$) were used. The equilibrium density of the all-atom system containing the fullerene was determined by 400 ps simulation employing a Parrinello-Rahman barostat [123]. 4 ns where then simulated for the production runs in the NVT ensemble.

In order to characterize the orientational structure of the solvent an order parameter is introduced which resembles a nematic order parameter [125].

$$Q = \left\langle \frac{3}{2}\cos^2(\theta) - \frac{1}{2} \right\rangle \tag{4.1}$$
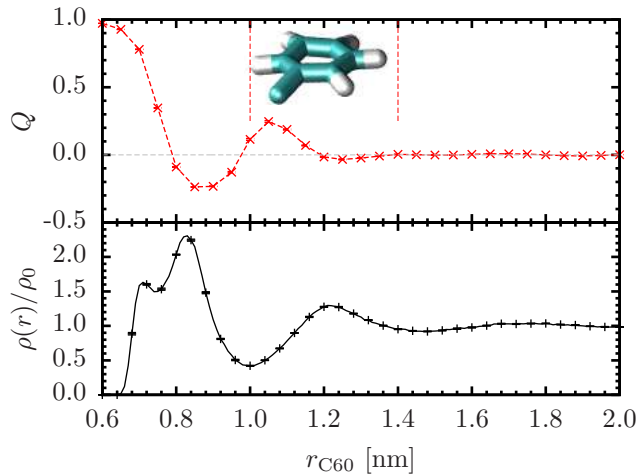
Figure 4.3: Spherical order parameter Q and spherical center of mass distribution for an all-atom simulation as a function of the relative distance r_{C60} to the center of mass of the C60 molecule. All values are averaged over the simulation trajectory (error bars are smaller than the line width). The inset shows the toluene molecule in the the same length scale.

where the angle $\cos(\theta) = \hat{\mathbf{n}} \cdot \hat{\mathbf{r}}$ describes the relative orientation of the smallest principal axis $\hat{\mathbf{n}}$ of a toluene molecule and the direction vector $\hat{\mathbf{r}}$ connecting C60 and toluene center of mass. The value of Q describes the orientation of the molecules relative to the C60 surface. The order parameter can be analyzed by averaging over molecules and time in spherical shells a distance r away from the fullerene. $Q = 1$ represents parallel orientation with respect to the surface, $Q = -0.5$ is perpendicular orientation. An average $Q = 0$ value means that there is no preferred orientation with respect to the surface. Figure 4.3 depecits the order parameter and the density as a function of the radial distance from the fullerene. The shell up to $0.9\,\mathrm{nm}$ consists of a double peak in density. The order parameter reveals that molecules in the first density peak prefer perpendicular and in the second peak orthogonal orientation relative to the surface. At $r = 1\,\mathrm{nm}$ a minimum in the density is visible in which the average order parameter is 0. The position of this minimum is in agreement with experimental studies using small-angle x-ray scattering [126, 127]. Another maximum in the density at $r = 1.2\,\mathrm{nm}$ shows only weak ordering perpendicular to the surface. Thus it can be concluded that approximately up to a distance of $1.2\,\mathrm{nm}$ the ordering of the molecules is affected by the presence of the C60. In the next section it will be discussed how this is affected by the adaptive simulation.

4.0.3   Adaptive simulation setup

The adaptive resolution simulations were performed as described in section 2.4, with the difference that the all-atom region is spherical. This was realized by choosing a weighting function $w(r)$ which depends on the absolute distance to the center of mass of the fullerene $r_\alpha = |\mathbf{X}_\alpha - \mathbf{X}_{\mathrm{C60}}|$. As was already discussed in section 3.4, the all-atom region is slightly perturbed by the presence of hybrid interactions $(w(r_\alpha)w(r_\beta) \neq 1)$ up to a cutoff distance away. To exclude an influence from this, a slighlty modified definition of all-atom region is used. The interpolation function (compare to eq. 2.73) then is

$$
w(r) = \begin{cases}
0 & : r > d'_{\mathrm{at}} + d_{\mathrm{hy}} \\
\cos^2\left(\frac{\pi}{2d_{\mathrm{hy}}}(r - d'_{\mathrm{at}})\right) & : d'_{\mathrm{at}} + d_{\mathrm{hy}} > r > d'_{\mathrm{at}} \\
1 & : d'_{\mathrm{at}} > r
\end{cases} \quad , \tag{4.2}
$$

where $d_{at'} = d_{at} + d_{skin}$ is the definition of the force interpolation region region as in section 2.4. The radius of the all-atom region is then defined by d_{at} and is hence simply shifted by the value of d_{skin}. The value of d_{skin} is chosen such that an pertubation from the hybrid interactions can be savely exluded.

For the simulations where the average structure of the solvent molecules is studied, the C60 molecule was restrained to its initial position by using harmonic restraints (i.e. the potential energy is extended by $V_{rest} = \frac{k}{2}(\mathbf{X}_{\mathrm{C60}} - \mathbf{X}_{ref})^2$ where $k = 1 \times 10^4 \frac{\mathrm{kJ}}{\mathrm{mol\,nm^2}}$ and $\mathbf{X}_{ref}$ was chosen as the center of the simulation box.

4.0.4   Thermodynamic force

As described in section 3.2, the adaptive scheme is known to produce a density depletion in the hybrid region. This artifact can be healed by the thermodynamic force approach, as described in the previous chapter. To this end, the thermodynamic force was parametrized in a smaller system containing 1000 molecules of pure toluene at the same density. Since in this simulation the all-atom region was spherical, the thermodynamic force is a function of the distance to the C60 center of mass $\mathbf{F}_{th} = F(r)_{th}\hat{\mathbf{r}}$ an points radially away from the the center of the all-atom zone. The size of the hybrid region was $d_{at} = 0.75\,\mathrm{nm}$ where a skin width $d_{skin} = 0.2\,\mathrm{nm}$ was found to give good results.

The thermodyamic force was calculated in 60 iterations of the scheme described in section 3.3, each iteration step consisting of 5×10^5 MD steps. The density profile of the initial (withouth thermodynamic force) and final step is compared in fig. 4.4. The iteration shows a flat density profile after 60 steps.

Also shown is the initial guess compared to the final converged integrated thermodynamic force $V_{th}^{(60)}(r)$. It is visible that a linear scaling of the inital guess is not sufficient to arrive at the converged thermodynamic force. This emphasizes the need for determining the thermodynamic force in a iterative procedure. It
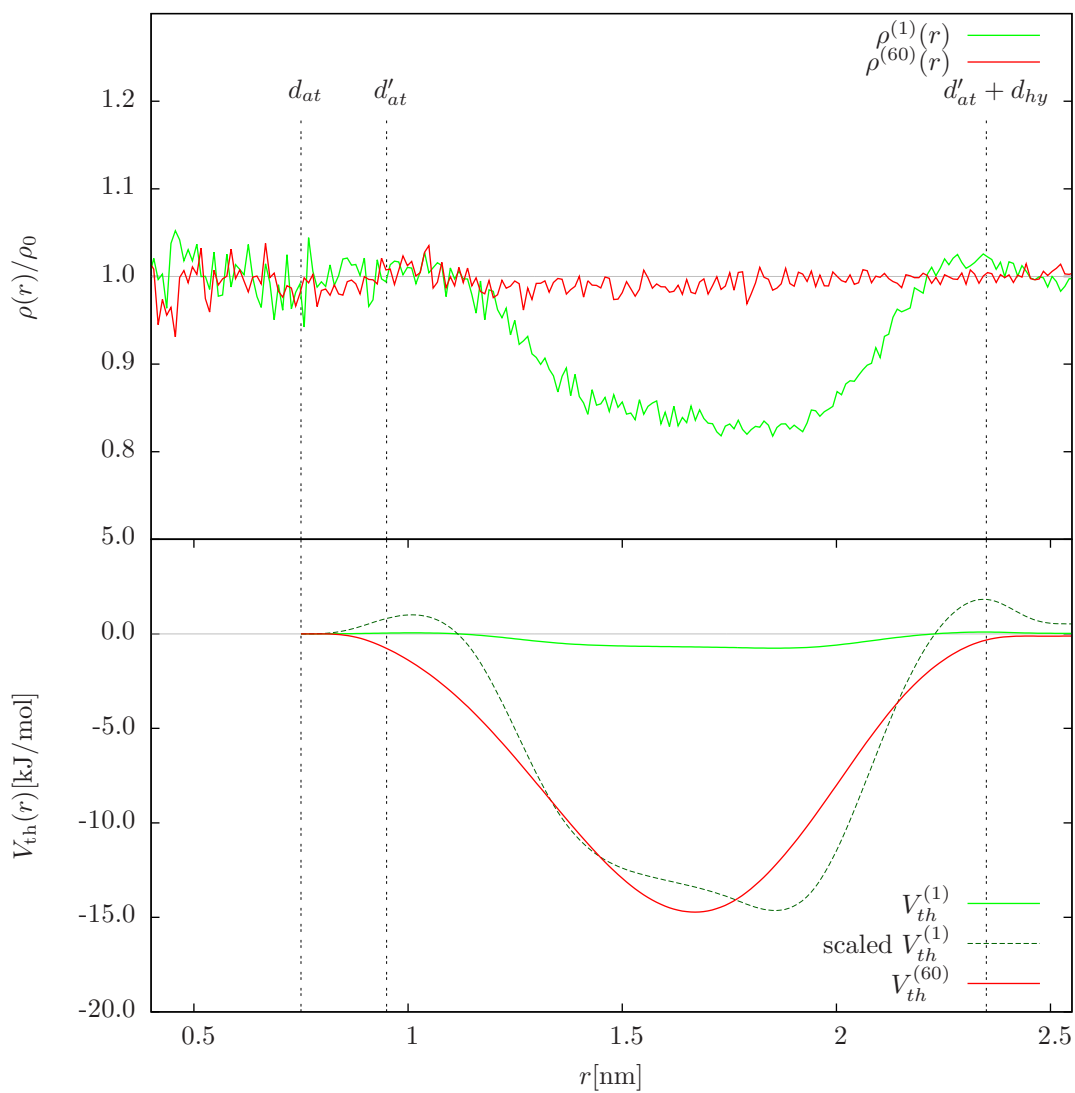
Figure 4.4: Thermodynamic force iteration in the spherical AdResS simulation. Shown are the density profile $\rho(r)$ and the integrated thermodynamic force $V_{th}(r)$ for the initial and final steps of the iteration. The initial thermodynamic force guess is shown scaled by a constant factor $kV_{th}^{(1)}(r)$ where $k = V_{th,\max}^{(60)}/V_{th,\max}^{(1)}$.

is important to stress a few differences to the thermodynamic force iteration in the previous chapter: First of all, as the volume of a thin spherical slab increases with r^2, the fluctuations in the radial density profile are larger for small r. This can cause difficulties as the density profile needs to be converged for deriving the thermodynamic force. Secondly, more iterations where necessary to converge the $V_{th}(r)$ compared to the AdResS simulation where the resolution was change along the x-axis. A possible explanation is that correlations in the liquid have a larger effect here since the volume of two subsequent spherical shells is different, which may invalidate the ansatz of linearizing the pressure in terms of the density. Another difference compared to chapter 3 is that here the pressure corrected CG potential has been used. As discussed before this alters the compressibility in the CG model, which was found to be increased by approximatly a factor of 7 (see section 4.0.1). The ansatz in the thermodynamic force iteration eq. 3.14 however assumes constant compressibility. As a consequence of this, preliminary simulations employing the all-atom compressibility were found to lead to an diverging thermodynamic force. This was overcome by inserting an increased value $\kappa_T = 3\kappa_T^{\mathrm{CG}}$ for the compressibility, which leads to a smaller prefactor to the density gradient. The iteration was then stable, however also required a larger number of iterations to converge.

4.0.5 Results of the adaptive simulations

Now being able to run adaptive simulations at constant density throughout, it is interesting to ask how the ordering of the solvent molecules is affected. In the CG part of the simulation the order parameter Q is zero by definition as the internal degrees of freedom are decoupled from the rest and thus cannot be ordered. How does this influence the all-atom order?

By varying the size of the all-atom region, the influence of the resolution change on the local ordering can be studied. The adaptive simulations were performed using three different radii of the all-atom region ($d_{at} = 0.75\,\mathrm{nm}$, $d_{at} = 1.0\,\mathrm{nm}$, $d_{at} = 1.5\,\mathrm{nm}$), while the size of the hybrid region was kept fixed $d_{hy} = 1.4\,\mathrm{nm}$.

First, the density is compared for different all-atom region sizes (fig. 4.5). The density profile matches in the all-atom region for all simulations, however in the hybrid region artifacts are visible. As discussed above, without the thermodynamic force there is a small depletion in the hybrid region, which also affects the solvation structure in the C60. The thermodynamic force, which was parameterized in the small system of pure toluene, was then used for the solvated C60. As can be seen from fig. 4.5, using the thermodynamic force, it is possible to recover the all-atom density profile. This is an interesting result as it is not clear a priory, but very desirable, if the thermodynamic force is transferable from the bulk to the solute system where strong local density variations are present.

The order parameter Q in the adaptive simulation is shown in fig. 4.6 for the simulation employing the smallest all-atom region (a=0.75 nm). It is visible that
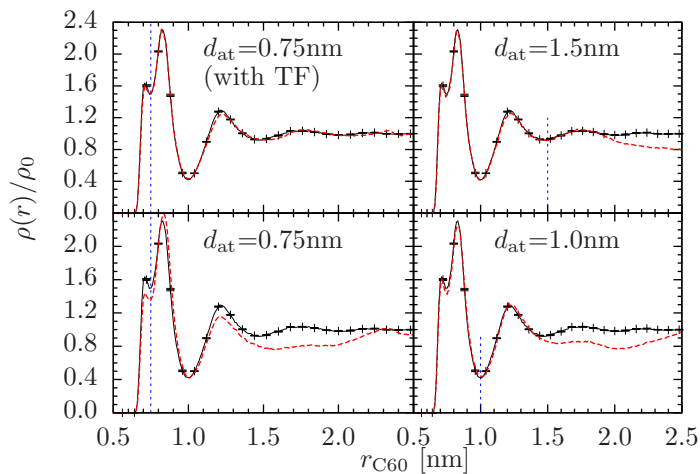
Figure 4.5: Spherical center of mass distribution relative to the center of mass of the C60. The distribution (red line) is plotted for different sizes of the all-atom region and with the thermodynamic force applied to the hybrid region in comparison to the purely all-atom distribution (black line). The vertical lines indicate the size of the all-atom region.

the ordering remains very close to the all-atom ordering in the simulation where no thermodynamic force is applied (even in the hybrid region, where the density is perturbed). In the hybrid region, the all-atom order is still present because the internal degrees of freedom are not yet fully decoupled. The simulation with the density corrected by the thermodynamic force, however, shows a slight deviation in the order parameter Q. The reason for this deviation is that the thermodynamic force does not take into account the orientation of the molecules. The fact that there is almost no deviation in the hybrid region, indicates that the ordering is a local effect. It would be interesting to reduce the size of the hybrid region further, but this is limited by a drawback in the current scheme: The width of the hybrid region cannot be reduced below the interaction cutoff in the simulation. If the hybrid region is smaller then the cutoff, strong layering and eventually freezing of the liquid was observed.

4.0.6 Freely diffusing all-atom region

Another interesting aspect of this are the dynamical properties. It is a well known fact that the dynamical properties in the CG model can show severe deviations with respect to the all-atom model. This is due to a much smoother energy landscape as a consequence of the reduced degrees of freedom which enhances the diffusion rates [54]. The diffusion constant found for the CG model was increased by approximately a factor of 4 (see tab. 4.1 for the exact values). Certainly this also affects the self diffusion constant in the CG region of the adaptive simulation. But what
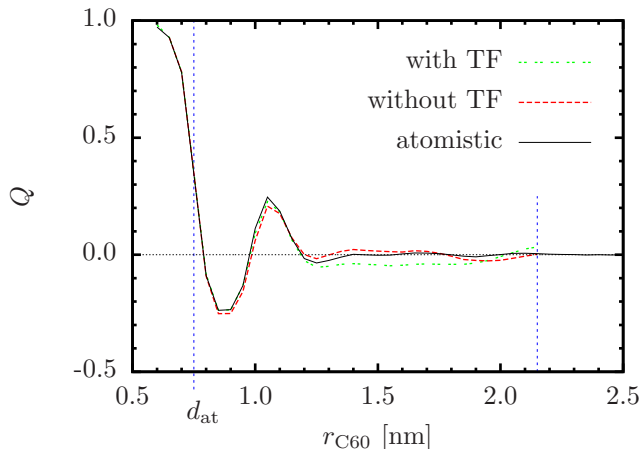
Figure 4.6: Comparison of order parameter with and without thermodynamic force for an all-atom region with size $d_{at} = 0.75$ nm. The all-atom value is shown for reference. All values are averaged over the simulation trajectory and normalized to the respective average number of molecules in spherical slabs with a thickness of 0.05 nm. The verticals lines indicate the boundaries of the hybrid region.

about the all-atom part? Is the effect fully localized to the CG region? This effect was studied here by removing the restraint on the C60 molecule thus allowing it to diffuse freely. The all-atom region was coupled to the C60 such that it diffuses along with its center of mass (i.e. allowing the center of the all-atom region to be a time dependant vector $\mathbf{X}_{C60}(t)$). This study has to be performed with some care as, due to the non-conservative forces occurring in the hybrid region, a thermostat has to be used. Thermostatting however alters the dynamics of the system due to the friction with the heat bath. To probe the influence of this, a series of initial test simulations with varying relaxation times were performed and the diffusion constant was measure as a function of the thermostat relaxation time (see section 2.1.2.1). It was found that for a value of $\tau_{Langevin} = 3$ ps the diffusion constant reaches a plateau, such that a weak influence of the heat reservoir can be assumed.

Using the weak coupling, adaptive simulations were performed for the different region sizes. The diffusion of the C60 center of mass is measured from the mean-squared displacement of the center of mass given by the Einstein relation [29]

$$6Dt = <|\mathbf{X}(t) - \mathbf{X}(0)|^2> \tag{4.3}$$

where D is the diffusion constant, and t is the time. The resulting diffusion constants are shown in tab. 4.1 and the mean-square displacement in fig. 4.7. For long times, the diffusive regime is reached where the mean-squared displacement is linear in time (coressponding to horizontal curves in fig. 4.7). The adaptive simulation shows enlarged diffusion rates compared to the all-atom simulation which

Table 4.1: Diffusion constants for the C60/toluene center of mass

C_{60}	Diffusion constants $[\mathrm{nm^2/ps}]$
All-atom	$D = 6.3 \times 10^{-4} \pm 6 \times 10^{-5}$
AdResS	
$d_{\mathrm{at}} = 0.75\,\mathrm{nm}$	$D = 9.0 \times 10^{-4} \pm 7 \times 10^{-5}$
$d_{\mathrm{at}} = 1.0\,\mathrm{nm}$	$D = 8.2 \times 10^{-4} \pm 2 \times 10^{-5}$
$d_{\mathrm{at}} = 1.5\,\mathrm{nm}$	$D = 7.9 \times 10^{-4} \pm 3 \times 10^{-5}$
AdResS with TF	
$d_{\mathrm{at}} = 0.75\,\mathrm{nm}$	$D = 7.3 \times 10^{-4} \pm 8 \times 10^{-5}$
Pure toluene	
All-atom	$D = 2.216 \times 10^{-3} \pm 3 \times 10^{-6}$
Coarse-grained	$D = 8.629 \times 10^{-3} \pm 1 \times 10^{-5}$

decrease with increasing all-atom region sizes. This could be explained considering that for the smallest all-atom region sizes, the CG influence is largest as the C60 feels least resistance from the mostly CG surroundings.

Part of the effect is however also due to the density variations which are present without the thermodynamic force. With applying the thermodynamic force, the diffusion constant is recovered within 15% error, even for the smallest all-atom region. This result indicates that for most practical applications, even the smallest all-atom region can reproduce the diffusion properties with good accuracy.

4.1 Implementation details

Within the projects of this thesis, the adaptive resolution scheme was implemented into the GROMACS [103] software package. Though AdResS is in principle easy to implement, an efficient computer code, which enables speed-ups compared to the all-atom simulation is more difficult. In the following the most critical aspects necessary for the GROMACS implementation are discussed.

4.1.1 Force-kernels

In a MD simulation, the most time consuming part is the force calculation, mostly dominated by the sum over non-bonded pairs (in some simulations also the long-range electrostatics plays a major role which were not considered here). Therefore, in GROMACS the sum over pairs is split into the different possible interaction types (Coulomb, Lennard-Jones, etc...) and optimized routines are used for each of these pair loops (these are called "kernels"). The optimizations make use of special hardware instructions (SSE, etc.), which can process multiple instructions
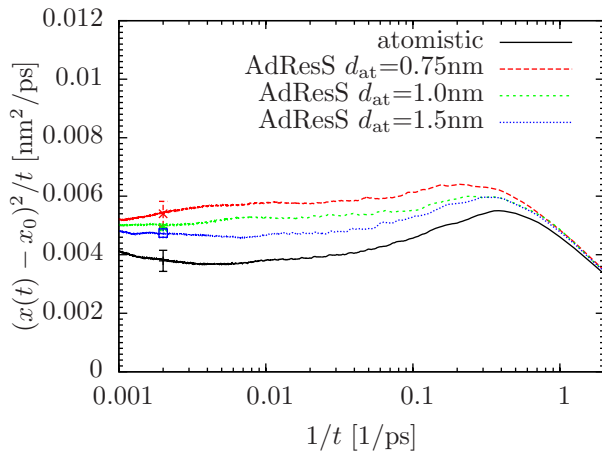
Figure 4.7: Diffusion behavior of the fullerene simulated in all-atom and in adaptive resolutions using different all-atom region sizes d_{at}.

in a single CPU cycle. Since water is the most abundant molecule used in these simulations, special routines exist which are optimized for the water-water interactions.

The AdResS simulation requires that for each molecule pair within the interaction range, the product of the weighting function is calculated in order to obtain the forces. For the adaptive scheme, a new set of kernels was needed to include the calculation of the force weights. According to the force interpolation eq. 2.84, each pair is weighted depending on the position of both molecules $w(X_\alpha)w(X_\beta)$. Hence, these hybrid kernels require more floating point operations for each pair and are slower than the "standard" kernels.

4.1.2   Dynamic load balancing

The systems studied using MD are often not homogeneous, which leads to different computational load in different regions of the simulation. This is of great importance in parallel computing codes (like GROMACS), where multiple CPUs are used to perform the simulation. The most common form of splitting the workload between processors is the domain-decomposition, where each processor calculates the interactions for a given spatial region. Since after each step, the new updated positions need to be communicated to each processor, the simulation speed is limited to the slowest processor (with the highest interaction density).

This is an absolutely crucial aspect for an AdResS simulation since the imbalance there is much greater compared to conventional simulations. The reason is that the coarse-grained interactions can be calculated with less effort. GROMACS includes a dynamical load balancing algorithm, which adapts the size of the region
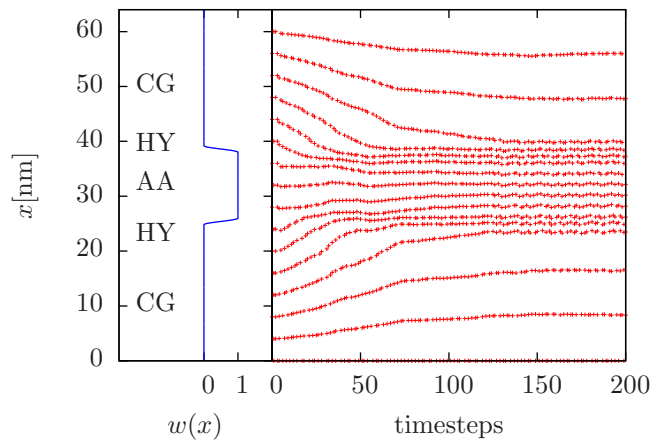
Figure 4.8: Dynamic load balancing in the AdResS simulation of water with the linear resolution change along the x-axis of the simulation box. For this, an adaptive simulation with a box size $64 \times 4 \times 4$nm and using 16 processors was run. The right hand plot shows the evolution of the domain boundaries between subvolumes assigned to different processors. Lines being close together indicate a high computational load. As time evolves, the processors assigned to the CG domain take a larger volume. Most computational effort is spent in the hybrid region, where the number of interactions per volume is larger and the force weights have to be calculated.

each processor is assigned to dynamical during the simulation. The resulting domain decomposition as a function of time is shown for the water system discussed in chapter 3. As can be seen from fig. 4.8, the algorithm adapts in a few steps to a situation where most processors are assigned to the hybrid region while, a few can calculate the interactions for the coarse-grained region.

4.1.3 Computational performance

What is the actual gain in computational speed of the adaptive scheme? In order to answer this, simulations with different sizes of the CG regions were prepared. The test system consisted of pure toluene molecules coupled to the CG model introduced in section 4.0.1. The molecular density and the size of the all-atom and hybrid region were kept fixed. It is expected that the AdResS gains most speed compared to a conventional MD simulation when the CG part is large compared to all-atom and hybrid regions. This behavior was indeed found in the scaling plot fig. 4.9. For small systems, AdResS simulations are slower then the conventional simulation. The reason is that the gain from simulating only CG interactions is outweighed by the additional cost of the hybrid interactions. For this system, a maximal speed up of a factor 3 was observed.

It is illustrative to calculate a maximal possible speed up using the current scheme. The total computer time needed consist of the time spent on the pairwise
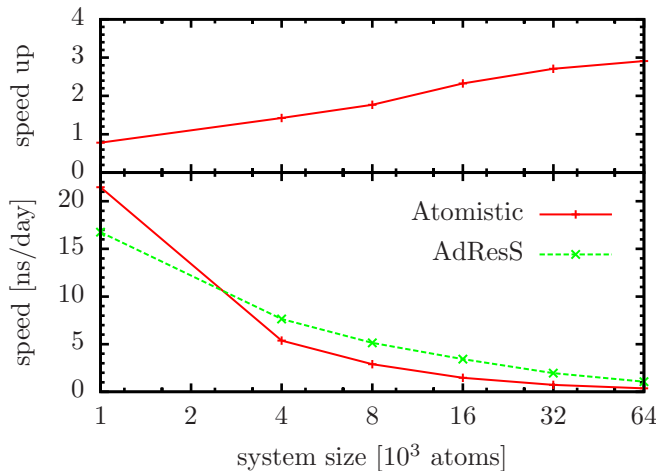
Figure 4.9: Computational scaling as function of total system size. For the AdResS simulation, the volume of the all-atom and hybrid regions are kept fixed ($d_{at} = 1\,\text{nm}$, $d_{hy} = 1.4\,\text{nm}$).

force t_{force} and the time for the rest (neighbor search, bonded interactions, integration) t_{rest}. In the corresponding all-atom conventional simulation of the largest system $(64,000)$ molecules, 73% of the computer time were spent on calculating the non-bonded forces. In this system, the hybrid and all-atom volume makes up only 0.5% of the simulation volume and hence its contribution can be neglected. If t_t denotes the total time in the all-atom and t'_t the time needed for the AdResS simulation, the speed up would be $k = \frac{t}{t'_t} = \frac{t_{force}+t_{rest}}{t'_{force}+t'_{rest}}$ The theoretically best possible speedup would be reached if the CG model would be an ideal gas (no non-bonded interactions) and thus $t'_{force} = 0$. Assuming that $t'_{rest} \approx t_{rest}$ this would yield a speed up of $k = \frac{t_{force}}{t_{rest}} + 1 \approx 4$. In this sense, the measured speed up of ≈ 3 for big systems is satisfactory.

The second biggest part ($\approx 9\%$) of the computational effort is spent for the neighbor search. In GROMACS the neighbor search is based on charge-groups (groups of atoms which are in interaction range). This concept was used here to make the neighbor search based on (CG) molecules (one molecule is represented by one charge-group). Hence, the cost of this part cannot be easily reduced further.

The third largest contribution is due to the cost of calculating the constraints for the internal interactions and integration (approximately 8% of the simulation time where spent on this for the largest system). Note that this is a consequence of the decision to integrate all degrees of freedom, even the internal degrees in the coarse-grained region. Thus further improvements are possible by considering an algorithm which freezes the internal motion entirely (with the drawback at having to re-introduce the velocities of the internal degrees of freedom when molecules enter the hybrid region).

4.2 Summary

In this part of the thesis the adaptive resolution scheme was applied to model a C60 fullerene in toluene solution. The embedding of the the spherical all-atom region into a reservoir of CG molecules offers several possibilities not available in a conventional simulation. One possibility is to study how the lack of chemical details in the CG region affects the local properties. In this case it was found that the ordering is only weakly perturbed. This is important for applications of the adaptive resolution in which the CG model is used to reduce the computational time needed to perform the simulation. Since the all-atom structure and diffusion constant could already well be reproduced by the smallest all-atom region, future studies could rely on this to save computational resources by replacing the bulk solvent with the efficient CG model. The technical challenges and actual measured speedups for the simulation are discussed in the following section.

It has to be said however, that there are also several pitfalls the user has to keep in mind. One pitfall is that, the extend to which coarse-graining will affect local properties is likely to be system dependent. Thus prior to using the adaptive scheme as a tool to speed up the simulation, test have to be performed to ensure that the physical properties important to the study are preserved. The concept of the thermodynamic force can help with this as it ensures the removal of pressure gradients coming from the mismatch in the all-atom and CG models. However, the thermodynamic force is also not cheap to compute and thus the costs of calculation the correction has to be compared to the gain from the coarse-graining. Thus it would be desirable to find a cheaper way of estimating the thermodynamic force or alter the adaptive scheme such that the pressure gradients do not appear.

Chapter 5

Path integral description of water

This chapter of the thesis connects several techniques discussed so far. It is demonstrated that a multi-scale model, which bridges several levels of description can be constructed using a combination of force-matching (FM), iterative Boltzmann inversion (IBI) and the adaptive resolution scheme (AdResS). A schematic illustration of the work flow is given in fig. 5.1.

The focus is on the inclusion of nuclear quantum effects in classical MD simulations of water. The path integral description of the nuclear quantum effect in water has been an active field of research since the pioneering simulations of Rossky et al. in the mid 80s [128]. Nuclear quantum effects are speculated to play a non-trivial role for the hydrogen bonding network even at room temperature[80]. Connected to this, is the observation of isotope effects that occur when water is replaced by heavy water, which can affect stability and thermodynamic properties of solvated biomolecules[80].

The first part deals with finding effective potentials for water based on a molecular dynamics simulation which does not rely on experimental quantities. In this "ab initio" simulation, electronic degrees of freedom are included explicitly by making use of the density functional theory (DFT)[9]. The nucleii are treated as classical point particles. The goal is to derive effective potentials based on the ab initio reference simulation which are suitable for path integral MD (PIMD) (see section 2.5), where nuclear qunatum effects are included. This way, double counting of nuclear quantum effects, as it is present when using empirical water models [92, 129], can be avoided. Thereby, use is made of force-matching to capture the essential physics of this description in the effective potentials, keeping the water molecule polarizable by allowing intra-molecular flexibility.

Using the FM potential in the PIMD simulation assumes that the nuclear quantum effect does not strongly alter the effective electron-electron interaction and thus the effective forces from the classical configurations are still valid as interactions in the PI picture. The advantage lies in the fact that the effective potentials can be simulated at much lower computational cost compared to the ab initio simulation
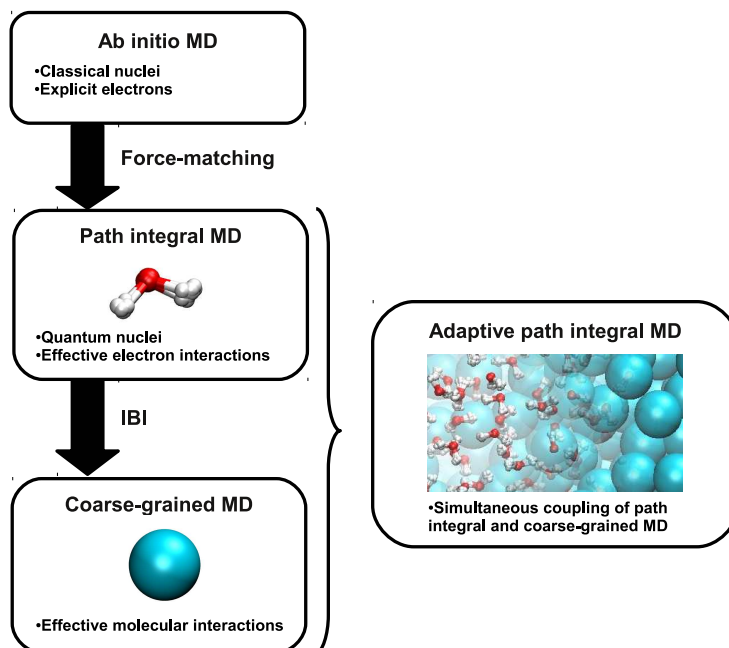
Figure 5.1: Schematical work flow of the path integral adaptive approach.

based on DFT. The simultaneous treatment of ab initio level electronic interactions and nuclear quantum effects is currently only possible for very small systems [130].

In the second part, it is demonstrated that based on this one can go to even larger scales using the adaptive resolution scheme. For this, an effective potential is derived which describes the path integral water model at the level of center of mass interaction, as it has been done in chapter 3. This is applied to simulating a water-vacuum slab where only the interface region is treated with full path integral MD interactions. It is demonstrated that a stable water-vacuum interface can be achieved and the nuclear qunatum effects are preserved close to the surface.

The two parts together thus connect models at different resolutions which is especially favorable for situations when the detailed model is needed only in a sub-region and the rest of the system can be described in a coarser way [85, 86, 87].

5.1 Test case: SPC/Fw

5.1.1 Setup

As a preliminary test, the ability of coarse-graining techniques to recover a known interaction potential based only on knowledge of the trajectory and total force per atom was investigated. The SPC/Fw [131] model was chosen because it de-
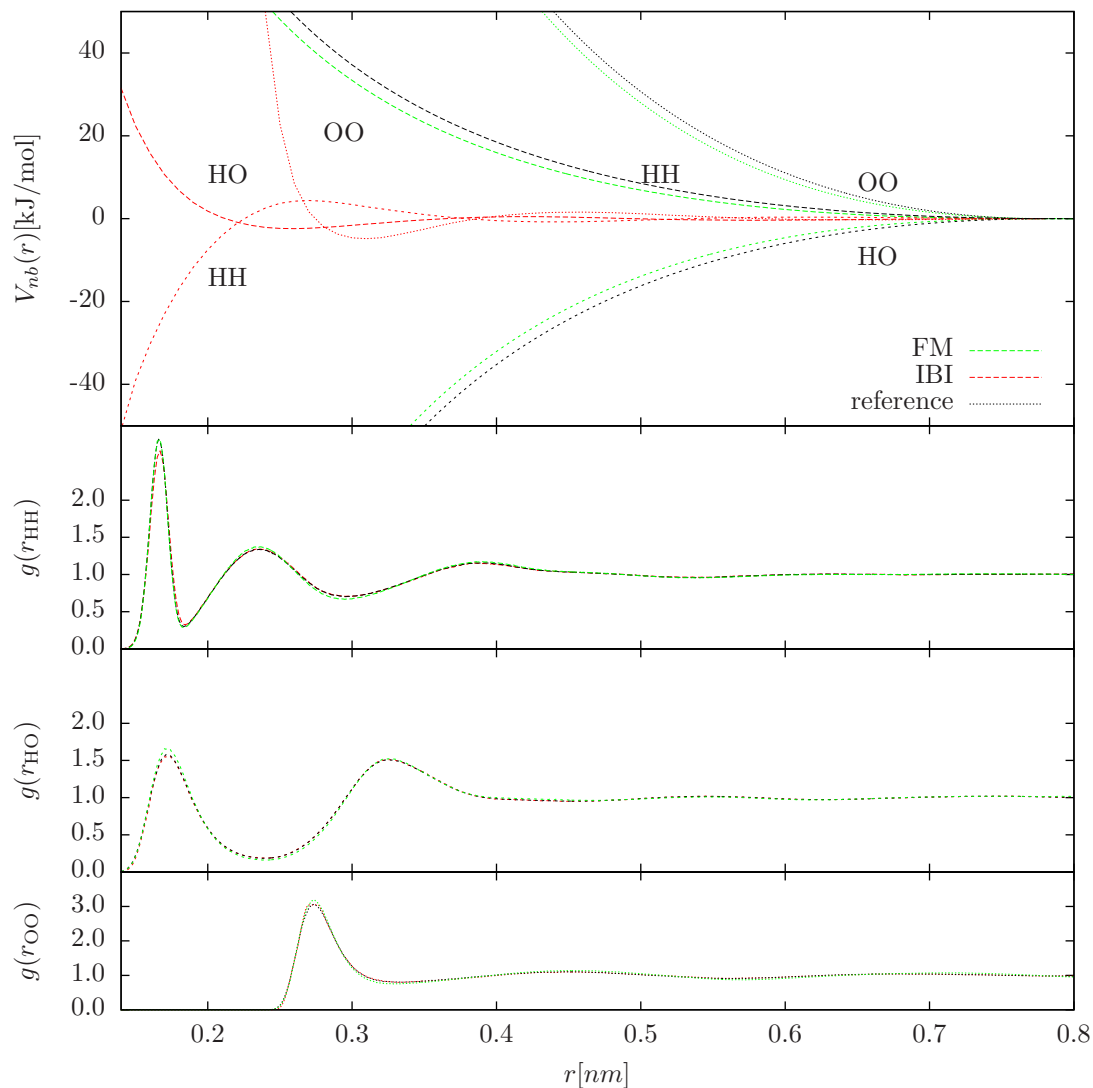
Figure 5.2: Results of testing the ability to recover the interaction potentials using FM and IBI in a test system of SPC/Fw water. The top figure shows the non-bonded interaction potential compared to the reference reaction-field Coulomb interaction ($r_{\text{cutoff}} = 0.78\,\text{nm}$) in SPC/Fw. In the bottom plot, the resulting structures from MD simulations using these potentials are shown.

| Hydrogen	Oxygen
$q_H = 0.41e$	$q_O = -0.82e$
$\sigma = 0$	$\sigma = 0.316557\,\mathrm{nm}$
$\epsilon = 0$	$\epsilon = 0.650194\,\mathrm{kJ/\,mol}$
Bond	Angle
$r^0_{OH} = 0.1012\,\mathrm{nm}$	$\theta^0_{HOH} = 113.24°$
$k_{OH} = 445153\,\mathrm{kJ/(\,mol\,nm^2)}$	$k^0_{HOH} = 317.565\,\mathrm{kJ/(\,mol\,rad^2)}$

Table 5.1: Parameters of the SPC/Fw water model [131]

scribes the bond and angular interaction using a flexible, harmonic potential (the parameters are given in tab. 5.1). Thus also the ability to recover intra-molecular potentials can be tested. First, a reference classical MD simulation of bulk water using the SPC/Fw model was performed and the configurations saved together with the total force on each atom. Then FM and IBI were applied on this reference trajectory in order to find non-bonded interactions potentials for oxygen and hydrogen, as well as intra-molecular (bond- and angular) interactions. This "one-to-one" mapping represents a test case to which the answer is known beforehand.

As a reference simulation, 324 water molecules at a density of $1.002\,\mathrm{g/\,cm^3}$ (corresponding to the equilibirum density of the SPC/Fw model at $T = 300K$ and $p = 1\,\mathrm{bar}$) were simulated with a timestep of $\Delta t = 0.5\,\mathrm{fs}$ for a total length of $10\,\mathrm{ns}$. The total force on each atom was saved every 100 steps. In order to be better comparable to the results in the previous chapters, long-range electrostatics were not considered. Instead, the reaction-field approach with an dielectric constant $\epsilon_{RF} = 80$ was used (see also section 2.1.4), where the interactions where truncated at $r_{cut} = 0.78\,\mathrm{nm}$.

In the case of IBI, the bond- and angular interaction potentials were calculated from simple Boltzmann inversion as described in section 2.3.3. For the non-bonded interactions, iterative Boltzmann inversion with the partial radial distribution functions $g(r_{OO})$, $g(r_{OH})$, $g(r_{HH})$ as target distributions. The CG potentials were calculated up to $r_{max} = 0.78\,\mathrm{nm}$ which corresponds to the cut-off of the interaction in the reference simulation. 300 iteration steps were carried out, each step consisting of 10^5 MD steps. The potentials were updated from the partial RDFs in each step based on eq. 2.44, where the partial RDFs were assumed to be independent of each other.

For force matching (FM), a smaller part of the reference trajectory consisting of 320×10^3 MD steps had to be considered due to the larger computational cost. The FM algorithm was used as implemented in the VOTCA software package [57]. Therein, cubic splines represent the set of unknown forces. For the χ^2 minimization, the reference trajectory is sub-divided into sets of 100 configurations and the minimization carried out for each set separately [62]. Subsequently, the final forces

	$r_{\min}[\,\mathrm{nm}]$	$r_{\max}[\,\mathrm{nm}]$	$\Delta r[\,\mathrm{nm}]$	$r_{min}^{fit}[\,\mathrm{nm}]$	$r_{max}^{fit}[\,\mathrm{nm}]$
$V_{OO}^{non-bonded}$	0.24	0.78	0.01	0.28	-
$V_{HO}^{non-bonded}$	0.14	0.78	0.01	0.17	-
$V_{HH}^{non-bonded}$	0.16	0.78	0.01	0.20	-
V_{HO}^{intra}	0.09	0.11	0.001	0.093	0.107
V_{HH}^{intra}	1.57 rad	2.1 rad	0.01 rad	full range	

Table 5.2: Ranges of tabulated potentials used for FM

are calculated from the average over all sets.

It has to be stressed, that for each interaction the range has to be adapted as the FM equations are ill-defined for regions where $g(r)$ is zero. The ranges used are shown in tab. 5.2. The forces from the converged χ^2 minimization were extrapolated into the regions not sampled in the references simulation. This was done by fitting a power law of the form $f(r) \propto \frac{q}{r^k}$ for the H-H and H-O interactions and Van-der-Waals+reaction field for the O-O interaction $f(r) \propto \frac{q}{\epsilon_{RF}}(\frac{1}{r^2} - 2k_{RF}) + (\frac{12A}{r^{13}} + \frac{6B}{r^7})$. For the fit only points in the range from r_{min} to r_{min}^{fit} were considered because no functional form for the entire range could be found. For the bonded interactions, second-order polynomials were fit in order to extrapolate into the unsampled regions $r < r_{min}$ and $r > r_{max}$. The exact fit ranges are specified in tab. 5.2.

5.1.2 Comparison

The resulting potentials and structures calculated from MD simulations[1] using these potentials are shown in fig. 5.2. It is visible that FM recovers the coulombic form of the potentials used in the reference simulation, while IBI fails to do so. On the other hand, it may be quite surprising to see that both IBI and FM very accurately recover the reference structure in terms of $g(r)$. The reason for this difference are the 3-body terms which are present in FM (as discussed in section 2.3.5.1) whereas IBI does only take into account 2-body correlations through the $g(r)$. This effect is best explained at the example of the O-O interaction potential, which shows a stable minimum at $r \approx 0.3\,\mathrm{nm}$ when calculated with IBI. The two-body O-O PMF at this position is attractive trough the "mediating" function of the hydrogen bond, which makes the oxygen atoms attract each other on average. In FM, the 3-body terms associated with the O-H-O configuration are subtracted from the resulting forces (see eq. 2.60). Thus FM recovers the repulsive nature of the reference O-O interaction. Similar arguments can be made for all other interactions.

---

[1]The parameters used for these simulations were identical to the reference MD simulation.

Overall, the potential energies in IBI are drastically lower compared to the FM energies. This is due to the effective 2-body interaction being much weaker in the case of IBI compared to the reference. The IBI thus gives the correct structure, but the pairwise forces deviate significantly from the reference. The FM interactions however are only slightly softened compared to the reference. This is most likely due to the effect of higher order correlations (e.g. 4-body and higher order correlations), which are not included. Note that one common source of error in FM, the lack of representability of the "coarse-grained" potentials, is excluded here in this simple test case, as no degrees of freedoms are lost. The pressure for the FM potentials was $p_{\mathrm{FM}} \approx 670\,\mathrm{bar}$, while IBI showed a significantly increased pressure of $p_{\mathrm{IBI}} \approx 4100\,\mathrm{bar}$.

The situation is thus somewhat different compared to the case of coarse-graining in terms of an effective center-of-mass interaction between water molecules (as discussed in chapter 3). For this case, the comparison between FM and IBI in [57] found that the potentials are not dramatically different, but FM fails to reproduce the structure.

As a consequence, the FM algorithm was chosen here for the effective atomistic interaction potential from the ab initio simulation. The IBI algorithm was then used in section 5.5 for the molecules center-of-mass interaction for which it is known that FM does not reproduce the pairwise strucutre [57].

5.2 Ab initio reference simulation

In the previous section, it was shown that the FM technique can be used to recover pairwise interaction potentials. Now, FM will be used to fit effective interaction potentials from an "ab initio" simulation as was initially suggested in [63]. In contrast to the simple test case described in the previous section, this is more similar to coarse-graining in the sense that the electronic degrees of freedom are removed. Their interaction is incorporated in the effective potentials, which can then be used to run classical MD simulation.

As a reference ab initio simulation, DFT-based ab initio molecular dynamics (AIMD) simulations was performed using the CP2K-QuickStep package[2]. This approach employs a mixed localized and plane waves basis set approach. In real space the electronic Kohn-Sham wavefunctions are expanded on a TZV2P localized basis set [132], while an auxiliary planewave basis set up to 280 Ry is used for the density in dual space. The PBE generalized gradient approximation is chosen for the DFT exchange and correlation (XC) functional [133]. Even if this XC functional gives an excellent description of the energetics of the water dimer [134], PBE water at 300 K is reported to be overstructured and less fluid than real water, with a relatively high shear viscosity [135], however it is still debated how much of

---

[2]http://cp2k.berlios.de

such overstructuring comes from neglecting quantum effects in the dynamics of the ions. In the abinitio simulation, the dynamics was evolved by integrating Newtons equations of motion with a timestep of $\delta t = 0.3$ fs. A system containing 128 heavy water molecules at the density of 1.1 g/cm^3 was simulated in the microcanonical NVE ensemble for 10 ps, after equilibration at $T = \sim 320$ K at constant V and T. The average temperature of the NVE run was found to be $T = 325 \pm 0.8$ K.

Heavy water was chosen as to reduce the computational resources needed: Because the bending and stretching modes occur at a smaller frequency compared to H-O, the timestep of the integration can be significantly increased and a larger fraction of the phase space can be sampled. The use of heavy water instead of light water is justified as the nuclear quantum effect is not included in the ab initio simulation. In this sense, heavy water represents a "more classical" case then light water due to the larger mass of the deuterium. The nuclear quantum effect is added in the path integral MD simulations (see section 5.3) using the FM effective potentials, wherby it is assumed that the effective interaction potential representing the elelctronic interactions of heavy water are also representative for light water.

The ab initio trajectory was then used to find the effective interaction via the FM approach. Non-bonded interactions were considered for D-D, D-O and O-O interactions, for the internal degrees of freedom the O-D bond and D-O-D angular potentials were used (see section 2.1.3). The force for these interactions were represented by cubic splines and the same ranges considered as specified in tab. 5.2. In the ab initio reference trajectory, the first 2.5 ps were considered equilibration and discarded, of the remaining 14.5 ps configurations at an interval of 5 fs where used for the force-matching. All other parameters were identical to the ones in the previous section. These forces were then integrated to yield the corresponding potentials, which are plotted in fig. 5.3.

5.2.1 Properties of the FM model

The interaction potentials obtained by FM were calculated with the aim of effectively incorporating the properties of the ab initio simulation. In order to put this to the test, simulations using the newly generated potentials were performed. Classical MD simulations (the nuclear quantum effect will be discussed in section 5.3) were run employing the effective bonded and non-bonded potentials. For this a, box containing 324 heavy water molecules at the same density of the ab-inito reference simulation was prepared. In order to be directly comparable to the ab initio simulation, a temperature of $T = 325$ K was enforced using a Langevin thermostat ($\tau_{\text{Langevin}} = 0.5$ ps). Simulations were then run for a total length of 1 ns using a timestep of $\Delta t = 0.5$ fs.
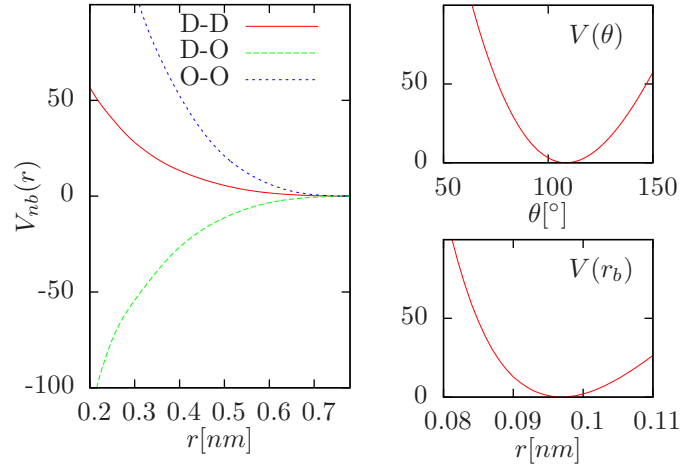
Figure 5.3: (left) Potential resulting from integration of the non-bonded pair forces from force-matching of the ab initio reference. (right) Internal interaction potential for angle- and bond interactions.

300K NVT	SPC/Fw	ab initio FM
$< V_{\mathrm{bond}} >$	1400 kJ/mol	1897 kJ/mol
$< V_\theta >$	918 kJ/mol	582 kJ/mol
$< V_{\mathrm{non\text{-}bonded}} >$	3259 kJ/mol	-17516 kJ/mol
$< V_{\mathrm{coulomb}} >$	-18939.5 kJ/mol	-
$< V >$	-13587 kJ/mol	-15037 kJ/mol
Pressure	$100 \pm 20\,\mathrm{bar}$	$-2450 \pm 21\,\mathrm{bar}$
Density ($p = 1\,\mathrm{bar}$)	1.002 g/cm^3	1.075 g/cm^3
$< q_4 >$	0.6336	0.7560

Table 5.3: Comparison of energy terms in bulk SPC/Fw and the FM system. Averages are taken over 900 ps from MD simulations of 324 water molecules.
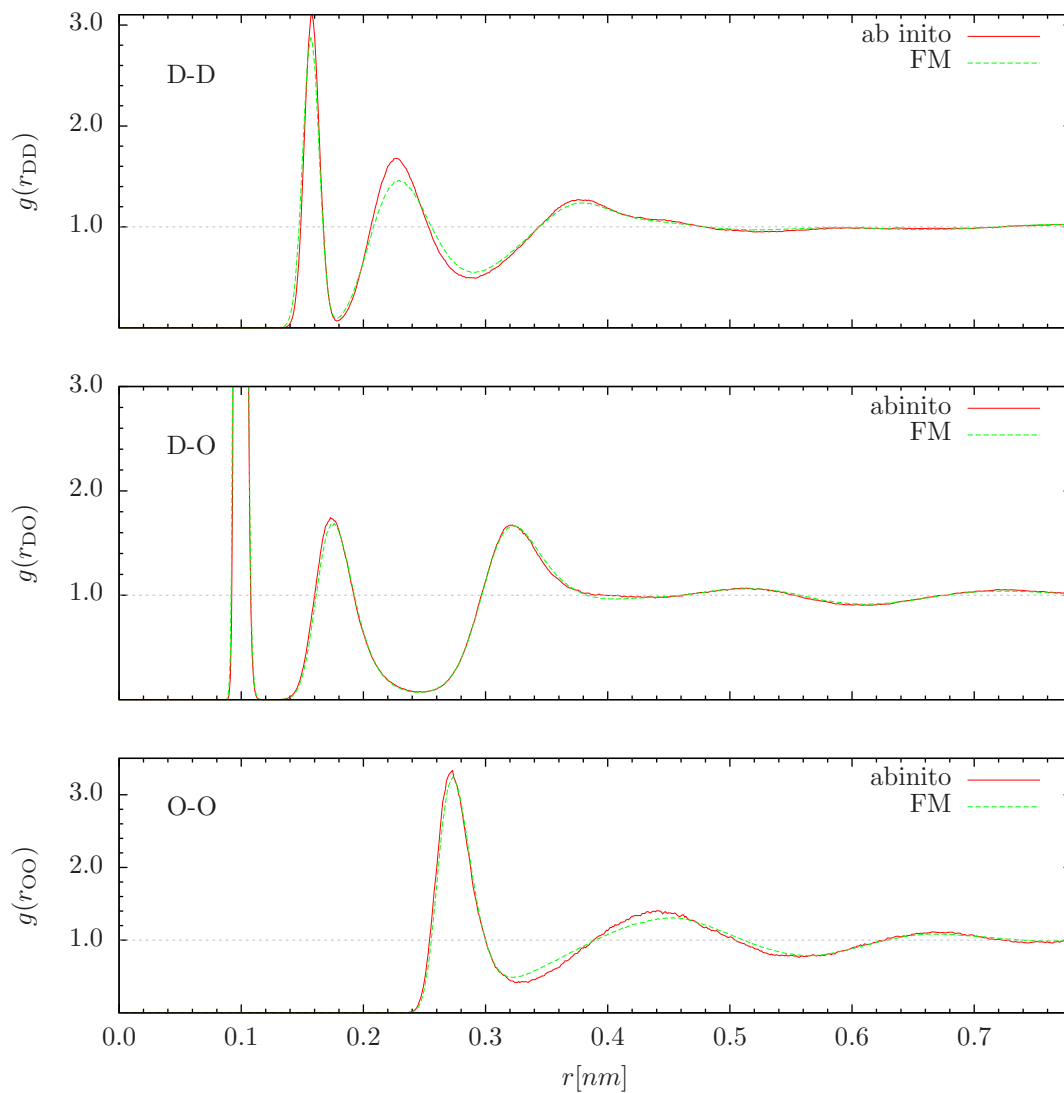
Figure 5.4: RDFs in the ab initio reference simulation and an MD simulation at $T = 325\,\mathrm{K}$ employing the FM effective potentials. The intra-molecular correlations are included (first peaks in the D-O and D-D distributions).

5.2.2 Structure

The resulting $g(r)$ is compared to the ab initio reference structure in fig. 5.4. The structure of the FM model closely recovers the structure found in the ab initio reference simulation. There is some deviation visible for the first non-bonded peak in the D-D distribution. This deviation will be discussed further by aid of the vibrational spectrum (see section 5.2.5). The intra-molecular structure, which is included in the distribution functions shown in fig. 5.4 (first peaks of D-D and D-O), is also well reproduced.

5.2.3 Thermodynamic properties

The dimer energy, i.e. the total energy for the energy-minimized structure of two-water molecules in vacuum was calculatead by optimizing the dimer geometry using the steepest descent technique [136]. For the FM model, a value of $E_{\text{dimer}} = -23.7 \frac{\text{kJ}}{\text{mol}}$ was found. In an ab initio simulation using the PBE XC by Zhang et al. [134], the dimer energy was found to be $E_{\text{dimer}} = -21.0 \frac{\text{kJ}}{\text{mol}}$. The FM potential (fitted in the bulk phase) hence gives a dimer energy deviating by only 12% from the value obtained in the ab initio simulation of a dimer in vacuum. The SPC/Fw model has a lower dimer energy of $E_{\text{dimer}} = -29.2 \frac{\text{kJ}}{\text{mol}}$.

The bulk thermodynamic properties are compared to the SPC/Fw model, where it has to be noted that the parameters for SPC/Fw were derived through the very different route of fitting selected experimental properties [131], thus discrepancies are expected. For this the classical simulations were re-rerun using the mass of the hydrogen instead of deuterium (and all other parameters as in section 5.1.1).

Table 5.3 compares several properties of the FM to the SPC/Fw model.The potential energies in the bulk were found to be lower compared to the SPC/Fw model. The sum of Coulomb and Van-der-Waals energies is comparable to the non-bonded potential energy in the FM model.

The presure in the FM model, measured at the equilibrium density of light water, was found to be $(p_{FM} \approx -2400\,\text{bar})^3$.There is two sources of error which can explain this deviation: First, the pressure in the ab initio simulation is not easiliy available and was therefore chosen based on the experimentally known equilibrium pressure of heavy water [134]). Secondly, it also known (see test case in section 5.1.1) that the FM technique does not preserve the pressure. Since only the total force on an atom is reproduced, but not the virial it can not be expected for the pressure to be identical. As a solution to this, it has been suggested to included the virial pressure in the χ^2 minimization [14]. This however requires the knowledge

---

[3]It may seem surprising that the FM model, which is a coarse-grained model, gives a negative pressure, wheras in chapter 2 it was found that the CG model of water has an pressure increased by 6kbar. However as the pressure does not enter the derivation in the CG algorithms, no clear trend for over or undererstimating can be expected.
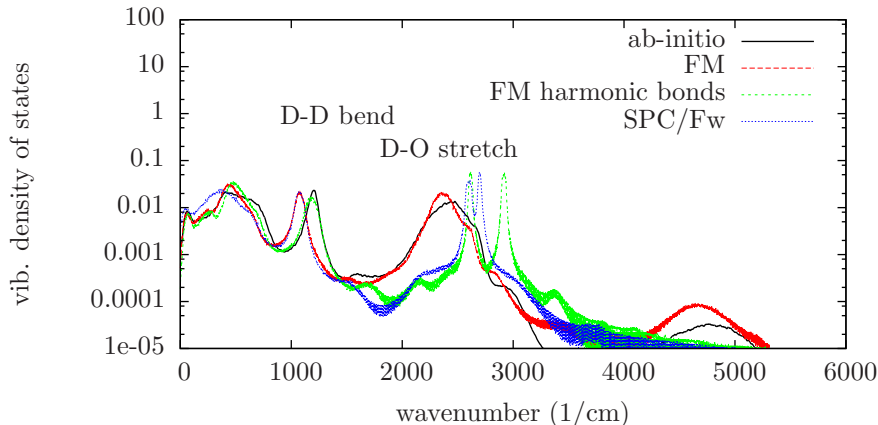
Figure 5.5: Vibrational density of states calculated from eq. 5.1. For the FM model, also results from replacing the internal interaction by harmonic potentials are shown (see the main text for an explanation).

of pairwise forces in the reference trajectory, a quantity not easily available from ab initio simulations. The q_4 order parameter (as defined in section 3.1.1) shows a significant increase, indicating a more ordered liquid. This can be explained as a consequence of more structured liquid with respect to the the experiment, which will be discussed in section 5.4.

5.2.4 Dynamical properties

Another interesting property is the vibrational spectrum. To facilitate the comparison with the ab initio reference simulation, short simulations (20 ps) of heavy water were performed (using the same setup as in the previous section 5.2.1). The vibrational spectrum was calculated by taking the Fourier transform of the velocity-autocorrelation function

$$I(\omega) = \frac{1}{N_D} \int_{-\infty}^{\infty} d\omega e^{-i\omega t} \sum_{i=1}^{N_D} (< v_i^D(t) v_i^D(0) >) , \qquad (5.1)$$

where the bracket indictes the average different initial times and the sum runs over all N_D deuterium atoms. The resulting vibrational spectrum shown in fig. 5.5 can be compared to experimental infrared absorption data[4]. The position of these bands matches well with experimental results from infra-red absorption spectroscopy [137] .

---

[4]In order to compute the infra-red spectrum, the autocorrelation of the dipoles has to be computed. This requires the assignment of charges to the FM atoms which cannot be done easily in a unique way an hence was left for future works.

For comparison, the procedure was also repeated for the SPC/Fw model. In the spectrum, the peaks corresponding to the D-D bend mode at $1100 - 1300 \, \text{cm}^{-1}$ and the D-O stretch mode at $2000 - 2700 \, \text{cm}^{-1}$ are clearly visible. The low frequency regime $< 1000 \, \text{cm}^{-1}$, is dominated by the hindered rotations (librations) and diffusion of the atoms. The FM model matches the reference ab initio spectrum very well. In particular, in the range of the D-O stretch mode, one single broad peak is visible. This feature cannot be reproduced by the flexible SPC/Fw model, which shows two separate peaks.

To further understand this effect, additional simulations using the FM potential for the non-bonded interactions, but simple harmonic potentials for the bond were carried out. Assuming that it is possible to describe bond and angle motions by classical harmonic oscillators $k = m_D \omega^2$, the force constants can be calculated from the ab initio spectrum, which yields $k_{DO} = 422.3 \times 10^3 \, \text{kJ}/(\, \text{mol} \, \text{nm})$ for the D-O $k_{DD} = 102.8 \times 10^3 \, \text{kJ}/(\, \text{mol} \, \text{nm})$ for the D-D peak. The resulting spectrum is also shown in fig. 5.5. It is found that with these harmonic internal interactions, the single broad D-O stretch band splits into two single peaks (which are separated further compared to the SPC/Fw model). A possible explanation is that the balance between bonded and non-bonded D-O forces is perturbed by the assumption of a simple harmonic potential. As a consequence intra-molecular and inter-molecular D-O vibrational modes oscillate at different frequencies. It can be concluded that the broad D-O peak cannot be reproduced by harmonic internal interactions. This is supported by the fact that the FM potentials show some anharmonicity, which can be seen from fig. 5.3.

The D-D bending mode at $1100 - 1300 \, \text{cm}^{-1}$ is slightly shifted to lower frequencies in the FM model, as well as as in the SPC/Fw model. For what concerns FM, this can be explained as follows: As can be seen in fig. 5.8, the distribution for D-D, the internal (first peak) and non-bonded configurations overlap. Since in FM only the total force enters the minimization at the position of the overlap, the bonded and non-bonded force cannot be distinguished. As a consequence, the internal force contains contributions from the non-bonded interactions and vice-versa. This can lead to an effective softening or strengthening of the bending mode. In this case, the reduced frequency of the D-D bend mode seems to suggest a weakening of internal interaction. This is supported by the $g(r_{DD})$ (fig. 5.4), where a broadening of the distribution distribution was found.

Interestingly, both in the ab initio as well as in the FM simulation, a weak band in the region of $4000 - 5200 \, \text{cm}^{-1}$ is visible, which is roughly twice the wavenumber of the DO-stretch band. It is thus likely that this peak is due to the first overtone of the DO-stretching vibration. This feature is lost by using the harmonic approximation FM or the SPC/Fw model. Overall, in contrast to the finding in a similar study [63], it has been found here that the spectrum can be reasonably well reproduced by the FM using tabulated internal interactions.
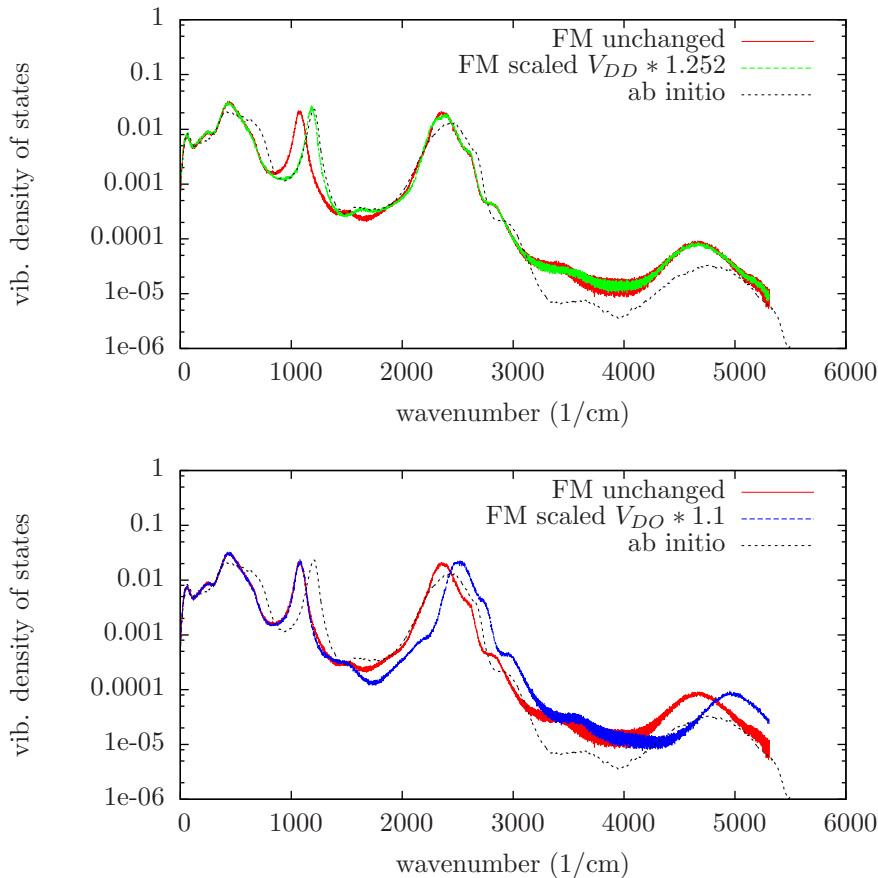
Figure 5.6: Vibrational density of states measured in short (20 ps) simulations with the bonded potential slightly modified as described in the main text. (top) scaling applied to the D-D angular interaction with $k' = 1.251$; (bottom) scaling applied to the D-O bond interaction with $k' = 1.1$.

5.2.5 Modifying the spectrum

The spectrum, although reproducing remarkably well the reference, shows some deviations in the bend frequency which peaks approximately at $\omega_{\mathrm{DD}}^{\mathrm{ab\,initio}} = 1203\,\mathrm{cm}^{-1}$ in the ab initio simulation. This behavior can be further analyzed using a simple rescaling of the intra-molecular interactions. To that end, the FM D-O-D angular potential was altered $V_\theta^*(\theta_{DOD}) = kV_\theta(\theta_{DOD})$ with $k' = \omega_{\mathrm{ab\text{-}initio}}^2/\omega_{\mathrm{FM}}^2$. This scaling assumes that the bond can be treated as approximately harmonic and thus $\omega = \sqrt{k/m}$. Since in FM model the D-D bend frequency was $\omega_{\mathrm{DD}}^{\mathrm{FM}} = 1075\,\mathrm{cm}^{-1}$, the scaling factor is $k' = 1.251$. The spectrum of a simulation with the modified potential is shown in fig. 5.6. It can be seen that the harmonic correction is sufficient to shift the D-D bend position such that it matches the reference. The rest of the spectrum is unaffected indicating that there is no strong coupling between

the internal modes.

The D-O stretch band is well reproduced in the FM potential, it is however still useful to see the effect of scaling the potential with a small, arbitrary factor $k = 1.1$. As fig. 5.6 shows, the D-O peak is independent of the other modes except for the peak at $4000 - 5200 \, \text{cm}^{-1}$, which also shifts to higher frequencies. This proves that this peak is the D-O overtone.

These modifications were not kept for the following simulations, they however indicate a possible direction on how to improve the force-matching results: As discussed above, there is some overlap in the D-D bonded and non-bonded distributions and thus the forces belonging to bonded and non-bonded modes cannot be assigned exactly to the appropriate interactions. The scaling factor obtained for the D-D stretch in the harmonic approximation implies that the bonded forces where underestimated by a factor k during the FM. This information could be included into the force-matching matrix by weighting the forces according to k.

However since the spectrum was already reasonably close to the reference, this approach was not followed further.

5.3   Path integral MD

In this section, path integral MD (PIMD) simulations using the newly derived FM water model are discussed. The electronic interactions are treated as an effectively classical potential, derived by FM. This assumes that the electronic and quantum nuclear effect can be "separated" in the sense that the effective potentials are still valid for the system with the quantum nuclear treatment. The reason for this approximation is that ab initio simulations based on DFT come at a much larger computational cost and a concurrent treatment of electronic and nuclear quantum nature would increase the costs by at least one order of magnitude.

In case this methodology is successful, it suggests a computational efficient way to improve the development of new ab initio based models in which nuclear quantum effects are explicitly included. In the next section, first a brief overview on the current state of resrach on modelling the nuclear quantum effect in liquid water is given.

5.3.1   Overview of results from literature

In MD simulation, the path integral description has been used in connection with several different water models. These models can be divided into two classes: empirical and ab initio based models.

The empirical models, such as SPC/Fw [131] or TIP4P/F [138] are based on fitting the parameters (e.g partial charges, bond length, etc.) as to match experimental quantities. Using these parameters in PIMD simulations has the difficulty

that the quantum nuclear effect is already effectively included, because the experimentally measured quantities do include the nuclear quantum effects. PIMD simulations based on these models thus "double-count" the nuclear quantum effects. This has been suggested to be overcome by a two step procdure [92, 129]: First simulations using the empirical model in PIMD are carried out. From this simulation, selected thermodynamic quantities are calculated, which also have to be known experimentally. The set of paramters in the model is then modified in a series of PIMD simulations until the experimental quantities are matched.

The ab initio models are based on accurate quantum calculations of small clusters. One example is the MCDHO model [139] or the more recent TTM3-F model [140, 141, 142]. The description of the bulk phase using these models might however be plagued by finite size effects of the clusters which are not representative of the bulk state [7]. the over-structuring in ab initio simulations is partially due to treatment of the nucleii as classical particles. For what concerns the structure, a significantly less structured liquid with respect to the classical simulations was found in most studies. A study by Stern et al [78] using the MCDHO model concluded that quantum effects lead to a less structured hydrogen bonding network: The agreement with intra-molecular structure measure from experiment improved, while the inter-molecular structure remained largely unchanged.

Another class of simulations includes the nuclear quantum effects directly in an ab inito simulation using Car-Parrinello MD (CPMD) (where the electrons are treated explicitly with plan wave basis sets at the DFT level)[143]. An early simulation using this technique found a hardening of the PIMD O-O distribution with respect to the classical equivalent[98]. In another study by Morrone et al. [130] a significant softening of inter- and intra molecular structure compared to CPMD using classical nucleii was found, which leads to closer agreement with experiment. This suggest a significant improvement of the ab initio description of water when including the nuclear quantum effect. The simultaneous treatment of nuclear and electronic qunatum effect however comes at very high computational cost, which limits the simulations to tiny system sizes (typically 64 water molecules) simulated for a few ps.

These difficulties show the need for an computationally effective model which can provide a realistic description of the bulk phase without having to rely on experimental data. This approach is attempted here by using the FM model which is DFT (Born-Oppenheimer MD [45]) based and does not rely on clusters. Since the ab initio simulation does not take account of any nuclear quantum effect, the double counting is avoided.

An exact quantitative description of the nuclear quantum effect in terms of measurable quantities is still not available in the literature. For the self-diffusion coefficient for example an increase in 50% has been reported [92] (using the re-parametrized SPC/Fw), while a much smaller increase of 15% has been reported for the re-parametrized TIP4P/F [129]. This was claimed to be a result of compe-
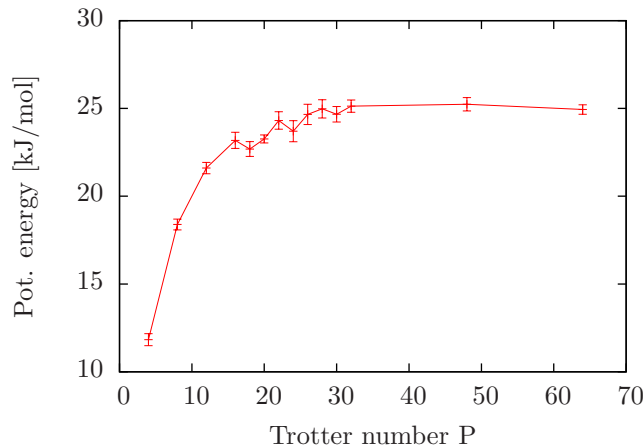
Figure 5.7: Total potential energy in a PIMD simulation of a single water molecule in vacuum. The trotter number P is varied to find a value for which the energy is reasonably converged.

tition effects: The intra-molecular zero point fluctuations increase the H-O bond length (and thus the dipole moment) which increases the hydrogen bond strength, while the inter-molecular quantum fluctuations lead to a less structured hydrogen bonding network.

5.3.2 Numerical results

To probe the nuclear quantum effect in the FM water model, PIMD simulations were performed, employing the path integral Hamiltonian of eq. 2.112. The only free parameter that has to be chosen is the the Trotter number P, which corresponds to the number of discrete imaginary time slices. The trotter approximation is exact in the limit of $P \to \infty$ (also see section 2.5). A higher P is however associated with higher computational costs, which increase approximately linearly with P.

In order to choose an appropriate Trotter number P, PIMD simulations of a single molecule were first carried out. The total potential energy as a function of P was computed and is shown in fig. 5.7. A trotter number of $P = 16$ was found to converge the potential energy to approx. 92%, thus representing a good compromise between accuracy and computational cost.

A simulation consisting of 324 water molecules (with P=16 this corresponds to a PIMD simulation using 15552 point particles) at the same density of the ab initio reference system was prepared and run for a simulated time of 750 ps. Note that now the FM potential is used for the effective interaction of hydrogen, although it was parametrized from simulations of heavy water (also see section 5.2). The initial configuration was prepared by starting from an equilibrated classical configuration and replacing each atom by its ring-polymer counterpart. For this, the positions of

the ring-polymer beads where placed in a circle (with a radius $r^0 = 0.01\,\text{nm}$) in the x-y plane centered in the classical position of the atom. The initial velocities where taken from a Maxwell-Boltzmann distribution at $T = 300\,\text{K}$. Short simulations where then performed using strong Langevin thermostatting ($\tau_{\text{Langevin}} = 0.2ps$) in order to relax the ring polymers to their equilibrium values.

For the production runs, the timestep was set to be $\Delta t = 0.5\,\text{fs}$, the temperature $T = 300\,\text{K}$ was controlled with a Langevin thermostat (coupled to all imaginary time beads) using a relaxation time of $\tau_{\text{Langevin}} = 2ps$.

The resulting structures measured from classical and PIMD simulations using the FM potential are shown in fig. 5.8. Compared to the classical case, in the PIMD simulation, a softening of the intra-molecular distribution is visible, which is most pronounced in the H-O intra-molecular bond distribution. The H-O and H-H bonded peaks are clearly broadened. The overall inter-molecular structure, as visible from the O-O distribution however remained mostly unaltered, although the H-O and H-H first non-bonded peaks (second peaks in the fig. 5.8) are also slightly softened.

Experimental results from neutron diffraction [144] are also shown in fig. 5.8 (adapted after [129]). For the intra-molecular H-O and H-H distribution, very good agreement with the experimental results is now reached in the PIMD description. In the O-O distribution, an over-structuring with respect to the experimental $g(r)$ is visible for both the classical as well as the PIMD. Note that this over-structuring was already present in the O-O distribution measured directly in the ab initio simulation (fig. 5.4). This is a well-known artifact of the treatment of water using local DFT functionals [145]. It has been attributed to deficiencies of the exchange correlation functional (which is approximated using local functionals) and to the lack of a proper description of the dispersion forces. As a consequence of this, the DFT water was found to be in a supercooled state at room temperature [146, 147].

As discussed above, Morrone et. al [130] attribute the overstructuring to the lack of nuclear quantum effects. This would mean that the inclusion of nuclear quantum effects leads to softening of the O-O distribution for DFT based models, which improves the agreement with experiment. Here, however, no such improvement for the inter-molecular degrees of freedom was found, which indicates that the effect has been overestimated by Morrone et al.[130].

A comparison of pressures in classical and PIMD is listed in tab. 5.4. The pressure in the PIMD simulation was measured using the primitive pressure estimator (chapter 1, eq. 2.121). The inclusion of nuclear quantum effects in PIMD has a significant effect on the pressure, visible in an increase of approximately $500\,\text{bar}$. As discussed above, the FM potential in the classical MD does not reproduce the reference pressure. A correction of the pressure was attempted by adding, as in the pressure correction in IBI (see section 2.3.4), a small linear term to the potential energy of the O-O non-bonded interaction (while keeping both the IBI potential update and the pressure update terms, as in eq. 3.2). The correction, obtained
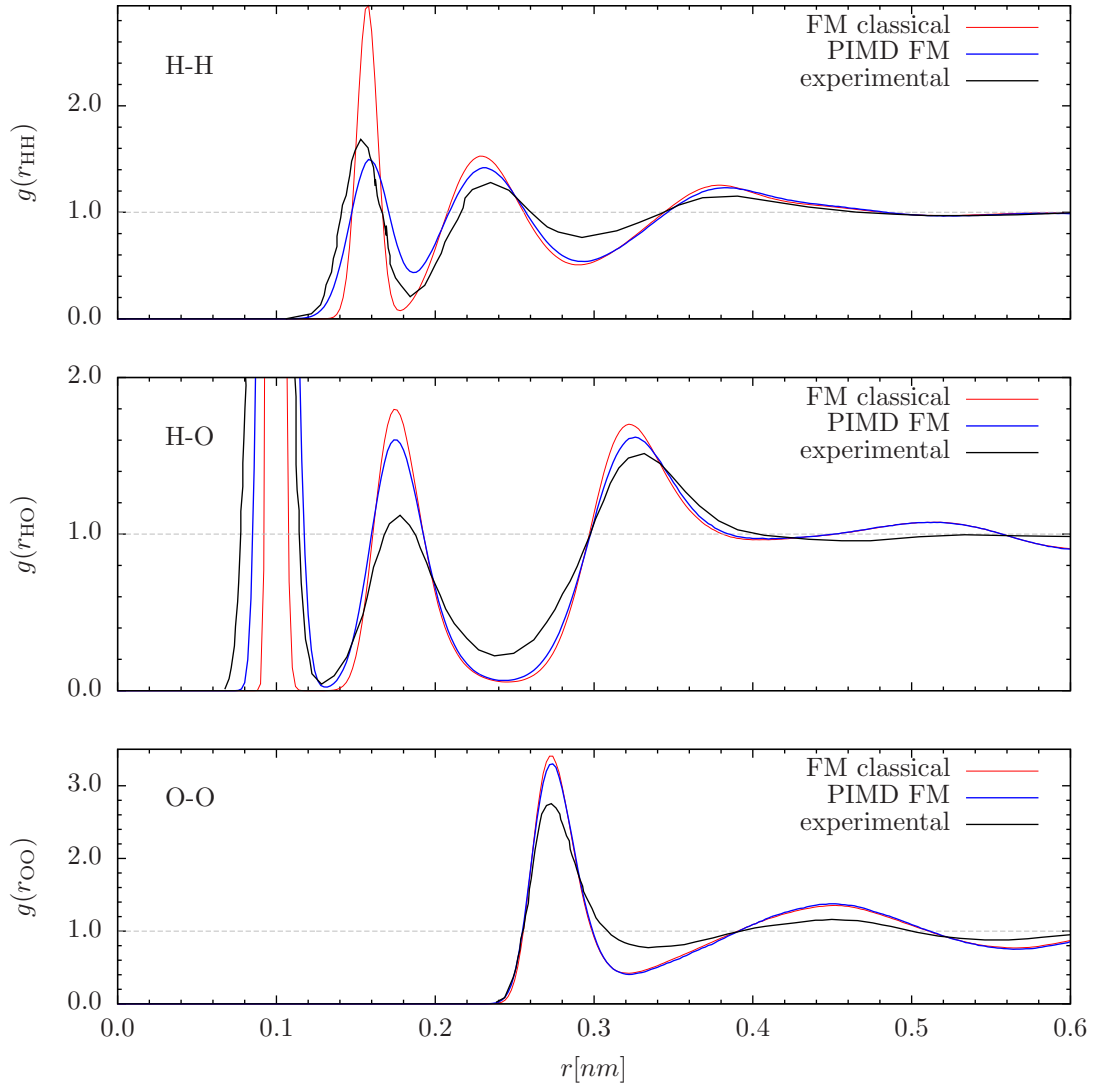
Figure 5.8: Radial distribution function of PIMD at $T = 300K$ using the FM potentials. Intra molecular interactions are included in the calculation of the $g(r)$. The experimental data is taken from [144].

FM model		
	classical	PIMD
Pressure ($\rho = 1.001\,\mathrm{g/\,cm^3}$)	$-2450 \pm 21\,\mathrm{bar}$	$-1816 \pm 100\,\mathrm{bar}$
FM-pc (pressure corrected)		
	classical	PIMD
Pressure ($\rho = 1.001\,\mathrm{g/\,cm^3}$)	$4 \pm 30\,\mathrm{bar}$	$500 \pm 30\,\mathrm{bar}$
Equilibrium density ($p = 1\,\mathrm{bar}$)	$1.001\,\mathrm{g/\,cm^3}$	$0.9902 \pm 0.0001\,\mathrm{g/\,cm^3}$

Table 5.4: Pressure measured in classical and PIMD simulations using the FM model

after 5 steps of iteration, introduced only a minor difference to the O-O interaction potential, while the pressure matched the ab initio reference (denoted as FM-pc in fig. 5.9 and tab. 5.4).

Since the pressure in the PIMD increases by approximately $500\,\mathrm{bar}$, one might argue that the comparison to the experiment (where $p = 1\,\mathrm{bar}$) should be done at the same pressure instead of the same density. In order to probe if the over-structuring can be explained by this difference of pressures, the PIMD system using the FM-pc potential was relaxed to the equilibrium density corresponding to a pressure of $p = 1\,\mathrm{bar}$ (using the Berendsen barostat [104]). The $g(r)$ was then measured again at the corresponding equilibrium density in the NVT ensemble (shown in fig. 5.9). Only a minor effect on the O-O distribution was visible. Thus the increase in pressure in the PIMD simulations cannot explain the remaining over-structuring.

5.4 Discussion of results

The nuclear quantum effect in the FM model leads to a considerable softening of intra molecular structures, but leaves the inter-molecular distributions largely unaffected. This finding is thus consistent with the results from studies with ab initio potentials parametrized from calculations of small water clusters [78] and results from an early study employing pathintegrals in CPMD simulations of water [98]. The improvement in terms of inter-molecular structure for the O-O distribution (as in Morrone et al. [130]) could not be reproduced. However the direct comparison is difficult due to the different ab initio techniques, exchange correlation functionals and systems sizes used.

There exists a number of possible sources of error in the FM approach itself. One source of error might stem from the assumption that the electronic and nuclear quantum effect can be separated. A way to test this assumption would be to perform PIMD with calculating the interactions using DFT at every timestep. This approach is out of reach with present day computational resources for all but
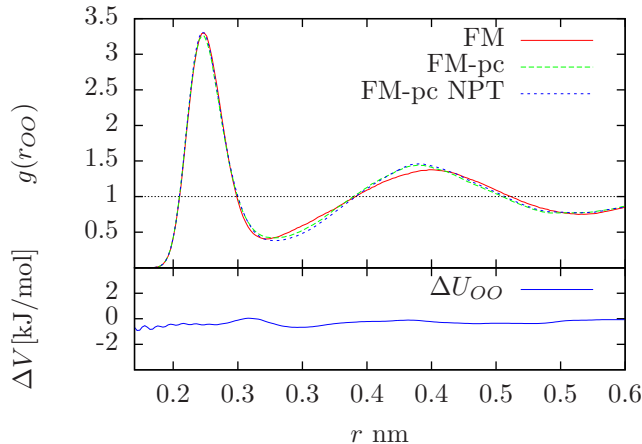
Figure 5.9: Effect of applying a pressure correction to the FM potential. The $g(r)$ as measured from PIMD simulations is shown in the top plot. FM denotes the unaltered potential. FM-pc is the pressure corrected potential, where the pressure correction to a target pressure of $p = 1\,\mathrm{bar}$ was applied in the classical simulation. FM-pc NPT was measured by calculating the equilibrium density at $p = 1\,\mathrm{bar}$ using the Berendsen barostat and then measuring the $g(r)$ from an NVT simulation at this density. The difference in the non-bonded interaction potential introduced by the pressure correction, $\Delta V(r) = V^{OO}_{corrected}(r) - V^{OO}(r)$ is shown below.

tiny systems due to the scaling of the computational load of DFT techniques with $\mathcal{O}(N^3)$ [53].

Another source of error comes from the fact that pairwise potentials (with the exception of the intra-molecular angle term) were used. The forces found by the FM algorithm are a projection of the multi-body potential of mean force onto the pairwise potentials. If these cannot represent the full interactions of the ab initio simulation, important physical details might be lost.

The force extrapolation made in order to extend the FM potentials into regions not sampled in the ab initio simulation represent another approximation. This can mainly affect the intra-molecular interactions as these are more delocalized in the PIMD simulation. Although the description of the internal structure was found improved, the extrapolation might affect the balance between non-bonded an bonded interactions. A solution to this might be to use enhanced sampling techniques, such as parallel tempering , in the ab initio simulation.

In summary, the FM approach is promising, the effective potentials accurately reproduce the classical structure and vibrational density of states while reducing the computational costs. This technique might therefore play a role in the future to build models from bulk ab initio simulations as an alternative to empirical models and cluster based ab initio methods. Including the nuclear quantum effect in the FM model leads to an improved agreement with the experimental structure for the internal degrees of freedom. Finally, it has to be noted that also the experimentally

measured structure has a significant error [144]. Further improvements on the experimental side as well as in the ab initio models are necessary to arrive at consistent picture of the nuclear quantum effect in water.

5.5 Adaptive path integral

The PIMD method allows to include nuclear quantum effects in MD simulations at the expense of an increase in computational cost of about one order of magnitude. This increase is due to the fact that each atom is represented as a ring polymer consisting of P beads which increases the cost for evaluting non-bonded interactions by a factor of P.

The AdResS approach is therefore attractive in order to reduce the computational cost. Thereby a high-detail quantum region using the full PI Hamiltonian can be employed, while the rest of the system is described using a CG potential [85]. This CG potential can be obtained by IBI and represent each molecule in terms of a center of mass interaction. In this approach, the CG potential is an effective, classical representation of the molecule-molecule interaction in the PI picture[5].

The coupling of the quantum to an classical region is possible if the quantum effect is sufficiently local, such that coupling of the CG representation does not perturb properties of the high-detail path integral region. This has been used to change the viewpoint and use AdResS as a tool to probe the locality of quantum effects [85, 86, 87]. In this approach the size of the quantum (PI) region is varied as to asses up to which size of the quantum region the quantum properties are preserved.

Here, the example of a water slab in vacuum is presented and it is demonstrated that the bulk water can be treated in a CG fashion while for the interface region the full PI interaction are maintained.

5.5.1 Classical adaptive system

Preliminary simulations using the classical FM-pc (without path integral) model described in the previous section were first carried out.

The CG model was obtained using IBI with a cutoff $r_{\text{cutoff}} = 0.8\,\text{nm}$, the other parameters were identical to the ones used in section 3.1. As a target distribution, the classical center-of-mass $g(r)$ from simulations of the FM-pc model was used. A pressure correction, as described in section 2.3.4.2 was performed in order to match the reference pressure $p = 1\,\text{bar}$. In order to set up the water-vacuum interface, a system of dimensions $12.093 \times 4.031 \times 4.031\,\text{nm}$ was first equilibrated

---

[5]The AdResS scheme also allows in principle for a coupling of the PI representation to a 3-bead molecule instead of the 1-bead CG model
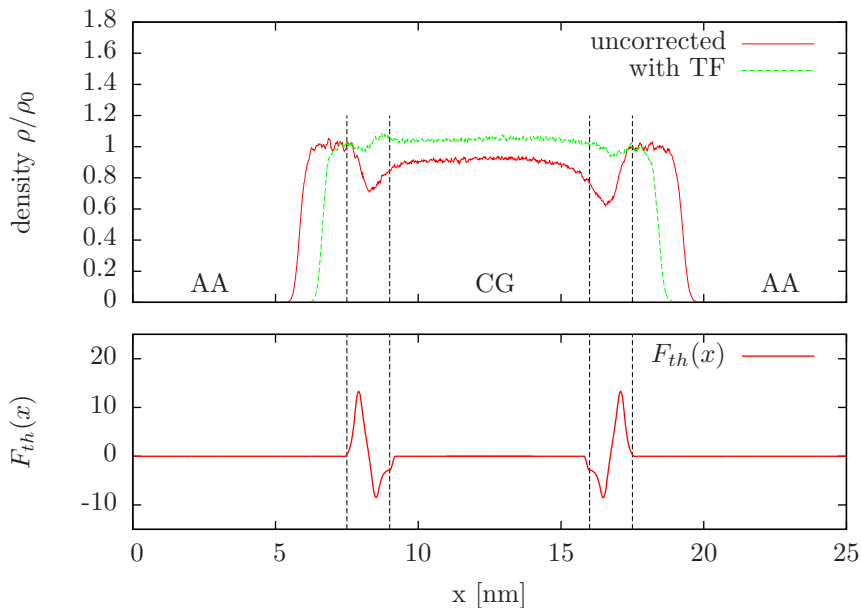
Figure 5.10: Local density profile along the x-axis in the water-vacuum slab (classical MD simulation). Shown is the unperturbed profile in the adaptive simulation (AdResS) where the center of the water slab is described using the IBI based CG potential. The x-component of the thermodynamic force, is shown below (y and z components are 0 by definition). Using the thermodynamic force, the artifact from the mismatch of pressures in the hybrid region is removed an the bulk density recovered.

at the classical density of $1.002\,\mathrm{g/cm^3}$ using periodic boundary conditions. Then, the simulation volume was extended in x-direction to $25\,\mathrm{nm}$, and the equilibrated bulk configurations were placed in the center.

For AdResS, the reference coordinate $\mathbf{x}_0$ of the resolution change was set to be at $(0, 0, 0)$, where the width of the explicit region was $a = 7.5\,\mathrm{nm}$ and the width of the hybrid region $b - a = 1.5\,\mathrm{nm}$. Note that due to the periodic boundary conditions, using this definition there are two hybrid regions symmetrically from the center of the simulation volume (also see fig. 5.10).

A classical AdResS simulation was then run for a length of $250\,\mathrm{ps}$. The resulting density profile measured along the x-axis is shown in fig. 5.10. Clearly a depletion is visible in the hybrid region. This is due to the fact that even though the pressure in the pure simulations of CG potential and FM-pc match, the pressure in the hybrid region is locally different as an artifact of the interpolation (see also section 3.2.1). The effect is more pronounced in the case of the interface (compared to the bulk simulation in section 3.2) since the liquid can respond to the increased pressure by expanding further into the vacuum.

In order to compensate for this, the thermodynamic force iteration as discussed in chapter 3 was applied in the range $7.3\,\mathrm{nm} < l < 9.2\,\mathrm{nm}$. For this it was assumed

that the hybrid region is far enough away from the interface and the local density is the same as in the bulk. After 4 steps of iteration, the flat bulk density profile was recovred within an error of 5%. The density profile is shown together with the converged thermodynamic force in fig. 5.10.

5.5.2 Adaptive PIMD simulation

Now, the adaptive PIMD simulation is discussed, where the interface is simulated using the PI interactions, while in the bulk the CG interactions are used. For the setup, the equilibrated classical configuration (as described in the previous section 5.5.1) was used as a basis for the PI system. In the initial configuration, atoms were replaced by their ring polymer representation using a Trotter number of $P = 16$. The thermodynamic force obtained in the classical case was used for the PIMD simulation. This approximation seems reasonable as the nuclear quantum effect was found to mostly affect intra-molecular properties (despite a difference in pressures, see tab. 5.4). The validity of this can be checked by analyzing the density profile in the adaptive PIMD system. 240 ps of adaptive PIMD simulation were then performed, wherby the identical AdResS parameters as in the classical case where used.

The resulting density profile is shown in fig. 5.11. The density profile at the interface is very similar to the classical density profile, indicating that the nuclear quantum effect does not affect the density. The density profile in the coarse-grained region is however slightly increased compare to the classical simulation. The reason for this may be the use of the thermodynamic force parameterized for the classical simulation, which represents only a first approximation. The nuclear quantum effect is analyzed in terms of the radius of gyration of the hydrogen ring polymer

$$r_{\mathrm{gyr}}^{\mathrm{H}} = \frac{1}{P} \sum_{i=1}^{P} (\mathbf{x}^{(i)} - \mathbf{x}^{C})^2 \tag{5.2}$$

where $\mathbf{x}_C = \frac{1}{P} \sum_{i=1}^{P} \mathbf{x}^{(i)}$ denotes the centroid of the ring polymer. This quantity describes the delocalization of the nucleii in the PI picture: A high value corresponds to a smeared out distribution when taking the ensemble average, while a low value represents a collapsed polymer with a localized distribution. The radius of gyration was measured in slabs along the x-direction of the simulation and normalized by the number of ring polymers in each slab. For comparison, the values of $r_{\mathrm{gyr}}^{\mathrm{H}}$ measured for a the hydrogen of a water molecule in vacuum and in bulk water at the same temperature are shown. It is visible that in the AdResS simulation, close to the CG region, $r_{\mathrm{gyr}}^{\mathrm{H}}$ approaches the vacuum value. This is due to the AdResS force interpolation which slowly fades out the non-bonded interaction forces. In the CG region, $r_{\mathrm{gyr}}^{\mathrm{H}}$ is close to the constant vacuum value. Note that in the CG region, the ring polymers do not have a physical meaning as they are completely
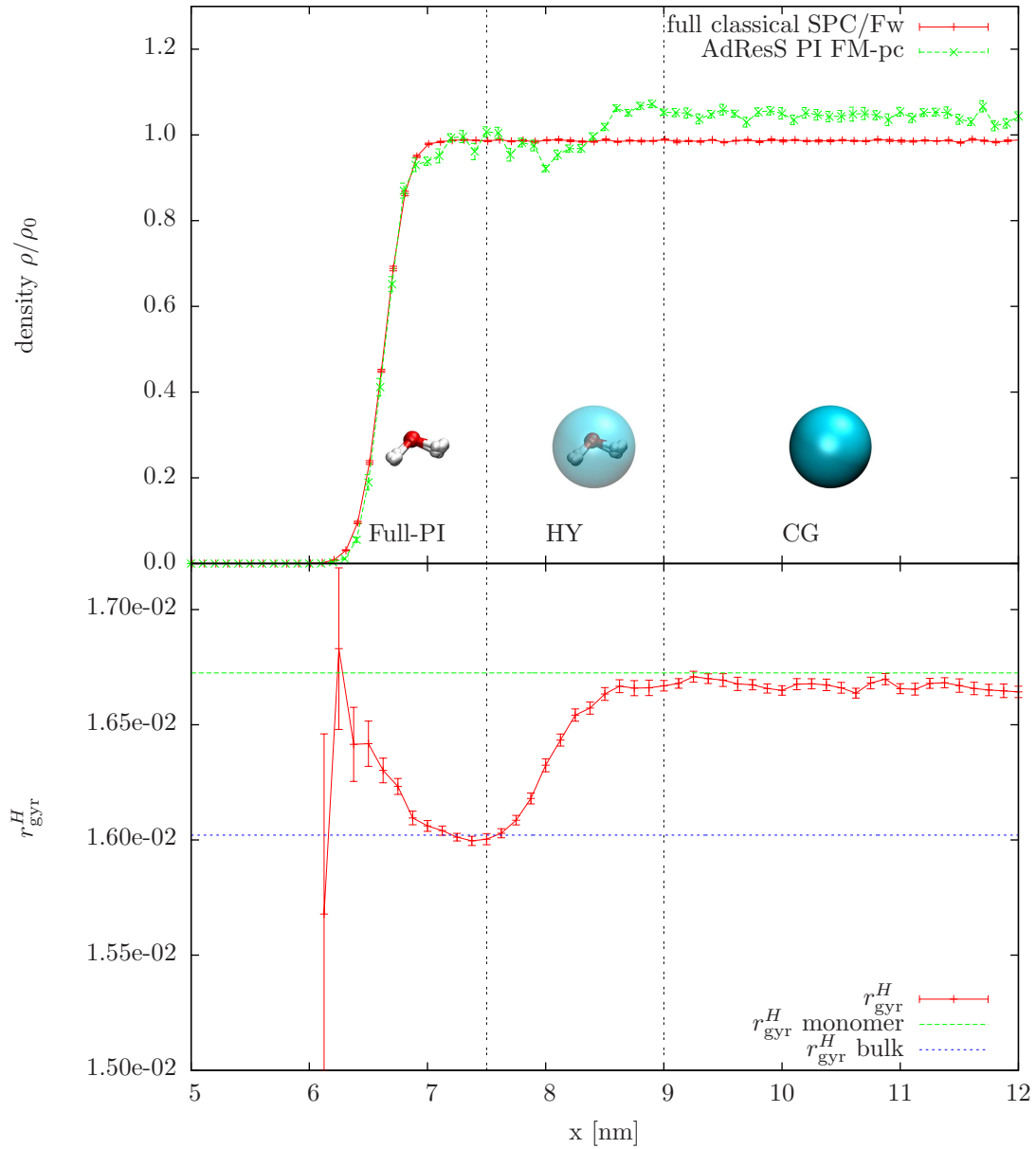
Figure 5.11: (top) Density profile of the water (FM-pc model) slab in adaptive PIMD using a Trotter number $P = 16$ compared against a fully classical simulation using the SPC/Fw model (this profile was shifted to the right by 0.21 nm for alignment). The thermodynamic force obtained from the classical equivalent was used. (bottom) Radius of gyration of the imaginary time ring polymer defined in eq. 5.2. This quantitiy was averaged over time and all- hydrogen atoms and binned according x-position of the ring centroid.

decoupled from the non-bonded interactions. Measurements should hence be only performed on the center of mass of molecules. The technical reasons for this have been discussed in detail in section 2.4.2.

The internal structure is not affected by the interpolation and hence the ring polymers approach the structure of isolated molecules. In the center of the hybrid zone, the bulk value of $r_{\mathrm{gyr}}^{\mathrm{H}}$ is recovered. This indicates that at this position the interface is already sufficiently bulk-like. Close to the vacuum, in the full-PIMD region, $r_{\mathrm{gyr}}^{\mathrm{H}}$ is found to be increased with respect to the bulk value. This can be explained considering that the liquid assumes a less dense structure and thus the hydrogen atom is more delocalized.

The fact that the bulk value of $r_{\mathrm{gyr}}^{\mathrm{H}}$ is reached in the adaptive scheme indicates that the small interface region simulated with PI interactions does not significantly perturb the properties of the PI description. Hence it is possible to use AdResS as a potential way to lower the computational costs for PIMD simulations while avoiding finite size effects that would be present in smaller systems.

5.6 Summary

In this chapter a route for bridging multiple models from the level of explicit electronic interactions up interactions at the molecular scale was presented. Within the FM approach it is possible to find a effective potentials which reproduce structure and vibrational properties of the reference ab initio simulation. These can be used as a basis for investigating the nuclear quantum effect in bulk water as double-counting is avoided.

The nuclear quantum effect in the water-vacuum interface was found to be sufficiently local to be unperturbed by the coupling to an CG bulk region. This result is promising as it shows the way for further applications of the AdResS scheme in situations where the nuclear quantum effect is of importance. Given that inaccuracies, possibly arising from insufficient ab initio exchange correlation functionals can be eliminated, this represents a promising step towards a new ab initio based model for use in PIMD simulations. This can be of interest to study for example the solvation of biomolecules where it could be used to treat the first solvation shell in the PI picture, while treating the remaining solvent in the efficient CG model.

Chapter 6

Hamiltonian based adaptive resolution

In this chapter, an adaptive simulation scheme is discussed which is based on a Hamiltonian formulation. Although the approach is different, the scheme shares key concepts with AdResS and is therefore refereed to as H-AdResS, the H indicating the Hamiltonian framework. As described extensively in chapter 1, the requirement behind AdResS was to satisfy Newtons third law instantaneously, leading to non-conservative forces. Here this requirement is loosened to be fulfilled only on average.

The advantage is clearly that now micro-canonical simulations (NVE ensemble) are possible. Although the AdResS scheme is well defined in terms of thermodynamic observables such as temperature, pressure and density, the Hamiltion formulation allows for the straightforward application of all statistical mechanical concepts. This paves the way for relating the thermodynamic force concept described in chapter 3 to well known thermodynamical expressions as will be described in the following chapter.
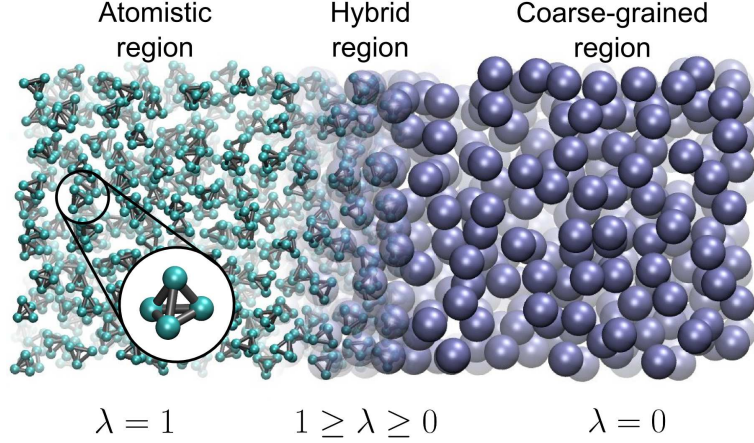
Figure 6.1: Setup showing a part of the H-AdResS test system composed of tetrahedral molecules. In the all-atom region on the left, molecular interactions are treated in full detailed, while in the coarse-grained region on the right, a center of mass description is used. The two regions are interfaced by a hybrid region.

6.1   Derivation

The goal is to achieve a Hamiltonian picture of coupling the all-atom (AA) and coarse-grained (CG) system. The coupling should be such that when the AA region is set to have zero size, the pure CG Hamiltonian is recovered and vice-versa. First the all-atom reference Hamiltonian needs to be defined, where the same notation as in chapter 3 was used. For N molecules (Greek indices) composed of n atoms (Latin indices) the Hamiltonian is written as

$$H^{AA} = \sum_{\alpha,i} \frac{p_{\alpha_i}^2}{2m_{\alpha_i}} + U^{int} + \sum_{\alpha} U_{\alpha}^{AA} \tag{6.1}$$

$$U_{\alpha}^{AA} \equiv \frac{1}{2} \sum_{\beta,\beta\neq\alpha} \sum_{i\in\alpha,j\in\beta} U^{AA}(|\mathbf{r}_{\alpha_i} - \mathbf{r}_{\beta_j}|)$$

,

where U^{int} denotes the intra-molecular interactions. For simplicity, the inter-molecular interactions are again assumed to be spherical symmetric and pairwise. The CG representation, as for example obtained by IBI is a function of the molecular center of mass and is denoted as

$$U_{\alpha\beta}^{CG} \equiv U^{CG}(|\mathbf{R}_{\alpha} - \mathbf{R}_{\beta}|) \ . \tag{6.2}$$

The total CG potential energy on molecule α is given by $U_{\alpha}^{CG} \equiv \sum_{\beta,\beta\neq\alpha} U_{\alpha\beta}^{CG}/2$.

In contrast to AdResS, in the H-Adress the total potential energy of a molecule α is weighted with a function $\lambda(\mathbf{R}_{\alpha})$. The idea behind this is that the "local nature"

(i.e. how much a molecule is coarse-grained or atomistic) defines the contribution to the potential energy. Now a Hamiltonian for the adaptive resolution system is constructed where the resolution changes based on the position of a molecule:

$$H = \sum_{\alpha,i} \frac{p_{\alpha_i}^2}{2m_{\alpha_i}} + \sum_{\alpha} \left\{ \lambda(\mathbf{R}_\alpha)U_\alpha^{AA} + (1 - \lambda(\mathbf{R}_\alpha))U_\alpha^{CG} \right\} + U^{\text{int}} . \qquad (6.3)$$

Taking the negative gradient in order to find the forces acting between atoms one arrives at

$$\mathbf{F}_{\alpha_i} = -\sum_{\alpha} \left\{ \lambda_\alpha \nabla_{\alpha_i} U_\alpha^{AA} + (1 - \lambda_\alpha)\nabla_{\alpha_i} U_\alpha^{CG} + \nabla_{\alpha_i} U^{\text{int}} + \nabla_{\alpha_i} \lambda_\alpha (U^{AA} - U^{CG}) \right\}$$

$$(6.4)$$

$$= \sum_{\beta,\beta \neq \alpha} \left\{ \frac{\lambda_\alpha + \lambda_\beta}{2} \sum_{j \in \beta} \mathbf{F}_{\alpha_i|\beta_j}^{AA} + \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2} \right) \mathbf{F}_{\alpha_i|\beta}^{CG} \right\}$$
$$+ \mathbf{F}_{\alpha_i}^{int} - \left[U_\alpha^{AA} - U_\alpha^{CG} \right] \nabla_{\alpha_i} \lambda_\alpha ,$$
$$(6.5)$$

where $\mathbf{F}_{\alpha_i|\beta_j}^{AA}$ is the force exerted by atom j of molecule β on atom i of molecule α, and $\mathbf{F}_{\alpha_i|\beta}^{CG}$ is the CG force of molecule β on atom α_i. The CG force is redistributed according the center of mass, as in AdResS eq. 2.83.

The weighting function $\lambda(\mathbf{R}_\alpha)$ can be chosen according to the desired geometry, where $\lambda = 1$ marks the all atom and $\lambda = 0$ the CG region. It is important to stress one conceptual difference compared to AdResS, which is that in principle, the atoms receive hybrid forces even in the CG zone. This is because the pairwise force is given by the average of the molecular weights (instead of the product). Thus the atomistic interaction can be non-zero if one of the two molecules is in the hybrid region. In practice the extend of this effect is limited by the interaction cut-off if no long-range method (e.g. Ewald summation, also see section 2.1.4) is used. The presence of hybrid interactions requires that the internal degrees of freedom are present and integrated also in the CG region (as was done for technical reasons in the previous AdresS implementations). In the CG region however, they are uncoupled from the neighboring molecules since the all-atom contribution is zero if both molecules are in the CG region.

The forces originating from the ansatz above have three contributions:

- The first term (curly brackets of eq. 6.5) contains the weighted sum of pairwise forces and is anti-symmetric under exchange of particle labels and hence respects Newtons third law.

- The second term is due to intra-molecular forces, which do not contribute to forces between molecules.
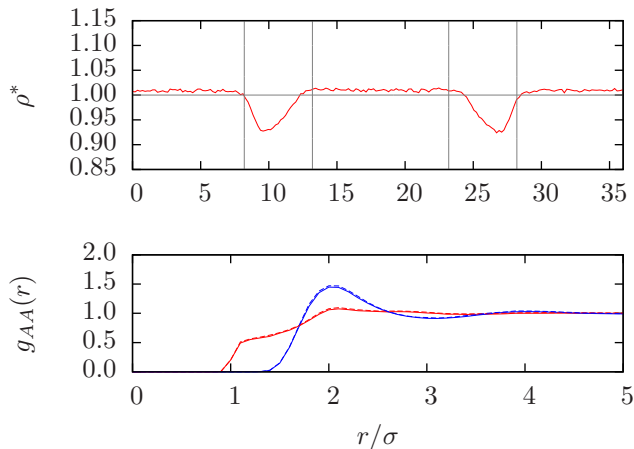
Figure 6.2: (top) Normalized density profile in the H-AdResS simulation of the tetrahedral model. (bottom) atomistic (blue) and center of mass (red) RDF calculated in the all-atom region ($13.4 < x < 23.4\sigma$) for a pure all-atom simulation (solid lines) and in the H-AdResS simulation

- The third term can be interpreted as an external field acting only in the hybrid zone. The integral of this term can be related to the difference of free energy between the AA and CG model.

How the third term (drift term) affects the simulation will be clarified in the following section. Here it is just important to note that the ansatz of a well-defined Hamiltionan is made, which can be used to generate a microcanonical trajectory of the adaptive resolution system.

6.2 Numerical validation

The H-AdresS approach was tested using a modified version of GROMACS. The drift term (eq. 6.5) requires the storage of the per-atom potential energy, which slightly increases the implementational effort. After the calculation of all non-bonded forces, the drift force is added to the force on each atom according to eq. 6.5.

In order to compare the results to the original AdResS implementation, the same system as in as in [70] was simulated. This system consists of a fluid of tetrahedral molecules, depicted in fig. 6.1. In the following, reduced units, i.e. $m = k_B = \epsilon = \sigma = 1$ are used since this is more convenient in the tetrahedral model. The tetrahedral molecule consists of 4 identical atoms connected by anharmonic bonds. The intra-molecular part is given by a quartic bond potential:

$$U^{int}(r_{int}) = \frac{1}{4}k(r_{int}^2 - b^2)^2 \tag{6.6}$$

Figure 6.3: Total energy drift during a microcanonical NVE H-AdResS simulation for different timesteps Δt. AA stands for the pure all-atom and H for the H-AdResS hybrid simulation.

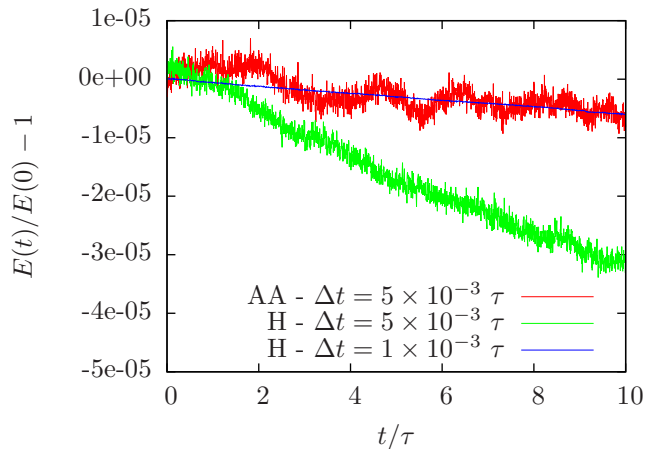with $r_{int} = |\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|$, $(\alpha = \beta)$ denoting the intra-molecular distance. For this interaction $b = 1\sigma$ and $k = 7500\epsilon/\sigma^2$ were used. The non-bonded interaction employed here is a Weeks Chandler Andersen [148] (WCA) potential

$$U^A(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 + \frac{1}{4} \right] & : r \leq 2^{1/6}\sigma \\ 0 & : r \geq 2^{1/6}\sigma \end{cases} \tag{6.7}$$

where $r = |\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|$, $(\alpha \neq \beta)$ denotes the inter-atom distance.

The simulation was prepared to contain 798 tetrahedral molecules in a box of dimensions $36.845\sigma \times 15\sigma \times 15\sigma$, corresponding to a molecular density of $\rho = 0.0962$. To equilibrate the system, all-atom simulations using a Langevin thermostat with $\tau_L = 2\tau$ were carried out, where the Temperature was set to $T = 1$. The atomistic pressure was measured to be $p_{AA} = 0.355\sigma^3/\epsilon$.

For the CG interaction U^{CG}, 240 steps of IBI where carried out, with the range of the effective potential being 4.5σ. No pressure correction was used, since the CG pressure $p_{CG} = 0.385\sigma^3/\epsilon$ was sufficiently close to the all-atom pressure.

The density and pairwise structure measured from H-AdResS simulations are plotted in fig. 6.2. The density profile shows a depletion in the hybrid region, but no overall shift of densities between the AA and CG region. This profile is qualitatively comparable to the profile observed in chapter 3 obtained in the AdResS simulation using the pressure-corrected water model (however, no asymmetry in the density profile is observed which is due to the CG model of the tetrahedral molecule being a well-behaved test case, where both pressure and compressibility are close).

The pairwise atomistic structure was calculated locally by considering only the atoms in the AA region and atoms up to a distance r in the hybrid region, which

is defined as

$$g(r)_{AA} = \frac{1}{h_{id}(r)} \left\langle \frac{1}{N_{AA}(t)} \sum_{i \in V_{AA}} \sum_{j \neq i} \delta(r - r_{ij}) \right\rangle , \qquad (6.8)$$

where $h_{id}(r) \approx 4\pi\rho r^2 \delta r$ is the (unnormalized) RDF of an ideal gas at the same density, $N_{AA}(t)$ is the number of atoms in the AA region at a time t and V_{AA} the volume of the all-atom region. When considering $V_{AA} = V$ the conventional definition of the RDF is recovered (see section 2.2.

The RDFs were calculated in this way for atoms and the centers of mass of the molecules (fig. 6.2, bottom). In both cases, the local RDFs in the AA match the all atom reference RDFs. Thus it can be concluded that H-AdResS preserves the local structure at least in the all-atom region.

In order to show the energy conservation H-AdResS, NVE simulations starting from the equilibrated configuration were performed. The energy, monitored for different time-steps Δt is shown in fig. 6.3. Within 10τ (corresponding to 2000 integration steps) the relative drift in energy is less then 4×10^{-5}. Compared to an all-atom simulation, the error at the same Δt is slightly higher. However, the relative drift further decreases if the timestep is reduced, indicating that it is simply due to the discretization of the equations of motion. Thus, the H-AdResS approach conserves the energy in an NVE simulation.

6.3 Helmholtz compensation

Now it will be demonstrated how the drift term (third term of eq. 6.5) can be associated with the free energy difference between the AA and the CG model. The connection is made through the long known technique of thermodynamic integration, also known as Kirkwood integration [149]. Within this approach, free energy differences of two model Hamiltonians H_A and H_B can be computed.

In Kirkwood integration, an interpolation between potential energies is made by defining

$$U(\lambda) = \lambda U_A + (1 - \lambda)U_B , \qquad (6.9)$$

where λ is a coupling parameter that takes values between 0 and 1. The kinetic energy does not appear in this expression since no change in degrees of freedom is performed, and thus the contributions to the free energy difference cancel. The free energy difference between the two systems is then calculated in the canonical ensemble as:

$$F(\lambda) = k_B T \ln Z(\lambda) , \qquad (6.10)$$

and

$$\frac{\partial F}{\partial \lambda} = k_B T \frac{1}{Z} \frac{\partial Z}{\partial \lambda} = \left\langle \frac{\partial U}{\partial \lambda} \right\rangle = \left\langle \frac{\partial (U^A - U^B)}{\partial \lambda} \right\rangle . \qquad (6.11)$$

The eq. 6.11 can be integrated with respect to λ to yield the free energy difference between model A and B

$$F_A - F_B \equiv \Delta F = \int_0^1 \mathrm{d}\lambda' \left\langle U^A - U^B \right\rangle_{\lambda'} \; , \qquad (6.12)$$

where the bracket $\langle \rangle_{\lambda'}$ denotes the canonical average at a particular λ'. In practice, the free energy difference is thus calculated by performing a series of simulations at constant λ' and integrating numerically according to eq. 6.12. Alternatively it is also possible to define λ as a function of time and switch from 0 to 1 in a dynamical sense, a method sometimes referred to as "slow growth method"[150]. If the switching is performed slowly enough, the average taken over small time windows will be close to canonical average at fixed λ.

The connection to the drift term in H-AdResS is apparent when considering a λ which further depends on the position of a molecule $\lambda(\mathbf{R}_\alpha)$. The free energy per molecule would then give rise to a force

$$-\nabla_{\alpha_i} \frac{\Delta F(\lambda)}{N} = -\frac{1}{N} \frac{\partial \Delta F}{\partial \lambda} \nabla_{\alpha_i} \lambda(\mathbf{R}_\alpha) = -\frac{\left\langle U^A - U^B \right\rangle}{N} \nabla_{\alpha_i} \lambda(\mathbf{R}_\alpha) \; , \qquad (6.13)$$

which has the same form as the drift term in H-AdResS (third term in eq. 6.5) averaged over time and all molecules, if the model A is associated with the all-atom (AA) and B with the CG model. This implies that, any free energy difference of the two models creates a drift force acting only in the hybrid region. The direction of the force is towards the region of the lowest free energy.

This effect is not desired for most applications as one would like a coarse-grained model where the free energy matches the atomistic reference. Since this is not always possible, one way to overcome this is to add the known free energy difference as a compensation term to the H-AdresS Hamiltonian, as to cancel the drift force on average. Therby, an approximation can be made by calculating $\Delta F(\lambda)$ from bulk simulations (making use of eq. 6.12) at constant λ throughout. The total Hamiltonian then reads

$$H_{\Delta F} = H - \sum_{\alpha=1}^{N} \frac{\Delta F(\lambda(\mathbf{R}_\alpha))}{N} \; . \qquad (6.14)$$

Since this compensation only acts like a position dependent external field, the forces are still conservative and the scheme conserves the total energy.

6.4 Gibbs compensation

The compensation introduced in the previous section removes the effect of the free energy difference, but as in AdResS also pressure differences can occur. In

analogy to the reasoning made in chapter 3, the chemical potentials μ of the two subsystems have to match. As before, the grand canonical ensemble is appropriate since the two subregions can exchange particles. In the grand-canonical ensemble the chemical potential is:

$$\mu N = G = F + pV \ , \tag{6.15}$$

where F is the (Helmholtz) free energy, G the Gibbs energy, V the volume and p the pressure. This suggest a compensation according to

$$\frac{\Delta G(\lambda)}{N} = \frac{\Delta F(\lambda)}{N} + \frac{\Delta p(\lambda)}{\rho} \ , \tag{6.16}$$

where ρ is the all-atom reference density. The pressure difference $\Delta p(\lambda) = p(\lambda) - p_A$ can be computed in the Kirkwood integration. The total Hamiltonian employing this compensation is then

$$H_{\Delta G} = H - \sum_\alpha \frac{\Delta G(\lambda(\mathbf{R}_\alpha))}{N} \ . \tag{6.17}$$

In the following, $\Delta F(\lambda)$ will be referred to as "Helmholtz-" while $\Delta G(\lambda)$ will be called "Gibbs" compensation. The role of the Gibbs compensation is similar to the thermodynamic force introduced before in the AdResS approach. The aim is to cancel drift forces which arise from the difference in chemical potential. In contrast to the thermodynamic force however, the compensation is calculated from Kirkwood integration where the whole system interacts at a constant λ. This represent an approximation, which can be expected to be reasonable if λ only changes little on the length-scale of the correlations between molecules. In the following, the features of the compensations and the differences to the thermodynamic force will be discussed further based on numerical examples. The effect of the different compensation terms is shown in figure fig. 6.4.

6.4.1 Numerical tests

The IBI-based CG model (discussed in section 6.2) was found to have almost the same pressure without any pressure correction term. To have a CG model which behaves similar to CG models of more realistic liquids (with a pressure higher then the reference), a modified WCA interaction was used for U_{CG}. To this end $\sigma' = 1.7\sigma$ was used and the potential shifted to the right by $r' = r - 0.05\sigma$. This yields a pressure in the pure CG simulation $p_{CG} = 0.438\sigma^3/\epsilon$ which is slightly higher then the atomistic value. Note that this CG interaction was not parametrized with any of the standard coarse-graining methods and consequently the pairwise structure is not expected to be reproduced.

The different compensation terms were then calculated in a "slow-growth" simulation [150] in a cubic box of side length 15σ at the reference all-atom density.
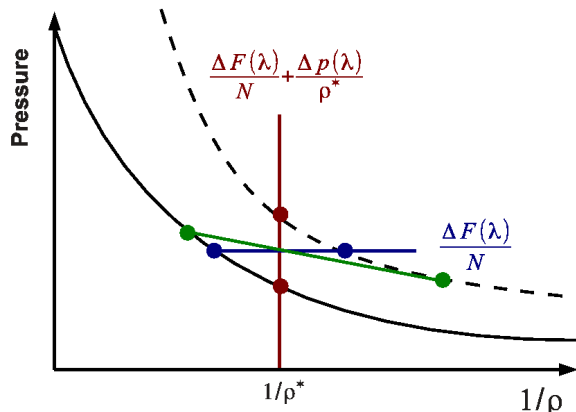
Figure 6.4: Cartoon illustrating the role of the different compensation terms by their effect on the state point of the atomistic (AA) and CG subsystems. The circles represent the state points of atomistic and coarse-grained sub-systems on the respective isotherms. In the system without compensation (green curve), both subsystems are at different pressure and density. The Helmholtz compensation (blue) removes the free energy component and the subsystems therefore are on horizontal equal pressure lines. The Gibbs compensation also removes the pressure difference and therefore the subsystems reside on a vertical, constant reference density (ρ^*) line.

The simulation was run for a total length of 10^6 steps, the coupling parameter λ was decreased at a rate of -10^{-6}/step from 1 to 0. Prior to starting the growth, the system was equilibrated at constant $\lambda = 0$. The different compensations obtained in this way are shown as a function of λ in fig. 6.5 (top panel).

 Then, H-AdResS simulations were run employing the different comensation terms. As is visible from fig. 6.5 (middle), the density profile without any compensation shows a depletion in the CG region and an increased density in the AA region. This is due to the different pressure and different free energies in the CG and AA model. The hybrid region acts as an "active membrane" pushing molecules into the AA region according to the free energy and pressure gradients. Adding the Helmholtz compensation slightly improves the imbalance by removing the drift coming from the Helmholtz free energy difference. The Gibbs compensation, on the other hand, removes almost completely the overall shift in density between AA and CG region and a flat density profile is reached. In the hybrid region there are however fluctuations on the order of 5%. A possible explanation for this is as follows: In the derivation of the compensation term it was assumed that all molecules interact at the same constant λ. In the H-AdResS setup however, a molecule in the hybrid region is surrounded by molecules with different λ values. The compensation does not take account of this and thus the drift forces are not exactly canceled, resulting in a small local variation in the density is produced. A possible way to overcome this would be computing the drift term and pressure locally during the H-AdResS simulation and deduce a compensation from it or alternatively use the
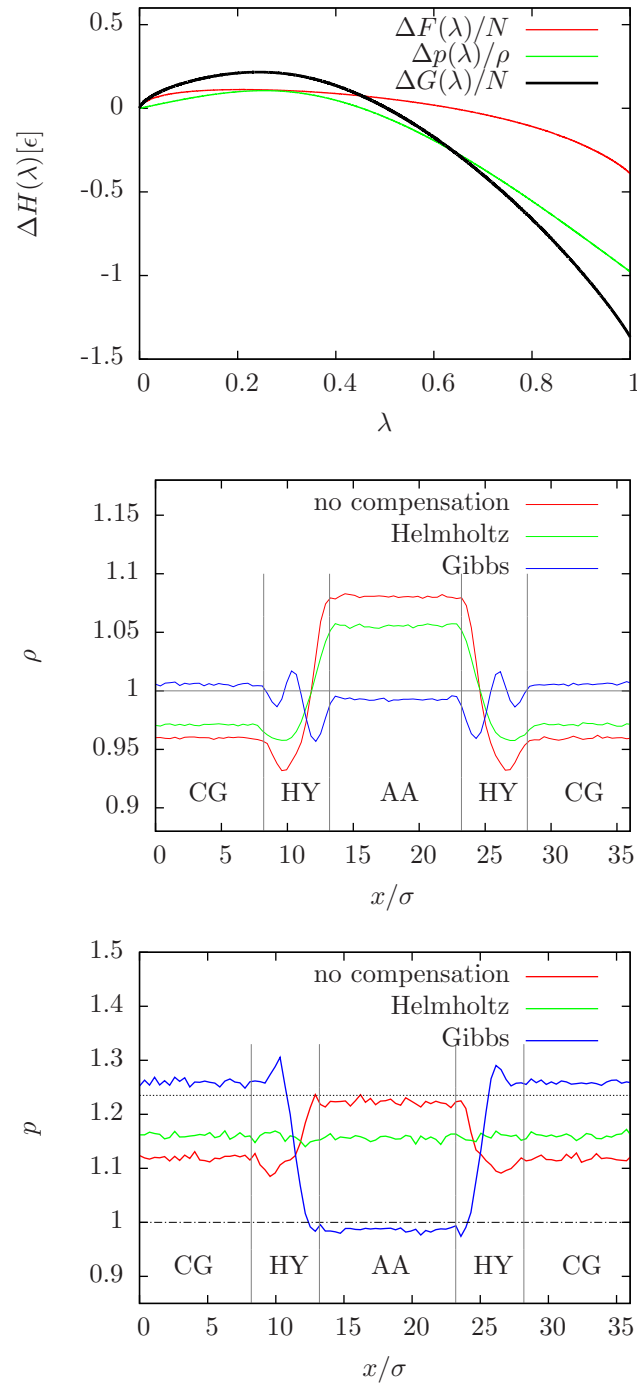
Figure 6.5: (top) Components of the Gibbs compensation term calculated by Kirkwood integration. (middle) Density profiles in an H-AdresS simulation employing the different compensation terms. (bottom) Local pairwise pressure as a function of the box x-coordinate.

thermodynamic force iteration introduced in chapter 3 to refine the compensation term.

Another interesting property is the local pairwise pressure in the simulation. This quantity is calculated locally in slabs as explained in appendix A.1. Note that only internal forces (and not the compensation forces) are taken into account. The resulting local pressure profile is shown in fig. 6.5 (bottom). In the uncorrected simulation, the pressure in the AA region was found to be larger compared to the CG region. The reason is that in equilibrium, the density adjusts until the forces coming from both free energy and pressure differences are canceled. The Helmholtz compensation shows a flat local pairwise pressure profile throughout the entire simulation volume, the value of which lies between the pure AA and CG pressures. This can be explained considering that Helmholtz construction was designed to cancel the drift term. Thus, on average no net force is acting on the system, and consequently the density adjusts such that the local pairwise pressures are equal. In the case of the Gibbs compensation, the local pairwise pressure are close to the respective pure reference values. The hybrid region in between shows an approximately linear interpolation between the two pressures. This is to be understood as follows: The Gibbs compensation is constructed to remove both the differences from the free energy and pressure gradients. Since the density profile is very well restored to the original density in the whole box, the pairwise molecular pressures returns to the reference bulk values for AA and CG.

As a consequence, it might be advantageous to use either compensation depending on the purpose of the adaptive simulation. Through the Gibbs compensation the density profile is flat while through the Helmholtz compensation the pairwise pressure profile remains flat. For applications where hydrodynamic properties are investigated, it could be more favorable to work at constant molecular pressure. When structural properties are of interest, most likely the Gibbs compensation will be favored.

6.5 Comparison to AdResS thermodynamic force

To compare the compensation terms directly to the thermodynamic force approach presented in chapter 3, the SPC/E water coupled to the non-pressure corrected CG model studied in section 3.4 was simulated again with the H-AdResS approach. For the sake of being comparable, these simulations were done in the NVT ensemble with identical parameters.

The compensation terms were computed using the slow growth method, where λ was decreased linearly from 1 to 0 within 10^5 MD steps. The contributions to the compensation terms $\Delta p(\lambda)$ and $\Delta F(\lambda)$ are shown in fig. 6.6 (top).

From the density profile it is visible that H-AdResS produces a larger imbalance
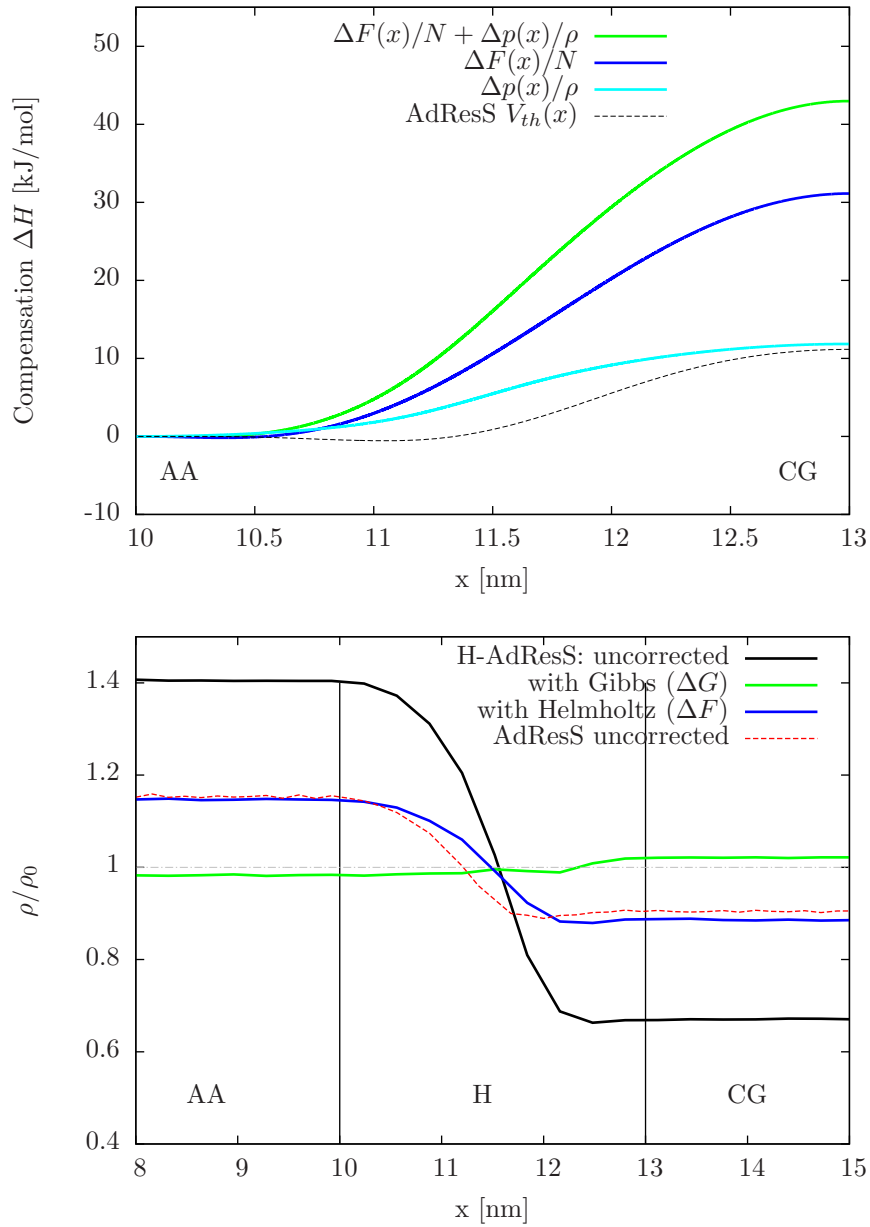
Figure 6.6: (top) Different contributions to the compensation term calculated from slow growth simulations in the SPC/E water system. For comparison, the thermodynamic force in the AdResS approach is shown for the same system. (bottom) Density profile along the axis of resolution change in simulations (H-AdResS and AdResS) employing the different compensations.

in the density compared to AdResS if no compensation is applied. This is expected, since H-AdResS includes the contribution of the drift term, and no equivalent term is present in AdResS. As a consequence, in H-AdresS, the difference in free energy drives an increased density imbalance. If the effect of the drift term is removed, as it is done within the Helmholtz compensation, the H-AdResS system shows a density profile which is very similar to the uncorrected AdResS profile. This implies that the only remaining force driving the imbalance is due to the difference in pressure. Indeed this is further verified as including the pressure compensation in H-AdResS the system relaxes to an almost perfectly flat density profile.

The comparison of the thermodynamic force (corresponding potential $U_{th}(x)$) with the pressure compensation term $\Delta p(\lambda)/\rho$ reveals that the value at the end-points is very similar, the shape being different. This is most likely due to the fact that the thermodynamic force was parametrized directly in the adaptive simulation, where molecules interacting at different λ contribute. Conversely, the $\Delta p(\lambda)$ was obtained from the slow-growth simulations which use a constant λ. This approximation can be speculated to be the reason for the deviations from the perfectly flat density profile which was provided by the thermodynamic force approach.

As a consequence, these findings imply that a good initial guess for the thermodynamic force can be obtained in slow growth simulations. In case the reference density needs to be matched more closely, additional iterations with the thermodynamic force approach can be carried out based on this initial guess.

A potential application is the possibility to predict free energy differences of models from the density profile calculated in H-AdResS simulations. To test this, an H-AdResS simulation of SPC/E water coupled to a CG region employing the pressure corrected CG water model discussed in section 3.2.1 was carried out. The density profile is shown in fig. 6.7 (bottom). In contrast to the result in chapter 3, the density in the all-atom region does not match the equilibrium density. The reason is the drift term (eq. 6.5 which is proportional to the free energy difference of the models. Since there is no difference in pressure (at least for the "pure" cases of $\lambda = 1$ and $\lambda = 0$), the density profile can be used to estimate the free energy difference based on the thermodynamic force. To this end, the zeroth order thermodynamic force (eq. 3.9) was calculated based on the atomistic compressibility ($\kappa_T = 5.1 \times 10^{-10} \, \mathrm{Pa}^{-1}$). The corresponding potential shows a free energy difference, which is within 15% accurate compared to the result from Kirkwood integration. The difference may be due to the fact that eq. 3.9 assumes a constant compressibility while it is known that the compressibility is significantly increased in the CG model. Using this technique, it is therefore possible to obtain an estimate for the free energy difference of two models by coupling them in a H-AdResS simulation and simply calculating the density profile.

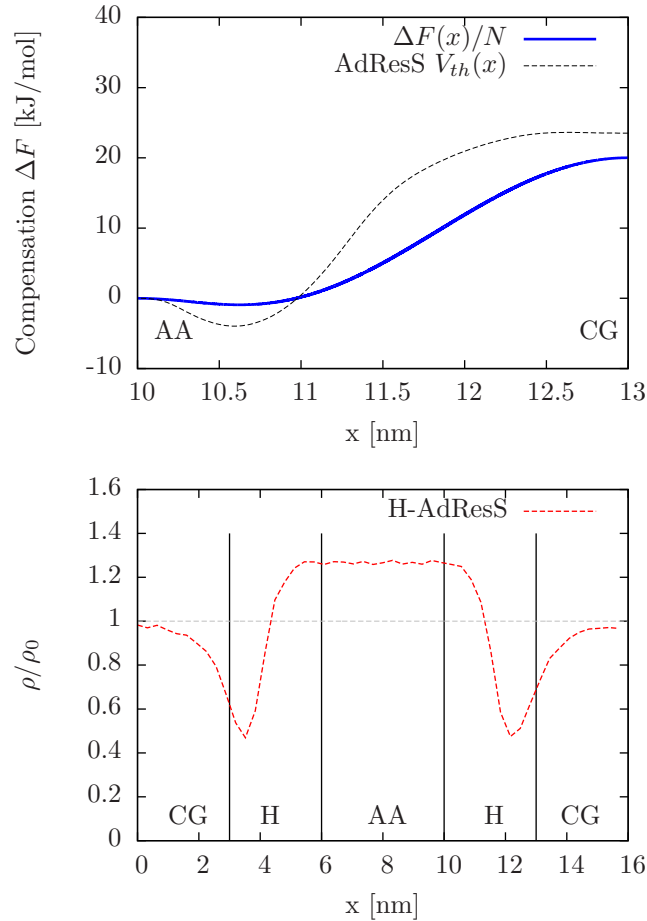Figure 6.7: (top) Free energy term calculated from slow growth simulations in the SPC/E water system coupled to the pressure corrected CG model. For comparison the integrated thermodynamic force $V_{th}(x)$ (zeroth order) calculated from H-AdResS simulations is shown. (bottom) Density profile along the axis of resolution change in H-AdResS simulations using the CG water model with pressure correction discussed in chapter 2.

6.6 Summary

In this chapter, an alternative approach for adaptive simulations was presented. It was shown that within this approach it is possible to run microcanonical simulations where the total energy of the system is conserved. Furthermore, Monte Carlo simulations would be possible with the H-AdResS approach, overcoming a limitation of the AdResS approach.

The forces derived from the Hamiltonian include a drift term which acts like an external field. By introducing a compensation term, the effect of this term can be removed on average, while instantaneously Newtons third law is not preserved. In the Hamiltonian description, however the application of concepts from thermodynamics is straightforward and thus promises a deeper conceptual understanding of the adaptive method.

The new approach may also enable optimization of the coarse graining procedure itself. Since the optimal CG potential does also match the reference free energy, it would be possible to use H-AdResS as a way to parametrize the CG potentials. One may think of an iterative scheme, starting out with a system as simple as an ideal gas (which should be treatable within the H-AdResS approach) and perform iterations of H-AdResS simulations to refine the CG model. There might even be applications where the ideal gas is sufficient as a CG model once the proper compensation terms are included.

Another interesting extension may be the calculation of free energies without performing Kirkwood integration. As was demonstrated in this chapter, the Gibbs compensation (which is related to the chemical potential by $\mu = G/N$) was shown to lead to an flat density profile between the two system. If the CG model is taken to be the ideal gas, its absolute free energy is know analytically. This suggests a way of calculating the free energy of any atomistic model by coupling it to an ideal gas model and simply finding the external field which leads to a flat density profile.

Chapter 7

Summary & Conclusion

In this work, different approaches for performing hierarchical as well as simultaneous multi-scale techniques were applied. The simultaneous approaches promise to be useful for many problems in soft-matter science where the system can be partitioned in a high- and low resolution region. However, special challenges are posed which are due to the simplifications made in the low-resolution models. Several strategies to address these problems were discussed.

7.1 Adaptive resolution simulations

The AdResS approach was studied in chapters 2-4, where it was shown to work for different systems. As was demonstrated in chapter 4, a considerable computational speed-up of the simulation can be achieved with an efficient implementation in the GROMACS software package. However, some effort has to be spent on implementing the method in a manner which takes care of the inhomogeneities in computational load in a domain-decomposition. The internal degrees of freedom in this implementation are kept throughout the entire simulation volume and are decoupled in the CG region. The speed-up is therefore only due to savings on the non-bonded force calculation, which usually is the most time consuming part in MD simulation. Future implementations could improve on this by removing internal degrees of freedom in the CG region and re-inserting them using a stochastic scheme once a molecule diffuses into the hybrid region.

For the study of a C60 fullerene in liquid toluene (chapter 4), it was found that already a small all-atom region around the fullerene is sufficient to retain structural and even diffusional properties. This can also be seen as a method for investigating the influence of chemical details, which are lost in the CG model: Since the all-atom properties could be recovered with only small all-atom regions, the chemical details missing in the bulk can be seen as "unimportant" for the local solvent structure around the C60.

The force interpolation in AdResS yields a stable simulation scheme with well defined pressure and temperature, but requires a thermostat to be used due to the non-conservative forces in the hybrid region. An alternative approach, named H-AdresS, was presented in chapter 6, where a Hamiltonian interpolation scheme was constructed. The gain is that standard statistical concepts can be applied straight-forwardly and micro-canonical MD as well as Monte Carlo simulations are possible. In comparison to the AdResS scheme, however, a few conceptual differences have to be noted: In H-AdResS, Newton's third law is not valid in every instance due to the presence of a drift term in the forces. It is however possible to remove the average effect of the drift term, hence preserving Newton's third law on average. Furthermore, the internal degrees of freedom in the CG region do contribute to the non-bonded forces (at least in the range of the interaction cut-off).

7.2 Thermodynamic force

In chapter 3 it was shown (in agreement with other studies [54]) that a pairwise interacting CG model can only reproduce either pressure or compressibility. This poses special challenges for the adaptive resolution simulations, where the mismatch causes and non-uniform density profile.

The thermodynamic force (introduced in chapter 3) represents one method to remove the mismatch of pressures based on the requirement that the grand potentials of the pure systems have to be equal. To compute the thermodynamic force, a practical iterative procedure was presented which is based on the gradient of densities. It was shown that it is sufficient to apply the thermodynamic force in the (ideally small) hybrid region, where it provides the amount of work necessary to compensate the pressure gradient coming from the interpolation of different models. Therefore also adaptive resolution simulations can be performed with a CG model which preserves the compressibility. This scheme has already found application in the calculation of Kirkwood-Buff integrals in mixtures [106, 107].

In chapter 4 it was shown that this concept can also be used for a spherical all-atom region embedded in a reservoir employing the pressure correct CG model. The all-atom region was coupled to the center of a C60 fullerene and the near ordering in the surrounding toluene solvent was studied. It was found that the thermodynamic force is very well transferable from the bulk, i.e. when parametrized in the bulk solvent it could also remove the density artifact in the solvation shell of the C60. Therefore, the thermodynamic force can be seen as a property of the all-atom and CG model alone, and does not need to be re-parametrized for each new situation it is used in (at least for the homogeneous, one-component fluids). However, in this case the thermodynamic force iteration needed more steps to converge, most likely due to the complications introduced by the spherical region and the mismatch of compressibilities of the two models (the derivation assumed

a constant compressibility).

This shows that it is desirable to obtain the thermodynamic force or at least a first guess in a more efficient way. A first step in this direction was explored in chapter 6, within the H-AdResS approach, where different compensation terms were studied. Also in the Hamiltonian H-AdResS approach, a non-uniform density profile is visible. Different compensation terms were derived based on either compensating the free energy or chemical potential imbalance. The chemical potential ("Gibbs-") compensation can be seen as the analogue of the thermodynamic force. An efficient way to obtain the compensation terms was presented which is based on "thermodynamic integration" in the bulk system. This compensation therefore does not rely on the relaxation of the density as the thermodynamic force approach and is therefore cheaper to compute.

The direct comparison of compensation terms to thermodynamic force reveals that the pressure component of the compensation term is similar to the converged thermodynamic force. Although the compensation does not take the correlations in the liquid into account and hence cannot achieve a fully flat density profile, it could be used as a first estimate of the thermodynamic force and hence reduce the number of iterations needed.

7.3 Nuclear quantum effects in water

In chapter 5 a different example of scale-bridging techniques was present which deals with the construction of a water model suitable for use with path integral MD (PIMD) simulations of water. For this, force-matching (FM) was used to transfer the properties from a high level (quantum, DFT) description to a lower level (classical), where the electronic interaction is described by effective pairwise potentials between nuclei. This classical model is able to reproduce structural properties, such as radial distribution functions and also the vibrational density of states within good accuracy.

Due to the light mass of hydrogen, it is of interest to include nuclear quantum effects, even at room temperature. Using the FM model, PIMD simulations were performed which allowed to investigate the influence of the nuclear quantum effects while avoiding any double-counting which occurs in empirical water models. A softening of mostly intra-molecular structures was found, but almost no effect on the inter-molecular structures. This is in agreement with a number of studies [78, 98] (and contradicting some recent study by Morrone [130] et al. which found a considerable softening both intra- and inter-molecular structures).

In a last step, it was demonstrated that using the AdResS scheme it is possible to set up a stable water-vacuum interface where the interfacial region is treated with the full PIMD details and the bulk water described by a CG potential at the level of molecules. This suggest a route of including nuclear quantum effects in

"important" regions while treating the rest in a cheap effective manner.

7.4   Outlook

It was shown in this thesis that many of the problems occurring in the adaptive resolution simulations can be overcome and a coupling of regions employing different models at the desired thermodynamic boundary conditions is possible. The new H-AdResS scheme can be used for micro-canonical simulations and promises a fresh look at the problems occurring in adaptive simulations due to the availability of a Hamiltonian. This may also aid to understand coarse-grained models in general as the adaptive setup is very sensitive to features lost in the coarse-graining process.

The Gibbs- and Helmholtz-compensations also suggest a new way of calculating free energy differences in hybrid H-AdResS simulations. As the density imbalance is, in first approximation, related to the difference in chemical potential, the chemical potential difference can simply be read off the density profile. For a more accurate estimate one may use the thermodynamic force approach presented in chapter 2, within which the external potential which yields a flat density profile can be found. The difference between the end points of this potential then yields the difference in chemical potential of the two models. In principle, arbitrary models can be coupled and one may even consider using and ideal gas as a CG model, where the absolute free energy is available and therefore absolute free energies for arbitrary atomistic models can be calculated.

An efficient implementation of the AdResS scheme is now available in GROMACS release version 4.6 [1]. However, the efficiency can still be improved further, for which it is certainly favorable to design a MD code from scratch while having the difficulties of the adaptive schemes in mind. Such a design is currently explored for the Espresso++ software package [2].

An adaptive treatment could especially be of interest for including the nuclear quantum effects through PIMD. There, the computational speed-up can be significant due to the replacement of the imaginary time beads with a classical effective interaction, which is typically a factor of $P \approx 20 - 40$ faster. With this it would be possible to include the nuclear quantum effect in bio-molecular simulations, where larger macromolecules need to be simulated. This might, for example, be relevant for protein-ligand binding where only the binding region could be simulated with the full PIMD details. Along the same lines, the nuclear quantum effects at the water-vacuum interface deserve a more detailed study as well. For these applications, the development of models which do not double count the nuclear quantum effect are of great relevance. The FM water model presented in chapter 4 fulfills

---

[1] www.gromacs.org

[2] www.espresso-pp.de

this criterion, since it is based on a reference ab-initio simulation where the nuclei are treated classically.

Appendix A

Local pressure

The local pressure can be defined as the contribution to force $d\mathbf{F}$ across an infinitesimal area $d\mathbf{A}$, such that [113].

$$d\mathbf{F} = -p \, d\mathbf{A} \qquad (A.1)$$

There is some ambiguity in the convention which defines how to count fources "across" an area. Two different methods where considered here: The local virial pressure (valid in equilibrium) was used for the Hamiltonian AdResS in chapter 6, while in section 3.4 was calculated using the more general "method of the planes".

A.1 Local virial pressure

For pairwise interactions, the virial pressure is defined as [30]

$$p = \frac{1}{V} \left\langle \sum_i \frac{\mathbf{p}_i^2}{2m_i} + \sum_{\substack{i,j \\ i \neq j}} \mathbf{F}_{ij} \mathbf{r}_{ij} \right\rangle \qquad (A.2)$$

where V is the volume, $\mathbf{F}_{ij}$ the pairwise force between atom i and atom j. In order to measure this quantitiy locally, the simulation box is dived into thin slabs of width Δx along the axis of resolution change (x-axis). The pressure is the measure for a bin centered on x as

$$p(x) = \frac{1}{3A\Delta x} \left\langle \sum_{i \in \text{bin}} \frac{\mathbf{p}_i^2}{2m_i} + \sum_{\substack{(i \in \text{bin}),j \\ i \neq j}} \mathbf{F}_{ij} \mathbf{r}_{ij} \right\rangle \qquad (A.3)$$

where the sum runs over particles which are within a bin (denoted as $i \in \text{bin}$). This definiton of the pressure is equivalent to measureing the Irving-Kirkwood pressure for pairwise interacting, homogeneous systems [151].

A.2 Method of the planes

The local virial pressure is not valid for non-pairwise interactions or non-equilibrium situations. A more general expression for the local pressure tensor, which is also valid for inhomogeneous, non-equilibrium systems is given by the "Method of the planes" [113]. Thereby, the cross-sectional area is perpendicular to an axis of the simulation box.

Here, it was assumed that the pressure profile is calculated in x-direction. For the adaptive simulation in section 3.4 the pressure profile has to be measured based on molecular (instead of all-atom) force. This ensures that the pressure is well defined also in the hybrid and CG regions. The local pressure tensor then reads:

$$p_{\mu\nu}(x) = \frac{1}{A\Delta x} \left\langle \sum_{\alpha \in \text{bin}} (\mathbf{p}_\alpha \cdot \hat{\mathbf{e}}_\nu)(\mathbf{p}_\alpha \cdot \hat{\mathbf{e}}_\mu) \right\rangle + \frac{1}{A} \left\langle \sum_{\alpha=1}^{N} \mathbf{F}_\alpha \cdot \hat{\mathbf{e}}_\mu \text{sgn}(\mathbf{X}_\alpha \cdot \hat{\mathbf{e}}_x - x) \right\rangle$$

$$(A.4)$$

where A is a cross-sectional area perpendicular to the x-direction. The first term corresponds to the ideal gas contribution calculated in a small slab centered at x with thickness Δx, and the second represents the contribution of the forces. Both sums run over molecules, where $\mathbf{p}_\alpha$ denotes the momentum and $\mathbf{X}_\alpha$ is the position of molecule α.

For the adaptive simulation, this means that the ideal gas contribution of the center of mass is fully included in the sum regardless of the weighting function of the molecule, since those degrees of freedom fully contribute to the average in any representation. The internal degrees of freedom, however, were not considered. Note that the second sum runs over all molecules, however particles far away from the plane at x (at distances larger than the maximal cutoff of the non-bonded interactions) do not contribute to the sum.

Appendix B

GROMACS run parameters

The following section describes the AdResS implementation which was added to the GROMACS[1] software package and will be part of release version 4.6. Instructions on how to use the method were also added to the GROMACS manual. The parameters relevant for AdResS are as follows:

- `adress_type` defines the geometry of the atomistic zone, possibilities are `sphere`, `Xsplit`, `Const` or `Off`. `sphere` realizes the spherical all-atom cavity as described in chapter 4, `Xsplit` defines the linear resolution swith as used in chapter 3. The type `Const` sets constant value of the weight specified with `adress_const_wf` which can be used for debugging or calculating pure hybrid properties.

- `adress_ex_width` defines the width of the atomistic zone, a of eq. 2.73.

- `adress_hy_width` defines the width of the hybrid zone, $b - a$ of eq. 2.73.

- `adress_interface_correction` defines the type of correction to be applied in the hybrid zone, possibilities `thermoforce` and `no`.

- `adress_reference_coords` a vector defining the center of the atomistic zone.

- `adress_cg_grp_names` defines which energy group(s) should be treated as coarse-grained. At least one CG energy group has be defined using `energygrps` for the AdResS simulation.

As discussed in section 2.4, the all-atom details are kept everywhere in the simulation box. Only the non-bonded interactions are interpolated. In GROMACS, the "topology" therefore has to include both the all-atom and CG interactions (the CG interaction as a virtual site).

---

[1]www.gromacs.org

The virtual sites interact like normal atoms, but the forces are redistributed according to a predefined rule. In most cases, it is placed in the center of mass and therefore, the force from the CG integration is distributed onto the atoms according to:

$$\mathcal{F}_{\alpha_i,\beta} = \frac{m_i}{M_\alpha} \mathbf{F}^{CG}_{\alpha\beta} \; ,\tag{B.1}$$

where $\mathbf{F}^{CG}$ is the CG interaction force and m_i the mass of atom i of molecule α, which has the total mass M_α.

This is best illustrated on the simple example of an adaptive simulation of water, as performed in chapter 3. The topology file (.itp file) is:

```
[ molecule type ]
; molname       nrexcl
SOL             2
[ atoms ]
;   nr   type  resnr residue  atom   cgnr      charge        mass
     1     OW      1    SOL     OW      1        -0.82
     2     HW      1    SOL     HW1     1         0.41
     3     HW      1    SOL     HW2     1         0.41
     4     CG      1    SOL     CG      2         0
[ settles ]
; OW    funct   doh     dhh
1       1       0.1     0.16330
[ exclusions ]
1       2       3
2       1       3
3       1       2
[ virtual_sites3 ]
; Site from funct a d
4 1 2 3 1 0.05595E+00 0.05595E+00
```

The water molecule therefore is technically made up of oxygen, hydrogens and the virtual site of type 'CG' (with zero mass). The virtual site (atom nr 4) is constructed such that is placed in the center of mass. Here the special form for 3-particles was used 'virtual_sites3', in general (as for the toluene molecule in chapter 4), 'virtual_sitesn' can be used. The interactions for hydrogen and oxygen are taken from standard force-fields, but the CG model is a custom interaction and therefore needs to be defined. This is done in another topology (.top) file, for example:

```
#include "ffgmx.itp"
[ atomtypes ]
;name  mass        charge    ptype  C6            C12
 CG    0.00000     0.0000    V      0.000000E+00  1.000000E+00
#include "hybrid_spc.itp"
[ system ]
```

```
Adaptive water
[ molecules ]
SOL    8507
```

This defines an atomtype CG, where only the C12 coefficient (see eq. 2.22) is set to 1. The functional form of the C12 interaction is the read from a table (as explained in detail in the GROMACS manual).

The run parameter file (.mdp) then needs to define all relevant parameters for the AdResS simulation:

```
(...)
; AdResS relevant options
energygrps             = CG
energygrp_table        = CG CG

; Method for doing Van der Waals
vdw-type               = user

adress                 = yes
adress_type            = xsplit
adress_ex_width        = 1.5
adress_hy_width        = 1.5
adress_interface_correction = off
adress_reference_coords = 8 0 0
adress_cg_grp_names = CG
```

The energy groups are used for several technical reasons (all these procedures happen internally, without need for the user to interfere): First, it is required to define an energy group for each tabulated interaction potential (which is not the default one in table.xvg) in GROMACS. Secondly, also the neighbor-search (verlet list) is performed for each energy group separately. Then, for calculating the interactions different kernels (i.e. functions containing the loops over all pairs) can be called. This is of great advantage as therefore the distinction between all-atom and CG interaction can be made outside of the inner loop (and hence no `if` clause is necessary for each pair. Finally, the AdResS implementation uses the energy groups to disable interactions between the CG and all-atom interactions. For this, the energy groups representing each coarse-grained interaction need to be listed in `adress_cg_grp_names`.

For providing a thermodynamic force, all that needs to be done is to enable the corresponding option

```
adress_interface_correction=thermoforce
```

Then, GROMACS expects an input table file (default name 'tabletf.xvg'), which contains three columns:

1. r in nm

2. $V_{th}(r)$ in kJ/ mol

3. $-F_{th}(r)$ in kJ/(mol nm)

the variable r is thereby defined as the distance to the center of the all-atom zone (specified by 'adress_reference_coords'). In the simulation, the force is the applied to the virtual sites (and therefore to the entire molecule). The direction is given by the geometry chosen, where the unit vector in x-direction (`adress_type=xsplit`) or the radial vector for spherical (`adress_type=sphere`) geometry is used.

B.1 Thermodynamic force iteration

The thermodynamic force iteration (as described in section 3.3) was implemented in the VOTCA[2] software package (version 1.2), which can be used as a scripting engine to run a series of GROMACS (adaptive) simulations. The parameters are specified a .xml file, where first the thermodynamic force has to be selected as a method:

```
<method>tf</method>
```

For each interaction, the the geometry parameters are defined in the non-bonded section.

```
<non-bonded>
    <name>SOL</name>
    <min>1.4</min>
    <max>3.1</max>
  <step>0.01</step>
    <tf>
        <spline_start>0.9</spline_start>
        <spline_end>3.6</spline_end>
        <spline_step>0.4</spline_step>
        <molname>SOL</molname>
        <prefactor>0.01382</prefactor>
    </tf>
    <inverse>
        <target>dens.SOL.xvg</target>
        (...)
    </inverse>
</non-bonded>
```

---

[2]www.votca.org

The `min` and `max` properties control the range in which the gradient of the density profile is computed. The `step` variable defines the bin width of the thermodynamic force. For sake of stability, it has been proven useful to smooth the density profile with a spline interpolation. The parameters `spline_start,spline_end and spline_step` control the range of the spline and the width of the grid used for a spline fit (i.e. a least-squares procedure for finding the spline coefficients which describe the data best). The `prefactor` variable is the constant factor $\frac{1}{\rho \kappa_T^2}$ appearing in eq. 3.13, which contains the compressibility. The variable `target` contains the file name used for the density profile. The number of iterations is the specified (as for the IBI method) as

```
<iterations_max>60</iterations_max>
```

and many of the remaining options (GROMACS run input parameters) apply in the same way as for standard iterative Boltzmann inversion of VOTCA. For more detailed instructions see the VOTCA manual[3]. A tutorial demonstrating the thermodynamic force iteration is provided with the VOTCA software package.

---

[3]www.votca.org

APPENDIX B. GROMACS RUN PARAMETERS

Appendix C

Acknowledgements

APPENDIX C.   ACKNOWLEDGEMENTS

140

Bibliography

[1] S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer. Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir. *Physical Review Letters*, 108(17):170602, 2012.

[2] S. Fritsch, C. Junghans, and K. Kremer. Structure formation of toluene around c60: Implementation of the adaptive resolution scheme (AdResS) into GROMACS. *Journal of Chemical Theory and Computation*, 8(2):398–403, 2012.

[3] R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio. Hamiltonian adaptive resolution simulation for molecular liquids. *(manuscript under review)*.

[4] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2002.

[5] J. Baschnagel, K. Binder, P. Doruker, A. Gusev, O. Hahn, K. Kremer, W. Mattice, F. Müller-Plathe, M. Murat, W. Paul, S. Santos, U. Suter, and V. Tries. Bridging the gap between atomistic and coarse-grained models of polymers: Status and perspectives. In *Viscoelasticity, Atomistic Models, Statistical Chemistry*, volume 152 of *Advances in Polymer Science*, pages 41–156. Springer Berlin / Heidelberg, 2000.

[6] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: a moleculardynamics simulation. *The Journal of Chemical Physics*, 92(8):5057–5086, 1990.

[7] 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.

[8] D. Kennedy and C. Norman. What don't we know? *Science*, 309(5731):75–75, 2005.

[9] R. G. Parr and W. Yang. Density-functional theory of the electronic structure of molecules. *Annual Review of Physical Chemistry*, 46:701–728, 1995.

[10] 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):3730–3737, 1995.

[11] W. Tschöp, K. Kremer, J. Batoulis, T. Burger, and O. Hahn. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polymerica*, 49(2-3):61–74, 1998.

[12] D. Reith, M. Pütz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *Journal of Computational Chemistry*, 24(13):16241636, 2003.

[13] C. F. Abrams and K. Kremer. Combined coarse-grained and atomistic simulation of liquid bisphenol APolycarbonate: liquid packing and intramolecular structure. *Macromolecules*, 36(1):260–267, 2003.

[14] S. Izvekov and G. A. Voth. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B*, 109(7):2469–2473, 2005.

[15] J. W. Mullinax and W. G. Noid. Generalized yvon-born-green theory for molecular systems. *Physical Review Letters*, 103(19):198104, 2009.

[16] 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, 2009.

[17] R. E. Rudd and J. Q. Broughton. Coarse-grained molecular dynamics and the atomic limit of finite elements. *Physical Review B*, 58(10):R5893–R5896, 1998.

[18] 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.

[19] G. Csnyi, T. Albaret, M. C. Payne, and A. De Vita. Learn on the fly: A hybrid classical and quantum-mechanical molecular dynamics simulation. *Physical Review Letters*, 93(17):175503, 2004.

[20] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, 103(2):227–249, 1976.

[21] J. Gao. Methods and applications of combined quantum mechanical and molecular mechanical potentials. In K. B. Lipkowitz and D. B. Boyd, editors, *Reviews in Computational Chemistry*, page 119185. John Wiley & Sons, Inc., 2007.

[22] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *Journal of Chemical Physics*, 123:4106, 2005.

[23] 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.

[24] 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.

[25] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu, and D. E. Shaw. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, 41(1):429–452, 2012.

[26] J.-P. Hansen and I. R. McDonald. *Theory of simple liquids*. Academic Press, 2006.

[27] H. Goldstein and C. Poole. *Classical Mechanics*. Addison-Wesley, San Francisci [u.a.], 2002.

[28] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, 1982.

[29] M. P. Allen and D. J. Tildesley. *Computer simulation of liquids*, volume 57 of *Oxford science publications*. Oxford University Press, 1989.

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

[31] L. E. Reichl. *A Modern Course in Statistical Physics, 2nd Edition*. Wiley-VCH, 2004.

[32] B. Dünweg. Langevin methods. In B. Dünweg, D. P. Landau, and A. I. Milchev, editors, *Computer Simulations of Surfaces and Interfaces*, volume 114 of *NATO Science Series*, pages 77–92. Springer Netherlands, 2003.

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

[34] P. H. Hünenberger. Thermostat algorithms for molecular dynamics simulations. In C. Dr. Holm and K. Prof. Dr. Kremer, editors, *Advanced Computer Simulation*, volume 173 of *Advances in Polymer Science*, pages 130–130. Springer Berlin / Heidelberg, 2005.

BIBLIOGRAPHY

[35] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics*, 72:2384–2393, 1980.

[36] H. Berendsen. Transport properties computed by linear response through weak coupling to a bath. in: Computer simulations in material science. pages 139–155. 1991.

[37] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *Journal of Chemical Physics*, 126:4101, 2007.

[38] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52:255–268, 1984.

[39] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

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

[41] 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):046702, 2003.

[42] J.-P. 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*, 23:327, 1977.

[43] 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):14631472, 1997.

[44] 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):952962, 1992.

[45] D. Marx and H. Jürg. *Modern Methods and Algorithms of Quantum Chemistry*, volume 3 of *NIC series*. John von Neumann Institute for Computing, Jülich, 2000.

[46] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

144

[47] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187217, 1983.

[48] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995.

[49] J. Barker and R. Watts. Monte Carlo studies of the dielectric properties of water-like models. *Molecular Physics*, 26(3):789–792, 1973.

[50] O. Steinhauser. Reaction field simulation of water. *Molecular Physics*, 45(2):335–348, 1982.

[51] L. Onsager. Electric moments of molecules in liquids. *Journal of the American Chemical Society*, 58(8):1486–1493, 1936.

[52] B. Dünweg and A. Ladd. Lattice boltzmann simulations of soft matter systems. In C. Holm and K. Kremer, editors, *Advanced Computer Simulation Approaches for Soft Matter Sciences III*, volume 221 of *Advances in Polymer Science*, pages 89–166. Springer Berlin / Heidelberg, 2009.

[53] 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, 2009.

[54] 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*, 28(2):221–229, 2009.

[55] W. G. Noid, J.-W. 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.

[56] 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.

[57] 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.

BIBLIOGRAPHY

[58] D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, 1987.

[59] R. L. Henderson. A uniqueness theorem for fluid pair correlation functions. *Physics Letters A*, 49:197–198, 1974.

[60] F. Ercolessi and J. B. Adams. Interatomic potentials from first-principles calculations: The force-matching method. *Europhysics Letters (EPL)*, 26(8):583–588, 1994.

[61] S. Izvekov and G. A. Voth. Multiscale coarse graining of liquid-state systems. *Journal of Chemical Physics*, 123:4105, 2005.

[62] W. G. Noid, P. Liu, Y. Wang, J.-W. 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.

[63] 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, 2004.

[64] W. G. Noid, J.-W. 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. (PMID: 17394308).

[65] J. W. Mullinax and W. G. Noid. A generalized-YvonBornGreen theory for determining coarse-grained interaction potentials. *The Journal of Physical Chemistry C*, 114(12):5661–5674, 2010.

[66] C. R. Ellis, J. F. Rudzinski, and W. G. Noid. GeneralizedYvonBornGreen model of toluene. *Macromolecular Theory and Simulations*, 20(7):478–495, 2011.

[67] H. M. Cho and J.-W. Chu. Inversion of radial distribution functions to pair forces by solving the YvonBornGreen equation iteratively. *The Journal of Chemical Physics*, 131(13):134107, 2009.

[68] D. Fritz, C. R. Herbers, K. Kremer, and N. F. A. v. d. Vegt. Hierarchical modeling of polymer permeation. *Soft Matter*, 5(22):4556–4563, 2009.

[69] D. Fritz, K. Koschke, V. A. Harmandaris, N. F. A. van der Vegt, and K. Kremer. Multiscale modeling of soft matter: scaling of dynamics. *Physical Chemistry Chemical Physics*, 13(22):10412, 2011.

146

[70] M. Praprotnik, L. D. 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.

[71] 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):066701, 2006.

[72] 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.

[73] M. Praprotnik, K. Kremer, and L. D. 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.

[74] M. Praprotnik, S. Poblete, L. Delle Site, and K. Kremer. Comment on Adaptive multiscale molecular dynamics of macromolecular fluids. *Physical Review Letters*, 107(9):099801, 2011.

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

[76] 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.

[77] J. J. Sakurai. *Modern quantum mechanics*. Addison-Wesley, 1985.

[78] H. A. Stern and B. J. Berne. Quantum effects in liquid water: Path-integral simulations of a flexible and polarizable ab initio model. *Journal of Chemical Physics*, 115:7622–7628, 2001.

[79] T. F. Miller and D. E. Manolopoulos. Quantum diffusion in liquid water from ring polymer molecular dynamics. *The Journal of Chemical Physics*, 123(15):154504–154504–10, 2005.

[80] F. Paesani and G. A. Voth. The properties of water: Insights from quantum simulations. *The Journal of Physical Chemistry B*, 113(17):5702–5719, 2009.

[81] R. P. Feynman. Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics*, 20:367–387, 1948.

BIBLIOGRAPHY

[82] R. P. Feynman and A. R. Hibbs. *Quantum mechanics and path integrals.* McGraw-Hill, New York, NY, 1965.

[83] H. F. Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959. (ArticleType: research-article / Full publication date: Aug., 1959 / Copyright 1959 American Mathematical Society).

[84] D. M. Ceperley. Path integrals in the theory of condensed helium. *Reviews of Modern Physics*, 67(2):279–355, 1995.

[85] A. B. Poma and L. Delle Site. Classical to path-integral adaptive resolution in molecular simulation: Towards a smooth quantum-classical coupling. *Physical Review Letters*, 104(25):250201, 2010.

[86] A. B. Poma and L. D. Site. Adaptive resolution simulation of liquid parahydrogen: testing the robustness of the quantum-classical adaptive coupling. *Physical Chemistry Chemical Physics*, 13(22):10510–10519, 2011.

[87] R. Potestio and L. Delle Site. Quantum locality and equilibrium properties in low-temperature parahydrogen: A multiscale simulation study. *The Journal of Chemical Physics*, 136(5):054101, 2012.

[88] R. W. Hall and B. J. Berne. Nonergodicity in path integral molecular dynamics. *Journal of Chemical Physics*, 81:3641–3643, 1984.

[89] M. E. Tuckerman, B. J. Berne, G. J. Martyna, and M. L. Klein. Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals. *Journal of Chemical Physics*, 99:2796–2808, 1993.

[90] R. Valladares, A. Fisher, and W. Hayes. Path-integral simulations of zero-point effects for implanted muons in benzene. *Chemical Physics Letters*, 242(12):1–6, 1995.

[91] M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello. Efficient and general algorithms for path integral CarParrinello molecular dynamics. *The Journal of Chemical Physics*, 104(14):5579–5588, 1996.

[92] F. Paesani, W. Zhang, D. A. Case, T. E. Cheatham, and G. A. Voth. An accurate and simple quantum model for liquid water. *Journal of Chemical Physics*, 125:4507, 2006.

[93] M. F. Herman, E. J. Bruskin, and B. J. Berne. On path integral Monte Carlo simulations. *The Journal of Chemical Physics*, 76(10):5150–5155, 1982.

[94] G. Norman and V. Filinov. Investigations of phase transitions by a Monte Carlo method. *High Temp.(USSR)*, 7:216—222, 1969.

[95] D. Adams. Chemical potential of hard-sphere fluids by Monte Carlo methods. *Molecular Physics*, 28(5):1241–1252, 1974.

[96] D. Adams. Grand canonical ensemble Monte Carlo for a lennard-jones fluid. *Molecular Physics*, 29(1):307–311, 1975.

[97] G. S. Heffelfinger and F. v. Swol. Diffusion in LennardJones fluids using dual control volume grand canonical molecular dynamics simulation (DCVGCMD). *The Journal of Chemical Physics*, 100(10):7548–7552, 1994.

[98] B. Chen, I. Ivanov, M. L. Klein, and M. Parrinello. Hydrogen bonding in water. *Physical Review Letters*, 91(21):215503, 2003.

[99] P. G. Kusalik and I. M. Svishchev. The spatial structure in liquid water. *Science*, 265(5176):1219–1221, 1994.

[100] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253287, 1921.

[101] H. J. C. Berendsen, J. Postma, W. Van Gunsteren, and J. Hermans. Intermolecular forces. page 331. 1981.

[102] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987.

[103] 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.

[104] 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.

[105] J. R. Errington and P. G. Debenedetti. Relationship between structural order and the anomalies of liquid water. *Nature*, 409(6818):318–321, 2001.

[106] D. Mukherji, N. F. A. van der Vegt, K. Kremer, and L. Delle Site. Kirkwood-Buff analysis of liquid mixtures in an open boundary simulation. *Journal of Chemical Theory and Computation*, 8(2):375–379, 2012.

[107] D. Mukherji, N. F. A. van der Vegt, and K. Kremer. Preferential solvation of tri-glycine in aqueous urea: An open boundary simulation approach. *Journal of Chemical Theory and Computation*, 2012.

BIBLIOGRAPHY

[108] 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.

[109] A. Villa, C. Peter, and N. F. A. van der Vegt. Transferability of nonbonded interaction potentials for coarse-grained simulations: Benzene in water. *Journal of Chemical Theory and Computation*, 6(8):2434–2444, 2010.

[110] M. Praprotnik, S. Matysiak, L. D. Site, K. Kremer, and C. Clementi. Adaptive resolution simulation of liquid water. *Journal of Physics: Condensed Matter*, 19(29):292201, 2007.

[111] B. Widom. Some topics in the theory of fluids. *Journal of Chemical Physics*, 39:2808–2812, 1963.

[112] N. Köckler, Hans Rudolf Schwarz and N. Köckler. *Numerische Mathematik.* Springer DE, 2009.

[113] B. D. Todd, D. J. Evans, and P. J. Daivis. Pressure tensor for inhomogeneous fluids. *Physical Review E*, 52(2):1627–1638, 1995.

[114] R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik. Concurrent triple-scale simulation of molecular liquids. *The Journal of Chemical Physics*, 128(11):114110, 2008.

[115] 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.

[116] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley. C60: Buckminsterfullerene. *Nature*, 318(6042):162–163, 1985.

[117] D. Wynands, M. Levichkova, K. Leo, C. Uhrich, G. Schwartz, D. Hildebrandt, M. Pfeiffer, and M. Riede. Increase in internal quantum efficiency in small molecular oligothiophene: C[sub 60] mixed heterojunction solar cells by substrate heating. *Applied Physics Letters*, 97(7):073503, 2010.

[118] M. M. Villiers, P. Aramwit, and G. S. Kwon, editors. *Nanotechnology in Drug Delivery.* Springer New York, New York, NY, 2009.

[119] 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.

[120] W. L. Jorgensen, E. R. Laird, T. B. Nguyen, and J. Tirado-Rives. Monte Carlo simulations of pure liquid substituted benzenes with OPLS potential functions. *Journal of Computational Chemistry*, 14(2):206–215, 1993.

[121] W. L. Jorgensen and T. B. Nguyen. Monte Carlo simulations of the hydration of substituted benzenes with OPLS potential functions. *Journal of Computational Chemistry*, 14(2):195–205, 1993.

[122] U. Essmann, L. Perera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

[123] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182, 1981.

[124] 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.

[125] M. J. Stephen and J. P. Straley. Physics of liquid crystals. *Reviews of Modern Physics*, 46(4):617–704, 1974.

[126] T.-H. Wei, T.-H. Huang, T.-T. Wu, P.-C. Tsai, and M.-S. Lin. Studies of nonlinear absorption and refraction in c60/toluene solution. *Chemical Physics Letters*, 318(1-3):53–57, 2000.

[127] B. Ginzburg, S. Tuchiev, S. Tabarov, and A. Shepelevski. Structurization of a solvent interacting with fullerene c60. *Crystallography Reports*, 50(5):735–738, 2005.

[128] R. A. Kuharski and P. J. Rossky. A quantum mechanical study of structure in liquid H2O and D2O. *The Journal of Chemical Physics*, 82(11):5164–5177, 1985.

[129] S. Habershon, T. E. Markland, and D. E. Manolopoulos. Competing quantum effects in the dynamics of a flexible water model. *The Journal of Chemical Physics*, 131(2):024501, 2009.

[130] J. A. Morrone and R. Car. Nuclear quantum effects in water. *Physical Review Letters*, 101(1):017801, 2008.

[131] 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. (PMID: 16422607).

[132] J. VandeVondele and J. Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics*, 127(11):114105–114105–9, 2007.

BIBLIOGRAPHY

[133] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865–3868, 1996.

[134] C. Zhang, J. Wu, G. Galli, and F. Gygi. Structural and vibrational properties of liquid water from van der Waals density functionals. *Journal of Chemical Theory and Computation*, 7(10):3054–3061, 2011.

[135] T. D. Kuhne, M. Krack, and M. Parrinello. Static and dynamical properties of liquid water from first principles by a novel CarParrinello-like approach. *Journal of Chemical Theory and Computation*, 5(2):235–241, 2009.

[136] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Reviews of Modern Physics*, 64(4):1045–1097, 1992.

[137] S. Y. Venyaminov and F. G. Prendergast. Water (H2O and D2O) molar absorptivity in the 10004000 cm1Range and quantitative infrared spectroscopy of aqueous solutions. *Analytical Biochemistry*, 248(2):234–245, 1997.

[138] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *The Journal of Chemical Physics*, 123(23):234505–234505–12, 2005.

[139] H. Saint-Martin, J. Hernndez-Cobos, M. I. Bernal-Uruchurtu, I. Ortega-Blake, and H. J. C. Berendsen. A mobile charge densities in harmonic oscillators (MCDHO) molecular model for numerical simulations: The waterwater interaction. *The Journal of Chemical Physics*, 113(24):10899–10912, 2000.

[140] G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas. A quantitative account of quantum effects in liquid water. *Journal of Chemical Physics*, 125:1102, 2006.

[141] F. Paesani, S. Iuchi, and G. A. Voth. Quantum effects in liquid water from an ab initio-based polarizable force field. *Journal of Chemical Physics*, 127:4506, 2007.

[142] G. S. Fanourgakis and S. S. Xantheas. Development of transferable interaction potentials for water. v. extension of the flexible, polarizable, thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *The Journal of Chemical Physics*, 128(7):074506–074506–11, 2008.

[143] R. Car and M. Parrinello. Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, 55(22):2471–2474, 1985.

152

[144] A. Soper. The radial distribution functions of water and ice from 220 to 673 k and at pressures up to 400 MPa. *Chemical Physics*, 258(23):121–137, 2000.

[145] Y. A. Mantz, B. Chen, and G. J. Martyna. Structural correlations and motifs in liquid water at selected temperatures: ab initio and empirical model predictions. *The Journal of Physical Chemistry B*, 110(8):3540–3554, 2006.

[146] S. Yoo, X. C. Zeng, and S. S. Xantheas. On the phase diagram of water with density functional theory potentials: The melting temperature of ice ih with the perdew-burke-ernzerhof and becke-lee-yang-parr functionals. *Journal of Chemical Physics*, 130:1102, 2009.

[147] S. Yoo and S. S. Xantheas. Communication: The effect of dispersion corrections on the melting temperature of liquid water. *The Journal of Chemical Physics*, 134(12):121105, 2011.

[148] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *Journal of Chemical Physics*, 54:5237–5247, 1971.

[149] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *Journal of Chemical Physics*, 3:300–313, 1935.

[150] P. Kollman. Free energy calculations: Applications to chemical and biochemical phenomena. *Chemical Reviews*, 93(7):2395–2417, 1993.

[151] J. H. Irving and J. G. Kirkwood. The statistical mechanical theory of transport processes. IV. The equations of hydrodynamics. *The Journal of Chemical Physics*, 18(6):817–829, 1950.

BIBLIOGRAPHY

154

Appendix D

Curriculum vitae

Personal Information

Name	Sebastian Fritsch
Date of birth	29. 6. 1983
Place of birth	Speyer
Nationality	German
Address	Hinter der Kapelle 46
	55128 Mainz
	fritsch@mpip-mainz.mpg.de

Education

September 2009 – today	PhD student,
	Max Planck Institute for Polymer Research, Mainz, Germany
July 2009	Diploma in Physics,
	University of Heidelberg, Germany
March 2003	Abitur,
	Gymnasium am Kaiserdom, Speyer

Publications

1. "Ion Selectivity Mechanism in a Bacterial Pentameric Ligand- Gated Ion Channel"
   S. Fritsch, Ivaylo Ivanov, Hailong Wang and Xiaolin Cheng
   Biophys J. 2011 100(2) 390-398

2. "Adaptive resolution molecular dynamics simulation through coupling to an internal particle reservoir"
   S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site and K. Kremer
   Physical Review Letters, 108, 17, 170602, 2012

3. "Structure Formation of Toluene around C60: Implementation of the Adaptive Resolution Scheme (AdResS) into GROMACS"
   S. Fritsch, C. Junghans, and K. Kremer
   J. Chem. Theory Comput. 2012, 8 (2), pp 398-403

4. "Hamiltonian adaptive resolution simulation for molecular liquids"
   R. Potestio, **S. Fritsch**, P. Espanol, R. Delgado-Buscalioni, K. Kremer, R. Everaers and D. Donadio (manuscript under review)