Development and Application of Hamiltonian Adaptive Resolution Simulations for Systems having Long-range Interactions

Max Planck Graduate Center
mit der Johannes Gutenberg-Universität

Maziar Heidari

Dissertation
zur Erlangung des Grades eines
"Doktor rerum naturalium (Dr. rer. nat.)"
der Fachbereiche:
08 - Physik, Mathematik und Informatik
09 - Chemie, Pharmazie, Geographie und Geowissenschaften
10 - Biologie,
Universitätsmedizin
der Johannes-Gutenberg-Universität Mainz

Angefertigt am Max-Planck-Institut für Polymerforschung

Mainz, 22. Januar 2019

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

INCLUDED PUBLICATIONS

This PhD thesis is the compilation of the partial results of the doctoral studies that I have carried out at the Max Planck Institute for Polymer Research in Mainz since April 2015. My studies were conducted within the framework of the Max Planck Graduate Center in Mainz. The results have been published in several peer-reviewed scientific papers and the thesis has been submitted to the Johannes Gutenberg University of Mainz as a cumulative dissertation (thesis by publication). As listed in the appended section, the reprinted publications are included in the chapters 2, 3, 4 and 6; in chapter 5, I have included a draft that, at the time of submission of the present thesis, was under review for publication in Physical Review Letters. The contributions of the different authors are described in the *Contribution* section.

Chapter 2:
Maziar Heidari, Robinson Cortes-Huerto, Davide Donadio and Raffaello Potestio
Accurate and general treatment of electrostatics in Hamiltonian Adaptive Resolution Simulation
Eur. Phys. J. Spec. Top. (2016) 225: 1505
DOI: 10.1140/epjst/e2016-60151-6

Chapter 3:
Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio
Spatially Resolved Thermodynamic Integration: An Efficient Method To Compute Chemical Potentials of Dense Fluids
J. Chem. Theory Comput., 2018, 14 (7), pp 34093417
DOI: 10.1021/acs.jctc.8b00002

Chapter 4:
Maziar Heidari, Robinson Cortes-Huerto, Kurt Kremer, and Raffaello Potestio
Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations
Eur. Phys. J. E (2018) 41: 64.
DOI: 10.1140/epje/i2018-11675-x
This work is highlighted by EPJE news portal [1]

Chapter 5:
Maziar Heidari, Robinson Cortes-Huerto, Raffaello Potestio and Kurt Kremer
Nonequilibrium work relations in multiscale simulations of molecular liquids
Submitted to Phys. Rev. Lett.

Chapter 6:
Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto
Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: revisiting the spatial block analysis method
Entropy 2018, 20(4), 222
DOI: 10.3390/e20040222

ABSTRACT

Computer simulations have proven to be a powerful tool in soft matter research since they have helped to elucidate microscopic details of many phenomena observed in experiments that would otherwise have remained unclear. Therefore, the high demand for computer simulations on one hand, and the emergence of very fast computational units on the other hand, have led to development of a great variety of computational methods. These techniques have provided the possibility to investigate phenomena occurring within a wide range of length and time scales, from chemical reactions at the quantum scale to self-assembly at the macroscale.

However, the computational costs of studying these phenomena in a single, highly detailed resolution are often too expensive. Hence, provided the locality of the phenomenon, it is advantageous to develop multi-resolution techniques. In these approaches, the system is divided into a high resolution subregion, described by an accurate but computationally expensive model, and a low resolution region, where the rest of the system is treated by means of a coarse but computationally efficient model. One of such multi-resolution techniques is the Hamiltonian Adaptive Resolution Simulations (`H-AdResS`) method. In this approach, the two resolutions are smoothly coupled through a transition layer in which compensating forces are applied on the molecules, and a constant chemical potential throughout the resolutions is enforced.

In this work, we first explain the challenges of implementing long-ranged electrostatic interactions in `H-AdResS`. We then propose and validate the usage of a short-range modification of Coulomb potential, the Damped Shifted Force model, in the context of the `H-AdResS` scheme. We validate this approach by reproducing the structural and dynamical properties of liquid water. Next, we take advantage of the constant chemical potential inherent to `H-AdResS` to introduce a new and efficient method to compute the chemical potential of liquids and mixtures. The method has been named spatially resolved thermodynamic integration (`SPARTIAN`). Subsequently, we employ the same approach to compute the free energy of solids by coupling the real crystals with their corresponding ideal Einstein crystals. Afterwards, we use the Jarzynski equality to obtain the solvation free energy of molecules by using steered molecular dynamics to pull the molecule from the atomistic (solvated state) into the ideal gas (unsolvated state) region. Lastly, we discuss the spatial block analysis (`SBA`) method to efficiently extrapolate thermodynamic quantities such as bulk isothermal compressibility from finite-size computer simulations, and discuss different types of finite-size effects in the `SBA` context. This study is designed to target the problems involving sampling in grand canonical ensembles which is also crucial for the extension and development of the `SPARTIAN` method into a new grand canonical molecular dynamics framework.

Zusammenfassung

Computersimulationen haben sich als leistungsfähiges Werkzeug zur Erforschung weicher Materie erwiesen, da sie geholfen haben zur Aufdeckung mikroskopischer Details vieler experimentell beobachteter Phänomene beizutragen, welche sonst unaufgeklärt geblieben wären. Demzufolge haben sowohl der hohe Bedarf an Computersimulationen, als auch die Entwicklung sehr schneller Computer zur Entwicklung einer großen Vielfalt von Simulationsmethoden geführt. Diese Simulationstechniken haben die Möglichkeit geschaffen Phänomene zu untersuchen, welche auf einer großen Bandbreite von Längen und Zeitskalen geschehen, angefangen bei chemischen Reaktionen auf der Quantenskala bis hin zur Selbstanordnung auf der Makroskala.

Jedoch ist der Rechenaufwand, um diese Phänomene in einer einzigen hochdetaillierten Auflösung zu studieren oft zu groß. Daher ist es von Vorteil Simulationstechniken mit mehreren Auflösungen zu entwickeln, sofern die räumliche Abgegrenztheit des Phänomens gegeben ist. In diesen Methoden wird das System in einen hochaufgelösten Teilbereich, beschrieben durch ein genaues aber rechenaufwendiges Modell, und einen niedrigaufgelösten Teilbereich, beschrieben durch ein vergröbertes recheneffizientes Modell aufgeteilt, welcher den Rest des Systems beschreibt. Eine dieser Methoden mit mehreren Auflösungen ist die Methode der hamiltonisch adaptiv aufgelösten Simulationen (Hamiltonian Adaptive Resolution Simulations) (`H-AdResS`). In dieser Herangehensweise werden die beiden Teilbereiche verschiedener Auflösung glatt durch eine Übergangsschicht miteinander gekoppelt, in welcher Kompensationskräfte auf die Moleküle wirken. Dadurch wird ein durchgehend konstantes chemisches Potential erzwungen.

In dieser Arbeit beschreiben wir als erstes die Herausforderungen, langreichweitige elektrostatische Wechselwirkungen in `H-AdResS` zu implementierchemen. Dann schlagen wir die Verwendung einer kurzreichweitigen Modifikation des Coulomb Potentials vor und validieren diese im Kontext des `H-AdResS` Schemas: das gedämpfte verschobene Kraft-Modell (Damped Shifted Force). Wir validieren diese Herangehensweise durch Reproduktion der strukturellen und dynamischen Eigenschaften von flüssigem Wasser. Als Nächstes benutzen wir das durchgehend konstante chemische Potential von `H-AdResS`, um eine neue und effiziente Methode einzufüren das chemische Potential von Flüssigkeiten und Mischungen zu berechnen. Diese Methode wurde räumlich aufgelöste thermodynamische Integration (spatially resolved thermodynamic integration) (`SPARTIAN`) genannt. Anschließend wenden wir die gleiche Herangehensweise an, um die freie Energie von Festkörpern zu berechnen, in dem wir realistische Kristallmodelle mit den korrespondierenden idealen Einstein-Kristallen koppeln. Danach verwenden wir die Jarzynski-Gleichung, um die freie Lösungsenergie von Molekülen mittels Molekulardynamiksimulationen zu berechnen, in denen Moleküle gezielt vom atomistischen Teilbereich (gelöster Zustand) in den Teilbereich des ide-

alen Gases (ungelöster Zustand) gezogen werden. Schließlich diskutieren wir die räumlich aufgelöste Blockanalyse-Methode (spatial block analysis method) (`SBA`), um effizient thermodynamische Eigenschaften wie die isothermische Kompressibilität aus Computersimulationen von Systemen endlicher Größe zu extrapolieren. Wir diskutieren verschiedene Arten von Effekten endlicher Systemgrößen im Zusammenhang der `SBA`-Methode. Dieser Teil der Arbeit ist daraufhin konzipiert, Sampling-Probleme großkanonischer Ensembles anzugehen, was entscheidend ist für die Weiterentwicklung der `SPARTIAN`-Methode im Rahmen großkanonischer Molekulardynamik-Simulationen.

1. Introduction

1.1 Overview

It is difficult to imagine any modern scientific endeavour where computer simulations do not play a pivotal role: from assisting and sometimes guiding experimental investigations to identifying and predicting new materials and physical principles. The paradigm shift associated to the use of computational models for the study of complex systems and phenomena started more than fifty years ago [2]. The modern Monte Carlo (MC) method was invented in the late Forties, and almost ten years later the first molecular dynamics (MD) simulations were performed. In the early Fifties, Enrico Fermi, John Pasta, Stanislaw Ulam and Mary Tsingou used the Los Alamos computer MINIAC I to investigate the thermalization of a crystal. To this end, a vibrating string has been modeled as a chain of particles interacting via both linear and weak nonlinear terms. A first analysis of the problem using the equipartition theorem suggested that the energy of an initial normal mode is distributed among other modes due to the presence of the nonlinear interaction. Surprisingly, they observed that the system exhibits a quasi-periodic behavior where, after some time, the initial state was seemingly recovered [3, 4]. The phenomenon was later dubbed as *Fermi-Pasta-Ulam-Tsingou* problem and it has been a center of focus in the field of nonlinear and chaotic dynamics [5, 6].

The insight provided by this early computer experiments sparked the interest of the physics community. Now a useful tool was available to investigate the connection between the microscopic components and interactions of a system and its thermodynamic properties as established in the foundations of statistical mechanics. Moreover, the study of the dynamic evolution of a physical system was now a real possibility. A pioneering work by Berni Alder using MD simulations of a small system (~ 100 particles) of hard spheres, within the possibilities of the UNIVAC computer, demonstrated the feasibility of such a method by comparing the obtained equation of state of the system with previous MC results [7]. Consequently, accompanied by the emergence of new and efficient algorithms and the boost in computing power, scientists have been able to study large-scale and complex systems, leading to significant achievements in both fundamental and applied studies. Further examples include the discovery of new and efficient organic electronics materials [8, 9], the assesment of different chemical compounds useful for drug delivery applications [10], the characterization of protein functions [11, 12] or structural and dynamical properties of polymers in polymer melts [13, 14].

As it is shown in Figure 1.1, depending on the time and length scales of the phenomena under examination, the approaches and methods are classified differently. At the electronic scale, *ab initio* methods incorporate the quantum mechanical effects by solving the Kohn-Sham equations [15,16]. At the atomic scale where the dynamics of the system is described by length and time scales much larger than the ones describing electronic degrees of freedom, the molecular dynamics simulation method is used. In this approach, Newton's equations of the motion are solved by incorporating classical force fields frequently obtained from *ab initio* calculations [2,17]. At the mesoscopic length and time scale, coarse-grained molecular descriptions and force fields are employed [18,19]. Here, molecular internal degrees of freedom that do not significantly affect thermodynamic quantities of interest are integrated out. And ultimately, at the largest scale, the continuum level, the dynamics of the system is described by field equations which are attained by the local but macro-scale conservation of quantities. For instance, Navier-Stokes equation is the field approach toward describing the fluid flow using the conservation of local mass and momentum [20]. Other approaches include the theory of elasticity and Finite Element method used to study the continuum mechanics of solids [21]. In the following, I will discuss in further detail molecular dynamics methods since my research interests revolve around such a method.

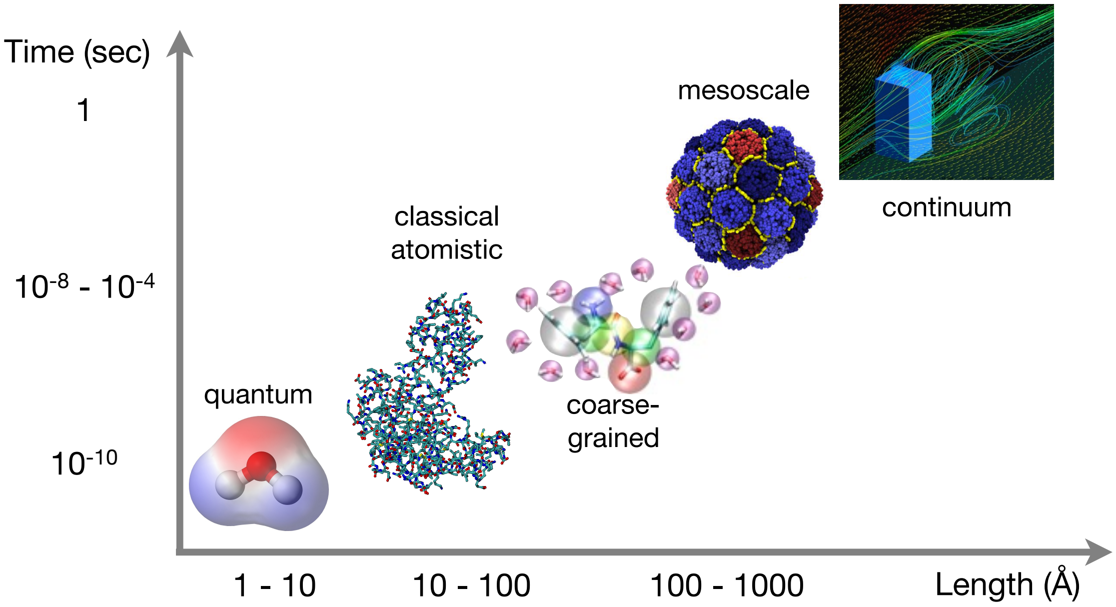


Figure 1.1. Multiscale nature of matter. The phenomena are described by different approaches depending on length and time scales of interest.

1.1.1 Atomistic molecular dynamics simulations

For the study of properties at length scales larger than those of electronic and nuclear structures, one can assume that quantum mechanical effects have negligible effects on the dynamics of the system. In this limit, one is allowed to use classical MD simulations in which the quantum mechanical effects are implicitly taken into account. In this approach, the atoms' nuclei are described by point-like particles that interact via classical force fields. These force fields are fitted to the data obtained from either experiments or *ab initio* methods. Then, the dynamics of the system is evolved by integrating Newton's equation of motion over successive finite time steps [2,17]. The mostly used integration scheme is the Velocity Verlet algorithm which has been proven to be symplectic (time evolution follows Hamiltonian's equations) and energy conserving [22]. Since in standard MD the integrator of the equations of motion is energy-conserving, the statistical ensemble of the system corresponds to the microcanonical ensemble. It is also possible to employ different statistical ensembles such as canonical or isobaric-isothermal ensembles [23–26]. In the canonical ensemble, the number of particles and the volume of the system are fixed, and to control the temperature of the system several strategies have been proposed [25–28]. In these approaches, the temperature of the system can be maintained constant locally or globally by coupling to a heat reservoir. In the former approach, each degree of freedom can be thermalized using Langevin or generalized Langevin equations [17,28] or by rescaling the particles' velocities [26]. In the latter, a new fictitious degree of freedom is introduced such that its dynamics controls the system temperature [25,27].

1.1.2 Coarse-grained models

Atomistic molecular simulations have provided profound and broad insights into molecular-scale phenomena. However, the length and time scales of many processes occurring in liquids and biomolecular systems are still far beyond current computational capabilities. To overcome such a difficulty, it is required to develop alternative approaches allowing one to access larger time and length scales. Coarse-grained (CG) simulation is one possible approach aiming at solving this problem. In such an approach, a CG representation of the system is introduced by clustering groups of atoms into CG sites. Figure 1.2 shows the molecular structures of water and urea molecules and their corresponding coarse-grained single particle representations. In the new CG scale, the molecular degrees of freedom are reduced and the CG sites interact through more computationally efficient interactions thus allowing access to large spatial and temporal scales. CG simulations have shown to be efficient and effective approaches for bridging the gap between the atomistic and mesoscopic scales as it has been demonstrated for molecular liquids [29], polymer elastic networks [30,31], lipid membranes [32,33] and other biomolecular systems [34,35].

In the context of statistical mechanics, the relation bridging the atomistic and the effective CG interactions is generally introduced by the following formula [36]

$$\begin{aligned} \exp(-F/k_BT) &= C_1 \int d\mathbf{x}^{AT} \exp\left[-V^{AT}(\mathbf{x}^{AT})/k_BT\right] \\ &= C_2 \int d\mathbf{x}^{CG} \exp\left[-V^{CG}(\mathbf{x}^{CG})/k_BT\right] \end{aligned} \tag{1.1}$$

Here, F represents the Helmholtz free energy, $V(\mathbf{x}^{AT})$ is the interaction potential at the atomistic level which is a function of the atomistic coordinates, $\mathbf{x}^{AT}$. T is the temperature of the system and k_B is the Boltzmann constant. C_1 and C_2 are constants. As it is expressed in the second line of equation 1.1, the effective CG potential $V(\mathbf{x}^{CG})$ is obtained by equating the free energy of the system at atomistic and coarse-grained levels. The procedure of finding CG interaction is not straightforward, and several different techniques have been proposed. The most commonly used coarse-graining techniques are Force Matching [37, 38], Reverse Monte Carlo [39, 40] and Relative Entropy [18, 41].

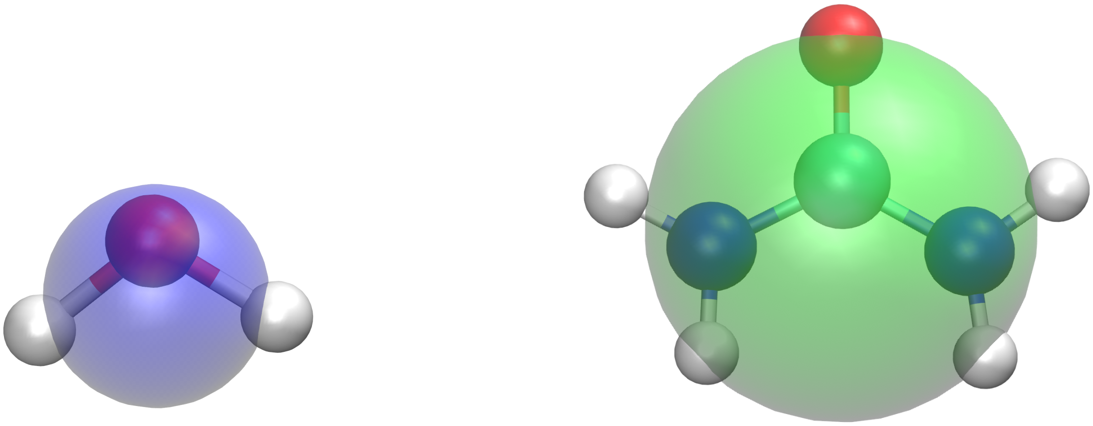


Figure 1.2. Molecular structures of water (left) and urea (right) molecules. The corresponding coarse-grained models are shown with transparent single beads.

1.1.3 Dual-resolution simulations

Although using a coarse-graining technique makes it possible to characterize the relevant properties of a system at a cheaper computational costs, it is not able to answer the questions related to the systems in which the chemical details of a small region have major effects on the system behavior. For example, the active site of a large enzyme or the electrostatic screening of ions by water molecules close to the charged residues of a protein. In these cases, on one hand, the computational cost increases by employing high resolution simulation (atomistic model) of all regions and, on the other hand, the system's properties at the region of the interest are largely distorted or cancelled by simulating this part with low resolution (coarse-grained model). The solution to this dilemma is given by the dual resolution models in which both atomistic and coarse grained resolutions are concurrently employed.

One of the mostly known concurrent multiscale scheme is the quantum mechanics/molecular mechanics (QM/MM) method [42–45]. In this scheme, a small region is described with *ab initio* resolution while being connected to a larger region which is treated using classical atomistic forcefields. These techniques have been widely employed in studying enzymatic chemical reactions [46, 47].

Another class of multiscale simulation methods allow to bridge the atomistic and CG models concurrently [48–51]. The coupling strategy of these techniques relies on a smooth spatial interpolation on the atomistic and CG force fields. Several methods have been established and can be classified depending on the interpolation strategy: (*i*) methods that interpolate forces acting on the particles (*ii*) methods that interpolate the interaction potentials. The most used and popular technique of the former class is Adaptive Resolution Simulation (`AdResS`) [48] while to the latter class belongs the Hamiltonian Adaptive Resolution Simulation (`H-AdResS`) [49]. Both methods have proven to be successful for dual resolution simulation of soft matter systems, for instance, solvated proteins, fullerene, DNA and macromolecules [49, 52–60] when compared to fully atomistic simulations. Since this thesis focuses on the `H-AdResS` method, a brief description is presented in the following section.

1.2 Hamiltonian Adaptive Resolution Simulation

In the `H-AdResS` scheme, the simulation domain is subdivided into two parts of high and low resolution [49, 50]. In the high resolution region, the accurate and more computationally expensive model is employed, while in the low resolution region a simpler description featuring a lower computational cost is used. The resolution of each molecule in the system is determined by its position in space. The existence of a transition layer (also dubbed hybrid region), which couples two resolutions through thermodynamically consistent forces and energies, allows the molecules to diffuse freely and change model, i.e. resolution, on the fly. Thus, the molecules in such dual resolution scheme smoothly vary their resolutions, adapting their representations to one model or the other in a continuous manner.

The description of the interactions between the particles in the `H-AdResS` scheme, is given in terms of a global Hamiltonian function $\mathcal{H}$, which has the following form [49,61]:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \left\{ \lambda_\alpha V_\alpha^{AT} + (1 - \lambda_\alpha) V_\alpha^{CG} \right\} \tag{1.2}$$

The term $\mathcal{K}$ is the atomistic kinetic energy, and V^{int} consists of all the intramolecular bonded interactions (e.g. covalent bonds or angle bending terms). The resolution of particle α is specified by the transition function $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$, which is computed on the center-of-mass coordinates $\mathbf{R}_\alpha$ of the molecule. The switching function should be continuous and smooth and it is bounded within the domain [0, 1]. Depending on the shape of the AT region, the switching function can be defined in slab, spherical or cylindrical coordinate systems. Figure 1.3 shows an `H-AdResS` simulation setup in which the shape of AT region is rectangular.

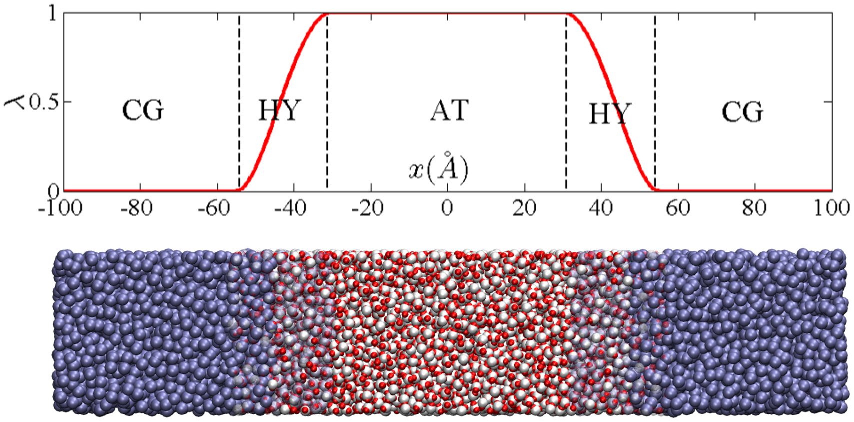


Figure 1.3. In Hamiltonian Adaptive Resolution Simulation, the periodic box involves three different regions: Coarse-grained (CG), Hybrid (HY), and Atomistic (AT). (Upper panel) The switching function λ is piecewise function between 0 (CG) and 1 (AT), defining the resolution of a molecule. (Lower panel) A snapshot of the `H-AdResS` setup showing the corresponding subdomains [61].

A molecule interacts with its neighboring particles through coarse-grained V^{CG} and atomistic V^{AT} potentials. The functional form of these potentials is arbitrary, as well as the order of the interaction (two-body, three-body...). In the Hamiltonian Eq. 1.2 the non-bonded potential energy contribution of each molecule α is given by a weighted sum of two terms V_α^{CG} and V_α^{AT}, defined as:

$$V_\alpha^{AT} \equiv \frac{1}{2} \sum_{\beta,\beta\neq\alpha}^{N} \sum_{ij} V^{AT}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|), \tag{1.3}$$

$$V_\alpha^{CG} \equiv \frac{1}{2} \sum_{\beta,\beta\neq\alpha}^{N} V^{CG}(|\mathbf{R}_\alpha - \mathbf{R}_\beta|).$$

The AT and CG terms of each molecule are weighted by λ_α or $(1-\lambda_\alpha)$, respectively. V_α^{AT} and V_α^{CG} have a factor 1/2 to account for the double counting.

As it is shown schematically in Figure 1.4, the equation of state of each resolution follows different isotherm curves in the pressure-density plane. As it has been thoroughly investigated [49, 50, 62, 63], if the dynamics of the system is solely described by equation 1.2, two states with different density and pressure are linked together (red line). However, coupling of two resolutions at either the same pressure (blue line) or the same density (black line) requires introducing new terms known as compensating energies (ΔH) into the Hamiltonian [49, 61]:

$$H_\Delta = H - \sum_{\alpha=1}^{N} \Delta H(\lambda(\mathbf{R}_\alpha)). \tag{1.4}$$

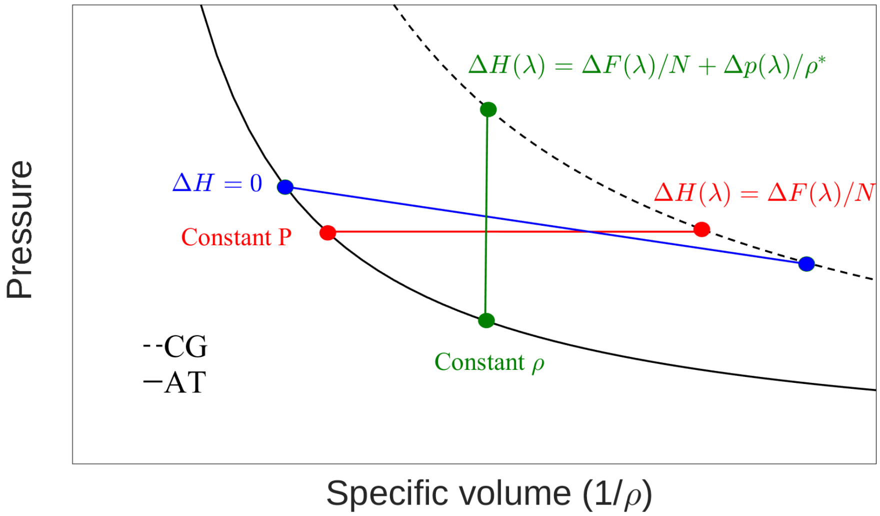


Figure 1.4. Isotherm curves of equation of states for atomistic (solid line) and corresponding coarse-grained (dashed line) model in the pressure-density plane. The blue line connects two thermodynamic states of AT and CG when the Hamiltonian of the system is not compensated ($\Delta H = 0$). The blue and red lines represent respectively the cases when the pressure and density of the two resolutions are equal [49].

To induce a uniform pressure across the resolutions, the compensation term $\Delta H(\lambda)$ has to account for the Helmholtz free energy difference ($\Delta F(\lambda)$) which is given by

$$\Delta H(\lambda) = \frac{\Delta F(\lambda)}{N} = \frac{1}{N} \int_0^\lambda \left\langle \left[V^{AT} - V^{CG} \right] \right\rangle_\lambda \tag{1.5}$$

While in the constant-density route, instead of imposing uniform pressure, the two resolutions coexist at the same density which is given by the reference density ρ^*. In this case, the compensating term is related to the Gibbs free energy difference $\Delta G(\lambda)$ or to the chemical potential difference $\Delta\mu(\lambda)$ across the transition layer and it is expressed by

$$\begin{aligned} \Delta H(\lambda) \quad &= \frac{\Delta G(\lambda)}{N} \equiv \Delta\mu(\lambda) \\ &= \frac{\Delta F(\lambda)}{N} + \frac{\Delta p(\lambda)}{\rho^*} \end{aligned} \tag{1.6}$$

It has been demonstrated that the compensation functions can be computed by preforming Kirkwood thermodynamic integration (KTI) [49, 64]. In this approach, the potential $\Delta H(\lambda)$ can be approximated in a mean field fashion by the Helmholtz free energy in the constant-pressure route or Gibbs free energy difference in the constant-density route between a system with hybrid Hamiltonian at resolution λ and the reference (CG) system with $\lambda = 0$ [49, 62]. However, the accuracy of this procedure can be limited when there are strong correlations within the hybrid region.

A more effective strategy is to compute the compensation energies locally and parametrize the compensation *on the fly* within an iterative scheme [61, 62]. In this respect, the hybrid region (HY) is discretized in a number of bins and corresponding compensating energies are computed and averaged over the simulation time. The running average update continues until the compensation energies have converged to a stable value in each bin; after this point, the update is interrupted and the resulting compensation is given.

Having developed and generally established the `H-AdResS` scheme for all types of short-ranged interactions of high and low resolutions, it is necessary to investigate the accurate implementation of the long-ranged electrostatic interaction which is crucial in the simulation of charged complex systems. In the following sections, we will briefly describe the method of EWALD summation followed by an alternative short-ranged potential known as Damped Shift Force. Afterwards, we will explain briefly the notions of chemical potential and and non-equilibrium work-free-energy relation on which we focused for their incorporation in the context of `H-AdResS` within the next chapters of this thesis.

1.3 Coulomb interaction

The electrostatic interaction in molecular simulations is one of the most computationally demanding tasks and it is essential to treat it properly. This is due to the long-range decay $\sim r^{-1}$ of the Coulomb interaction with respect to the charged molecules' pair distance r. There have been several techniques to handle the electrostatic interactions properly and accurately [65–68] which are categorized into implicit and explicit methods. The electrostatic interaction between the macromolecules and solvent molecules in the methods of former category such as continuum dielectrics method [66], is approximated by a continuum field and the resulting continuum dielectric of the medium is used to simulate the macromolecules. However, in the methods of latter category, such as Ewald summations [68], or interaction shifting and truncation [67], the electrostatic interaction between all molecules is explicitly computed. Additionally, there are methods that combine the two approaches, for instance the reaction field [69–71] method. In the strategy of these schemes, the electrostatic interaction is assumed short-ranged and while the solvent molecules are explicitly simulated, a mean-field approximation is also used to improve the accuracy of the pairwise interaction. In the following sections, we will present briefly the Ewald summation method and its alternative short-ranged electrostatic potential known as Damped Shift Force (DSF).

Ewald Summation

Suppose that we have a cubic simulation box of length L in a vacuum that is composed of N ions with charges $q_1, q_2, \dots,\ q_N$ and positions $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$,. For the periodic boundary conditions, we can express the total Coulomb interaction energy between all ion pairs (i, j) and their periodic images as [17]

$$U = \frac{1}{4\pi\epsilon_0} \sum_S \sum_{(i,j)} \frac{q_i q_j}{|\mathbf{r}_{ij,S}|} \tag{1.7}$$

Here, $\mathbf{r}_{ij,S} = \mathbf{r}_j - \mathbf{r}_i + S$ and S accounts for the periodic images of the ions in the system $S = mL$ with $m = \{\dots, -1, 0, 1, \dots\}$. The vacuum permittivity is $\epsilon_0 = 8.854 \times 10^{-12} C^2 N^{-1} m^{-2}$. The sum goes over all pairs of ions, i.e. $N(N-1)/2$.

One can separate the Coulomb interaction into short-range and long-range parts:

$$\frac{1}{r} = \frac{\mathrm{erfc}(\alpha r)}{r} + \frac{\mathrm{erf}(\alpha r)}{r} \tag{1.8}$$

where erf(r) and erfc(r) are the error and complement error functions defined by

$$\mathrm{erfc}(x) = 1 - \mathrm{erf}(x) = \frac{2}{\pi} \int_x^{\infty} dt e^{-t^2} \tag{1.9}$$

The parameter α controls the range of the short term interaction.

Using the aforementioned range separation scheme, the electrostatic short-range interaction is given by

$$U_{\mathrm{short}} = \frac{1}{4\pi\epsilon_0} \sum_S \sum_{(i,j)} \frac{\mathrm{erfc}(\alpha r_{ij,S}) q_i q_j}{|\mathbf{r}_{ij,S}|} \tag{1.10}$$

and the long-range interaction is given by

$$U_{\text{long}} = \frac{1}{4\pi\epsilon_0} \sum_{S} \sum_{(i,j)} \frac{\text{erf}(\alpha r_{ij,S}) q_i q_j}{|\mathbf{r}_{ij,S}|} \tag{1.11}$$

The parameter α determines how fast the short-range interaction is decaying and its value is determined by the cut-off radius r_c. It is shown that $\alpha = 3.5/r_c$ is an appropriate value to have a good balance between short- and long-range parts. Since $\text{erf}(\alpha r)/r$ behaves as $1/r$ for large r, evaluation of the long-range behavior is important to accurately calculate forces and energy. However, it is computationally inefficient to evaluate the long-ranged interaction in the real space. In this case, one can take advantage of Fourier or reciprocal space in which a long-range function is converted into a short-ranged one and, given the periodicity of the simulation box, it is consistent to use Fourier expansion of the function $\text{erf}(\alpha r)/r$. For the cubic simulation box of size L and volume $V = L^3$, the reciprocal-space vectors is given as $\mathbf{g} = 2\pi \text{n}/L$, where $\mathbf{n}$ is a vector of integers. The Fourier series of the error function is

$$\sum_{S} \frac{\text{erf}(\alpha\, |\mathbf{r} + S|)}{|\mathbf{r} + S|} = \frac{1}{V} \sum_{\mathbf{g}} C_{\mathbf{g}} e^{i\mathbf{g}.\mathbf{r}} \tag{1.12}$$

The Fourier coefficients are

$$C_{\mathbf{g}} = \sum_{S} \int_{V} d\mathbf{r} \frac{\text{erf}(|\mathbf{r} + S|)}{|\mathbf{r} + S|} e^{-i\mathbf{g}.\mathbf{r}} = \frac{4\pi}{|\mathbf{g}|^2} e^{-|\mathbf{g}|^2/4\alpha^2} \tag{1.13}$$

As it is evident from the last relation in equation 1.13, $C_{\mathbf{g}}$ coefficients decay in the Fourier space and one can assume a truncation in the reciprocal space. Thus the long-range interaction is obtained by evaluating $C_{\mathbf{g}}$ coefficients.

The computational cost of the original Ewald summation is $\mathcal{O}(N^2)$ where N is the number of ions in the system [17, 65]. However, by appropriate choice of α, the computational cost reaches $\mathcal{O}(N^{3/2})$ [72]. Several methods such as particle-particle particle-mesh (P3M) and particle-mesh Ewald (PME), have been introduced to optimize the reciprocal summation by utilizing the fast Fourier transform [73–75], leading to reduction in the computational cost to $\mathcal{O}(N \text{log} N)$.

The optimizations and developments of the Ewald summation method is based on the intrinsic three-dimensional periodicity of the system. However, in certain cases, for example, liquid-vapour interface and biological membranes, it is essential to reformulate the EWALD summations for two-dimensional (2D) systems [76–78]. The existence of the system's replicas using Ewald sum has also been shown problematic for the 3D systems. For example, the compact folded state of the solvated protein might be artificially stabilized by the periodic replicas introduced by Ewald summations [79]. The existence of system's charge neutrality is also essential for the convergence of Ewald summation. Several studies have shown that for a system having a net charge, a compensating charge has to be considered in order to avoid spurious effects on the system dynamics [80]. The discussed limitations on utilizing the Ewald summation has led us to use an alternative short-ranged but accurate potential known as Damped Shifted Force in the `H-AdResS` scheme.

Damped Shifted Force potential (DSF)

The original idea of short-ranged electrostatic potential was introduced by Wolf *et al.* [81, 82]. Based on two key features of condensed phase systems that are (i) the electrostatic interaction is effectively short-ranged and (ii) the charge neutrality is essential for accumulation of the electrostatic interaction in a pairwise fashion within a cut-off sphere with radius r_c, they proposed a pairwise summation method

$$V_{\mathrm{Wolf}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0}\left[\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} - \lim_{r_{ij}\to r_c}\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}}\right] \tag{1.14}$$

where $r_{ij} \leq r_c$ and r_c is the cut-off radius and α is a damping parameter with dimension of inverse length. They showed that employing a distance-dependent decay function (first term in equation 1.14) which is identical to that obtained in the real-space portion of EWALD summation method, and the neutralizing image charges on the cut-off sphere (second term in equation 1.14) is sufficient to obtain excellent estimates of Madelung energies of many crystals [82]. However, because of force discontinuities at the cutoff radius, the use of the Wolf potential is problematic in MD simulations [65]. To overcome this problem, the potential function was modified so to give continuous potential and forces everywhere [65], thus becoming a valuable short-range alternative to EWALD summation method. In the new potential known as Damped Shifted Force (DSF), the electrostatic potential between two charges q_i and q_j separated by a distance r_{ij} is given by the following expression:

$$\begin{aligned} V_{\mathrm{DSF}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0}\Bigg[&\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\mathrm{erfc}(\alpha r_c)}{r_c} \\ &+ \left(\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}}\frac{\exp\left(-\alpha^2 r_c^2\right)}{r_c}\right)(r_{ij} - r_c)\Bigg] \, , \end{aligned} \tag{1.15}$$

The DSF method scales linearly with the system size $\mathcal{O}(N)$ similar to the other pairwise cut-off methods [65]. Additionally, in situations where the periodicity of the system has to be avoided or the system is not neutrally charged, for example the interfacial systems [76–78, 80] and dual resolution simulations (`H-AdResS`) [49], the standard Ewald sum is problematic and it requires corrections and reformulations. However, the DSF technique can be employed normally without modifications [65].

1.4 Chemical potential

The chemical potential of a substance determines many of its physical properties, for instance, the thermodynamic states of the coexisting phases and crystallization [83,84]. Another example is non-equilibrium diffusio-osmotic flow in liquid mixtures where a gradient of chemical potential of one substance in proximity to a solid interface causes a flow [85,86].

In statistical mechanics, the chemical potential of a species is defined as a thermodynamic quantity showing the free energy changes upon variation of the number of that species. To calculate the chemical potential, consider a liquid containing N particles of the same type [2]. The introduction of a new particle into the system can be done either at constant volume or constant pressure. By calculating the change in corresponding free energy of the system, Helmholtz free energy (F) in the former or Gibbs free energy (G) in the latter, one can define the chemical potential by

$$\mu \equiv \left(\frac{\partial G}{\partial N}\right)_{P,T} = \left(\frac{\partial F}{\partial N}\right)_{V,T} \tag{1.16}$$

To obtain the free energy of the system, one has to compute the partition function. For a system containing N particles interacting with potential energy function $U(r)$, volume V and thermal de Broglie wavelength Λ, the classical canonical partition function is given by

$$Q(N,V,T) = \frac{1}{\Lambda^{3N}N!}\int d\mathbf{r}_1 \cdots d\mathbf{r}_N \, \exp\left[-\beta U(\mathbf{r}_1,\cdots,\mathbf{r}_N)\right] \tag{1.17}$$

Where $\mathbf{r}_i$ is the position of the i^{th} particle. The partition function can be expressed in terms of the coordinates which are rescaled by the system size, $\mathbf{s}_i = \mathbf{r}_i/L$

$$Q(N,V,T) = \frac{V^N}{\Lambda^{3N}N!}\int_0^1 ... \int_0^1 d\mathbf{s}_1 \cdots d\mathbf{s}_N \, \exp\left[-\beta U(\mathbf{s}_1 \cdots \mathbf{s}_N; L)\right] \tag{1.18}$$

The Helmholtz free energy in the canonical ensemble is given by $F = -k_B T \ln Q(N,V,T)$ and, in the limit of large number of particles, the chemical potential can be derived by $\partial F/\partial N$ as following

$$\begin{aligned} \mu &= -k_B T \ln\left(\frac{Q(N+1,V,T)}{Q(N,V,T)}\right) \\ &= -k_B T \ln\left(\frac{V/\Lambda^3}{N+1}\right) - k_B T \ln \int d\mathrm{s}^{N+1}\langle \exp\left[-\beta \Delta U\right]\rangle_N \\ &= \mu_{id}(\rho) + \mu_{ex} \end{aligned} \tag{1.19}$$

Here, the potential energy difference between $(N+1)$- and N-particle system is $\Delta U = U(\mathbf{s}_1, \cdots \mathbf{s}_{N+1}; L) - U(\mathbf{s}_1 \cdots \mathbf{s}_N; L)$. Thus, the chemical potential is ultimately expressed by two terms, the ideal gas contribution, $\mu_{id}(\rho)$ which can be analytically calculated for the system of density ρ; and the excess chemical potential:

$$\mu_{ex} = -k_B T \ln \int d\mathrm{s}^{N+1} \langle \exp\left[-\beta \Delta U\right] \rangle_N \tag{1.20}$$

Correct calculation of the integral in the highly dense liquid is challenging because the energy cost of inserting a new particle into the system is high leading to vanishingly small contribution $\langle \exp\left[-\beta \Delta U\right] \rangle_N$. Thus, it is required to use enhanced sampling techniques to correctly and efficiently calculate the integral in the dense regime [87–90]. However, as it was stated in equation 1.6, in the `H-AdResS` method, the chemical potential difference between two resolutions is systematically computed by inducing the uniform density profile across the system. We will describe the procedures to compute the chemical potential using `H-AdResS` more in detail in chapter 3.

1.5 Nonequilibrium work-free-energy relation

In the previous section, we explained the importance of the chemical potential and described the challenges of its accurate calculation. In this section, we describe a way to compute the free energy difference between two equilibrium states of the system. As it will be explained thoroughly in chapter 5, this approach will be incorporated in the context of `H-AdResS` in order to compute the chemical potential of mixtures in a very dilute concentration regime.

From the second law of thermodynamics, we know that the average amount of work $W_{\mathcal{AB}}$ to take the system from the state $\mathcal{A}$ to the state $\mathcal{B}$ is equal to or higher than the free energy difference $\langle W_{\mathcal{AB}} \rangle \leq \Delta F_{\mathcal{AB}}$. The equality holds only if the process is performed reversibly. That means that, during the procedure, the system remains close to equilibrium at the intermediate states. However, any dissipation leads to entropy production and deviation of the work from the free energy difference.

In 1993, Chrsitopher Jarzynski introduced an ideal equality between work and free energy which holds for out-of-equilibrium processes [91],

$$\langle e^{-\beta W_{\mathcal{AB}}} \rangle_{\mathcal{A}} = e^{-\beta \Delta F_{\mathcal{AB}}}, \tag{1.21}$$

where $\beta = 1/k_B T$ and the ensemble average is defined over the initial conditions $\mathbf{x}_0$ which are canonically sampled at state $\mathcal{A}$,

$$\langle e^{-\beta W_{\mathcal{AB}}} \rangle = \frac{C_N}{Q_{\mathcal{A}}(N, V, T)} \int d\mathbf{x}_0 e^{-\beta H_{\mathcal{A}}(\mathbf{x}_0)} e^{-\beta W_{\mathcal{AB}}(\mathbf{x}_0)}. \tag{1.22}$$

Here, the prefactor $C_N = 1/N! h^{3N}$ is determined by Planck's constant h and the number of the particles N. The canonical partition function of the canonical ensemble is given by $Q_{\mathcal{A}}(N, V, T) = \int dx_0 e^{-\beta H_{\mathcal{A}}(\mathbf{x}_0)}$.

The equation 1.21 which is referred to as *Jarzynski equality* together with a succeeding but more general relation known as *Crook's fluctuation theorem* [92] are considered as the foundation for the nonequilibrium free energy techniques. In the following, we will describe how *Jarzynski equality* can be derived, specifically for the isolated system.

The Hamiltonian of a N-particle system which is time-dependent due to a dynamic interacting potential is expressed by [17]

$$H(\mathbf{r}, \mathbf{p}, t) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \ldots, \mathbf{r}_N, t) \tag{1.23}$$

The phase space vector at time t is defined as $\mathbf{x}_t = (\mathbf{r}_1, \ldots, \mathbf{r}_N, \mathbf{p}_1, \ldots, \mathbf{p}_N)$ where $\mathbf{r}_i$ and $\mathbf{p}_i$ are the position and momentum vectors of the i^{th} particle, respectively. The integration of the time derivative of the Hamiltonian over an interval $[0, T]$ is given by

$$\int_0^T dt\, \frac{dH}{dt} = \int_0^T dt\, \frac{\partial H}{\partial t} + \int_0^T dt\, \left(\nabla_{\mathbf{x}} H\right) \cdot \mathbf{x}_t \tag{1.24}$$

It is worth mentioning that the equation 1.24 is the microscopic version of the first law of thermodynamics, wherein the system's change in internal energy during the time interval ($U_T = \int_0^T dt\, dH/dt$) is equal to summation of work ($W_T = \int_0^T dt\, \partial H/\partial t$) and transferred heat ($Q_T = \int_0^T dt\, \nabla_{\mathbf{x}} H \cdot \mathbf{x}_t$). Let us assume that the initial trajectories $\mathbf{x}_0$ are evolved in isolation, i.e. $Q = 0$. Thus, equation 1.24 becomes

$$W_T(\mathbf{x}_0) = \int_0^T dt\, \frac{dH(\mathbf{x}_0, t)}{dt} = H(\mathbf{x}_T, T) - H(\mathbf{x}_0, 0) \tag{1.25}$$

If we define the system's canonical ensemble of the initial conditions as state $\mathcal{A}$, then the ensemble averages 1.21 can be written as

$$\langle e^{-\beta W_{\mathcal{AB}}} \rangle = \frac{C_N}{Q_{\mathcal{A}}(N, V, T)} \int d\mathbf{x}_0 e^{-\beta H_{\mathcal{A}}(\mathbf{x}_0)} e^{-\beta [H(\mathbf{x}_T, T) - H_{\mathcal{A}}(\mathbf{x}_0)]} \tag{1.26}$$

$$= \frac{C_N}{Q_{\mathcal{A}}(N, V, T)} \int d\mathbf{x}_0 e^{-\beta H(\mathbf{x}_T, T)} \tag{1.27}$$

As the trajectories are evolved by the Hamiltonian's equations of motion, according to the Cauchy-Lipshitz theorem on the uniqueness of solution of the first order differential equation, for each initial condition $\mathbf{x}_0$, there is a unique solution $\mathbf{x}_T$. Additionally, by Lioville's theorem on the incompressibility of the phase space, the differential volume along the evolving path is preserved, i.e. $d\mathbf{x}_0 = d\mathbf{x}_T$. Hence, recognizing $H(\mathbf{x}_T, T) = H_{\mathcal{B}}(\mathbf{x}_T)$ we can change variables $\mathbf{x}_0$ to $\mathbf{x}_T$ and ultimately reach the *Jarzynski equality*,

$$\begin{aligned}\langle e^{-\beta W_{\mathcal{AB}}}\rangle &= \frac{C_N}{Q_{\mathcal{A}}(N,V,T)}\int d\mathbf{x}_T e^{-\beta H_{\mathcal{B}}(\mathbf{x}_T)} \\ &= \frac{Q_{\mathcal{B}}(N,V,T)}{Q_{\mathcal{A}}(N,V,T)} \\ &= e^{-\beta\Delta F_{\mathcal{AB}}}\end{aligned} \tag{1.28}$$

It is worth noting that in equation 1.28 we assumed that final configurations of the system at time T are sufficiently sampled in the phase space such that the integral $C_N \int d\mathbf{x}_T e^{-\beta H_{\mathcal{B}}(\mathbf{x}_T)}$ is equal to the canonical partition function of the system at state $\mathcal{B}$, i.e. $Q_{\mathcal{B}}(N,V,T)$.

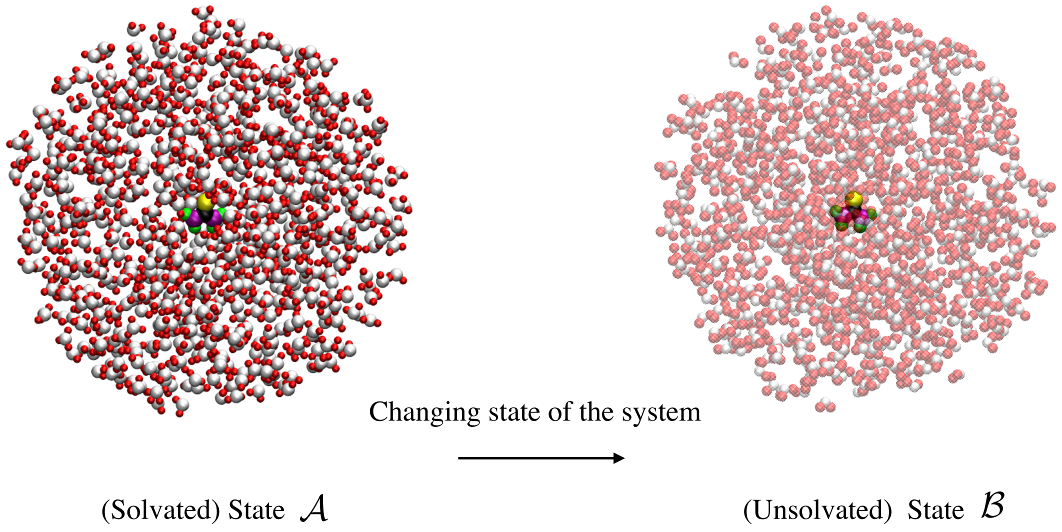


Figure 1.5. Presentation of a two-state system involving a urea and water molecules. In the solvated state $\mathcal{A}$, the urea molecule is interacting with the water molcules and forms hydration shells while in unsolvated state $\mathcal{B}$, there is no interaction between the molecules and all molecules behave like ideal gas particles.

This approach can be employed to compute the solvation free energy of molecules. As it is schematically shown in Figure 1.5 for the urea molecule solvated in water molecules for instance, the two equilibrium states $\mathcal{A}$ and $\mathcal{B}$ are associated to, respectively: a solvated state in which the urea interacts with the water molecules and forms solvation shell; and an unsolvated state in which the urea loses the interaction with water molecules and all molecules behaves like ideal gas particles. The work required to force the system to change its state between the states $\mathcal{A}$ and $\mathcal{B}$ is related to the free energy difference by means of equation 1.21 which is essentially equal to the solvation free energy of the urea molecule.

1.6 Outline

In the following we provide a brief overview over the chapters 2-6.

In chapter 2 we propose and validate the usage of a short-range modification of Coulomb potential, the Damped Shifted Force (DSF) model, in the context of the `H-AdResS` scheme. This approach, which is validated here on bulk water, ensures a reliable reproduction of the structural and dynamical properties of liquid water. The resulting dual-resolution setup is implemented in the `LAMMPS` simulation package, and its customized version employed in the present work is made publicly available.

Following the accurate treatment of electrostatics in the `H-AdResS` scheme, in chapter 3 we explain how to couple the target atomistic system (AT) to an ideal gas (IG) bath of point-like particles. In this case, by enforcing a uniform density profile across the simulation box, a single-molecule external potential is computed. The external potential is identified with exactly the excess chemical potential of the target system. The analogy with thermodynamic integration becomes evident and therefore the method has been named spatially resolved thermodynamic integration, or SPARTIAN for short. In the highly dense regime, the SPARTIAN method surpasses many other popular methods which rely on the insertion of test particles into the target system. This is because in the SPARTIAN method, increasing the density/concentration of the liquid/components, enhances the statistics of sampling in the hybrid region and accordingly, improves the precision of the results. From a computational efficiency point of view, SPARTIAN surpasses particle insertion methods applied to complex molecular systems because the ideal gas representation contributes negligibly to the computational cost of the simulation, thus allowing one to make use of large reservoirs at minimal expenses.

In Chapter 4 we extend the previous results to the study of solids, namely, an atomistic crystal coupled to an ideal representation (Einstein crystal) of the system. we show that in virtue of such a coupling, the excess Helmholtz free energy of a crystal can be efficiently computed. This result completes the study done for liquids and presents a promising tool for the study of solvation properties of materials.

In chapter 5 we use the Jarzynski equality to obtain the solvation free energy of molecules in the context of Hamiltonian adaptive resolution (`H-AdResS`) [93]. In this method, the solvent is concurrently simulated in the atomistic (SPC/E water molecule) and coarse-grained (ideal gas of water molecules) resolutions. The resolutions are coupled in a thermodynamically consistent manner such that the chemical potentials of both become equal. Then, we use steered molecular dynamics to pull the solute molecule from atomistic (solvated state) into the ideal gas (unsolvated state) region. The free energy difference between the two states is obtained using either equilibrium (umbrella sampling) or nonequilibrium (Jarzynski) method.

In chapter 6 we discuss the spatial block analysis (SBA) method [94,95] to efficiently extrapolate thermodynamic quantities from finite-size computer simulations. This is accomplished by subdividing the simulation box into blocks of increasing size and calculating volume dependent fluctuations of the number of particles. Then we show that it is possible to extrapolate the bulk isothermal compressibility and Kirkwood-Buff integrals in the thermodynamic limit. Furthermore, we discuss two types of finite-size effects in the context of the SBA method: i) the statistical ensemble and ii) the finite integration domains used in computer simulations. We consider prototypical Lennard-Jones liquids and liquid mixtures to illustrate the aforementioned effects. This study is relevant in the context of the `SPARTIAN` method as its approach will be used to test and characterize the results of ongoing projects linked to the extended and developed `SPARTIAN` method in the grand canonical simulations.

2. Accurate and general treatment of electrostatics in Hamiltonian Adaptive Resolution Simulation

This chapter is a research article that has been published in the European Physical Journal Special Topics.

Maziar Heidari, Robinson Cortes-Huerto, Davide Donadio and Raffaello Potestio
Accurate and general treatment of electrostatics in Hamiltonian Adaptive Resolution Simulation
Eur. Phys. J. Spec. Top. (2016) 225: 1505
DOI: 10.1140/epjst/e2016-60151-6
©EDP Sciences and Springer 2016

Abstract

In adaptive resolution simulations the same system is concurrently modeled with different resolution in different subdomains of the simulation box, thereby enabling an accurate description in a small but relevant region, while the rest is treated with a computationally parsimonious model. In this framework, electrostatic interaction, whose accurate treatment is a crucial aspect in the realistic modeling of soft matter and biological systems, represents a particularly acute problem due to the intrinsic long-range nature of Coulomb potential. In the present work we propose and validate the usage of a short-range modification of Coulomb potential, the Damped shifted force (DSF) model, in the context of the Hamiltonian adaptive resolution simulation (`H-AdResS`) scheme. This approach, which is here validated on bulk water, ensures a reliable reproduction of the structural and dynamical properties of the liquid, and enables a seamless embedding in the `H-AdResS` framework. The resulting dual-resolution setup is implemented in the `LAMMPS` simulation package, and its customized version employed in the present work is made publicly available.

2.1 Introduction

The definition of soft matter encompasses a broad spectrum of different systems, from liquids composed by single atoms or simple polymer molecules [96–99] to large and complex biomolecular assemblies [30, 100–110]. The structural and dynamical properties of this ample variety of systems spans an equivalently wide range of length and time scales, the interplay of which gives rise to a wealth of different properties.

This multi-scale nature, however, poses substantial challenges to modeling and simulation. *In silico* experiments are limited by the size of the system and the duration that can be reached by the simulation. It is often the case that the process of interest takes place in a typical time interval that is not within the reach of present day computers. An alternative to overcome such limitation is to employ a simpler description of the system, where the level of detail is reduced, for example replacing a fully atomistic description and representing a group of different atoms as a single interaction site. These coarse-grained (CG) models [111–116] have provided an extraordinary tool to made mesoscale systems accessible by simulations over time scales that would not be viable through an atomistic description. Nonetheless, there are several circumstances in which it is not possible to investigate a certain system or process by means of a CG model, for example when chemically specific interactions play a substantial role.

In recent years several strategies have been developed to find a compromise between model accuracy and computational efficiency. A prominent example of these approaches are adaptive, dual-resolution simulation schemes [48–51, 54–57, 62, 117, 118]. These methods identify a specific, typically small region of the system that necessitates modeling at full detail, e.g. atomistic (AT). This subregion is indeed described with the high-resolution model required, while in the remainder of the system a simpler, effective CG representation is employed. In modeling liquid systems, a crucial feature is that the two regions at different resolutions are connected by an open boundary, which allows the diffusion of molecules (e.g. solvent particles) from one subdomain to the other. The instantaneous location of a molecule specifies the interactions with the neighboring molecules, thus allowing the resolution of a molecule to change on the fly.

Adaptive resolution methods thus enable the simulation of a system whose description is not bound to be the same everywhere, rather it is more accurate where strictly necessary, and simpler and computationally more efficient in the rest. Two advantages can be envisaged: on the one hand, there is an obvious gain in terms of simulation time, due to the reduced number of degrees of freedom and the simpler interactions that are employed in the low-resolution subdomain. On the other hand, these setups can be employed to characterize the physical properties of the system by systematically changing the size of the high-resolution region, so to effectively probe the locality of physical phenomena and finite size effects. This strategy enables the controllable decoupling between the internal degrees of freedom of a chosen subregion of the system from the rest, yet without modifying the thermodynamical equilibrium in the high resolution domain [52, 119].

Among the methods that have implemented this strategy, a notable place is occupied by the adaptive resolution simulation (`AdResS`) [48, 55–57] and the Hamiltonian adaptive resolution simulation (`H-AdResS`) [49, 50, 62, 118] schemes. In these setups, schematically represented in Fig. 4.5, the resolution of a molecule is determined by the value of a function, usually dubbed *switching function* λ, that is equal to 1 in the high-resolution or atomistic (AT) subregion, 0 in the low-resolution or coarse-grained (CG) subregion, and smoothly interpolates between these values in an interface region,

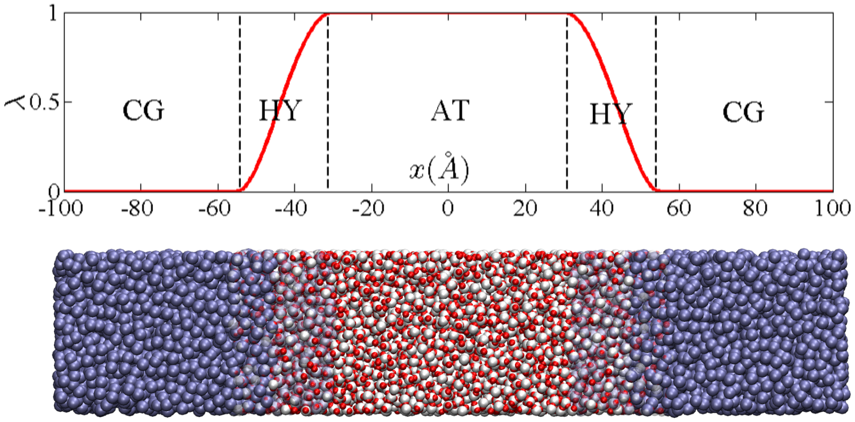


Figure 2.1. Setup of a Hamiltonian Adaptive Resolution Simulation. The periodic box is partitioned into three different regions, namely: Coarse-grained (CG), Hybrid (HY), and Atomistic (AT). Upper panel: the switching function λ takes values between 0 (CG) and 1 (AT), thus defining the resolution of a molecule (here water). Lower panel: simulation snapshot explicitly showing the various subdomains.

dubbed hybrid region (HY). When a molecule is in the AT or CG domains it is treated as fully atomistic or fully coarse-grained, respectively; in the HY region the interactions are obtained by interpolation of the AT and CG ones.

The `AdResS` and `H-AdResS` methods have been validated on various systems and in different contexts, and have been shown to effectively and efficiently provide an accurate description of the system in the AT region at a lower computational cost with respect to an equivalent simulation employing the high-resolution model everywhere. However, a fundamental problem affects these schemes and limits the range of systems that are amenable by them, namely the treatment of long-range interactions, such as electrostatics. Devising an efficient implementation of electrostatics in computer simulations of systems with periodic boundary conditions is still an open problem, due to the long-range nature of Coulomb potential. Nonetheless, well-established techniques are presently available to deal with these interactions in a variety of different physical contexts. In particular, the Ewald summation (ES) [68] method and its subsequent variations provide a theoretically sound procedure to decompose the interaction in a local term, that is treated as a conventional short-range potential, and a non-local, long-range term that is efficiently computed in Fourier space.

This second term is the critical one, as it is intrinsically incompatible with the space-dependent decomposition of interactions that characterizes dual-resolution simulations. In fact, the latter approaches rely on the short-range nature of interaction potentials in order to provide a local definition of molecular resolution. In the presence of a non-local term, which requires the simultaneous knowledge of the position of each particle to compute the forces, this position-based separation of the interactions is not anymore possible.

The strategy that has been employed so far to circumvent this problem is to locally approximate the Coulomb interaction by the Reaction Field (RF) [69–71] potential. The assumption underlying this approach is that beyond a fixed (short) cutoff distance the effective long range electrostatic contribution is equivalent to that of a uniform and homogeneous dielectric medium. The standard form of the potential is thus replaced by a mean-field function, thereby reducing the interaction to a short-range one. The RF method has been successfully employed in dual-resolution setups [52, 120], however it suffers from two substantial limitations. The first one is that the assumption of a uniform and homogeneous medium beyond the cutoff distance is not always valid: this would be the case for heterogeneous interfaces, for example, polar molecules in proximity of a metal surface, or large biomolecules (protein, DNA filament) in solution [121]. The second limitation is that the parametrization of the RF potential necessitates the previous knowledge of the relative dielectric constant of the medium, which is not always available *a priori* and would then have to be computed in an independent simulation. It may also be undesirable to introduce as a parameter of the model a quantity that is indeed an emergent property of the system. Additionally, it has been shown that an accurate modeling of the system under examination sometimes necessitates a specific parametrization not only of the dielectric constant, but also of the underlying force field [71].

Here we approach the problem of electrostatic interaction in dual-resolution, adaptive simulations making use of an alternative formulation of Coulomb potential, namely the damped shifted force (DSF) potential [65, 82] method. This strategy allows us to rephrase the electrostatic interaction in terms of finite-ranged, two-body analytical potentials, as in the case of the RF. Albeit the computational cost of RF and DSF would be identical if the same cutoff range is employed, the DSF allows us to circumvent the limitations intrinsic in both the ES [122–124] and RF methods.

It is generally the case, when making use of adaptive resolution strategies, that one saves simulation time at the cost of performing an accurate (and computationally cumbersome) parametrization of the setup, in terms of CG model, approach-specific parameters, and, as in the case just discussed, a very limited treatment of electrostatic interaction in the high-resolution model. Here we make use of elegant solutions to these problems that have been already developed and successfully validated, and can therefore be seamlessly implemented and employed. In the case of DSF, for example, the use of this technique would prove to be even advantageous over Ewald summations for some cases of interest to the computational biochemistry community. Furthermore, we present and discuss the DSF approach to the `H-AdResS` scheme in the comprehensive framework of the implementation of the latter in the `LAMMPS` [125] simulation package. The goal of this work is thus twofold: to present and validate the DSF approach in the context of `H-AdResS` simulations, and to present the implementation of these methods in the `LAMMPS` package, together with some advanced features whose practical implementation is here discussed in detail for the first time.

The manuscript is organized as follows. In Sect. 2.2 we discuss in detail the `H-AdResS` scheme, with a particular focus on the implementation features; then we briefly report the formulation of the RF method and the DSF method. In Sect. 2.3 we list and define the quantities that have been employed to validate the efficacy and accuracy of the proposed model. In Sect. 2.4 we report the details of the setups and the simulations, while in Sect. 2.5 we present the results of our study. The conclusions and perspectives are discussed in Sect. 4.6. The Appendix, Sect. 2.7, provides a list of technical details on the LAMMPS implementation of the `H-AdResS` method and its usage.

2.2 Methods

In this Section we review the different computational techniques employed in the present work. The first part is devoted to the `H-AdResS` scheme. Most of the fundamental and conceptual aspects of this method have been thoroughly discussed in previous publications [49, 50, 62, 118]. The focus is here given to the computational aspects of the scheme and, in particular, of the specific implementation in the `LAMMPS` [125] simulation package.

Subsequently, we summarize two of the three methods here used to treat electrostatic interaction, specifically the Reaction Field and the DSF methods. The Particle-Particle Particle-Mesh (PPPM) [73] Ewald summation method, not discussed in detail here, is taken as the golden standard, against which the results of the other two strategies are compared.

2.2.1 `H-AdResS`

In the `H-AdResS` scheme, the description of the interactions within a system of particles is given in terms of a global Hamiltonian function $\mathcal{H}$, which has the following form:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \left\{ \lambda_\alpha V_\alpha^{AT} + (1 - \lambda_\alpha) V_\alpha^{CG} \right\} \tag{2.1}$$

The term $\mathcal{K}$ is the atomistic kinetic energy, and V^{int} consists of all the intramolecular bonded interactions (e.g. bond stretching). The resolution of particle α is specified by the transition function $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$, which is computed on the center-of-mass coordinates $\mathbf{R}_\alpha$ of the molecule.

The value of the switching function is determined by the sizes d_{at} and d_{hy} of the AT and HY regions, respectively, and of the specific geometry of the AT region. If the latter is defined as a slab of the simulation box, for example as it is represented in

Fig. 4.5, d_{at} will correspond to the width of the atomistic subdomain; if a spherical geometry is employed, d_{at} will correspond to the AT region diameter. In all cases, the value of d_{hy} indicates the width or thickness of the HY layer embedding the AT region.

In the present work we employ a rectangular simulation box, and the AT region is a slab located in the middle of it. The resolution of a molecule is then determined through the following piece-wise λ function:

$$\lambda(x) = \begin{cases} 1 & |x| \leq d_{at}/2 \\ \cos^2\left(\frac{\pi(x-d_{at}/2)}{2d_{hy}}\right) & \frac{d_{at}}{2} < |x| \leq \frac{d_{at}}{2} + d_{hy} \\ 0 & |x| > d_{at} + d_{hy} \end{cases} \tag{2.2}$$

The mid point of the simulation box is set in the origin of the coordinate system. As it is shown in Fig. 4.5, in all simulations the width of the AT and HY region is set to $d_{at} = 60$Å and $d_{hy} = 25$Å, respectively.

A molecule interacts with its neighboring particles through coarse-grained V^{CG} and atomistic V^{AT} potentials. The functional form of these potentials is arbitrary, as well as the order of the interaction (two-body, three-body...) as long as the extension of the interaction is finite and short-ranged. For simplicity, in the following we shall restrict the discussion to the most common case of pairwise interactions. In the Hamiltonian of Eq. 4.1 the non-bonded potential energy contribution of each molecule α is given by a weighted sum of two terms V_α^{CG} and V_α^{AT}, defined as:

$$\begin{aligned} V_\alpha^{AT} &\equiv \frac{1}{2}\sum_{\beta,\beta\neq\alpha}^{N}\sum_{ij} V^{AT}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|) \\ V_\alpha^{CG} &\equiv \frac{1}{2}\sum_{\beta,\beta\neq\alpha}^{N} V^{CG}(|\mathbf{R}_\alpha - \mathbf{R}_\beta|) \end{aligned} \tag{2.3}$$

The AT and CG terms of each molecule are weighted by λ_α or $(1-\lambda_\alpha)$, respectively. As the total non-bonded potential is given by the sum of this linear combination over all molecules, V_α^{AT} and V_α^{CG} contain a factor 1/2 to account for the double counting. The corresponding force acting on atom i of molecule α is given by:

$$\begin{aligned} \mathbf{F}_{\alpha i} &= \mathbf{F}_{\alpha i}^{int} \\ &+ \sum_{\beta,\beta\neq\alpha}\left\{\frac{\lambda_\alpha+\lambda_\beta}{2}\mathbf{F}_{\alpha i|\beta}^{AT} + \left(1-\frac{\lambda_\alpha+\lambda_\beta}{2}\right)\mathbf{F}_{\alpha i|\beta}^{CG}\right\} \\ &- \left[V_\alpha^{AT} - V_\alpha^{CG}\right]\nabla_{\alpha i}\lambda_\alpha \end{aligned} \tag{2.4}$$

The first term, $\mathbf{F}_{\alpha i}^{int}$, is due to the interactions with atoms in the same molecule, and is not subject to resolution-dependent reweighting; the second term is the sum, over all other molecules β in the interaction range, of the pairwise atomistic and coarse-grained forces, weighted by the average resolutions of the two molecules. This term is antisymmetric under molecule exchange, and satisfies Newton's Third Law by construction. The last term emerges as a consequence of the non-uniformity of space in the dual-resolution simulation setup, that is, the fact that different interactions are present in different parts of the system. Because of this, translational invariance is locally broken, and a force emerges in the hybrid region (where $\nabla\lambda \neq 0$) and acts on the molecules pushing them in one of the two subdomains, depending on the sign of the prefactor $(V_\alpha^{AT} - V_\alpha^{CG})$.

The terms V_α^{AT} and V_α^{CG} contain the different potentials acting on the molecules, and in most cases they represent an atomistic potential that takes into account chemical specificity, and and effective, coarse-grained potential, which acts on collective degrees of freedom of the molecules (e.g. the center of mass). The CG potentials are normally potentials of mean force that enclose entropic contributions and are parametrized over specific thermodynamic properties, e.g. the radial distribution function (RDF), at a specific state point. These CG interactions and the corresponding reference system follow different equations of state, and, once coupled together via an open boundary as in the case of `H-AdResS`, they exchange particles to balance the differences in equilibrium pressure and chemical potential. As it has been already thoroughly investigated [49, 50, 62], models that, for the same temperature and density, attain different pressure, will determine in the dual-resolution setup a non-homogeneous density, as the region where the pressure is higher will relax by pushing molecules in the other region. Furthermore, in the `H-AdResS` setup the aforementioned *drift force* term $\mathbf{F}_{\alpha i}^{dr} = -\left[V_\alpha^{AT} - V_\alpha^{CG}\right]\nabla_{\alpha i}\lambda_\alpha$ contributes to determine an imbalance in the pressure across the HY region, as it pushes molecules in the subdomain where Helmholtz free energy is locally lower [49, 50].

To overcome these effects and enforce a uniform density profile, it is possible to introduce a new term in the Hamiltonian:

$$H_\Delta = H - \sum_{\alpha=1}^{N} \Delta H(\lambda(\mathbf{R}_\alpha)) \tag{2.5}$$

This term acts separately on each molecule in the system and plays two roles: it removes, on average, the drift force, and enforces a uniform density profile by imposing, in each subdomain, the pressure at which each model has, separately, the correct density. In the following we discuss the computational techniques employed to parametrize the term $\Delta H(\lambda)$.

Compensation of the drift force (pressure route)

In order to remove, on average, the effect of the drift force, the compensation term $\Delta H(\lambda)$ has to satisfy the relation:

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_\alpha} = \left\langle \left[V_\alpha^{AT} - V_\alpha^{CG}\right] \right\rangle_{\mathbf{R}_\alpha} \tag{2.6}$$

If this is the case, the total drift force resulting from the modified Hamiltonian reads:

$$\hat{\mathbf{F}}_\alpha^{dr} = \left(V_\alpha^{AT} - V_\alpha^{CG} - \left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_\alpha} \right) \nabla\lambda\left(\mathbf{R}_\alpha\right) \tag{2.7}$$

and by construction $\langle \hat{\mathbf{F}}_\alpha^{dr} \rangle = 0$. It has been shown [49, 62] that by compensating the drift force the hydrostatic pressure is uniform across the whole simulation domain, while in each of the two regions the densities may differ, as they are the equilibrium ones at that pressure according to the equation of state of each model. To compute the appropriate value of the compensation function it is possible to perform a Kirkwood thermodynamic integration (KTI) [64], since, as it has been demonstrated [49, 62], the potential $\Delta H(\lambda)$ can be approximated in a mean field fashion by the Helmholtz free energy difference between a system with hybrid Hamiltonian at resolution λ and the reference (CG) system with $\lambda = 0$. However, this procedure requires a free energy calculation just to parametrize the compensation term, and its accuracy can be limited when the correlations within the hybrid region are too strong.

A more effective strategy is to compute and balance the drift force locally and parametrize the compensation *on the fly* within an iterative scheme [62]. The HY region is discretized according to the resolution λ in a number N_b of bins of width $\delta\lambda = 1/N_b$. For each molecule α in a given bin $i = \texttt{floor}[\lambda(\mathbf{R}_\alpha)/\delta\lambda]$ of the HY region, the V_α^{CG} and V_α^{AT} terms are computed and accumulated in separated local variables V_i^{CG} and V_i^{AT}, respectively; also variables N_i^{CG} and N_i^{AT} are defined to keep track of the number of molecules present in the bin. These computations are performed simultaneously and in the same routine where the CG and AT forces are calculated (See 2.7.3 and 2.7.4).

This procedure is carried out for all molecules within the HY region and continues for a time interval of duration Δt. At the end of the n-th interval the average AT and CG potential terms are computed, defined as $\bar{V}^R[i, n] = V_i^R/N_i^R$, where the index $R \in \{AT, CG\}$ specifies the resolution. The variables V_i^R, N_i^R are emptied and the averaging procedure continues. The average values calculated at the end of the n-th interval Δt are employed to compute the running average $\mathcal{V}_{i,n}^R$ of the terms, that is:

$$\mathcal{V}_{i,n+1}^R = \frac{n\ \mathcal{V}_{i,n}^R + \bar{V}^R[i, n]}{n+1} \tag{2.8}$$

where n is initialized at 0 and $\mathcal{V}_{i,0}^R = 0$.

For $n > 0$, the running average terms are employed to compute separately the different components of the force needed to compensate the drift force. Specifically, at time t such that $t_0 + n\Delta t < t \leq t_0 + (n+1)\Delta t$ a molecule located in bin i of the HY region will experience the following compensation forces:

$$\mathbf{F}^R_{\alpha,i} = s\ \mathcal{V}^R_{i,n} \nabla\lambda\left(\mathbf{R}_\alpha\right) \tag{2.9}$$

where $s = +1$ if $R = AT$ and $s = -1$ if $R = CG$. At each time step this force is spread to the atoms of the molecule proportionally to the relative mass of the atom over the mass of the molecule (see equation 2.23).

The running average update continues until the compensation forces have converged to a stable value in each bin i; after this point, the update is interrupted and the resulting compensation is given by a time-independent, resolution-based force field that can be integrated to compute the corresponding contribution $\Delta H(\lambda)$ to the total energy of the system.

Compensation of the density imbalance (density route)

The application of the compensation of the drift force in the HY region enforces a uniform pressure profile across the whole system. However, due to the different equations of state of the AT and CG models, at equilibrium a density gradient between the two main subdomains will arise. When one needs a uniform density profile in the simulation box, it is necessary to modify the compensation term ΔH in order to establish, in each subregion, the appropriate pressure at which the different models attain the same density.

This correction can be obtained in an iterative scheme dubbed *thermodynamic force* calculation [51], consisting in successively applying to the molecules in the HY region a force proportional to the density gradient:

$$\mathbf{F}^{th}_{n+1} = \mathbf{F}^{th}_{n} + \frac{c\nabla\rho_n(x)}{\rho^*} \tag{2.10}$$

where the prefactor c has the units of energy and scales the magnitude of the force, ρ^* is the reference density, and ρ_n is the density profile computed at step n of the iterative procedure. The calculation and application of this force has to be iterative because a single step will not be sufficient to flatten the density profile; however, the convergence to a uniform density is guaranteed by the fact that the scheme has a fixed point when $\nabla\rho = 0$. We note, in passing, that the sum of the contributions obtained from Eqs. 4.9 and 4.10 provides a force whose integral corresponds to the difference of the chemical potentials between the AT and CG domains. This method not only ensures the same density in the two subdomains, but also leads to a flat density profile also in HY region [49, 51, 62].

In general, the procedure to compute the thermodynamic force consists in an equilibration phase of the simulation setup where no compensation is applied (with the possible exception of the drift force compensation), followed by a production run during which an accurate density profile is computed. The latter has to be sufficiently smooth so to employ its numerical gradient as a force in the following simulation, which will provide the new density profile and so on. When the density is deemed to be uniform within a pre-established tolerance, the iterations are interrupted, and the compensation force is given by the sum of the terms computed up to that point.

As already mentioned, this scheme has the advantage of "working by default", since the new terms of the force systematically reduce the density imbalance and the amplitude of the next correction with it. However, this simple approach necessitates a possibly very large number of full simulations employed to compute the density profiles at each iteration stage with sufficient accuracy. Here we make use of an improved strategy to compute the appropriate density compensation, which is iterative as the regular one but is performed *on the fly*, and requires substantially less time.

Also in this approach the thermodynamic force is iteratively computed as the numerical gradient of the density and applied to the molecules in the HY region. The difference lies in the fact that the measurement of the density profile is performed on a very short time interval $\Delta T = \nu \delta t$, where δt is the integration time step and ν is an integer number of the order of $\sim 10^2 - 10^3$. The force is thus obtained according to Eq. 4.10.

The advantage of this scheme is that the small deviations of the density from a uniform profile are immediately corrected for, and the system has no time to equilibrate into a state of substantial density imbalance. However, it is obvious that the density profile computed in the small time interval ΔT would be too noisy to make any use of its numerical gradient. The solution to this problem is to convolute the position of the center of mass of a molecule with a Gaussian function with a half-width $\sigma/2$ comparable with the typical excluded volume radius of the molecules, so that the coordinates are spread on a wider range of bins rather than a single one. One thus has that the density in the bin i covering the coordinate range $[x_i, x_{i+1}]$ in a specific simulation frame is computed as:

$$\begin{aligned} \hat{\rho}_i &= \sum_{\alpha} \frac{1}{A} \int_{x_i}^{x_{i+1}} dy \ \exp\left[-\frac{(y - x_\alpha)^2}{2\sigma^2} \right], \qquad (2.11) \\ A &= \int_{-l}^{l} dy e^{-\frac{y^2}{2\sigma^2}} \end{aligned}$$

The parameter l, whose appropriate value depends on the specific system under examination, controls the range of the Gaussian function; a sensible choice is to set $l = 2.5\sigma$.

2.2.2 Reaction Field

An alternative method to the Ewald summation scheme, aiming at treating electrostatics interactions, assumes a homogeneous polar fluid beyond a cutoff sphere enclosing an atom i. The charge distribution within the cavity polarizes such a fluid, and this polarization in turn influences the charge in the sphere (reaction field) [70,126]. The Coulomb potential is modified as:

$$V_{RF}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \left[1 + \frac{\epsilon_{RF} - 1}{2\epsilon_{RF} + 1} \left(\frac{r_{ij}}{r_c} \right)^3 \right] , \tag{2.12}$$

where r_{ij} is the inter-atomic distance, q the electric charge, ϵ_0 vacuum permittivity and r_c the cutoff distance. This expression depends on knowing beforehand macroscopic information of the system, namely, its permittivity ϵ_{RF}. Discontinuous jumps in energy occur when particles enter/leave the sphere of another particle. To tackle this problem, expression (2.12) has to be attenuated to zero near the cutoff radius. In particular [127]:

$$V_{RF}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \left[1 + \frac{\epsilon_{RF} - 1}{2\epsilon_{RF} + 1} \left(\frac{r_{ij}}{r_c} \right)^3 \right] - \frac{q_i q_j}{4\pi\epsilon_0 r_c} \frac{3\epsilon_{RF}}{2\epsilon_{RF} + 1} . \tag{2.13}$$

Finally, the force acting on atom i, derived from eq. (2.13), reads:

$$\mathbf{F}_{RF}(\mathbf{r}_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0} \left[\frac{1}{r_{ij}^2} - 2\frac{1}{r_c^3} \frac{\epsilon_{RF} - 1}{2\epsilon_{RF} + 1} r_{ij} \right] \frac{\mathbf{r}_{ij}}{r_{ij}} . \tag{2.14}$$

The RF method has been extensively used and both its advantages and drawbacks have been widely recognized (for a review see Ref. [121]). So far, it has been the method of choice for adaptive resolution simulations, mostly for practical reasons. However, we find that conditions such as the implicit homogeneity of the system required to describe the neighborhood of every atom in terms of a dielectric function, or the fact that it might be necessary to modify the force field to reach the desired accuracy, limit substantially the number of systems we are able to simulate. For such reasons, we turn our attention to a different method to deal with electrostatic interactions.

2.2.3 Damped shifted force potential (DSF)

The idea behind the DSF method was introduced in Refs. [81,82], where it was demonstrated that for a perfect ionic crystal the effective Coulomb interactions are short ranged. Moreover, when comparing a straight cutoff method with Ewald calculations, electrostatic energies are very accurate for characteristic system-dependent cutoff distances. This tendency is due, in addition to a damped oscillatory behavior, to an almost exact charge neutrality for such particular cutoff spheres.

By combining short-range nature and charge neutrality, a pairwise summation method was introduced [82] and shown to give comparable results to standard ES. However, the use of this method is dubious for molecular dynamics simulations, in particular because of force discontinuities at the cutoff radius. To solve this problem, the DSF method was modified so to give continuous potential and forces everywhere [65], thus becoming a valuable short-range alternative to ES. In DSF, the electrostatic potential between two charges q_i and q_j separated by a distance r_{ij} is given by the following expression:

$$\begin{aligned} V_{\mathrm{DSF}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0} \Bigg[& \frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\mathrm{erfc}(\alpha r_c)}{r_c} \\ & + \left(\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp\left(-\alpha^2 r_c^2\right)}{r_c} \right) (r_{ij} - r_c) \Bigg] , \end{aligned} \tag{2.15}$$

where $r_{ij} \leq r_c$, ϵ_0 is the vacuum permittivity, r_c is the cut-off radius, and α is a damping parameter with dimension of inverse length. $\mathrm{erfc}(r)$ is the complementary error function that takes into account the damping proposed in [82]. The gradient of potential (4.15) gives the force acting on atom i

$$\begin{aligned} \mathbf{F}_{\mathrm{DSF}}(\mathbf{r}_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0} \Bigg[& \frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp(-\alpha^2 r_{ij}^2)}{r_{ij}} \\ & - \frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} - \frac{2\alpha}{\pi^{1/2}} \frac{\exp(-\alpha^2 r_c^2)}{r_c} \Bigg] \frac{\mathbf{r}_{ij}}{r_{ij}} . \end{aligned} \tag{2.16}$$

We emphasise here that the short range character of electrostatic interactions has been confirmed by *ab initio* simulations of water [128]. Moreover, DSF has been successfully applied in simulations where the Ewald method can introduce spurious electrostatic effects [129]. This is the case for polarized systems when dipole-dipole interactions in the simulation box and its replicas are artificially introduced.

2.3 Computed quantities

Our aim is to validate the effectiveness of the DFS method to accurately reproduce the electrostatic interaction experienced by water molecules in the liquid phase, and to show that this approach is perfectly suited to be employed in the framework of an adaptive dual-resolution simulation.

To this end, we will perform a number of analysis of different structural, thermodynamical, and dynamical properties of the liquid, namely: radial distribution functions (RDFs), tetrahedral orientation order parameter, fluctuations of the number of molecules, and velocity autocorrelation functions (VACF). The results obtained in the dual-resolution setup are compared with the same quantities computed in fully atomistic benchmark simulation. In the following, a succinct description of the quantities under examination is provided.

The orientational order parameter is defined for oxygen atoms in water as [130]:

$$q = 1 - \frac{3}{8}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos\psi_{jk} + \frac{1}{3}\right)^2 , \tag{2.17}$$

where for a given oxygen atom i one identifies its four nearest neighbors, and computes the angles ψ_{jk} with vertex i and segments ij and ik. For a single molecule $-3 < q < 1$. By contrast, for a collection of molecules $0 \leq \langle q \rangle < 1$, with 0 corresponding to the ideal gas case and 1 to a perfect tetrahedral network.

Fluctuations of the number of molecules are calculated using the expression:

$$\left.\frac{\Delta^2(N)}{\langle N \rangle}\right|_x = \left.\frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle}\right|_x \tag{2.18}$$

where the subscript x indicates that the simulation box has been divided in slabs of width 10 Å along the X axis. The average $\langle N \rangle$ and standard deviation $\Delta^2(N)$ in the number of molecules have been calculated for such slabs.

Finally, the velocity autocorrelation function (VACF) is defined as [126]:

$$C_{vv}(t) = \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle , \tag{2.19}$$

where $\mathbf{v}_i(t)$ is the velocity of molecule i at time t. To integrate numerically Newton's equations of motion, molecular dynamics simulations rely on discrete time algorithms. Therefore, Eq. (4.17) should be estimated to take into account such discretization. Here we use the discrete estimator described in Ref. [131] where the VACF for the t_m-th time step takes the form:

$$C_{vv}(t_m) = \frac{1}{N_{AA}}\sum_{i=1}^{N_{AA}} \frac{1}{M-m}\sum_{n=0}^{M-m-1} \mathbf{v}_i(t_{n+m}) \cdot \mathbf{v}_i(t_n) , \tag{2.20}$$

with M the total number of time steps and $t_m = m\delta t$, where δt is the integration time step. This expression is constructed in such a way that it includes all possible contributions $\mathbf{v}(t + n\delta t) \cdot \mathbf{v}(t)$ that result from shifting the time origin by m steps. The normalization factor $1/(M - m)$ ensures that the estimator is unbiased. We implemented Eq. (4.18) following the protocol reported in Ref. [132]. Finally, N_{AA} is the number of molecules that always remain within a predefined region of the simulation box. In the case of fully-atomistic simulations, $N_{AA} = N$, the total number of molecules. The error in the calculation of the VACF is given by $2t_{corr}/N_{AA}t_{tot}$ [126], with t_{corr} the correlation time and t_{tot} the total time of the simulation.

2.4 Simulation details

In all case studies unless otherwise stated, there are 10240 water molecules in the simulation box. The time step for the `H-AdResS` as well as the fully atomistic simulations is set to $\delta t = 1fs$. The initial configuration for every simulation setup is extracted from the simulation results of a similar fully atomistic system which is equilibrated for $50ps$ in the isothermal-isobaric ensemble at a temperature $T_0 = 300K$ and pressure $P_0 = 1bar$. The atomistic interactions between the atoms of water molecules are based on the SPC/E model [133–135].

Three different methods have been used to simulate the atomistic Coulomb interactions: particle-particle-particle-mesh (PPPM) Ewald summation [73], reaction field (RF) [69, 70], and damped shifted potential (DSF) [65, 82]. For the latter, everywhere in the present work we used the following parameters, for which the DSF potential best reproduces the RDFs of the reference ES simulations: damping parameter $\alpha = 0.2\text{Å}^{-1}$; cut-off radius $r_c = 12\text{Å}$.

The Weeks-Chandler-Andersen (WCA) [136] potential is used for the interaction of water molecules in the coarse-grained domain:

$$V^{CG}(r) = \begin{cases} 4\epsilon\left[(\sigma/r)^{12} + (\sigma/r)^{6}\right] & r \leq 2^{1/6}\sigma \\ 0 & r > 2^{1/6}\sigma \end{cases} \tag{2.21}$$

where $\epsilon = 1.0$ kcal/mol and $\sigma = 2.2\text{Å}$, which is roughly the excluded volume diameter of the water molecules in the fully atomistic simulations.

In all `H-AdResS` cases, the initial equilibrated configuration is simulated for 100 ps in the Canonical (NVT) ensemble; A uniform temperature profile at 300 K is maintained throughout the system (data not shown) via a Nosé-Hoover thermostat with a damping parameter of 0.1 ps. In this step, the equations of motion are integrated according to the Hamiltonian in Eq. 4.1 in absence of the compensation terms. The presence of the drift force in the HY region leads to a pressure imbalance between two resolutions (see solid blue curve in Fig. 2.5). After the initial 100 ps, the *on-the-fly* calculation of the drift force correction is applied. The resolution range is subdivided in 20 bins of size $\Delta\lambda = 0.05$. The drift force compensation is updated every 1 ps for 150 ps. At the end of this step, the converged compensation force ultimately cancels out the drift force and leads to a uniform pressure across the simulation domains (see red dashed line in Fig. 2.5), while a density gradient persists across the HY region. To enforce a uniform density the *on-the-fly* density balancing method is applied for the next 300 ps. In this step, the length of the simulation box is uniformly discretized into slabs of size $\Delta x = 1.5\text{Å}$ and the thermodynamic force is updated every 0.5 ps. We employed values of $c = 2.0$, $\sigma = 6\text{Å}$ and $l = 12\text{Å}$ for smoothing and scaling the thermodynamic force.

All simulations are performed with the `LAMMPS` simulation package, with the exception of the RF runs, that have been performed on the `GROMACS` [127] platform.

2.5 Results and discussion

In this section we report the results of the validation of the DSF method for the electrostatic interaction in the context of a `H-AdResS` simulation. The first part is devoted to the comparison of the two short-range modifications of Coulomb potential (RF and DSF) with the Ewald summation PPPM scheme, that we take here as our golden standard. Subsequently, we focus on the DSF method and compare the properties of water in the AT region of the `H-AdResS` setup with those measured in an equivalent domain of a reference fully atomistic simulation.

2.5.1 All atom simulations

The three methods here under exam to reproduce electrostatic interactions in MD simulations have been extensively investigated in the past. However, we consider useful to include this simulations here to provide a self-contained validation of their performance within the framework of the `H-AdResS` scheme.

From the point of view of structural properties, PPPM, DSF and RF give identical results. In particular, RDFs for the three cases can be seen in Fig. 2.2 where the three data sets overlap perfectly.

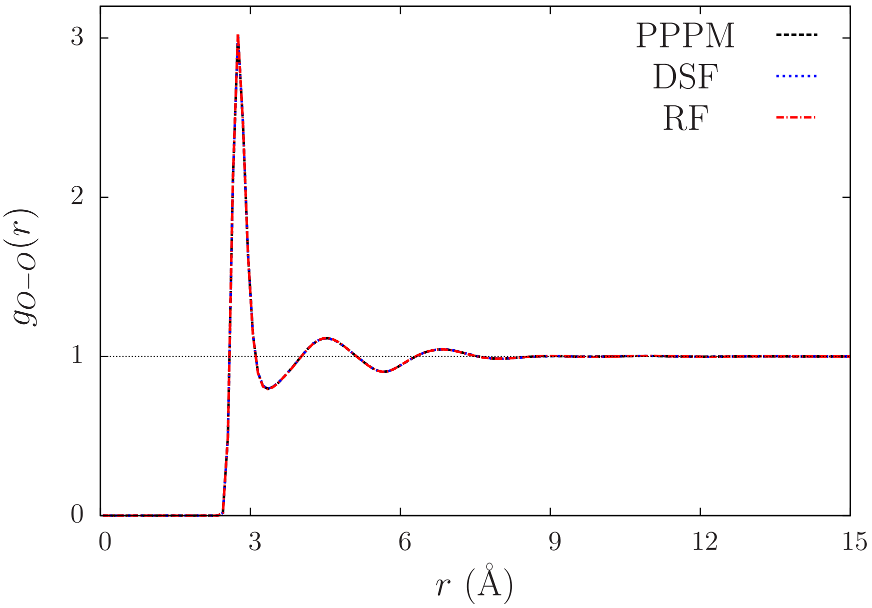


Figure 2.2. Oxygen-Oxygen radial distribution function (RDF), $g_{O-O}(r)$, for SPC/E water using three different approaches to compute electrostatic interactions. Namely, particle-particle particle-mesh (PPPM), damped shifted force potential (DSF) and reaction field (RF).

This is also the case for the orientational order parameter q. Fig. 2.3 shows the normalised distribution of q for the three cases considered where, as expected, a bimodal character is observed [130]. In such a distribution, low values of q, related to angular distorsions, indicate local disorder of water molecules. By contrast, since q is associated to the angular ordering of first neighbours and ignore their radial ordering, high values of q do not necessarily demonstrate a tetrahedral ordering [137].

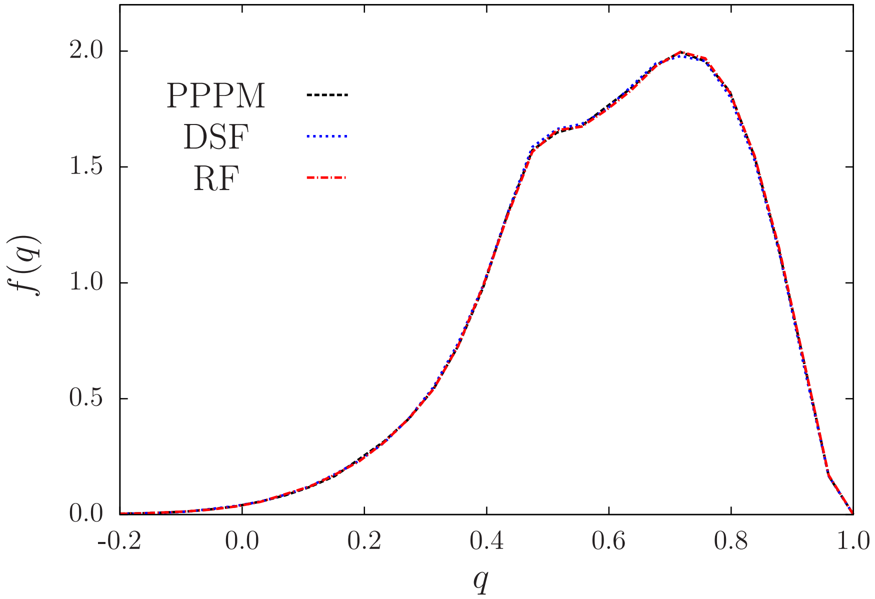


Figure 2.3. Normalised distribution of the orientational order parameter. The fraction of molecules with a given value q is given by $f(q)dq$.

The short-time dynamics of water molecules has been investigated by means of the oxygen VACF. In all three cases under examination we have run a 2 ps long simulation in absence of the thermostat, with a time step of 1 fs and recording velocities every 10 fs. The error in the VACF is $\sim 1\%$, and it is estimated using $2t_{corr}/N_{AA}t_{tot}$ with $t_{corr} = 1$ ps [126]. Consistently with the structural results, also the dynamical (equilibrium) behavior of the system is not affected by the different method employed to treat Coulomb interaction. A subtle deviation of the RF from the other two setups is to be attributed to numerical discrepancies due to the different software (`GROMACS` [127]) employed in the former case.

The reported analysis shows that the DSF scheme is capable of reproducing quantitatively accurate structural and dynamical properties of liquid water, in addition to ionic liquids and other complex systems [65, 138–142]. The method is thus a strong candidate to replace the RF as the “short-range alternative” to Ewald summation in dual-resolution simulation schemes.

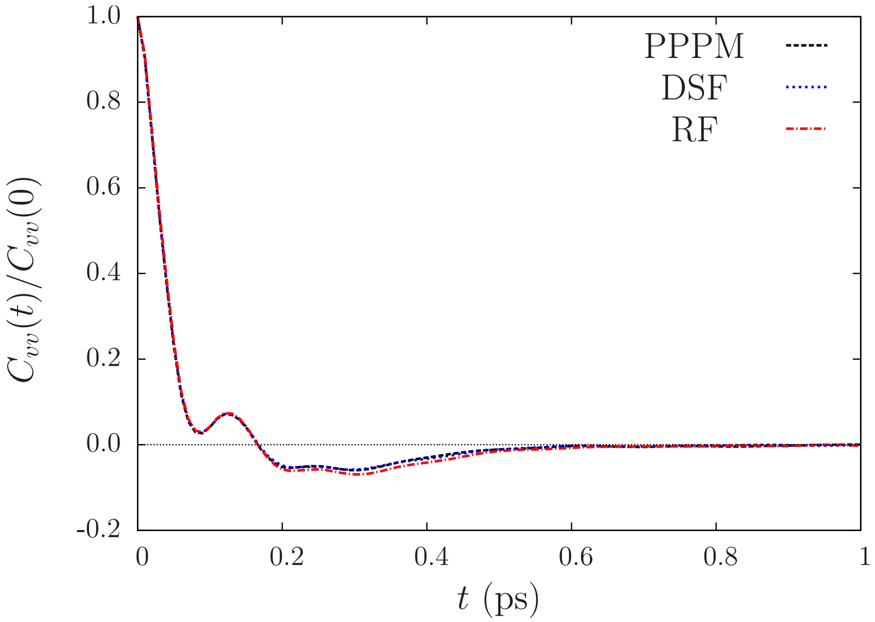


Figure 2.4. Normalized velocity autocorrelation function, $C_{vv}(t)/C_{vv}(0)$, for SPC/E water using three different approaches to compute electrostatic interactions, as in Fig. 2.2.

2.5.2 H-AdResS simulations

The most basic requirement of an adaptive dual-resolution simulation is that the compensation applied to the molecules in the HY region is sufficient to enforce a uniform density profile of the fluid across the simulation domain. That this is the case in the water model under examination is confirmed by the data reported in Fig. 2.5, which provide a clear picture of thermodynamic properties of the `H-AdResS` setup in terms of pressure and density profiles.

Without any compensation (solid blue line Fig. 2.5) the system equilibrates in such a way that both pressure and density (upper and lower panel, respectively) are different in the AT and CG regions, and therefore different from the reference values. The situation changes if the drift force is counterbalanced, or in other words, a pressure correction is included which removes on average the drift force. The profiles in this case show that the pressure is now the same in both subregions; the density has improved thanks to the removal of the extra pressure exerted on the molecules in the HY region, however it is still higher in the AT region (red dashed line Fig. 2.5).

Finally, after the application of the thermodynamic force to compensate the density imbalance (dotted green line Fig. 2.5), we observe that in this setup the pressure is different in the two bulk subdomains whereas the density is uniform, with a deviation of one percent from the reference value. In the following, we analyze our results for the case where both compensations are applied.

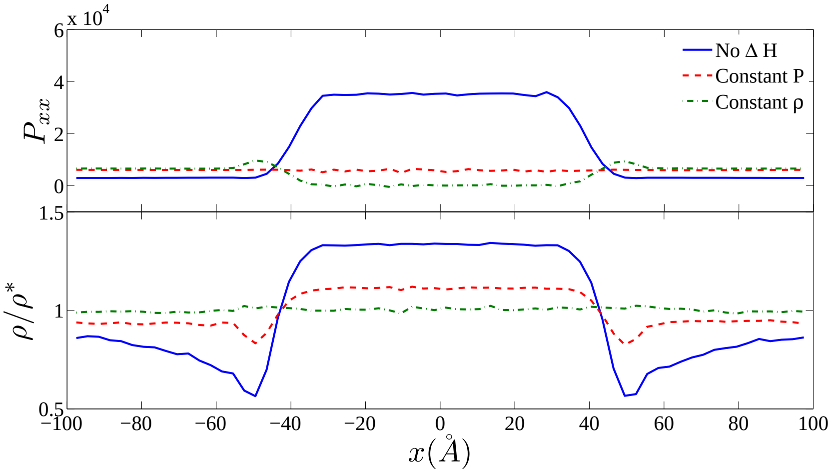


Figure 2.5. Pressure (top) and density (bottom) profiles for different `H-AdResS` setup. The blue (solid) curves represent `H-AdResS` setup with no compensation in the Hamiltonian. The red dashed curves illustrate the setup with a constant-pressure route, and the results of the setup with constant-density route are shown in green dash-dotted curves.

From the structural point of view, the fully atomistic and `H-AdResS` simulations provide perfectly compatible results. This is evident from Fig. 2.6, where we report both sets of multicomponent RDFs which exhibit an excellent overlap. The normalized distributions of the orientational order parameter for the fully atomistic and dual-resolution cases are shown in Fig. 2.7, and they clearly overlap with great accuracy. The bimodal profile, observed in both cases, indicates that the tendency of water molecules to form ordered structures is well preserved in `H-AdResS` simulations.

A relevant property that has to be correctly reproduced in the AT region in order to guarantee that the thermodynamics of this subdomain is representative of the reference simulation is the profile of fluctuations of the number of particles [51, 143]. Profiles of fluctuations for fully atomistic and `H-AdResS` simulations, reported in Fig. 2.8, coincide in the AT subdomain. Beyond the hybrid region, as expected, the profile of fluctuations increases, due to the different isothermal compressibility of the CG model.

Finally, concerning short-term dynamical properties, we confirm that fully atomistic and `H-AdResS` simulations display consistent VACFs. The measurement has been performed only in the AT region in both cases, hence the error in the VACF ($\sim 3\%$) is higher than in the measurement performed for the fully atomistic cases with different

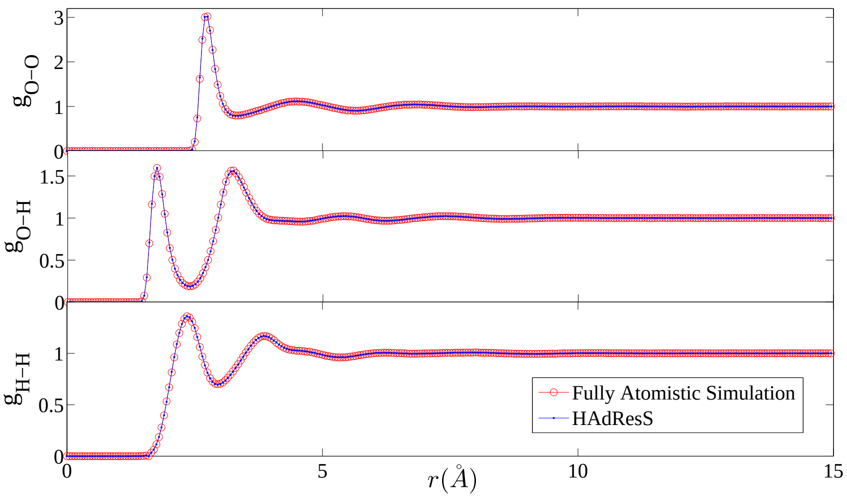


Figure 2.6. RDFs of water molecules at pressure $P_0 = 1bar$ and temperature $T_0 = 300K$ in two different simulation setups: fully atomistic simulation (red line with open circles) and `H-AdResS` (blue line with dots). From top to bottom, the plots show oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen RDFs. The DSF damping parameter is set to $\alpha = 0.2\text{Å}^{-1}$, and the cut-off radius is $R_c = 12\text{Å}$.

electrostatics (Fig. 2.4). Nonetheless, differences observed in both fully atomistic and `H-AdResS` cases are larger than error bars ($\sim 10\%$) in the 0.2 - 0.4 ps interval. To explain this aspect, let us recall that in `H-AdResS` simulations as presented here, the density in the atomistic region is approximately 1% above the reference density. Therefore, differences in VACF appear because we compare systems with slightly different densities, as indicated by further fully atomistic simulations performed at higher density (results not shown). Interestingly, this strong density dependence of the VACF is the matter of recent discussion [144].

2.6 Conclusions

The appropriate treatment of electrostatic interaction in computer simulations of soft and biological matter is still an open problem. One of the most challenging issues is the possibility to find a balance between the accurate description of the potential and a computationally economic implementation of the corresponding model. If, on the one hand, the original decomposition of the interaction in short- and long-range terms devised by Ewald has been substantially optimized, on the other hand some of the undesired artifacts due to the unphysical periodicity implicit in this treatment are

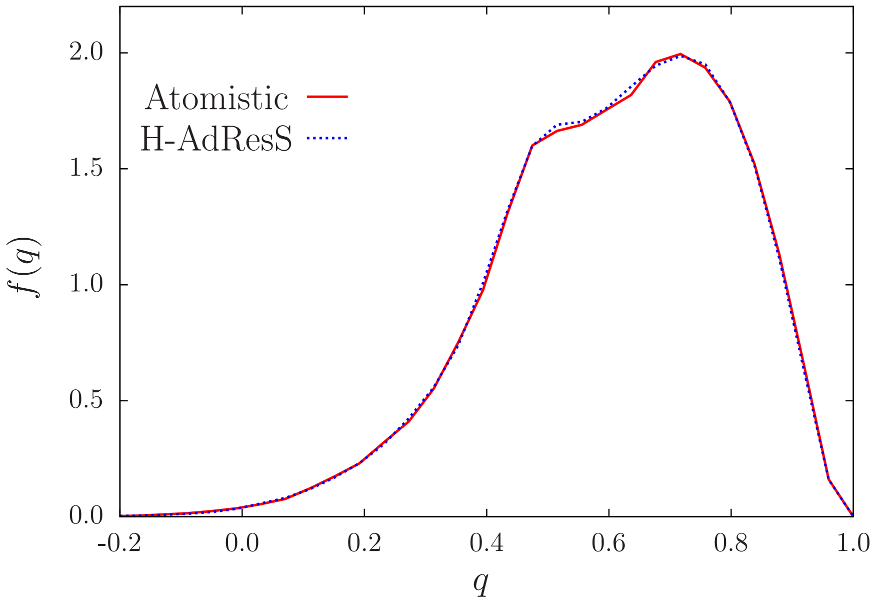


Figure 2.7. Normalized distribution of the orientational order parameter for fully-atomistic and `H-AdResS` simulations. Only water molecules in the interval $-25\mathring{A} < x < 25\mathring{A}$ were considered for the calculation.

still cause of concern in specific systems. In the context of adaptive, dual-resolution simulations Ewald summation-based schemes are in any case practically unviable, as the long-range term would have to be computed on models featuring substantially different physical properties. Alternative modifications of Coulomb potential, such as the reaction field approach, circumvent these problems and provide a computationally effective short-range interaction; also in this case, however, there are some limitations originating in the underlying assumption of a uniform medium, which do not always apply.

The DSF potential, on the other hand, has been shown to reproduce the physical (structural and dynamical) properties of many charged systems with high accuracy, albeit being short-range and without the necessity of a pre-parametrization based on emergent properties of the system, e.g. the dielectric constant. This method is thus ideally suited to be employed in the context of dual-resolution simulations, and its validity has been here demonstrated by means simulations of liquid water. Specifically, the region of the dual-resolution setup where the fluid was modeled with atomistic resolution showed quantitatively consistent properties compared to a reference, fully atomistic simulation.

The possibility to accurately reproduce the effects of Coulomb potential in the context of adaptive resolution simulations without the need to parametrize the atomistic

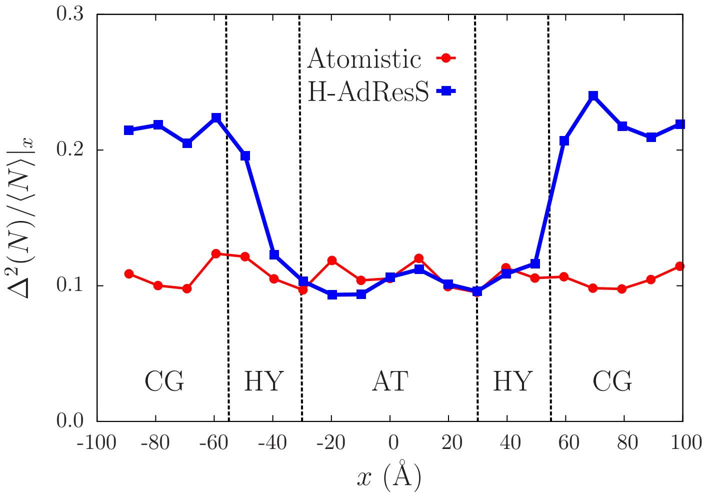


Figure 2.8. Profile of fluctuations of the number of molecules calculated along the x-axis. Red circles correspond to a benchmark fully atomistic simulation of water molecules. Blue squares correspond to an `H-AdResS` simulation of a slab of water molecules enclosed by a WCA system. Fluctuations are calculated using sub-volumes of size $10 \times 40 \times 40$Å. Vertical lines are guides to the eye and indicate the location of the AT, HY, and CG regions of the `H-AdResS` setup.

force field, as it would be the case when employing the reaction field, thus opens the way to the efficient modeling and simulation of complex systems in which electrostatic interaction is known to play a primary role, for example ions in solutions, ionic liquids, and nucleic acids. Additionally, the flexible and efficient implementation of the `H-AdResS` method in the `LAMMPS` simulation package, equipped with the DSF method for Coulomb potential, provides a broad community with an effective instrument to investigate soft and biological matter.

2.7 Appendix: LAMMPS Implementation

We report here the basic technical details of the `H-AdResS` implementation in the `LAMMPS` simulation package. The software as well as a more detailed documentation can be downloaded at the web page:
`http://www2.mpip-mainz.mpg.de/~potestio/software.php` or cloned from the Git repository `https://github.com/hadress/lammps/tree/USER-HADRESS`. The modified `LAMMPS` version featuring `H-AdResS` is currently under review for being merged in the main `LAMMPS` branch.

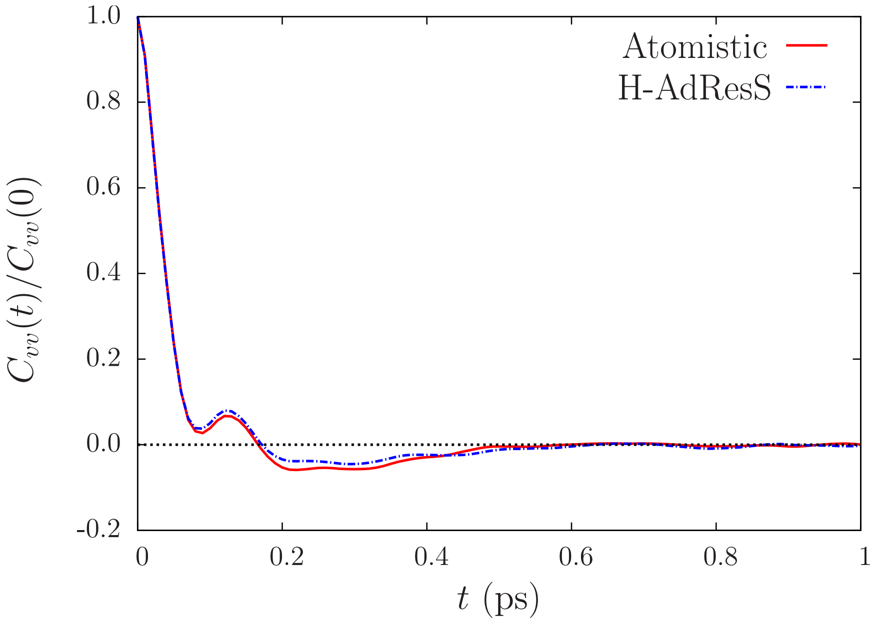


Figure 2.9. Normalized velocity autocorrelation function for fully-atomistic (full red line) and H-AdResS (dashed blue line) simulations of a 2 ps trajectory with a time step of 1 fs. Water molecules in the interval $-25\text{Å} < x < 25\text{Å}$ were considered for the calculation.

2.7.1 `H-AdResS` Atom style

We introduced an atom style called `full/hars` in which an atom i, in addition to `LAMMPS`'s indigenous atom properties (e.g. coordinates $\mathbf{x}_i$, velocities $\mathbf{v}_i$, charge $\mathbf{q}_i$), is provided the following `H-AdResS`-specific attributes:

- λ_i: the resolution of the atom in the system as determined by the value of the switching function computed on the center of mass coordinate of the molecule to which atom i belongs
- $\nabla\lambda_i$: the gradient of the switching function
- $\mathbf{x}_\alpha^{CG}$: the center of mass coordinate of molecule α to which atom i belongs
- $\texttt{Rep}_i$: the *representation* flag indicating which atom in the molecule carries the information pertaining the whole molecule
- $\texttt{MolType}_i$: the molecule type index specifying the CG model parameters of molecule when in the low resolution region

All these properties are assigned and set by the two files `atom_vec_full_hars.cpp/h`.

2.7.2 Fixing particle resolutions

As the atom passes through different resolutions, the resolution function λ_i and its gradient $\nabla\lambda_i$ have to be updated at each time step. This step is carried out within a fix file called `fix_lambdah_calc.cpp/h`: here, the center of mass coordinates of each molecule, $\mathbf{x}_\alpha^{CG}$, is calculated and then spread to all atoms of that molecule. Depending on the (user-specified) shape of the hybrid region, also the switching function and its gradients are computed based on the molecule position and transmitted to the corresponding atoms.

2.7.3 Calculating coarse-grained pairwise interaction

The generalized pairwise coarse-grained potentials and forces are computed based on Eqs. 4.3 and 5.8. Depending on the type of potential, a specific interaction file (with corresponding header) is introduced. For this study, the interaction between the water molecules in the CG region is given by a WCA potential which, in turn, is obtained by assigning a specific set of parameters to a Lennard-Jones potential; accordingly, the two files `pair_lj_cut_hars_cg.cpp/h` overseeing the computation of Lennard-Jones interaction in the CG and HY regions are employed. The computation of the forces among neighboring molecules is restricted to the representative atoms provided with the details about the molecule properties. Hence, for molecule α the coarse-grained force due to the interaction with all other neighboring β molecules is computed as:

$$\begin{aligned} \mathbf{F}_\alpha &= \sum_{\beta,\beta\neq\alpha} \left\{ \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2}\right) \mathbf{F}_{\alpha|\beta}^{CG} \right\} \\ &+ V_\alpha^{CG} \nabla_\alpha \lambda_\alpha \end{aligned} \tag{2.22}$$

The force $\mathbf{F}_{\alpha i}$ acting on atom i^{th} is obtained by scaling the molecular force $\mathbf{F}_\alpha$ by each atom's mass m_i divided by the whole molecule's mass M_α:

$$\mathbf{F}_{\alpha i} = \frac{m_i}{M_\alpha} \mathbf{F}_\alpha \tag{2.23}$$

2.7.4 Calculating atomistic pairwise interaction

A procedure similar to the one discussed for the CG potential is carried out for the atomistic part of the interaction. Depending on the specific atomistic force field, two files need to be written and added to `LAMMPS`'s source directory. For our study we employed Lennard-Jones and DSF Coulomb interactions, and the files `pair_lj_cut_coul_dsf_hars_at.cpp/h` have been accordingly created. The pairwise atomistic interactions between neighboring atoms are calculated through:

$$\mathbf{F}_{\alpha i} = \sum_{j\in\beta, i\in\alpha, i\neq j} \frac{\lambda_\alpha + \lambda_\beta}{2} \mathbf{F}^{AT}_{\alpha i|\beta j} \tag{2.24}$$

Since drift forces are acting on the molecules, an additional force contribution is added to each molecule:

$$\mathbf{F}_{\alpha} = -\left(\sum_{i\in\alpha} V^{AT}_{\alpha i}\right) \nabla_\alpha \lambda_\alpha \tag{2.25}$$

Afterwards, the computed molecular drift force is spread to the atoms of molecule α based on equation (2.23). For the intra-molecular interactions, such as bond and angle potentials, there is no need to modify the corresponding parts of the current `LAMMPS` implementation.

2.7.5 Speedup

The reduced number of interactions in the CG region and their shorter range enable a reduction of the computational cost of the simulation. In order to quantitatively demonstrate this gain, we have performed fully atomistic as well as dual-resolution simulations of water systems of increasing size, and compared their run time. In Fig. 2.10 we show the time necessary to these different setups to perform the same number of integration steps, namely 10^5, and the corresponding speedup, calculated as the ratio of the time of a fully atomistic run over the corresponding dual-resolution one. The simulated system is the same discussed in the Methods section; the AT interactions are given by DSF and Lennard-Jones potentials with a cutoff radius of 12Å, and the CG model is a purely repulsive WCA. The widths of the AT and HY regions are kept constant, while the size of the CG domains is systematically increased; the probed extensions of the CG domain are thus approximately 90, 180, and 270Å. By comparing the simulation time of fully atomistic and dual-resolution setups it is possible to appreciate that the latter has a very weak linear growth as a function of the CG domain size, indicating that the computational cost of the simulation is almost completely determined by the computation of AT interactions.

2.8 Acknowledgments

MH, DD and RP acknowledge financial support under the project SFB-TRR146 of the Deutsche Forschungsgemeinschaft. RCH acknowledges financial support from the

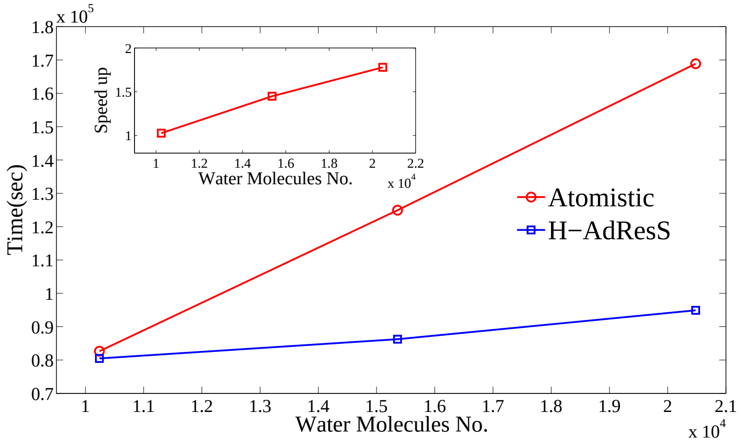


Figure 2.10. Computation time in seconds for fully atomistic simulations (red circles) and `H-AdResS` (blue squares) is shown for three different simulation box sizes. For the case of `H-AdResS`, the length of the atomistic and hybrid region are kept constant as the length of the simulation box increases. The speedup which is calculated by the computation time ratio of fully atomistic simulation to `H-AdResS` is shown in onset.

Alexander von Humboldt Foundation. The authors are indebted with Kostas Daoulas for an attentive reading of the manuscript and useful comments.

3. Spatially Resolved Thermodynamic Integration: An Efficient Method To Compute Chemical Potentials of Dense Fluids

This chapter is a research article that has been published in the Journal of Chemical Theory and Computation.

Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio
Spatially Resolved Thermodynamic Integration: An Efficient Method To Compute Chemical Potentials of Dense Fluids
J. Chem. Theory Comput., 2018, 14 (7), pp 34093417
DOI: 10.1021/acs.jctc.8b00002
©2018 American Chemical Society

Abstract

Many popular methods for the calculation of chemical potentials rely on the insertion of test particles into the target system. In the case of liquids and liquid mixtures, this procedure increases in difficulty upon increasing density or concentration, and the use of sophisticated enhanced sampling techniques becomes inevitable. In this work we propose an alternative strategy, spatially resolved thermodynamic integration, or SPARTIAN for short. Here, molecules are described with atomistic resolution in a simulation subregion, and as ideal gas particles in a larger reservoir. All molecules are free to diffuse between subdomains adapting their resolution on the fly. To enforce a uniform density profile across the simulation box, a single-molecule external potential is computed, applied, and identified with the difference in chemical potential between the two resolutions. Since the reservoir is represented as an ideal gas bath, this difference exactly amounts to the excess chemical potential of the target system. The present approach surpasses the high density/concentration limitation of particle insertion methods because the ideal gas molecules entering the target system region spontaneously adapt to the local environment. The ideal gas representation contributes negligibly to the computational cost of the simulation, thus allowing one to make use of large reservoirs at minimal expenses. The method has been validated by computing excess chemical potentials for pure Lennard-Jones liquids and mixtures, SPC and SPC/E liquid water, and aqueous solutions of sodium chloride. The reported results well reproduce literature data for these systems.

3.1 Introduction

An accurate estimation of the chemical potential (μ) is essential to understand many physical and chemical phenomena [145, 146]. Consider the study of nucleation processes at the nanoscale as an example: in this context, prototypical systems such as water–alcohol mixtures [147], mineral clusters [148–150] or ions in solution [151, 152] present a challenge to existing computational methods. Even the computation of μ for aqueous table salt is still the subject of intense discussion [153–158].

Because of this, there has been a continuous, decades long effort to compute free energy differences and, in particular, chemical potentials [159, 160]. Given a molecular liquid, the free energy difference between a state of N and one of $N + 1$ molecules yields the chemical potential of the substance. There are several methods that implement this strategy, which can be classified [160] in expanded ensembles [161], histogram-reweighting [162–164] and, more important for the present discussion, free energy perturbation methods [165–169] and thermodynamic integration (TI) [64].

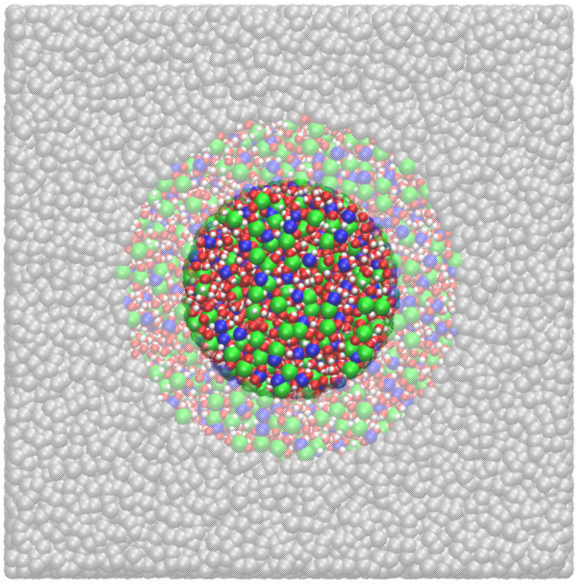


Figure 3.1. Simulation snapshot of a typical `H-AdResS` setup of a system composed of sodium chloride in aqueous solution. Sodium, chlorine, oxygen and hydrogen atoms are represented by blue, green, red and white spheres, respectively. The atomistic, hybrid and ideal gas domains of the system are separated radially from the center of the simulation box, and the following convention is used to distinguish them in the figure: opaque (atomistic), transparent (hybrid) and silver (ideal gas) regions. The resolution of the molecules is determined by the switching function $\lambda(R)$ (Eq. 3.2).

Free energy perturbation methods are based on computing the Zwanzig identity [170], which relates the target free energy difference to a canonical ensemble average over configurations generated by the N-particle Hamiltonian. A single stage application of Zwanzig identity results in the Widom method [165] where frequent test particle insertions are used to calculate free energy differences. To increase the sampling efficiency, multi-stage applications of the Zwanzig identity, e.g Bennett acceptance ratio method (BAR) [166–169], have been developed and are routinely used for simulations involving dense molecular fluids. Because of the necessity to sample a sufficiently large number of trial moves, these methods require a substantial computational effort that increases with density and/or concentration. Moreover, an adequate treatment of systems composed of complex molecules should include several intermediate states [160] or, in general, involve more sophisticated sampling techniques [87].

Thermodynamic integration [64] is among the most widely used methods to compute free energy differences, which allows one to treat rather challenging systems such as solids [171], molecular crystals [172], and molecular fluids [58]. TI relies on the connection between the reference and the target states through a continuum of intermediates, parametrised by a factor combining the Hamiltonians of the two systems. The difference in free energy is obtained from ensemble averages of the appropriate observables computed on such intermediate states. Also in this case, the accuracy of the results depends on generating a sufficient number of system configurations – most of them uninteresting – and on a fine grid of coupling parameter values, which, if binned in an insufficient number of windows, could introduce a systematic bias in the free energy estimates. These shortcomings thus hinder the overall efficiency of the method [169, 173].

It is thus fair to say that the calculation of chemical potentials of dense liquids and complex molecular mixtures remains a challenging task. For these systems, the particle insertion procedure is highly inefficient and in some cases becomes unfeasible. To fill this gap, in this work we introduce a method to compute directly the excess chemical potential of a target system based on the Hamiltonian adaptive resolution scheme `H-AdResS` [49, 50]. In this version of `H-AdResS`, the target atomistic system (AT) is coupled to an ideal gas (IG) bath of point-like particles [143, 174]. The excess chemical potential is obtained by integrating in space the compensation forces necessary to ensure a uniform density profile throughout the whole AT+IG system; therefore we dub the method *spatially resolved thermodynamic integration*, or `SPARTIAN`. `SPARTIAN` is already implemented in a local version of the LAMMPS simulation package [125] and is made freely available[1].

`SPARTIAN` is reminiscent of methods to compute the chemical potential in which inhomogeneities are imposed on the target system [88, 175, 176], or strategies in which the target and reference systems are physically separated by a semi-permeable membrane [177]. In contrast with such methods, thermodynamic equilibrium is

---

[1] A in-house version of the LAMMPS package featuring all method implementations discussed in the present work can be freely downloaded from the webpage `http://www2.mpip-mainz.mpg.de/~potestio/software.php`.

carefully monitored and identified with a uniform density profile across the simulation box. Moreover, finite size effects are made negligible when a substantially large reservoir is coupled to the atomistic region without increasing the computational cost – as it is the case with IG particles since they do not interact. A similar idea was proposed in the context of force-based adaptive resolution simulations [117] where the calculation of effective potential energies is based on the configurations generated by a non-conservative dynamics. Conversely, our method relies on the same Hamiltonian function for both the generation of the dynamics and the computation of the chemical potential, the latter naturally emerging from the formulation of the `H-AdResS` method, as discussed later on. Furthermore, `SPARTIAN` is particularly efficient, because it employs IG particles in the reservoir, and it is flexible because its extension to multicomponent systems is straightforward.

The paper is organised as follows: in the *Method* section, after shortly describing `H-AdResS`, the theoretical basis of `SPARTIAN` is introduced. In the *Results and Discussion* section, excess chemical potential calculations are presented for Lennard-Jones liquids and liquid mixtures, as well as for SPC and SPC/E [133–135] water and for the Joung and Cheatham (JC) sodium chloride force field in SPC/E water [178]. The *Conclusion* section recapitulates the presented work.

3.2 Method

One of the biggest challenges in computational soft matter physics is, arguably, to treat accurately and efficiently the wide range of time and length scales typically encountered when simulating complex molecular systems. One possibility to overcome this problem, in contrast to classical force-fields, consists of using coarse-grained models to access time and length scales usually out-of-range for fully atomistic simulations. However, there are situations in which it is convenient to keep a higher level of detail in a relatively small region within the simulation box (for example when the involved detailed chemistry is relevant) and simultaneously treat the neighbouring region using a computationally more efficient, i.e. coarse-grained, model. Adaptive resolution simulation methods [48–51, 54–57] implemented this strategy as schematically depicted in Fig. 3.1. Specifically, in the `H-AdResS` framework molecular interactions are treated either at the atomistic level in the AT region, fully coarse-grained in the CG region, or as an interpolation of the two in the HY region, in terms of a global Hamiltonian of the form:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \left\{ \lambda_\alpha V_\alpha^{AT} + (1-\lambda_\alpha) V_\alpha^{CG} \right\} \tag{3.1}$$

with $\mathcal{K}$ the total kinetic energy and V^{int} the sum of all intramolecular bonded interactions. The molecule α has resolution given by $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$ associated to the center of mass coordinates $\mathbf{R}_\alpha$. The resolution of a molecule is thus determined by the value of this position-dependent switching function λ_α, taking value 0 in the coarse-grained (CG) region and 1 in the fully atomistic (AT) region, which smoothly interpolates between such values in an intermediate hybrid (HY) region. As illustrated in Fig. 3.1, the geometry of the AT region corresponds to a sphere of radius r_{at}, centered at the origin of coordinates. The HY region is a spherical shell of thickness d_{hy} enclosing the AT region. The switching function plotted in Fig. 3.2 depends on such a geometry, and it is defined by a function of the form:

$$\lambda(r) = \begin{cases} 1 & r \leq r_{at} \\ \cos^2\left(\frac{\pi(r-r_{at})}{2d_{hy}}\right) & r_{at} < r \leq r_{at} + d_{hy} \\ 0 & r > r_{at} + d_{hy} \end{cases} . \tag{3.2}$$

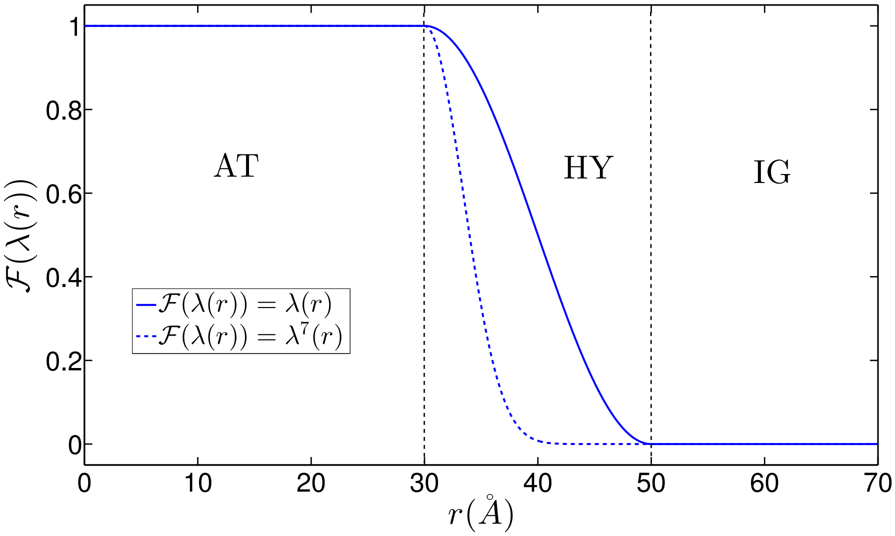


Figure 3.2. Plot of the switching function, re-defined as $\mathcal{F}(\lambda(r)) = \lambda^\nu(r)(\nu \geq 1)$ (see main text), as a function of the radial distance from the center of a cubic simulation box. Two different exponents, $\nu = 1$ (see Eq. 3.2) and 7, have been considered. In Eqs. 4.1-5.13 the function λ, hence with $\nu = 1$, is directly employed to lighten the notation. In this plot, the sizes of the atomistic and hybrid regions are $r_{at} = 30\text{Å}$, $d_{hy} = 20\text{Å}$, respectively.

Non-bonded molecular interactions are described at atomistic or coarse-grained level, and the resulting potential energy contribution for a given molecule α is the result of a weighted sum of two terms, V^{AT} and V^{CG}, defined as:

$$\begin{aligned} V_\alpha^{AT} &\equiv \frac{1}{2} \sum_{\beta,\beta\neq\alpha}^{N} \sum_{ij} V^{AT}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|)\,, \\ V_\alpha^{CG} &\equiv \frac{1}{2} \sum_{\beta,\beta\neq\alpha}^{N} V^{CG}(|\mathbf{R}_\alpha - \mathbf{R}_\beta|)\,. \end{aligned} \tag{3.3}$$

Note that there is no constraint regarding the use of arbitrary, e.g. many body, potentials. However, to lighten the notation, we carry out the following discussion making use of pairwise interactions. The total force acting on atom i of molecule α is given by:

$$\begin{aligned} \mathbf{F}_{\alpha i} &= \mathbf{F}_{\alpha i}^{int} \\ &+ \sum_{\beta,\beta\neq\alpha} \left\{ \frac{\lambda_\alpha + \lambda_\beta}{2} \mathbf{F}_{\alpha i|\beta}^{AT} + \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2}\right) \mathbf{F}_{\alpha i|\beta}^{CG} \right\} \\ &- \left[V_\alpha^{AT} - V_\alpha^{CG}\right] \nabla_{\alpha i}\lambda_\alpha \end{aligned} \tag{3.4}$$

with $\mathbf{F}_{\alpha i}^{int}$ the intramolecular total force on atom i of molecule α. The second term is the sum over all molecules $\beta \neq \alpha$ within cutoff distance of AT and CG forces weighted by the average resolution of the molecule pair (α, β). The origin of the last term in the sum can be traced to the fact that molecules interact depending on their position within the simulation box. This breaking of translational invariance generates a force that acts on molecules located in the HY region, where $\nabla\lambda \neq 0$, and pushes them towards the AT or CG regions depending on the sign of $(V_\alpha^{AT} - V_\alpha^{CG})$.

This drift force thus contributes to a pressure imbalance in the HY region. Furthermore, by joining AT and CG representations of a system using open boundaries, particles will diffuse to stabilise differences in pressure and chemical potential between the two representations. Hence, a non homogeneous density profile appears as a result of molecules being pushed to the region with lower molar Gibbs free energy [49, 50]. To impose a uniform density profile, an extra term is included in the Hamiltonian:

$$H_\Delta = H - \sum_{\alpha=1}^{N} \Delta H(\lambda(\mathbf{R}_\alpha))\,, \tag{3.5}$$

that compensates on average the drift force discussed above and imposes the pressure at which the AT and IG models exhibit the same density. To compensate the drift force, $\Delta H(\lambda(\mathbf{R}_\alpha))$ should satisfy the relation:

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_\alpha} \equiv \mathcal{V}(\lambda_\alpha) = \left\langle \left[V_\alpha^{AT} - V_\alpha^{CG}\right] \right\rangle_{\mathbf{R}_\alpha}\,, \tag{3.6}$$

in such a way that the total drift force becomes:

$$\hat{\mathbf{F}}_{\alpha}^{dr} = \left(V_{\alpha}^{AT} - V_{\alpha}^{CG} - \mathcal{V}(\lambda_{\alpha})\right) \nabla\lambda\left(\mathbf{R}_{\alpha}\right) \tag{3.7}$$

where $\langle\hat{\mathbf{F}}_{\alpha}^{dr}\rangle \equiv 0$. The strategy introduced in Ref. 61, 62 is used to compute $\mathcal{V}(\lambda_{\alpha})$. This method, whose most technical aspects are detailed in the SI, averages over short time intervals the drift force acting upon each molecule species in the hybrid region as a function of the resolution. In between intervals, the computed average is used to update the drift force compensation $\mathcal{V}(\lambda)$, in such a way that correlations between molecules with different instantaneous resolutions are explicitly taken into account.

As anticipated, the drift force is not the only source of density imbalance in the system. The models coupled together in the same setup naturally feature different pressures, and a non-uniform density profile emerges as a consequence of the tendency of the system to equate the pressure imbalance between the two subdomains. To compensate for this density gradient it is customary to introduce a force, dubbed thermodynamic force [51, 61], which, just as the aforementioned free energy compensation, acts only on the molecules in the HY region. This force is obtained through an iterative procedure, with updates proportional to the density gradient:

$$\mathbf{F}_{n+1}^{th} = \mathbf{F}_{n}^{th} + \frac{c\nabla\rho_n(x)}{\rho^*}\,. \tag{3.8}$$

The parameter c modulates the force strength and has units of energy, ρ^* is the reference density, and ρ_n is the density profile computed at the n-th step of the iteration. This procedure converges to a uniform density profile throughout the simulation box when $\nabla\rho = 0$. In Fig. 3.3, converged density profiles for `H-AdResS` simulations of sodium chloride with the JC force field in SPC/E water [178] are presented to illustrate this point.

The total force acting on the molecules in the hybrid region is the sum of the compensation needed to cancel the drift force plus the thermodynamic force, hence:

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_{\alpha}} = -\mathcal{V}(\lambda_{\alpha})\nabla\lambda(\mathbf{R}_{\alpha}) + \mathbf{F}_{\alpha}^{th} \tag{3.9}$$

from which we obtain:

$$\Delta H(R_b) = -\int_{R_a}^{R_b} dR\left[-\mathcal{V}(\lambda)\nabla\lambda(R) + \mathbf{F}^{th}(R)\right]\,, \tag{3.10}$$

where $R_a = r_{at} + d_{hy}$ and $R_b = r_{at}$. Eq. 5.13 allows one to compute the free energy compensation necessary to the system to attain a uniform density profile.

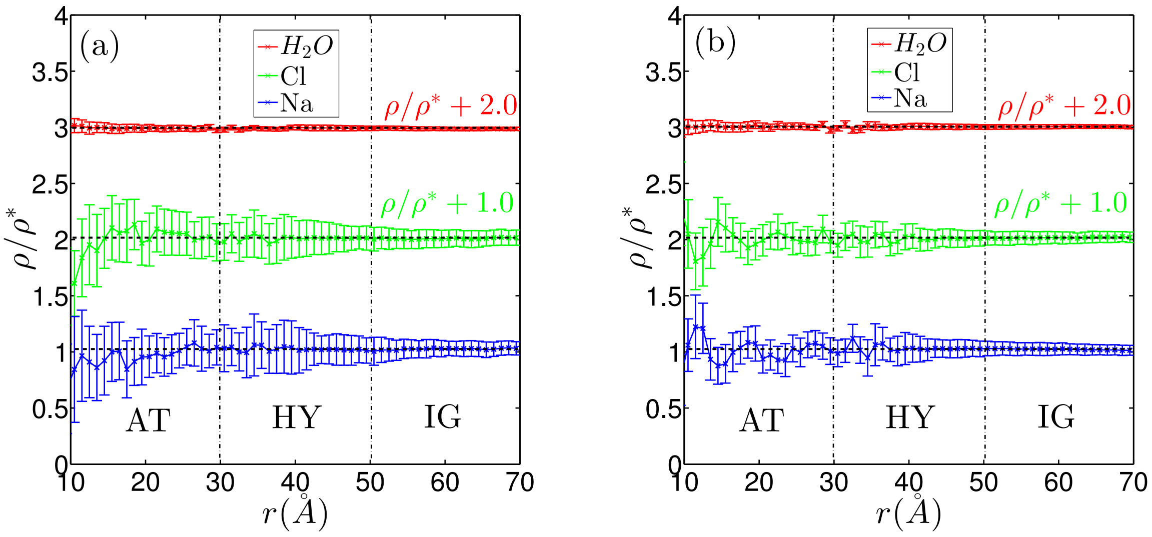


Figure 3.3. Normalized and shifted density profiles of sodium chloride solutions as a function of the radial distance from the origin of the spherical atomistic region, for molalities $m = 3.0$ (a) and $m = 7.0$ (b). The increasing error bar sizes for distances approaching the origin are consistent with the reduced number of molecules available in small spherical shells.

The compensation term to the Hamiltonian has a simple and fundamental physical meaning, that is, it is the difference in chemical potential between the AT and CG regions [49, 50]:

$$\Delta H(\lambda(R)) \equiv \frac{\Delta G(R)}{N} = \Delta\mu(R)\,, \tag{3.11}$$

with $\Delta G/N$ being the molar Gibbs free energy. Note that all quantities appearing in Eq. 5.2 are functions of the molecule's position R: indeed, all free energies and chemical potential differences are computed with respect to a reference given by $R = R_a \equiv \lambda = 0$, that is, with respect to a CG model of arbitrary nature and complexity. If these functions are computed for $\lambda = 1$, one obtains the free energy / chemical potential difference between AT and CG model.

This is precisely the core of the method proposed here: in a nutshell, `H-AdResS` is equivalent to a spatially resolved Kirkwood thermodynamic integration [49, 50]. Therefore, it is possible to calculate $\Delta\mu$ between a target (AT) and a reference (CG) system coexisting at the state (ρ^*,T) by varying λ, a coupling parameter of the global Hamiltonian of the system, across the interface. However, to compute the chemical potential of the target AT system, it is necessary to know the one of the reference CG system. To circumvent this extra step, we couple the AT target system to a bath of ideal gas (IG) particles [143]. In this case, the global Hamiltonian of the system reduces to:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \mathcal{F}(\lambda_\alpha) V_\alpha^{AT}, \tag{3.12}$$

since $V_\alpha^{CG} \equiv 0 \; \forall \alpha$. We have introduced, in Eq. 3.12, a modification to the switching function λ, which has been replaced by a function $\mathcal{F}(\lambda) = \lambda^\nu$ ($\nu \geq 1$). Similarly to λ, $\mathcal{F}(\lambda)$ takes values between 0 and 1, however it has a faster decay to zero as it approaches the IG region. This is required because it might happen that two molecules come extremely close to each other when they are both located near the IG/HY interface. The choice $\nu = 7$ is sufficient to smooth out such divergent interactions and avoid the systematic sampling of huge potential energy values which might affect the calculation of free energy compensations [179]. Furthermore, as it is depicted in Fig. 3.2, by increasing the exponent ν the effective boundary between hybrid and CG region moves deeper towards the AT domain, leading to a more stable and controlled thermodynamic force convergence (see Eq. 5.12). A more detailed discussion of the properties of the switching function in the context of Hamiltonian-based adaptive resolutions can be found in Refs. [61, 143, 180] as well as in the Supporting Information.

In addition to this, the potential of the high resolution region is capped at a distance $\hat{r}$ to avoid large forces resulting from overlapping molecules:

$$V^{AT}(r) = \begin{cases} V^{AT}(\hat{r}) - \left.\frac{\partial V^{AT}}{\partial r}\right|_{r=\hat{r}} (r - \hat{r}) & r < \hat{r} \\ V^{AT}(r) & r \geq \hat{r} \end{cases} . \tag{3.13}$$

For the systems and simulation conditions considered here, these overlapping events are anyways extremely rare, with the highest probability not exceeding $2 \cdot 10^{-4}$ (a plot showing the probability distribution of capping events as a function of the switching function value can be found in the Supporting Information). However, for the sake of stability, they need to be identified and removed from the simulation. We verified that this capping does not change appreciably the thermodynamical nor the structural properties of the system. In particular, we performed fully atomistic simulations of SPC/E water with and without capping potentials and we do not observe any significant difference in the RDFs (data not shown). More important for the present discussion, capping the potential does not affect the calculated values of free energies and chemical potentials. The high energy contributions resulting from the excluded volume are located in the tail of the energy distribution of the system and have an accordingly small effect.

From the Hamiltonian given by (3.12), the total force acting on the atom i of the molecule α whose resolution is λ_α is given by:

$$\begin{aligned}\mathbf{F}_{\alpha i} &= \mathbf{F}_{\alpha i}^{int} \\ &+ \sum_{\beta,\beta\neq\alpha}\left\{\frac{\mathcal{F}(\lambda_\alpha)+\mathcal{F}(\lambda_\beta)}{2}\mathbf{F}_{\alpha i|\beta}^{AT}\right\} \\ &- V_\alpha^{AT}\frac{\partial\mathcal{F}}{\partial\lambda}|_{\lambda=\lambda_\alpha}\nabla_{\alpha i}\lambda_\alpha\,. \end{aligned} \tag{3.14}$$

In addition to the obvious computational advantage of using an IG over a standard, i.e. interacting, CG model, by coupling an AT model with an IG bath of particles at a thermodynamic state with density ρ^* and temperature T it is possible to compute automatically the excess chemical potential from $\Delta H(\lambda(\mathbf{R}_\alpha))$. Moreover, in `H-AdResS` this result can be immediately extended to multicomponent systems [50], thus:

$$\mu_{ex}^{i} = \Delta H_i(\lambda(\mathbf{R}_{i,\alpha}))\,, \tag{3.15}$$

where the index i indicates the species in the mixture. This implies that in the case of a liquid mixture the compensations are computed for every species separately *yet simultaneously*, therefore in a single `H-AdResS` equilibration it is possible to compute all the μ_{ex}^{i}.

We conclude this section by pointing out that the strategy presented here to compute the chemical potential of dense liquids stems entirely from an implicit property of adaptive resolution approaches, and in particular of the `H-AdResS` method. As a matter of fact, the procedures to calculate the free energy compensations are a basic step and fundamental ingredient of this method, and are necessary in order to prepare the simulation setup with a uniform density profile.

3.3 Results and Discussion

Lennard-Jones fluid and comparison with the Widom method

We first validate our method by computing the excess chemical potential μ_{ex} of a Lennard-Jones liquid. We consider systems whose interaction potential is given by a (12,6) Lennard-Jones (LJ) potential truncated and shifted with cutoff radius 2.5σ. The units of energy, length and mass are defined by the parameters, ϵ, σ and m, respectively. The results for this section are expressed in LJ units with time $\tau = \sqrt{m\sigma^2/\epsilon}$, temperature $k_BT = 2\epsilon$ and pressure ϵ/σ^3. Simulations were carried out using LAMMPS [125] with a time step $5\times10^{-4}\tau$. Constant temperature was enforced by a Langevin thermostat with coupling parameter 100τ. A system of size $N = 1687500$ was considered in the density range $\rho\sigma^3 = 0.3\cdots1.0$, with a corresponding number of particles in the AT region ranging in the $2800\cdots9400$ interval. The radius of the atomistic region and the thickness of the hybrid shell are both 15 σ. We performed equilibration runs of 10^5 MD steps and production runs of 10^6 MD steps. Furthermore, to compare the results obtained with the method outlined here, we use the Widom method [165] for equivalent systems but of size $N = 1000$ and in the range of densities $\rho\sigma^3 = 0.3 \cdots 0.8$

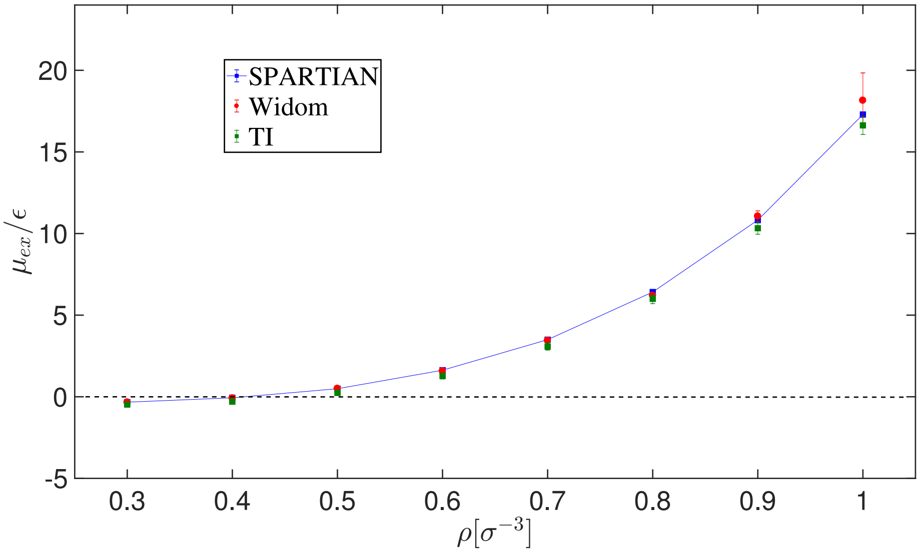


Figure 3.4. Excess chemical potential for the pure Lennard-Jones liquid computed with the `SPARTIAN` method (blue squares), the Widom method (red circles) and TI (green squares). In all cases, the error bars are smaller than the symbol size up to densities $\rho =\sim 0.8\sigma^{-3}$; beyond this value, the deviations between data obtained with the `SPARTIAN` method and the conventional approaches become more visible. In particular, Widom and TI sets show a drift from the `SPARTIAN` values as well as a clear increase in the error bars, which indicate the limits of applicability of the Widom and TI methods. The horizontal dashed line indicates the ideal gas case.

Results for μ_{ex} as a function of ρ are shown in Fig. 3.4 where a remarkable agreement between the Widom, TI and the `SPARTIAN` method results can be appreciated up to densities $\rho =\sim 0.8\sigma^{-3}$. Furthermore, and in contrast to the Widom method and TI, the adaptive resolution calculation of μ_{ex} provides consistent results for high densities. This result is expected since `SPARTIAN` takes advantage of the accurate sampling made possible by the large density of the system.

Lennard-Jones mixture

To test the range of applicability of `SPARTIAN`, we compute the excess chemical potential of simple molecular liquid mixtures. In particular, we simulate a glass-forming binary Lennard-Jones mixture using the interaction parameters of Ref. [181]. The mixture consists of 80% particle A and 20% particle B. All the results are expressed in LJ units with energy ϵ, length σ, mass m, time $\sigma(m/\epsilon)^{1/2}$, temperature ϵ/k_B and pressure ϵ/σ^3. The potential parameters are chosen as $\sigma_{AA} = 1.0\sigma$, $\epsilon_{AA} = 1.0\epsilon$, $\sigma_{BB} = 0.88\sigma$, $\sigma_{AB} = 0.8\sigma$, $\epsilon_{BB} = 0.5\epsilon$ and $\epsilon_{AB} = 1.5\epsilon$, and the cut-off radius and the temperature are set $r_c = 2.5\sigma$ and $k_BT = 0.75\epsilon$, respectively. In this case as well, the radius of the atomistic region and the thickness of the hybrid shell is 15 σ.

As discussed in the Method section, we can treat independently all the species in the mixture, i.e., there is a density profile associated to species A and B and thermodynamic forces are applied to every density profile. In this way, we can automatically extract the excess chemical potential for every species in the mixture.

Table 3.1.
Excess chemical potentials of both particle species of the Lennard-Jones mixture as of Ref. 181 at temperature $T = 0.75/k_B$, computed by the `SPARTIAN` method and compared to the values obtained *via* particle insertion enhanced by means of Metadynamics [87]. The unit of all values is ϵ_{AA}.

Component / model	`SPARTIAN`	Ref. 87
μ_{ex}^{A}	3.95 ± 0.02	3.99 ± 0.04
μ_{ex}^{B}	-4.61 ± 0.06	-4.65 ± 0.02

Results for this system are presented in Table I, where an excellent agreement with calculations based on Metadynamics [87] is apparent. Specifically, we observe the deviations from the values obtained by Perego *et al.* [87] by the same absolute amount (0.04ϵ) for both particle types. The sign of the excess chemical potential reflects that it is more favourable to insert the small B particles in the system due to their low concentration and relatively (with respect to A particles) weak interaction energy. An interesting behaviour can be observed in the errors: our method, in fact, provides more precise estimates –that is, a smaller statistical error– for the excess chemical potential of A particles rather than B particles. This behaviour differs from that of methods using test particle insertions, and stems from the fact that `SPARTIAN` substantially relies on – and takes advantage of – accumulating statistics, which improves as the mole fraction of solute molecules increases.

SPC/E Water

The calculation of the chemical potential for dense liquid water is a rather challenging task. Standard methods to compute free energy differences like the Widom insertion do not converge convincingly [159] and it is thus necessary to use more sophisticated methods even for this system composed of relatively small molecules.

We have computed the excess chemical potential of two rather popular water models, SPC and SPC/E [133–135]. Molecular dynamics simulations have been carried out for 117000 water molecules. The size of the cubic simulation box is 152Å, the radius of the AT region is 30Å and the thickness of the HY region is 20Å, which is larger than the Bjerrum length of pure water ($\lambda_B = 7.5$Å). A 0.5 ns equilibration run has been performed with time step $\delta t = 1$ fs in the NPT ensemble for a fully atomistic system. Temperature and pressure are enforced at T=298 K and P =1 bar using the Nosé-Hoover thermostat and barostat with damping coefficient of 100 fs and 1000 fs, respectively. This procedure provides the initial configuration for the subsequent `SPARTIAN` simulations, which have been performed in the NVT ensemble. Here we have used the same δt and enforced the same temperature T with a Langevin thermostat with coupling parameter 100τ.

Results are presented in Table II. For completeness, we have compared with results obtained using thermodynamic integration (TI) (SPC [182], SPC/E [117]) and two-stage particle insertion methods, i.e. Bennett acceptance ratio method (BAR) [166–169] (SPC [183], SPC/E [157]). Once again, our results agree reasonably well, approximately less than 5% difference, with the values reported in the literature in all cases. The discrepancy is particularly low for the chemical potential of the SPC/E water model. It should be noted, however, that the value of the chemical potential of SPC water computed *via* TI refers to a relatively old calculation which is thus likely to be outdated with respect to state-of-the-art results.

Table 3.2.
Excess chemical potential of water molecules at temperature T=298 K as computed in this work, compared to the values obtained with thermodynamic integration [117, 182] and two-stage particle insertion methods [157]. The experimental value is -26.46 kJ/mol [184]. The unit of all values is kJ/mol.

Water model	`SPARTIAN`	TI	BAR
SPC	-25.68 ± 0.02	-23.9 ± 0.6 [182]	-26.13 ± 0.05 [183]
SPC/E	-29.01 ± 0.09	-29.53 ± 0.03 [157]	-29.70 ± 0.05 [183]

Aqueous solution of sodium chloride

Lastly, we have computed the excess chemical potentials μ_{ex}^{NaCl} and $\mu_{ex}^{H_2O}$ for sodium chloride (NaCl) in water. For this prototypical electrolyte solution present in biological, geological, and industrial contexts, many computational studies have been devoted to calculate μ_{ex} using various different methodologies [153–158]. This wealth of results constitutes an excellent database to benchmark the method proposed here. Furthermore, strong electrostatic interactions present in salt solutions provide us with a challenging testing ground.

We have performed molecular dynamics simulations for NaCl aqueous solutions with 117000 water molecules and in the range of molalities $\mathrm{mol}_{\mathrm{solute}}/\mathrm{kg}_{\mathrm{solvent}} = 0 \cdots 10$. This interval includes substantially higher ion concentrations than the ones reported recently [154, 157]. We have used the force field parameters of the Na^+ and Cl^- ions from Ref. 178, truncated and shifted at $r_{\mathrm{cutoff}}^{LJ} = 10\AA$ (for the non–Coulombic terms), and the SPC/E [133–135] parameters for water. This combination provides the value of solubility closest to experimental measurements [153]. The cubic simulation box side is 154.5Å, the radius of the AT region is 30Å, and the thickness of the HY region is 20Å. As previously done for pure water, we first performed a 1 ns long equilibration run for the fully atomistic system with $\delta t = 1$ fs in the NPT ensemble. We kept temperature and pressure constant at T =298 K and P =1 bar using the Nosé-Hoover thermostat and barostat with damping coefficient of 100 fs and 1000 fs, respectively. The resulting equilibrated configurations have been employed as starting point for `SPARTIAN` simulations which have been performed in the NVT ensemble using the same δt and T. We have controlled the temperature with a Langevin thermostat with coupling parameter 100τ.

The `H-AdResS` method relies on the use of short-range potentials and forces to treat electrostatic interactions. In a previous study, we have implemented and validated the damped shifted force potential [65] (DSF) in Hamiltonian adaptive resolution simulations [61]. Following Ref. 61, the DSF parameters employed in the present study are $\alpha = 0.2\AA^{-1}$ and $r_{\mathrm{cutoff}}^{DSF} = 12\AA$. Since we expect electrostatics to influence the accuracy of the chemical potential calculations, it is necessary to assess how different the results might be when using either the DSF method or the standard Ewald summation method [185]. We have performed fully atomistic NPT simulations for the same setups previously discussed. We then compared the difference in electrostatic potential $V_{P3M} - V_{DSF}$ between the Ewald and DSF calculations for NaCl and water molecules. The results are presented in Fig. 3.5, where a nearly constant difference is observed for all salt concentrations considered (the fluctuations about the average value of $V_{P3M} - V_{DSF}$ across all molalities are of approximately 1% and 3% for NaCl and water, respectively). Since the difference in potential energies when using Ewald summation or DSF can be treated as constant, then $V_{P3M} - V_{DSF}$ for every species in the system amounts to a constant shift in the excess chemical potential. Furthermore, it is possible to avoid doing an extra simulation using the Ewald method if we investigate the theoretical origin of the mismatch with respect to simulations using the DSF

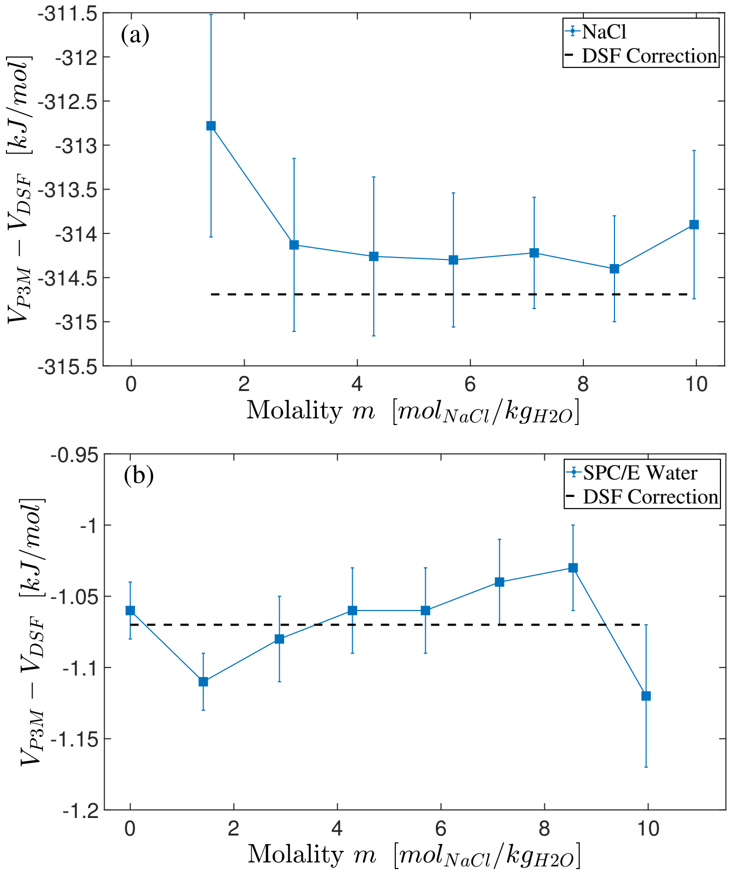


Figure 3.5. Difference between the total electrostatic potential of an aqueous solution of NaCl as computed *via* Ewald summation method and DSF method, as a function of molality. For both NaCl (a) and water (b), the difference is nearly independent of the salt concentration, and its dominant contributions can be rationalized in terms of the sum of self-interaction (V^{self}) and excluded electrostatic interactions (V^{excl}) of the DSF potential, whose theoretical value is reported as dashed lines. For a detailed discussion of these corrections see the SI, in particular Eqs. (SI3) and (SI4).

method. The DSF potential includes a contribution V^{self} that guarantees charge neutrality at a given cutoff radius, and a contribution V^{excl} that excludes, in the case of rigid molecules such as SPC/E water, intramolecular electrostatic interactions (see SI for a detailed discussion). The sum of the two contributions is force field– but not salt concentration–dependent, as evidenced by the black horizontal lines in Fig. 3.5, and it accounts precisely, within statistical error, for the chemical potential shift $V_{P3M} - V_{DSF}$.

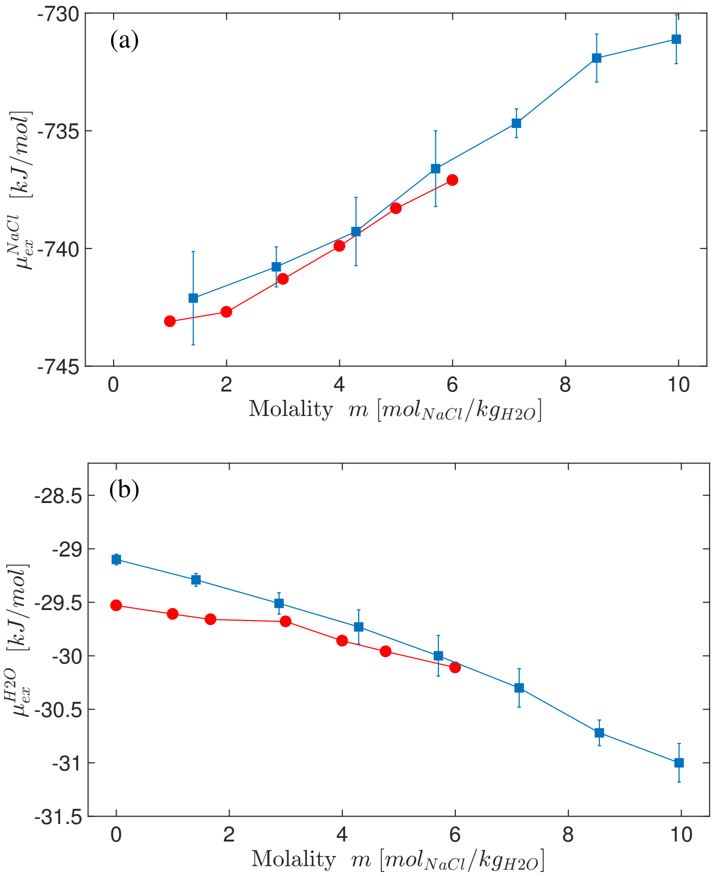


Figure 3.6. Excess chemical potential of water molecules μ_{ex}^{H2O} (a) and molecular NaCl μ_{ex}^{NaCl} (b) as computed for different salt concentrations. The results obtained with the `SPARTIAN` and BAR [157] methods are represented by blue squares and red circles, respectively.

Hence, after applying this corrections we have compared directly the results computed with our method and with those reported in Ref. 157 using BAR [166–169]. It is apparent from Fig. 3.6 that our results for μ_{ex}^{NaCl} and $\mu_{ex}^{H_2O}$ are both in excellent agreement with such reported values. Furthermore, we report here, for the first time to our knowledge, values of excess chemical potentials for this system for molalities larger than 7 [153]. These results show well–defined trends, indicating that upon increasing NaCl concentration the addition of a further solute molecule becomes energetically less favourable, roughly 10 kJ/mol in the range 6-10, whereas in the same range it is slightly more favourable to add another water molecule to the system (1 kJ/mol).

3.4 Conclusions

The chemical potential is of central importance for the comprehension of the physico-chemical properties of a substance and the capacity to manipulate it for scientific and industrial purposes. The vast majority of computational methods devised to calculate chemical potentials rely on periodic attempts to insert a test particle into the system. For dense liquids and highly concentrated liquid mixtures, this procedure is inefficient, perhaps in some cases unfeasible, and the use of enhanced sampling techniques or the design of alternative methods becomes crucial.

In this work we presented a method, spatially resolved thermodynamic integration, or `SPARTIAN`, which introduces a different perspective based on the Hamiltonian adaptive resolution framework. Here, the target system is physically separated from a reservoir of ideal gas particles by a hybrid region where molecules change resolution, from atomistic to ideal gas and vice versa, on the fly. To ensure a uniform density profile of the whole system, free energy compensations are parameterised and applied to the molecules present in the hybrid region. Under such conditions, the system reaches thermal equilibrium and the chemical potential of both target system and ideal gas reservoir equates. Therefore, the free energy compensations are identified with the difference in chemical potential between the two representations, which is precisely the excess chemical potential of the target system.

The method is efficient and of general applicability, as demonstrated by the reported results on pure and multi-component Lennard Jones liquids, pure water, and aqueous solutions of sodium chloride. The values of the excess chemical potential computed for the various species under examination are consistent with the data in the literature where available. For those regions of concentration that remain out of the scope for most established techniques, the proposed strategy has proven especially capable of providing results in line with the trend indicated by other methods. This observation suggests that the increased molecular density represents a vantage point for the method, which avoids the necessity to perform artificial particle insertions and profits of the large number of molecules to improve the convergence of statistical averages.

The `SPARTIAN` strategy reported in this work thus offers a novel, effective, and versatile instrument to compute the excess chemical potential of liquids and liquid mixtures. The method is particularly well-suited to use in cases where the density of the liquid or the concentration of solute in the mixture are high. This constitutes a significant advantage over already available techniques and paves the way for a broad range of applications where the accurate determination of chemical potentials is of central importance.

3.5 Supplementary Information

3.5.1 Technical aspects of the drift force calculation

Calculation of drift and thermodynamic forces requires discretisation of the hybrid region by bin sizes $\Delta\lambda$ and Δr respectively. Force each component of the system, the corresponding drift and thermodynamic forces are sampled at time intervals Δt_{smp}^{dr} and Δt_{smp}^{th} and averaged over time intervals Δt_{ave}^{dr} and Δt_{ave}^{th} respectively. At each time intervals, the program generates two separate output files: (1) the averaged atomistic potential, $\langle V^{AT}(\lambda)\rangle$ and (2) the modulus (with sign) of the thermodynamic force $F_{th}(r)$. In each file, for each component there exists corresponding column which are listed by coordinates of each bin. Thus, for the case of spherical hybrid region, the excess chemical potential of component i is computed as:

$$\begin{aligned} \mu_{\mathrm{exc}}^{i} &= \int_0^1 d\lambda \frac{\partial \mathcal{F}}{\partial \lambda} \langle V^{AT,i}(\lambda)\rangle \\ &+ \int_{r_{at}}^{r_{at}+d_{hy}} dr \mathbf{F}_{th}^{i}(r) \cdot \hat{n} \end{aligned} \tag{3.16}$$

where $\hat{n}$ is the normal vector outward the atomistic region. The time intervals and bin sizes of different systems are given in the Tables 3.3 and 3.4.

For all cases, we used $c = 2.0$ as the modulation parameter of the thermodynamic force, (see main text) and the number density profile ρ is convoluted with a Gaussian function of width three times larger than the bin size [61]. For the case of Lennard-Jones pure fluid and mixtures, the Lennard-Jones potentials are capped at $\hat{r} = 0.1\sigma$; for pure water and aqueous NaCl, the Van der Waals and the electrostatic potentials are capped at $\hat{r} = 0.5$Å.

The compensations to obtain μ_{exc}^{i} in Eq. (3.16) are calculated on the fly [61]. This implies that upon equilibration, the difference in chemical potential between the atomistic and ideal gas representations converges to the excess chemical potential of the atomistic model. Fig. 3.7 shows μ_{exc}^{i} as a function of computation time for NaCl in aqueous solution at different concentrations. Clearly μ_{exc}^{i} converges in all cases. For i=NaCl, the amplitude of the oscillations increases substantially as the ions concentration decreases. This verifies that the SPARTIAN method becomes more accurate as the density/mole fraction of the sample increases.

System	Δt^{dr}_{smp}	Δt^{th}_{smp}	Δt^{dr}_{ave}	Δt^{th}_{ave}	δt	Iterations
LJ	200	200	20000	50000	0.0005 τ	30
LJM	10	10	5000	10000	0.001 τ	100
SPC/E	1	1	5000	20000	1.0 fs	80
AqNaCl	1	10	5000	20000	1.0 fs	140

Table 3.3.
Sampling and averaging time intervals are listed for Lennard-Jones fluid (LJ), Lennard-Jones mixtures (LJM), SPC/E water (SPC/E) and NaCl in aqueous solution (AqNaCl). The values are presented in number of time steps. The second to last column shows the time step. The last column reports the duration of the production runs of each simulation, measured in numbers of iterations, each of which corresponds to Δt^{th}_{ave} simulation steps.

System	$\Delta\lambda$	ΔR
LJ	0.005	0.5 σ
LJM_A	0.005	0.5 σ
LJM_B	0.005	0.5 σ
SPC/E	0.001	1.0 Å
AqNaCl_{Na}	0.001	1.0 Å
AqNaCl_{Cl}	0.001	1.0 Å
AqNaCl_{H_2O}	0.001	1.0 Å

Table 3.4.
Bin sizes used for discretising the hybrid region are listed for pure Lennard-Jones fluid (LJ), type A (LJM_A) and type B (LJM_B) in Lennard-Jones mixtures, SPC/E water (SPC/E), Sodium ion (AqNaCl_{Na}), Chlorine ion (AqNaCl_{Na}) and water molecules (AqNaCl_{H_2O}) of NaCl in aqueous solution.

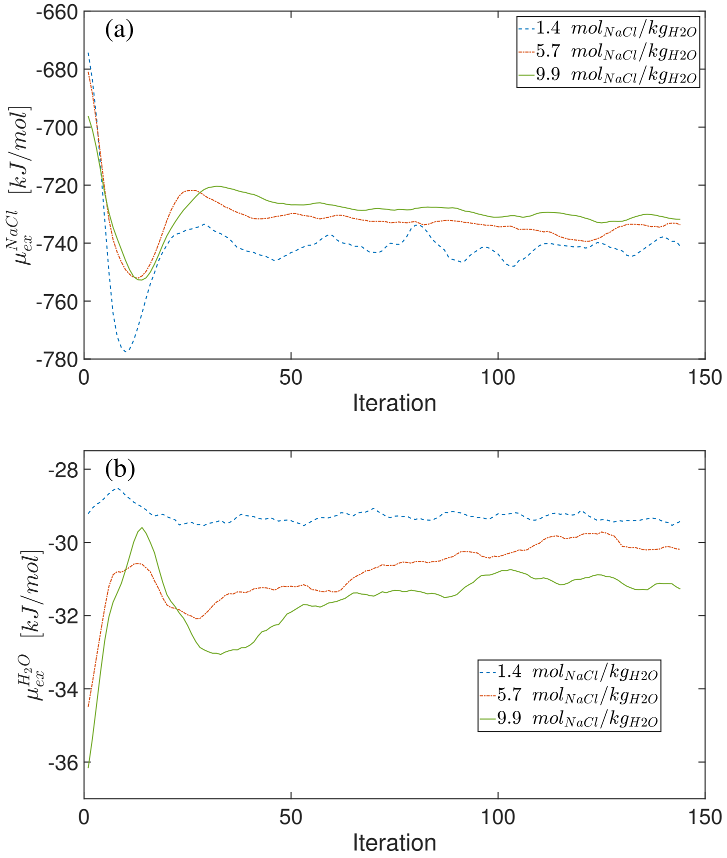


Figure 3.7. Excess chemical potential of water molecules μ_{exc}^{H2O} (a) and molecular NaCl μ_{exc}^{NaCl} (b), for various concentrations, as a function of computation time.

3.5.2 Correction to the chemical potential from the DSF treatment of electrostatic interaction

For a system having N charged particle, the short-ranged electrostatic DSF potential is given by:

$$
\begin{aligned}
V^{\text{DSF}} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq 1}^{N_i(r_c)} q_i q_j \left[\frac{\text{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\text{erfc}(\alpha r_c)}{r_c}\right. \\
&+ \left.\left(\frac{\text{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}}\frac{\exp\left(-\alpha^2 r_c^2\right)}{r_c}\right)(r_{ij} - r_c)\right] \\
&+ V_{\text{self}} - V_{\text{excl}}\,,
\end{aligned}
\tag{3.17}
$$

where $N_i(r_c)$, is the number of particles which are located within the sphere of radius r_c and the center of particle i. The decay of the potential is set by the parameter α whose dimension is an inverse length, and $\mathrm{erfc}(r)$ is the complementary error function [82]. Since the net charge of all $N_i(r_c)$ particles inside the truncating sphere is not zero, a correction to the potential is needed to compensate the lack of charge neutrality. This is done by approximating the net charge of the truncating sphere and placing the approximated "imaginary" charge on the surface of the sphere [186, 187]. For the DSF potential, the electrostatic interaction of particle i with the imaginary charge is expressed as [187]:

$$V^{\mathrm{self}} = -\left(\frac{\mathrm{erfc}(\alpha r_c)}{r_c} + \frac{\alpha}{\sqrt{\pi}} + \frac{\alpha}{\sqrt{\pi}}\exp(-\alpha^2 r_c^2)\right)\sum_{i=1}^{N} q_i^2. \tag{3.18}$$

For the case of rigid structural molecules, the intra-molecular electrostatic interaction has to be excluded. This is done by subtracting the corresponding Coulomb interaction from the DSF potential. Thus, for particles i and j separated by a distance r_{ij} in a rigid molecule, the excluded electrostatic interaction is computed as:

$$V_{ij}^{\mathrm{excl}}(r_{ij}) = \frac{q_i q_j}{r_{ij}}. \tag{3.19}$$

3.5.3 Computational gain of `SPARTIAN` for SPC/E water

Fig. 3.8 illustrates the computational gain of an `H-AdResS--SPARTIAN` calculation over a conventional all-atom simulation. This gain is evaluated as the ratio between the run time of a fully atomistic and a `SPARTIAN` simulations, as a function of the cubic box side. In all cases, the system is composed of SPC/E water molecules, in numbers ranging from 91,124 to 342,999. The box side ranges from ~ 140 to ~ 220 Å, while the spherical AT region radius of 30 Å and the HY region thickness of 20 Å are fixed. The simulations are run on a single cpu for 1000 time steps; the neighbour list is built every 10 time steps and the skin is set to 1 Å. The inset shows the absolute time (in minutes) required to run the simulations: the `SPARTIAN` runs grows weakly with system size, staying close to ~ 100 minutes for 1000 steps.

3.5.4 Effect of the potential capping

The probability of capping events, i. e. that two particles get closer than the capping radius, calculated for a LJ system at number density $\rho = 1.0\sigma^{-3}$ is reported in Fig. 3.9. This probability is calculated within the hybrid region during three different time periods: initial ($T \leq 1$ iteration), intermediate ($T \leq 10$ iterations), and final ($T \leq 60$ iterations). The capping events are rare for the resolutions $\lambda \leq 0.015$ and they are zero for $\lambda > 0.015$. The capping probability over the entire hybrid region is shown in the inset. This low probability combined with the weight provided by the

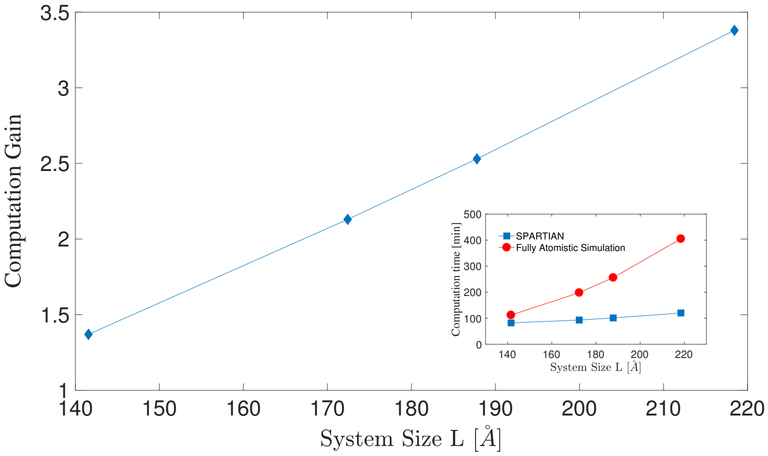


Figure 3.8. Computational efficiency of a `SPARTIAN` run of SPC/E water with respect to a reference all-atom simulation. Main figure: computational gain. Inset: absolute run time.

gradient of the switching function $\mathcal{F}$ ensures that the errors introduced by the capping events have a null contribution to the calculation of the excess chemical potential.

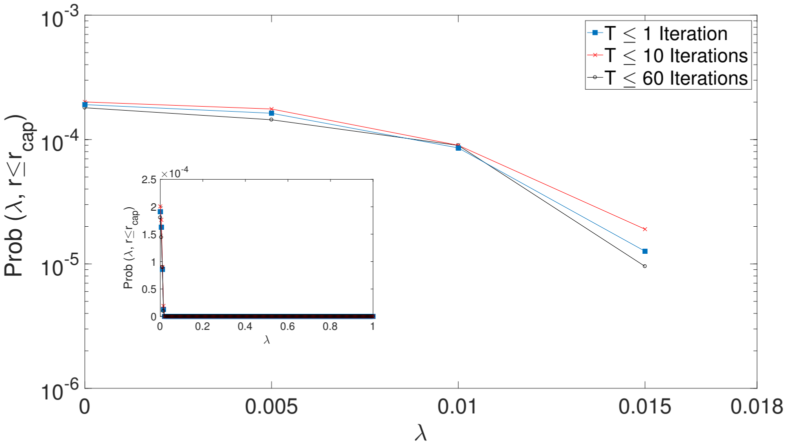


Figure 3.9. Probability that two particles get closer than the capping radius, calculated for a LJ system. Main figure: probability in logarithmic scale. Inset: The capping probability over the entire hybrid region.

3.5.5 Convergence of the chemical potential

Hereafter we provide details about the convergence of the chemical potential with respect to various parameters of the simulation setup, specifically: number of iterations and density (Fig. 3.10); simulation box size (Fig. 3.11); hybrid region size (Fig. 3.12); number of iterations and box size (Fig. 3.13). These data support the assumption that the values of μ_{exc} are insensitive, well within the statistical error, to variations of the chosen setup parameters, and provide a guideline for selecting the values of such parameters for the systems under examination.

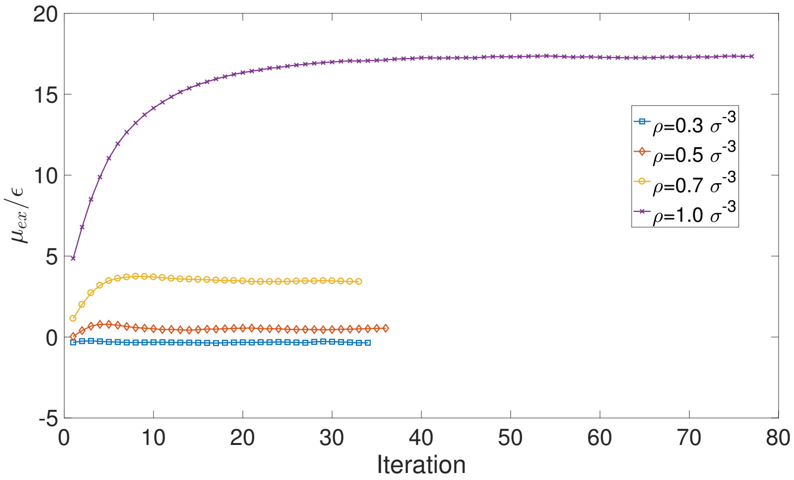


Figure 3.10. Excess chemical potentials of LJ fluids at different number densities as the function of `SPARTIAN` iterations.

3.5.6 Effect of the parameter ν

The choice of the switching function $\mathcal{F}(\lambda) = \lambda^{\nu}$, and more precisely the choice of the exponent $\nu = 7$, is motivated by the need to suppress the occurrence of molecule overlaps in the hybrid region. As illustrated by Fig. 3.9, a non-zero probability exists that particles near the HY–IG interface get arbitrarily close. This might induce a systematic sampling of enormous potential energy values that could make the calculation of free energy compensations unreliable. In practice, it is necessary to smoothen the divergence $1/r^{12}|_{r\to 0}$ implicit in the Lennard-Jones potential. Since the definition of the switching function includes a $\cos^2(r)$ term, and the free energy compensations are weighted by the gradient of $\mathcal{F}$, the exponent $\nu = 7$ rigorously smoothens the singularity in the potential. We stress that this problem, which also affects TI-based methods, has been already identified and solved using similar

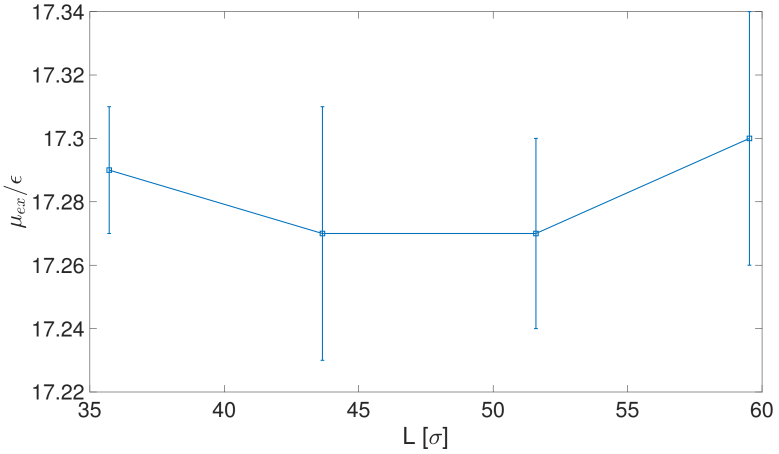


Figure 3.11. Excess chemical potentials of LJ fluid at number density $\rho = 1.0\sigma^{-3}$ for different simulation box sizes. In all simulations, the size of atomistic and hybrid regions are fixed at $d_{at} = 15\sigma$ $d_{hy} = 15\sigma$.

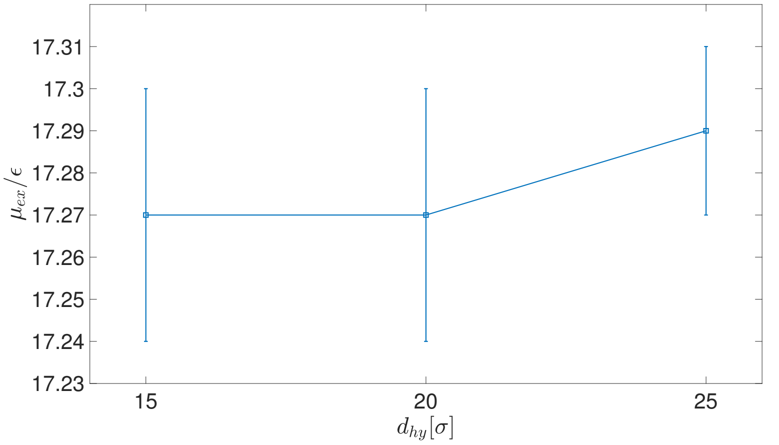


Figure 3.12. Excess chemical potentials of LJ fluid at number density $\rho = 1.0\sigma^{-3}$ for different sizes of hybrid region (d_{hy}). In all simulations, the size of atomistic region and simulation box size are fixed at $d_{at} = 15\sigma$ and $L = 51.59\sigma$, respectively.

arguments [179]. Furthermore, this choice of the switching function skews the sampling close to the AT–HY interface where this problem no longer exists. Fig. 3.14 shows μ_{exc} as a function of ν: this result suggests that the absolute values obtained improve for

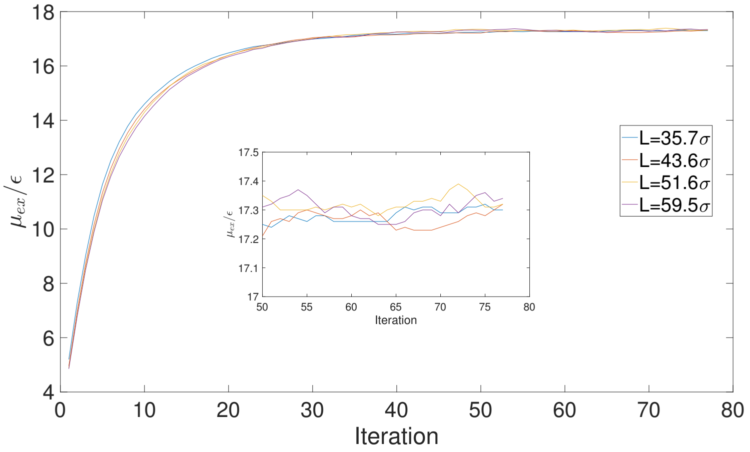


Figure 3.13. Excess chemical potentials of LJ fluid at number density $\rho = 1.0\sigma^{-3}$ as the function of the `SPARTIAN` iterations for different simulation box sizes. In all simulations, the size of atomistic and hybrid regions are fixed at $d_{at} = 15\sigma$ $d_{hy} = 15\sigma$, respectively.

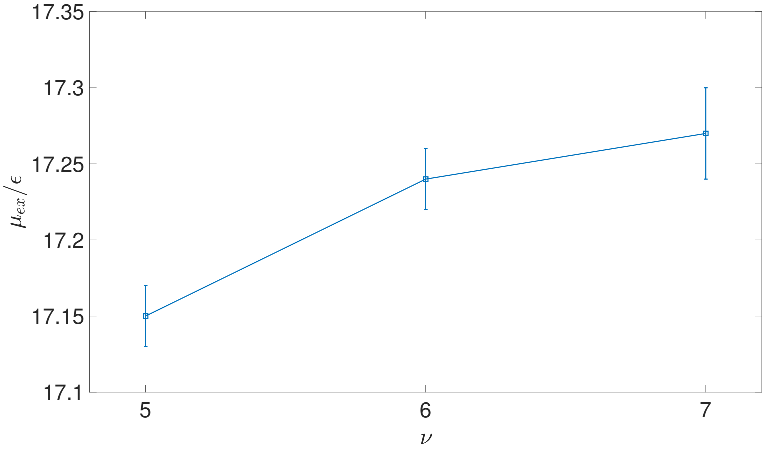


Figure 3.14. Excess chemical potentials of LJ fluid at number density $\rho = 1.0\sigma^{-3}$ for different switching function exponents (ν). In all simulations, the size of atomistic and hybrid regions and simulation box are fixed at $d_{at} = 15\sigma$, $d_{hy} = 15\sigma$ and $L = 59.53\sigma$, respectively.

$\nu = 7$. Values of μ_{exc} for $\nu < 5$ are not available because in this interval the singularity in the potential makes the numerical integration of the equations of motion unstable.

3.6 Acknowledgement

We thank Claudio Perego, Omar Valsson and Debashish Mukherji for a critical reading of the manuscript. MH, RP and KK acknowledge financial support under the project SFB-TRR146 of the Deutsche Forschungsgemeinschaft. RCH gratefully acknowledges the Alexander von Humboldt Foundation for financial support. MH thanks Paolo Raiteri and Julian D. Gale for fruitful discussions on the energetic comparison between Coulomb and DSF potentials. All authors are indebted with Douglas Murphy for insightful and valuable suggestions on the composition of the manuscript.

4. Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations

This chapter is a research article that has been published in the European Physical Journal E.

Maziar Heidari, Robinson Cortes-Huerto, Kurt Kremer, and Raffaello Potestio
Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations
Eur. Phys. J. E (2018) 41: 64.
DOI: 10.1140/epje/i2018-11675-x
©2018 by the authors

Abstract

To understand the properties of a complex system it is often illuminating to perform a comparison with a simpler, even idealised one. A prototypical application of this approach is the calculation of free energies and chemical potentials in liquids, which can be decomposed in the sum of ideal and excess contributions. In the same spirit, in computer simulations it is possible to extract useful information on a given system making use of setups where two models, an accurate one and a simpler one, are concurrently employed and directly coupled. Here, we tackle the issue of coupling atomistic or, more in general, interacting models of a system with the corresponding idealised representations: for a liquid, this is the ideal gas, i.e. a collection of non-interacting particles; for a solid, we employ the ideal Einstein crystal, a construct in which particles are decoupled one from the other and restrained by a harmonic, exactly integrable potential. We describe in detail the practical and technical aspects of these simulations, and suggest that the concurrent usage and coupling of realistic and ideal models represents a promising strategy to investigate liquids and solids *in silico*.

4.1 Introduction

Since the dawn of computer-aided research, when electronic calculators were employed to numerically solve analytical expressions, a steadily growing fraction of scientific investigation relies on the help of machines [188–191]. This is particularly true for the field of soft matter [2, 55, 111, 126], where computer simulations have enabled researchers to “infuse life” in models of increasing complexity and investigate the behaviour of systems ranging from hard spheres liquids [192] to melts of polymers [112], novel materials [193], or biomolecules such as proteins [101, 105, 109, 116, 194] and DNA [195]. The power of the computational approach lies especially in the flexibility and arbitrariness of the model design, where level of resolution, interaction potentials, particle-based or continuum description, thermodynamical conditions and many other features are left to the modeller, who has free hand in creating a world where her/his rules apply.

Computer simulations have thus broken the chains that kept us bound to exact results and simple approximations of even simpler models, allowing us to study strongly interacting systems with several different types of potentials and large sizes [196]. The determination of free energy differences between states of a given system, for example, does not require the exact calculation of its partition function or some multi-body expansion of the latter, rather can be (often) performed by means of standard techniques such as thermodynamic integration, umbrella sampling [2, 126], or any other enhanced sampling algorithm [102, 197].

And yet, the simplicity, exact solvability, or physical intuitiveness of the simplest models exert not only an intellectual attraction on the scientist fascinated by the universality of the concepts and properties that these models entail; rather, they continue to represent a pivotal element in the construction of our picture of reality, in that they serve as bedrock, reference, and gauge for many other, more complex systems. One example for all: the absolute values of fundamental thermodynamical properties, such as free energies and chemical potentials, can almost always be separated in two terms: an ideal part, originating from the exactly solvable ideal contribution, and an excess part, which contains the effect of the interactions present in the “real” system. Remarkable is, that these two terms are exactly additive (again, with exceptions) in a non-perturbative fashion [2, 17]. In general, then, the calculation of important system properties goes through the comparison between the most accurate model one can simulate on a computer and a simpler, more tractable representation much of which is already known. The additivity of the property under investigation does the rest.

In recent years, a peculiar class of computer simulations has emerged as a rich and versatile tool to perform model comparison, that is, adaptive resolution simulations [51, 56, 57, 62, 198]. These are setups in which the same systems, typically a fluid, is represented within the same simulation by means of two different models at different resolution. The simulation domain is subdivided in two parts: a high resolution region, where the most accurate and computationally expensive model is employed, and a low resolution region, where the simpler description featuring a lower computational cost is used. Depending on its position in space, a molecule is described by one model or the other; however, an open boundary, geometrically separating these two domains, allows the molecules to diffuse freely and change model, i.e. resolution, on the fly. Within a finite-sized layer located at the interface between the two main subregions, dubbed hybrid or transition region, the molecule smoothly varies its resolution, adapting its representation to one model or the other in a continuous manner.

The motivations behind this setup, in which the position in space dictates a molecule's model, are several. The simplest and most obvious is the computational gain: in the largest fraction of the simulation volume the computationally "cheaper" model is employed, thus reducing the amount of resources necessary to calculate forces and potentials; on the contrary, in a small volume –where the interesting things happen– the system is described with high accuracy, and the smooth coupling with the lower-resolution environment preserves its thermodynamical properties (e.g. density, temperature, particle number fluctuations...).

Another important *raison d'être* of these methods is the fact that the space-dependent coupling of two different models establishes a thermodynamical relationship between them from which nontrivial information can be obtained. In fact, in adaptive resolution simulations forces emerge, whose origin depends on the specific technique employed, that steer the system towards an equilibrium state where the local densities in the two subdomains differ. To attain the same density in all parts of the simulation one has to impose a single-molecule potential that can be traced back to the difference in Gibbs free energy between the two models concurrently employed to represent the same system [49, 50]. Hence, in the process of parametrising the setup so as to have a uniform density profile, one quantifies the liquid's chemical potential difference between the simple and the accurate representation [199].

It is at this point that the simplest, most fundamental models enter the scene. With a tool at hand that "compares" the chemical potentials of two systems (more precisely: two different representations of the same system) it is natural to think of employing, as low resolution model, one that is as inexpensive, simple, understood, and exactly solvable as possible. In the realm of fluids, such a model is the ideal gas, that is, a collection of noninteracting particles (representing atoms as well as molecules) fully described by temperature and density. All relevant thermodynamical properties of the ideal gas are known and can be calculated exactly, and represent the main contribution to those observables, such as free energies, that can be decomposed into the sum of the ideal part, determined indeed by the ideal gas, and the excess part due to the configurational partition function and the interaction potentials within it.

As adaptive resolution simulation methods enable the direct, smooth coupling of a given model to its corresponding ideal representation, the possibility opens to a new class of approaches to extract excess quantities. However, while for dense liquids such as water the coupling with an ideal gas has been already performed [200], a comparable matching has not been demonstrated in the case of a solid. In this case, the ideal reference model is given by the Einstein crystal, that is, a collection of point-like, noninteracting particles restrained in specified positions in space by means of harmonic potentials. As the ideal gas, also for this model the partition function (hence the full thermodynamics) can be computed analytically: it thus represents the reference starting point of a non-perturbative computation of free energies and other quantities of interest.

In this work we discuss in detail adaptive resolution simulations of liquid water, described at the all-atom level, coupled to an ideal gas, and a Lennard-Jones solid coupled to an Einstein crystal, the latter being presented here for the first time. Particular attention is given to the technical aspects of these simulations, with thorough descriptions of the computational and algorithmic characteristics.

The manuscript is organised as follows: in section 4.2 we review the specific approach employed here, that is, the Hamiltonian adaptive resolution simulation scheme, or `H-AdResS`; in section 4.3 we illustrate the significance of the external field required to obtain a uniform density profile, and describe in detail the algorithms developed to efficiently parametrise; in section 4.4 we present the coupling between an atomistic model of water and an ideal gas, and describe the treatment of the electrostatic interaction within the `H-AdResS` framework; in section 4.5 we concentrate on the Lennard-Jones crystal and its coupling with the ideal Einstein crystal model; finally, in section 4.6 we summarise our results and provide a brief account of the possible applications of the presented methods.

4.2 Theoretical Background

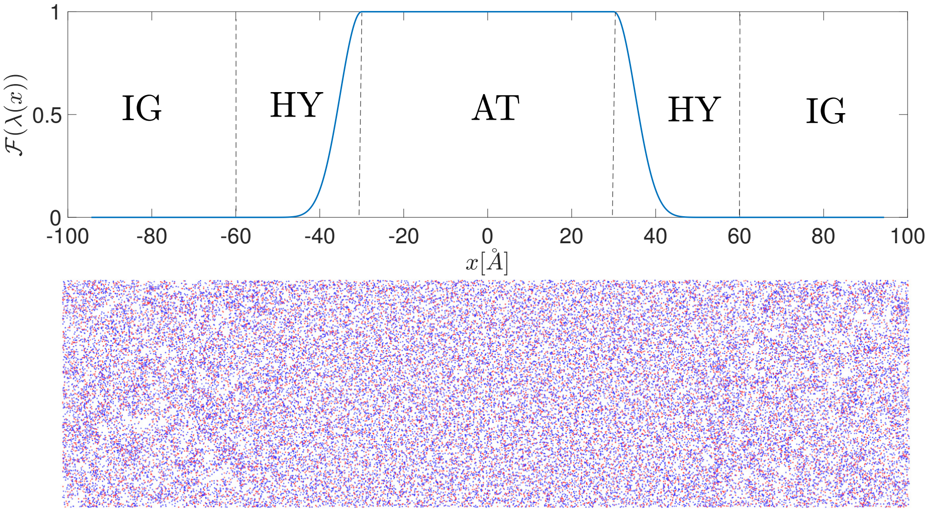


Figure 4.1. Setup of a Hamiltonian Adaptive Resolution Simulation. The periodic box is partitioned into three different regions, namely: Ideal gas (IG), Hybrid (HY), and Atomistic (AT). Upper panel: the switching function $\mathcal{F}(\lambda)$ takes values between 0 (IG) and 1 (AT), thus defining the resolution of a molecule (here water). Lower panel: simulation snapshot explicitly showing the various subdomains. The oxygen and hydrogen atoms of the water molecules are represented by red and blue dots respectively.

The `H-AdResS` scheme [49, 50] belongs to a family of adaptive resolution simulation methods [48, 56, 57] in which a small portion of the system, usually a fluid described with atomistic resolution (AT), is embedded in a reservoir of particles of the same system, modelled using a coarse grained (CG) representation; the coupling between the two resolutions takes place in an open boundary region, the so-called hybrid (HY) region. A snapshot of the simulation setup is presented in Fig. 4.1. The main feature of the `H-AdResS` method is that the whole system is described in terms of a global Hamiltonian function H of the form:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \left\{ \lambda_\alpha V_\alpha^{AT} + (1 - \lambda_\alpha) V_\alpha^{CG} \right\} . \tag{4.1}$$

The term $\mathcal{K}$ is the atomistic kinetic energy, and V^{int} includes all the intramolecular bonded interactions. The resolution of a particle α is specified by the transition function $\lambda_\alpha = \lambda(\mathbf{R}_\alpha)$, which is computed on the centre-of-mass coordinate $\mathbf{R}_\alpha$ of the molecule.

Without loss of generality, here we employ a rectangular simulation box where the AT is represented by a slab. The resolution of a given molecule is thus determined by the following piece-wise function:

$$\lambda(x) = \begin{cases} 1 & |x| \leq d_{AT}/2 \\ \cos^2\left(\frac{\pi(x - d_{AT}/2)}{2 d_{HY}}\right) & \frac{d_{AT}}{2} < |x| \leq \frac{d_{AT}}{2} + d_{HY} \\ 0 & |x| > d_{AT}/2 + d_{HY} , \end{cases} \tag{4.2}$$

with d_{AT} and d_{HY} the sizes of the AT and HY region, respectively. The mid-point of the simulation box is set in the origin of the coordinate system. Concerning non-bonded interactions, a molecule α interacts with its neighbouring particles through coarse grained V^{CG} and atomistic V^{AT} potentials, defined as:

$$\begin{aligned} V_\alpha^{AT} &\equiv \frac{1}{2} \sum_{\beta, \beta \neq \alpha}^{N} \sum_{ij} V^{AT}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|) \qquad (4.3) \\ V_\alpha^{CG} &\equiv \frac{1}{2} \sum_{\beta, \beta \neq \alpha}^{N} V^{CG}(|\mathbf{R}_\alpha - \mathbf{R}_\beta|) , \end{aligned}$$

where the coordinates of an atom or a molecule are represented by vectors $\mathbf{r}$ and $\mathbf{R}$, respectively. The factor 1/2 accounts for the double counting of particles in the sum since the total non-bonded potential of a molecule α is given by a sum of AT and CG contributions weighted by λ_α or $(1 - \lambda_\alpha)$, respectively. Albeit two-body interactions have been employed here for the sake of clarity, the extension to multi-body, short-ranged potentials is straightforward.

The total force acting on atom i of molecule α is given by:

$$\begin{aligned} \mathbf{F}_{\alpha i} &= \mathbf{F}_{\alpha i}^{int} \\ &+ \sum_{\beta,\beta\neq\alpha} \left\{ \frac{\lambda_\alpha + \lambda_\beta}{2} \mathbf{F}_{\alpha i|\beta}^{AT} + \left(1 - \frac{\lambda_\alpha + \lambda_\beta}{2}\right) \mathbf{F}_{\alpha i|\beta}^{CG} \right\} \\ &- \left[V_\alpha^{AT} - V_\alpha^{CG}\right] \nabla_{\alpha i} \lambda_\alpha \, . \end{aligned} \tag{4.4}$$

The forces due to intramolecular interactions, $\mathbf{F}_{\alpha i}^{int}$, are not resolution-dependent. The second term on the r.h.s. is the sum, over all other molecules β in the interaction range, of the pairwise atomistic and coarse-grained forces, weighted by the average resolutions of the two molecules. Each of the terms in the sum is antisymmetric under molecule exchange, and satisfies Newton's Third Law by construction. The last term emerges as a consequence of the non-uniformity of space in the dual-resolution simulation setup, that is, the fact that different interactions are present in different parts of the system. Because of this, translational invariance is broken, and a force emerges in the hybrid region (where $\nabla\lambda \neq 0$) and acts on the molecules pushing them in one of the two subdomains, depending on the sign of the prefactor $(V_\alpha^{AT} - V_\alpha^{CG})$. This *drift force* term $\mathbf{F}_{\alpha i}^{dr} = -\left[V_\alpha^{AT} - V_\alpha^{CG}\right] \nabla_{\alpha i} \lambda_\alpha$ induces a pressure imbalance across the HY region, as it pushes molecules in the subdomain where Helmholtz free energy is locally lower [49, 50].

In general, the AT and CG representations of the same physical system follow different equations of state. Once coupled together *via* an open boundary, the AT and CG regions exchange particles to balance the differences in equilibrium pressure and chemical potential. This results, as it has been already thoroughly investigated [49,50,62], in a non-homogeneous density profile. To overcome these effects and enforce a uniform density profile, it is possible to introduce a new term in the Hamiltonian:

$$H \to H_\Delta = H - \sum_{\alpha=1}^{N} \Delta H(\lambda(\mathbf{R}_\alpha)) \tag{4.5}$$

This term acts separately on each molecule in the system and plays two roles: it removes, on average, the drift force, and enforces a uniform density profile by imposing, in each subdomain, the pressure at which each model has, separately, the correct density. In the following we discuss the computational techniques employed to parametrise the term $\Delta H(\lambda)$.

4.3 Calculation of the free energy compensations

4.3.1 Compensation of the drift force

The free energy compensation term $\Delta H(\lambda)$ that neutralises, on average, the effect of the drift force, has to satisfy the relation:

$$\left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_\alpha} = \left\langle \left[V_\alpha^{AT} - V_\alpha^{CG}\right] \right\rangle_{\mathbf{R}_\alpha} , \tag{4.6}$$

in such a way that the total drift force resulting from the modified Hamiltonian reads:

$$\hat{\mathbf{F}}_\alpha^{dr} = \left(V_\alpha^{AT} - V_\alpha^{CG} - \left.\frac{d\Delta H(\lambda)}{d\lambda}\right|_{\lambda=\lambda_\alpha} \right) \nabla\lambda\left(\mathbf{R}_\alpha\right) , \tag{4.7}$$

and, by construction, $\langle \hat{\mathbf{F}}_\alpha^{dr} \rangle = 0$.

Thermodynamic integration [64] has been originally proposed to compute the compensation term (4.6). This procedure can be justified on the basis that the potential $\Delta H(\lambda)$ bears a strict relationship with the Helmholtz free energy difference between a system with a Hamiltonian $H(\lambda)$ and a reference system defined by the CG representation at $\lambda = 0$ [49, 62]. This procedure, apparently the most straightforward route to compute $\Delta H(\lambda)$, presents two main drawbacks: i) it requires an additional free energy calculation, and ii) the accuracy of the obtained $\Delta H(\lambda)$ can be limited when the system displays strong correlations within the hybrid region.

In a previous paper [61], some of us have shown that a more effective strategy is to compute and locally balance the drift force and parametrise the compensation *on the fly* by using an iterative scheme [62]. A given number N_b of bins is used to discretise the hybrid region in terms of λ in such a way that the width of the bins is $\delta\lambda = 1/N_b$. The index of the bins is given by $i = \texttt{floor}[\lambda(\mathbf{R}_\alpha)/\delta\lambda]$. For a molecule α in a bin i, the contributions V_α^{CG} and V_α^{AT} are computed and accumulated in the local variables V_i^{CG} and V_i^{AT}, respectively. Moreover, the variables N_i^{CG} and N_i^{AT} are defined to monitor throughout the simulation the number of molecules in the bin. These quantities are computed in the same routine and simultaneously with the AT and CG forces.

For all molecules in the hybrid region, this procedure is carried out for a time interval of duration Δt and successive n iterations are performed. At the end of the n-th interval, average AT and CG potentials, $\bar{V}^R[i,n] = V_i^R/N_i^R$ with $R = AT, CG$, are computed. Subsequently, the variables V_i^R, N_i^R are emptied and the average procedure continues. These average values calculated at the end of the n-th cycle are used to compute the running average $\mathcal{V}_{i,n}^R$, defined by:

$$\mathcal{V}_{i,n+1}^R = \frac{n\ \mathcal{V}_{i,n}^R + \bar{V}^R[i,n]}{n+1} , \tag{4.8}$$

where initially $n = 0$ and $\mathcal{V}_{i,0}^R = 0$.

This running average, for $n > 0$, is used to compute the different components of the compensation force. Specifically, for t such that $t_0 + n\Delta t < t \leq t_0 + (n+1)\Delta t$, a molecule α in the bin i of the HY region will feel compensation forces of the form:

$$\mathbf{F}^{R}_{\alpha,i} = s\ \mathcal{V}^{R}_{i,n} \nabla\lambda\left(\mathbf{R}_{\alpha}\right)\ , \tag{4.9}$$

with $s = +1$ for $R = AT$ and $s = -1$ for $R = CG$. At each time step, this force is distributed to the atoms in the molecule α with weights proportional to the atom/molecule mass ratio.

The update of the running average continues until the compensation forces converge to a steady value for every bin i. Finally, the update is interrupted and the resulting force compensation is integrated as a time-independent resolution-based force field, and the corresponding energy compensation $\Delta H(\lambda)$ can be easily computed.

The compensation of the drift force ensures that the hydrostatic pressure becomes uniform across the simulation domain [49, 62]. The AT and CG representations equilibrate at this reference pressure according to the equation of state of the model. This equilibrations implies that the densities in the two regions might differ, and an additional compensation must be applied to the system to ensure a flat density profile.

4.3.2 Compensation of the density imbalance

The uniform pressure enforced by the application of the compensation of the drift force does not necessarily guarantee a uniform density for the whole system. A density gradient might appear in the HY region as a result of the two representations, AT and CG, following different equations of state. A flat density profile can be attained if the compensation ΔH is modified in such a way that in each subregion the corresponding model attains a pressure that gives the same reference density.

This correction can be obtained via an iterative scheme dubbed *thermodynamic force* calculation [51], which consist in successively applying to the molecules in the HY region a force proportional to the density gradient:

$$\mathbf{F}^{th}_{n+1} = \mathbf{F}^{th}_{n} + \frac{c\nabla\rho_n(x)}{\rho^*}\ , \tag{4.10}$$

where the prefactor c has the units of energy and scales the magnitude of the force, ρ^* is the reference density, and ρ_n is the density profile computed at step n of the iterative procedure. The convergence to a uniform density is guaranteed by the fact that the scheme has a fixed point when $\nabla\rho = 0$. In addition to ensuring the same density in the two subdomains, this method also leads to a flat density profile in the HY region [49, 51, 62].

Usually, the procedure to compute the thermodynamic force consists in an equilibration phase of the simulation setup where no compensation is applied (with the possible exception of the drift force compensation), followed by a production run during which an accurate density profile is computed. The latter has to be sufficiently smooth so to employ its numerical gradient as a force in the following simulation. A new density profile is thus obtained and a new force is calculated. When the density is deemed to be uniform within a pre-established tolerance, the iterations are interrupted, and the compensation force is given by the sum of the terms computed up to that point.

The iterative calculation of the thermodynamic force relies on an accurate estimation of the density profile, which is performed on a very short time interval $\Delta T = \nu \delta t$, where δt is the integration time step and ν is an integer number of the order of $\sim 10^2 - 10^3$. The advantage of this procedure is that small deviations of the density from the reference value are immediately suppressed. However, it is evident that the density profile obtained from a small time interval ΔT is too noisy to reliably compute its numerical gradient.

To overcome this difficulty, the position of the centre of mass of the molecules is convoluted with a Gaussian function with a half-width $\sigma/2$, comparable with the typical excluded volume radius of the molecules. Thus, the density in the bin i, covering the coordinate range $[x_i, x_{i+1}]$ in a specific simulation frame, is computed as:

$$\begin{aligned} \hat{\rho}_i &= \sum_\alpha \frac{1}{A} \int_{x_i}^{x_{i+1}} dy \ \exp\left[-\frac{(y-x_\alpha)^2}{2\sigma^2}\right], \\ A &= \int_{-l}^{l} dy e^{-\frac{y^2}{2\sigma^2}} . \end{aligned} \tag{4.11}$$

The parameter l, whose appropriate value is system-dependent, controls the range of the Gaussian function. A sensible choice is to set $l = 2.5\sigma$.

4.4 Water–Ideal gas coupling

The reference, analytically solvable model of a liquid is the ideal gas (IG): in a computer simulation, this can be implemented as a collection of particles (with or without internal degrees of freedom) which do not interact one with the other. If subject only to the classical equations of motion and in absence of intermolecular potential, these particles would move along rectilinear trajectories with constant orientation and velocity (this is, assuming that the simulation box features periodic boundary conditions and not hard walls). To prevent this, and to be consistent with our goal of simulating a system in the canonical ensemble, a Langevin thermostat acts on the molecules in the ideal gas region, thereby imposing on them a stochastic, diffusive dynamics. On the other hand, the conservative part of the Hamiltonian becomes:

$$H = \mathcal{K} + V^{int} + \sum_\alpha \mathcal{F}(\lambda_\alpha) V_\alpha^{AT} , \tag{4.12}$$

where only the AT part of the interaction is present. $\mathcal{K}$ is the kinetic energy term and V^{int} is the potential resulting from intramolecular interactions. Since the interactions are modulated by the resolution λ, and $V_\alpha^{CG} = 0\,\forall\alpha$, particles can get extremely close at the HY/IG interface. Thus, a systematic sampling of huge potential energies might result in unphysical ensemble averages $\langle V^{AT}\rangle$ needed for the calculation of the compensations. To ensure a satisfactory statistical sampling, λ has been replaced by $\mathcal{F}(\lambda) = \lambda^\nu$ with $\nu \geq 1$, which also takes values between 0 and 1 (See Fig. 4.1). The value $\nu = 7$ is sufficient to smoothen out divergent interactions. Another advantage of using an exponent $\nu \geq 1$ is that the effective HY/IG interface moves closer to the AT domain. This results in a more stable and controlled thermodynamic force convergence. Furthermore, the AT potential is capped at a distance $\hat{r}$ to suppress large forces that might result from overlapping molecules. That is:

$$V^{AT}(r) = \begin{cases} V^{AT}(\hat{r}) - \left.\frac{\partial V^{AT}}{\partial r}\right|_{r=\hat{r}} (r-\hat{r}) & r < \hat{r} \\ V^{AT}(r) & r \geq \hat{r} \end{cases} . \tag{4.13}$$

In the case of liquid water at room temperature, the overlapping events are rare (one in 0.5 ns, approximately). Moreover, they do not affect thermodynamic or structural properties of the system, as we have verified in the case of SPC/E water model. Furthermore, they do not contribute significantly to the calculation of the compensations since high energy contributions, related to substantial overlaps of the molecules, are suppressed by the excluded volume and restricted to the tail of the configurational probability distribution.

The total force acting on the atom i of the molecule α can be obtained from the Hamiltonian (5.7) as:

$$\begin{aligned} \mathbf{F}_{\alpha i} &= \mathbf{F}_{\alpha i}^{int} \\ &+ \sum_{\beta,\beta\neq\alpha} \left\{ \frac{\mathcal{F}(\lambda_\alpha) + \mathcal{F}(\lambda_\beta)}{2} \mathbf{F}_{\alpha i|\beta}^{AT} \right\} \\ &- V_\alpha^{AT} \frac{\partial \mathcal{F}}{\partial \lambda}|_{\lambda=\lambda_\alpha} \nabla_{\alpha i}\lambda_\alpha \,. \end{aligned} \tag{4.14}$$

Finally, concerning electrostatic interactions, we have used an alternative to the Ewald summation method, i. e. the damped shifted potential (DSF) [65, 82]. In this approach, two charges q_i and q_j separated by a distance r_{ij} follow the electrostatic potential given by:

$$\begin{aligned} V_{\mathrm{DSF}}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0} &\left[\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\mathrm{erfc}(\alpha r_c)}{r_c} \right. \\ + &\left. \left(\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}} \frac{\exp\left(-\alpha^2 r_c^2\right)}{r_c} \right) (r_{ij} - r_c) \right] , \end{aligned} \tag{4.15}$$

where $r_{ij} \leq r_c$ and ϵ_0 is the vacuum permittivity. Only two parameters, the cutoff radius r_c and the damping parameter α, need to be specified. $\mathrm{erfc}(r)$ is the complementary error function that takes into account the damping proposed in [82]. The gradient of the potential (4.15) gives the force acting on atom i

$$\begin{aligned}\mathbf{F}_{\mathrm{DSF}}(\mathbf{r}_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0}\Bigg[&\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}^2} + \frac{2\alpha}{\pi^{1/2}}\frac{\exp(-\alpha^2 r_{ij}^2)}{r_{ij}} \\ &-\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} - \frac{2\alpha}{\pi^{1/2}}\frac{\exp(-\alpha^2 r_c^2)}{r_c}\Bigg]\frac{\mathbf{r}_{ij}}{r_{ij}}\,.\end{aligned} \tag{4.16}$$

This framework is employed to perform MD simulations in the canonical ensemble for 15615 SPC/E [133–135] water molecules in a simulation box of size $188 \times 50 \times 50$Å. The initial fully-atomistic equilibrated configuration has been obtained from a simulation of 100 ps in the NPT ensemble followed by a 0.1 ns equilibration run performed with a time step of $\delta t = 0.001$ ps. The temperature and pressure are enforced at T =298 K and P =1 bar using the Nosé-Hoover thermostat and barostat with damping coefficient of 0.1 ps and 1 ps, respectively. The parameters for the DSF electrostatic potential, damping coefficient $\alpha = 0.2$Å^{-1} and cut-off radius $r_c = 12$Å, were chosen to reproduce the RDFs of reference simulations using the Ewald summation method.

Once the initial configuration has been obtained, a `H-AdResS` simulation is performed using the Hamiltonian of Eq. 5.7 without any compensation terms. After 100 ps, the *on the fly* calculation of the drift force compensation is applied, with updates every 5 ps, during 1400 ps. The resolution interval is divided into 1000 bins of size $\Delta\lambda = 0.001$. The *on the fly* density balancing method is applied simultaneously to the drift force correction. In this case, the length of the simulation box is uniformly discretised into slabs of size $\Delta x = 1.0$Å and the thermodynamic force is updated every 10 ps. We employed values of $c = 8.368$ kJmol^{-1} Å^{-1}, $\sigma = 3$Å and $l = 6$Å for smoothing and scaling the thermodynamic force. All simulations are performed with the `LAMMPS` simulation package [61, 125], where the method is implemented, freely available and ready to use. The results presented in Fig. 4.2 for the compensation terms obtained with this method show a rather smooth behaviour that validates the present approach.

The successful coupling between a dense and strongly interacting fluid, such as water modelled at the atomistic level, and the ideal gas reference model has been previously demonstrated [143]: it is our scope, however, to include this verification in order to make the case for a sound and effective concurrent usage of atomistic and ideal models of different systems. The analysis of radial distribution functions (RDFs) and velocity autocorrelation functions (VACF) has been performed to check that the ideal gas reservoir bears no effect on structural, thermodynamical, and dynamical properties of the liquid in the all-atom subdomain. The results obtained in the dual-resolution setup are compared with the same quantities computed in fully atomistic benchmark simulation. In the case of the RDFs, Fig. 4.3 shows a remarkable agreement between the two cases.

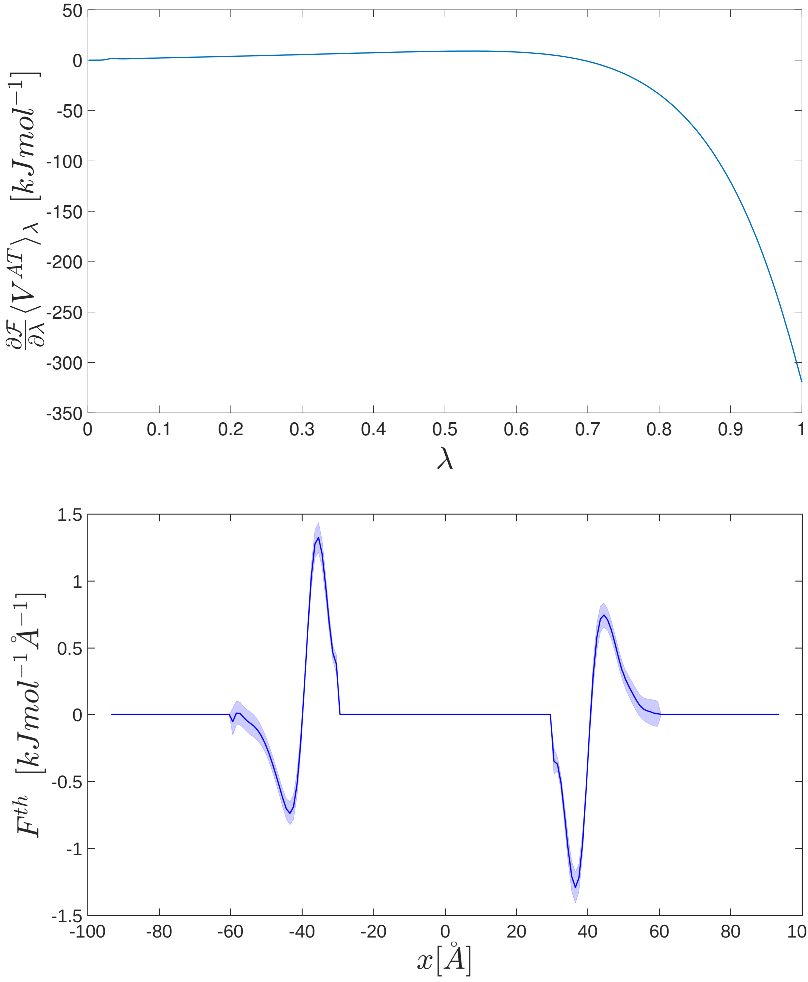


Figure 4.2. Force compensations obtained from the procedure described in the text. Upper panel: Compensation to the drift force as a function of λ. The use of the $\nu = 7$ coefficient guarantees that the contributions close to the IG/HY interface ($\lambda \sim 0$) are negligible compared to the contributions close to the AT region. Lower panel: Thermodynamic force as a function of the position within the simulation box.

The velocity autocorrelation function (VACF) is defined as [126]:

$$C_{vv}(t) = \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle \,, \tag{4.17}$$

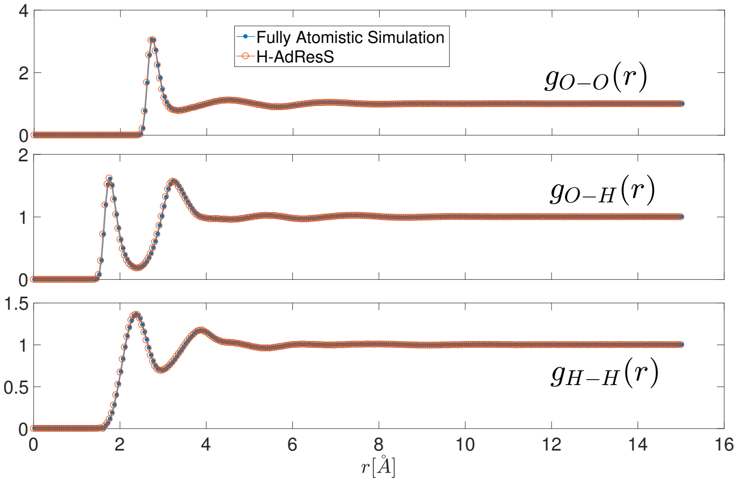


Figure 4.3. RDFs of water molecules at pressure $P_0 = 1bar$ and temperature $T_0 = 298K$ in two different simulation setups: fully atomistic simulation (blue line with dots) and `H-AdResS` (red line with open circles). From top to bottom, the plots show oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen RDFs. The DSF damping parameter is set to $\alpha = 0.2\mathring{A}^{-1}$, and the cut-off radius is $R_c = 12\mathring{A}$.

where $\mathbf{v}_i(t)$ is the velocity of molecule i at time t. To compute the VACF for the t_m-th time step the discrete estimator described in Ref. [131] was used:

$$\begin{aligned} C_{vv}(t_m) &= \\ &= \frac{1}{N_{AA}} \sum_{i=1}^{N_{AA}} \frac{1}{M-m} \sum_{n=0}^{M-m-1} \mathbf{v}_i(t_{n+m}) \cdot \mathbf{v}_i(t_n)\,, \end{aligned} \tag{4.18}$$

with M the total number of time steps and $t_m = m\delta t$, where δt is the integration time step. N_{AA} is the number of molecules that always remain within a predefined region of the simulation box. In the case of fully-atomistic simulations, $N_{AA} = N$, the total number of molecules. The error in the calculation of the VACF is given by $2t_{corr}/N_{AA}t_{tot}$ [126], with t_{corr} the correlation time and t_{tot} the total time of the simulation.

The results are reported in Fig. 4.4. These show that a perfect consistency exists between the observables computed in the all-atom reference simulations and those obtained from the `H-AdResS` runs.

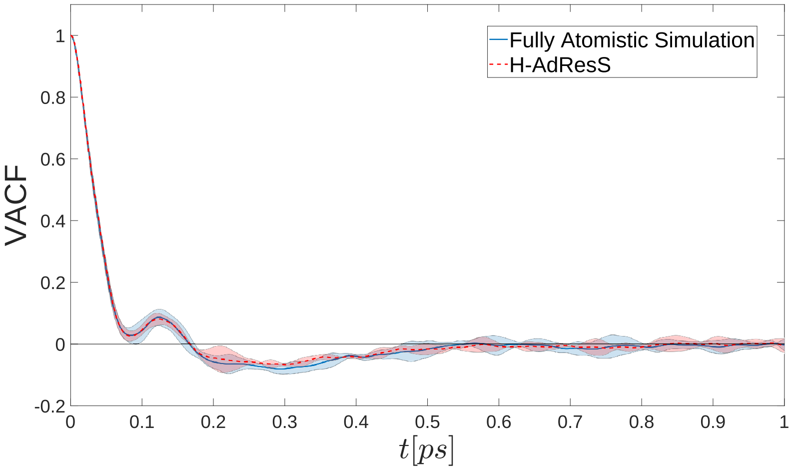


Figure 4.4. Normalized velocity autocorrelation function $(C_{vv}(t)/C_{vv}(0))$ calculated for a reference fully atomistic (blue curve) and for the `H-AdResS` (red curve) simulations. In both cases only the atoms with coordinates in the interval $-20\,\text{Å} < x < 20\,\text{Å}$ have been taken into account.

4.5 Solid–Einstein Crystal coupling

The possibility of modulating the resolution of a system as a function of the position in space of its parts can be easily extended from fluids to solids. Indeed, the field of material science has been the cradle of adaptive resolution simulation methods, since these were initially developed to perform computer simulations of e.g. crack propagation [201–206].

Here we describe the coupling of two different models of crystal within the framework of the `H-AdResS` method. Our objective, however, is not (only) to reduce the computational resources necessary to simulate large chunks of a solid of which only a tiny part, such as the expanding crack, is of interest; rather, we aim at two other goals: On the one hand, we want to describe how to practically perform an adaptive resolution simulation of a solid using, as a coarse-grained model, an extremely simple and exactly solvable representation, expanding the atomitstic-to-ideal gas coupling outside of the realm of liquids. On the other hand, we want to demonstrate that a substantial advantage can come from this coupling, namely the possibility to extract important thermodynamical information on the system's free energy.

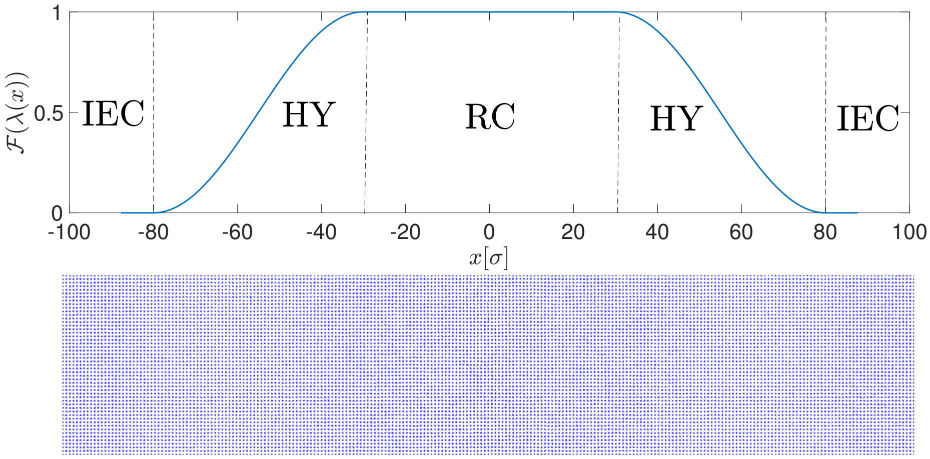


Figure 4.5. Setup of a Hamiltonian Adaptive Resolution Simulation for solids. The periodic box is divided into three different regions, namely: ideal Einstein crystal (IEC), hybrid (HY), and real crystal (RC). Upper panel: the switching function $\mathcal{F}(\lambda)$ takes values between 0 (IG) and 1 (AT), thus defining the resolution of a molecule. Lower panel: simulation snapshot explicitly showing the various subdomains. The system is composed of Lennard Jones particles at number density $\rho = 1.28\sigma^{-3}$ and the thermal energy $k_BT = 2.0\epsilon$.

Arguably, the most idealised particle-based representation of a solid is the ideal Einstein crystal (IEC) where N noninteracting particles are coupled to their lattice sites using harmonic potentials. The system's potential energy has the form:

$$V^E = \sum_{i=1}^{N} V_i^E = \frac{1}{2}\sum_{i=1}^{N} \kappa_i(\mathbf{r}_i - \mathbf{r}_i^0)^2 \,, \tag{4.19}$$

where each particle i fluctuates around its equilibrium position $\mathbf{r}_i^0$ with spring constant κ_i. To illustrate the method, as high-resolution model we employ a Lennard-Jones (LJ) potential V^{LJ}, whose well depth ϵ and excluded volume size σ set the energy and length scale, respectively. The corresponding `H-ADResS` Hamiltonian has the form:

$$H = \mathcal{K} + \sum_{i=1}^{N}\{\mathcal{F}(\lambda_i)V_i^{LJ} + (1 - \mathcal{F}(\lambda_i))V_i^E\} \,, \tag{4.20}$$

with $\mathcal{K}$ being the kinetic energy term and N the total number of particles in the system; the switching function is $\mathcal{F}(\lambda) = \lambda^{\nu}$ with $\nu = 1$. This Hamiltonian generates the following total force acting on the atom i:

$$\begin{aligned}\mathbf{F}_i = &\sum_{j\neq i}^{N} \left(\frac{\mathcal{F}(\lambda_i) + \mathcal{F}(\lambda_j)}{2} \mathbf{F}_{ij}^{LJ} \right) + (1 - \mathcal{F}(\lambda_i))\mathbf{F}_i^{E} \\ &- [V_i^{LJ} - V_i^{E}] \left. \frac{\partial \mathcal{F}(\lambda)}{\partial \lambda} \right|_{\lambda=\lambda_i} \nabla_i \lambda_i \, .\end{aligned} \tag{4.21}$$

Since, by construction, the density is uniform throughout the simulation box, the only energy compensation needed is the one required to counteract the drift force, i.e., the Helmholtz free energy difference between the LJ and the IEC models. Therefore, the Hamiltonian in Eq. (4.20) becomes:

$$H_{\Delta} = H - \sum_{i=1}^{N} \Delta H(\mathcal{F}(\lambda(\mathbf{r}_i))) \, , \tag{4.22}$$

with:

$$\left. \frac{d\Delta H(\mathcal{F}(\lambda(\mathbf{r}_i)))}{d\mathcal{F}(\lambda(\mathbf{r}_i))} \right|_{\lambda=\lambda_i} = \langle [V_i^{LJ} - V_i^{E}] \rangle_{\mathbf{r}_i} \, , \tag{4.23}$$

Analogously to the case of fluids, we can integrate the previous expression and obtain the Helmholtz free energy compensation:

$$\Delta H(\mathcal{F}(\lambda(\mathbf{r}_i))) = \int_0^{\lambda} d\lambda' \frac{d\mathcal{F}(\lambda')}{d\lambda'} \langle [V_i^{LJ} - V_i^{E}] \rangle_{\mathbf{r}_i} \, . \tag{4.24}$$

In the upper panel of Fig. 4.5 we provide a schematic representation of the simulation box and the different subregions: real crystal (RC), hybrid (HY), and Einstein crystal (IEC), together with $\mathcal{F}(\lambda(x))$. The lower panel of the same figure shows a snapshot of the simulation setup.

Eq. (4.24) is analogous to the difference in Helmholtz free energy as obtained from the thermodynamic integration (TI) method proposed by Frenkel and Ladd [171, 207–209]. The free energy of an IEC can be computed analytically, hence it can be used as the reference to calculate the Helmholtz free energy of a target crystal through a regular TI. To perform the latter, in our notation we write a Hamiltonian of the form:

$$H(\lambda) = \mathcal{K} + \lambda V^{LJ} + (1 - \lambda) V^{E} \, , \tag{4.25}$$

where in this case λ is a *global* coupling constant. The derivative of the free energy F with respect to λ gives:

$$\begin{aligned}\frac{\partial F}{\partial \lambda} &= -\beta^{-1}\frac{\partial}{\partial \lambda}\left[\ln \int d\mathbf{r}^N \exp(-\beta H(\lambda))\right] \\ &= \langle V^{LJ} - V^{E}\rangle_{\lambda}\,,\end{aligned} \tag{4.26}$$

and the free energy of a real crystal F^{RC} is related to the free energy of the Einstein crystal F^{IEC} by:

$$F(\lambda = 1) = F(\lambda = 0) + \int_0^1 d\lambda\, \langle V^{LJ} - V^{E}\rangle_{\lambda}\,, \tag{4.27}$$

with $F(\lambda = 1) = F^{RC}$ and $F(\lambda = 0) = F^{IEC}$.

To compare the results obtained employing the two methods, it is necessary to set the value of the spring constant κ in Eq. 4.19. A sound criterion to fix this parameter is to ensure that the mean squared displacement (MSD) of the particles in the fully atomistic simulation equals that of the ideal Einstein crystal, i. e. [171, 207]:

$$\frac{3}{\kappa\beta} = \langle (x - x_0)^2\rangle\,, \tag{4.28}$$

with $\beta^{-1} = k_B T$. For all cases, a simulation box of size $188.8\sigma \times 49.96\sigma \times 49.96\sigma$ is used. Initially, the LJ particles are placed on the fcc lattice structure with number density $\rho = 1.28\sigma^{-3}$ and their initial positions are set as the equilibrium positions (see Eq. 4.19). In `H-AdResS`, the sizes of the LJ crystal and hybrid regions are set to be 60σ and 50σ respectively (see Fig. 4.5). The temperature $k_B T = 2\epsilon$ is fixed by a Langevin thermostat with damping coefficient of 10τ. The particles are thermalised in the hybrid and IEC regions only, yet the temperature is uniform throughout the system (data not shown). In all simulations, the LJ potential is truncated at a cutoff radius $R_c = 2.7\sigma$ and not shifted [209]. The time step is $\delta t = 0.001\tau$. To obtain the drift force, the resolution interval is divided into 20 bins of size $\Delta\lambda = 0.05$, and every 50000 time steps the *on the fly* calculation of the drift force compensation is performed and then updated. The duration of each simulation run is at least 2×10^6 steps.

The time evolution of the MSD of the Lennard-Jones crystal is presented in Fig. 4.6. The curve shows the characteristic behaviour for a solid, where an initial diffusive regime is followed by a plateau, which in our case appears after 1 τ. From the asymptote at 0.024 σ^2 we obtain the spring stiffness $\kappa = 125 k_B T$: this value is employed in both the `H-AdResS` and the TI simulations.

To validate the consistency of the method, we computed the two components of the derivative of the Helmholtz free energy, namely $\langle V^{LJ}\rangle$ and $\langle V^{IEC}\rangle$, making use of the `H-AdResS` method and the TI, with and without the fixed centre of mass (CoM) constraint. These data are reported and compared in Fig. 4.7: the three data sets agree remarkably well, with the only exception of the point $\lambda = 1$ for the regular

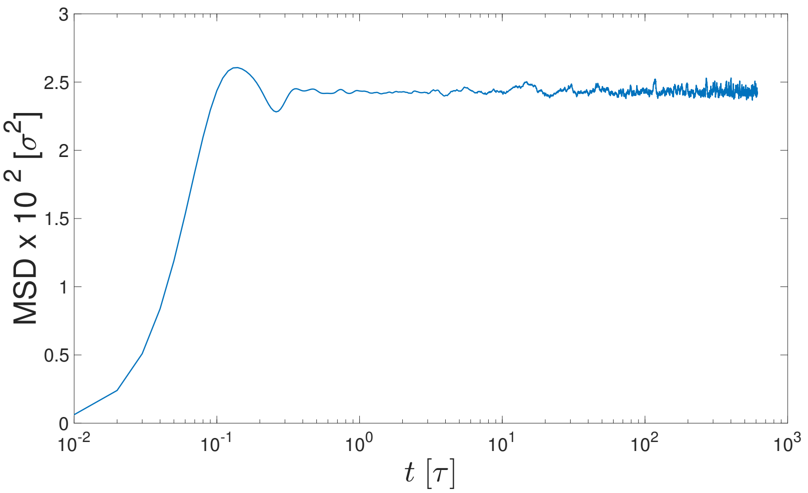


Figure 4.6. Mean squared displacement (MSD) of LJ particles obtained from a fully atomistic simulation. The curve reaches a plateau after 1 τ, and the asymptote is 0.024 σ^2.

TI. This is expected, because in the TI without restraints the LJ crystal is kept at a fixed position in space only through the coupling with the IEC. When $\lambda \rightarrow 1$ this constraint relaxes, and the LJ crystal gets asymptotically free to diffuse away from the IEC. For this reason, it is customary to include in the TI simulation a constraint to the LJ crystal centre of mass position. Conversely, in the `H-AdResS` simulation the LJ crystal cannot drift because of the explicit and permanent spatial coupling to the IEC. This is made evident by the excellent agreement for the full range of λ between the `H-AdResS` and the data obtained from a TI integration with fixed CoM.

As for the case of liquid–ideal gas coupling, also here we can monitor the consistency of the equilibrium dynamical properties of the crystal in the high-resolution region of the `H-AdResS` setup. This is done by calculating the VACF for the fully atomistic and the dual-resolution setup, whose results are presented in Fig. 4.8. In both cases the VACF shows the same behaviour within error bars, the latter being indicated by the shaded regions around the solid lines. As expected for a solid, the curves fluctuate around zero with amplitudes decreasing with time. These fluctuations result from the incoherent vibrations of the particles at their equilibrium positions. This is a remarkable result, since for an ideal Einstein crystal the VACF exhibits a sinusoidal behaviour: in the dual-resolution simulation, however, the dynamics of the LJ subdomain are not affected by the harmonic character of the IEC.

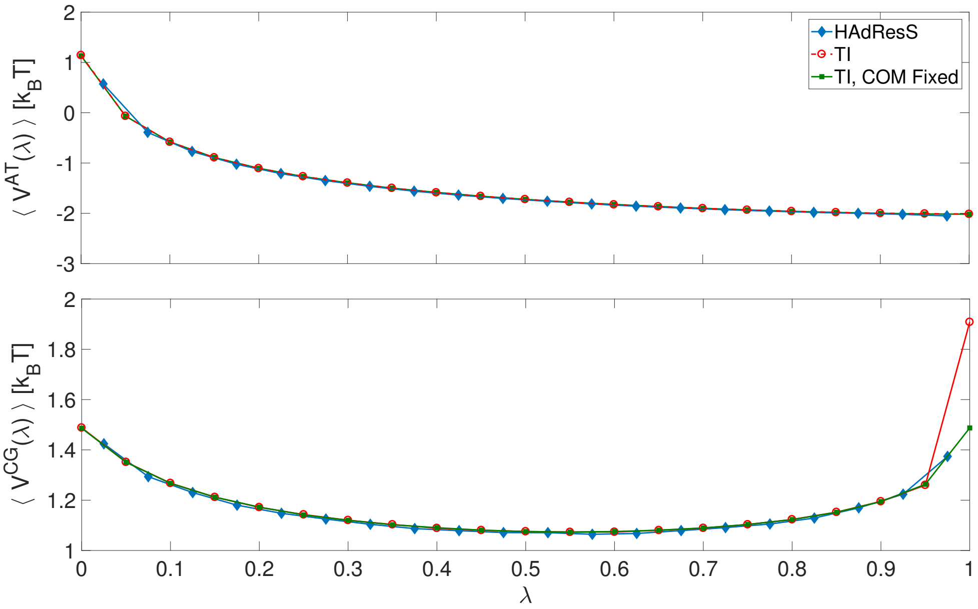


Figure 4.7. Atomistic potential V^{AT} (upper panel) and coarse-grained potential V^{CG} (lower panel) as a function of λ. In both panels, diamond, circle, and square data points are obtained from `H-AdResS`, thermodynamic integration, and fixed centre of mass (CoM) thermodynamic integration, respectively.

4.6 Conclusions and outlook

Ideal models, in the context of soft matter, are characterised by two elements: absence of interactions (or at least coupling) and exact integrability. This makes them historically pivotal in the comprehension of many physical systems, as they entail a great fraction of the complexity of the latter in spite of a bare-to-the-bones nature. Semi-perturbative corrections can take quite far from the initial simplicity and a lot closer to physical reality, as it is the case of van der Waals' equation of state for real liquids. When such a simple modification, which still leaves the integrability of the model substantially intact, is not possible, other routes can open up, such as the decomposition of physical quantities in ideal and excess terms: the problem is now flipped from a perturbative modification of the ideal case to a finite, often substantial yet additive contribution, whose calculation is made possible by analytical or computational tools.

Adaptive resolution simulations methods are the most suitable instruments to take advantage of the comparison between a realistic description of a system and the corresponding ideal models: this is because such comparison is as direct as possible,

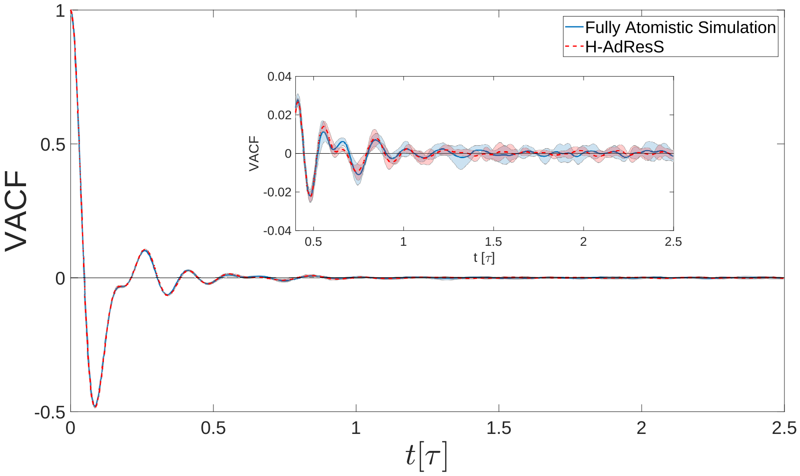


Figure 4.8. Normalized velocity auto-correlation function ($C_{vv}(t)/C_{vv}(0)$) of LJ solid. The result of the fully atomistic, reference simulation is shown with the solid blue curve. The result of the `H-AdResS` simulation is shown in red dashed curve. The simulation box has dimensions of $188.8\sigma \times 49.96\sigma \times 49.96\sigma$. In both cases, only those atoms with coordinates such that $-20\,\sigma < X < 20\,\sigma$ are considered for the calculation of the VACF.

with the two models coexisting in the same simulation and in direct contact, exchanging particles and energy. To enforce the equality of a given property in both subdomains at different resolution implies to compute the differences that exist between them, often in a simpler and more effective way than by means of well established techniques. In this work we have revised the application of the Hamiltonian adaptive resolution simulation scheme `H-AdResS` to the coupling of all-atom models of a liquid to an ideal gas, and have demonstrated how an analogous setup can be constructed for a solid. The subdomain where the system is described with an interacting, accurate model does not suffer from the coupling, as it is shown by all structural and dynamical quantities measured there, whose values are in excellent agreement with the fully high resolution reference. For the solid this is particularly relevant and remarkable, as any perturbation caused by the ideal model in the low-resolution domain cannot be expected to decay simply because of diffusion, as it is the case in liquids. The switch from the interacting system to the ideal one, that takes place in the hybrid region, is smooth enough as to decouple the fine details of each model's properties, yet preserving the same thermodynamics throughout the simulation box.

These results provide the conceptual and technical bedrock for applications with important consequences. The computational cost of these simulations, in fact, is reduced to that of the high-resolution domain, as the ideal gas/solid parts require a negligible amount of force calculations. The size of systems that can be treated with these technique can thus be substantially large, so as to minimise finite size effects and enable the simulation of complex molecules. As a particularly useful perspective application we envisage the calculation of chemical potentials of complex molecular crystals, a topic of increasing relevance for its industrial and pharmaceutical implications.

4.7 Acknowledgments

The authors thank R. Menichetti and G. Lattanzi for a critical reading and insightful comments. MH, KK and RP gratefully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG). RCH gratefully acknowledges the Alexander von Humboldt Foundation for financial support.

5. Nonequilibrium work relations in multiscale simulations of molecular liquids

This chapter is a draft of research paper that has been submitted in the Physical Review Letters during the submission of the thesis.

Maziar Heidari, Robinson Cortes-Huerto, Raffaello Potestio and Kurt Kremer
Nonequilibrium work relations in multiscale simulations of molecular liquids

Abstract

We verify the Jarzynski equality in multiscale simulations of soft matter by computing free energies of solvated molecules. These calculations are usually performed using particle-insertion methods where the coupling to nonequilibrium work relations poses major practical limitations. We approach this problem by combining steered molecular dynamics and spatially-resolved thermodynamic integration methods. Results for solvation free energies and chemical potentials of small molecules well agree with literature data and pave the way to systematic studies of arbitrarily large and complex molecules.

5.1 Introduction

Physical systems of reduced size often exhibit unexpected effects and phenomena that challenge our understanding of well-established physical principles. The second law of thermodynamics, for instance, imposes a strict lower bound to the amount of work necessary to bring a *macroscopic* thermalized system out of equilibrium (Clausius inequality). When a *microscopic* system is driven away from equilibrium, thermal fluctuations produce a distribution of work values that includes cases where the second law of thermodynamics breaks [210, 211]. Hence, in this context thermodynamic results must be interpreted statistically and the Clausius inequality is rewritten in terms of the average nonequilibrium work.

The intrinsic randomness in the thermodynamics of small systems reveals stronger statements relating this nonequilibrium work to equilibrium free energies. In particular, exact nonequilibrium relations such as the Jarzynski equality (JE) [91, 212], and the Crooks fluctuation theorem (CFT) [92], equate exponentials of free energy differences and fluctuations in the work performed to drive a physical system between equilibrium states, independently of the speed at which this process occurs.

This remarkable result found immediate application in various single molecule pulling experiments aiming at obtaining equilibrium free energy profiles [213–216] where many fast pulling events, in contrast to slow ones, can be easily realized. The JE has also captured the attention of the computer simulation community where methods to calculate potential of mean force [217], free energy of conformational changes [218] and even interconversion free energies [219] have been developed based on nonequilibrium work relations.

These results offer a tantalizing prospect for nonequilibrium computational methods aiming at calculating absolute free energies of solvated molecules. As a matter of fact, a large majority of the soft matter community's research efforts are devoted to such calculations [220–227]. However, bias and error [228] in the related estimators hinder the efficiency of nonequilibrium methods and persuade one to use optimized equilibrium methods instead [168, 229].

In this letter, we demonstrate the viability of using the JE to calculate absolute free energies in the context of multiscale simulations of solutions. We propose to take advantage of the smooth coupling present in the Hamiltonian adaptive resolution method (`H-AdResS`) [93] between a fully atomistic and an ideal representation of the solvent, and combine it with steered molecular dynamics simulations applied on the solute molecule.

To illustrate the method, we focus on the calculation of absolute solvation free energies (SFE). Nevertheless, we anticipate that the method can be easily extended to more complex multiple-solvation states and to compute other sought-after quantities such as binding free energies, relevant in view of drug discovery applications. The SFE is the difference in free energy resulting from considering the solute molecule in a solvent and in the gas phase at a given temperature and pressure. We directly compute the SFE as a function of the work necessary to pull the molecule from a region within the simulation box containing the solvent to another region containing an ideal gas representation of the solvent, thus effectively isolating the solute molecule.

This procedure resembles the thermodynamic integration method, and as such, the initial and final states needed to compute the change in free energy should correspond to identical systems but interacting via different potentials. Since it is required that these initial and final states are separated in space but smoothly merging one into the other, a uniform solvent density should be enforced throughout the simulation box. Thus, molecules can freely diffuse while changing their resolution, from interacting to ideal gas, *on the fly* which consequently guarantees a constant chemical potential.

5.2 Method

Recently [93, 143, 230], this simulation setup has been achieved within the `H-AdResS` [49, 50] formalism resulting in the Spatially Resolved Thermodynamic Integration (SPARTIAN) method [93, 230]. In the following and without loss of generality, we write the global `H-AdResS` Hamiltonian of a system consisting of N_s solvent molecules and one solute molecule σ in the form:

$$\begin{aligned} H = \mathcal{K} + V^{\mathrm{int}} + \sum_{\alpha \in N_\mathrm{s}} \left[\mathcal{F}(\lambda_\alpha) V_\alpha^{\mathrm{AT}} - \Delta H_\alpha \right] \\ + V_\sigma^{\mathrm{res}} + \mathcal{F}(\lambda_\sigma) V_\sigma^{\mathrm{AT}} - \Delta H_\sigma \, , \end{aligned} \tag{5.1}$$

where $\mathcal{K}$ is the kinetic energy and V^{int} include all the intramolecular interactions. The intermolecular interactions are described by the potential

$$V_\alpha^{AT} \equiv \frac{1}{2} \sum_{\beta, \beta \neq \alpha}^{N} \sum_{ij} V^{AT}(|\mathbf{r}_{\alpha i} - \mathbf{r}_{\beta j}|) \, ,$$

and $V_\sigma^{\mathrm{res}} = \kappa/2(x_{i\in\sigma} - x_{i\in\sigma}^0)^2$ is a restraining potential applied to the x-coordinate of the i-th atom of the solute molecule σ. For a given molecule with coordinates R [1], the function $\mathcal{F}(\lambda)$, which takes values between 0 and 1, specifies its resolution. That is, for $\mathcal{F}(\lambda) = 0$ intermolecular interactions are set to zero and the molecules behave as ideal gas (IG) particles and for $\mathcal{F}(\lambda) = 1$ the molecules interact fully atomistically (AT). In between these two cases, a hybrid (HY) region is defined where a smooth interpolation between 0 and 1 unambiguously defines the resolution of a molecule based on its position (schematically represented in Fig. 5.1).

The free energy compensation (FEC) terms $\Delta H_{\alpha,\sigma} \equiv \Delta H(\mathcal{F}(\lambda(\mathbf{R}_{\alpha,\sigma})))$ has been introduced in the Hamiltonian in Eq. (5.7) for two reasons: (i) to eliminate the spurious contribution to the dynamics of the system generated by forces proportional to the gradient of the switching function $\mathcal{F}(\lambda)$; and (ii) to guarantee that, at a given temperature T, the AT and IG representations of the solvent coexist in a thermodynamic state with equal density. In this particular example and for the case of solvent molecules, the FEC term has a simple and fundamental physical meaning: it is the difference in chemical potential between the AT and IG regions [49, 50], namely

$$\Delta H(\mathcal{F}(\lambda(R))) \equiv \frac{\Delta G(R)}{N} = \Delta\mu(R) \, , \tag{5.2}$$

with $\Delta G/N$ being the molar Gibbs free energy. Note that all quantities appearing in Eq. (5.2) are function of the molecule's position R: indeed, all free energies and chemical potential differences are computed with respect to a reference given by $R = R_a \equiv \lambda = 0$. The procedure to compute the FEC is described in detail in the section 5.5.

---

[1] R is a collective coordinate of the molecule, namely, the position of a given atom or the molecule's center of mass

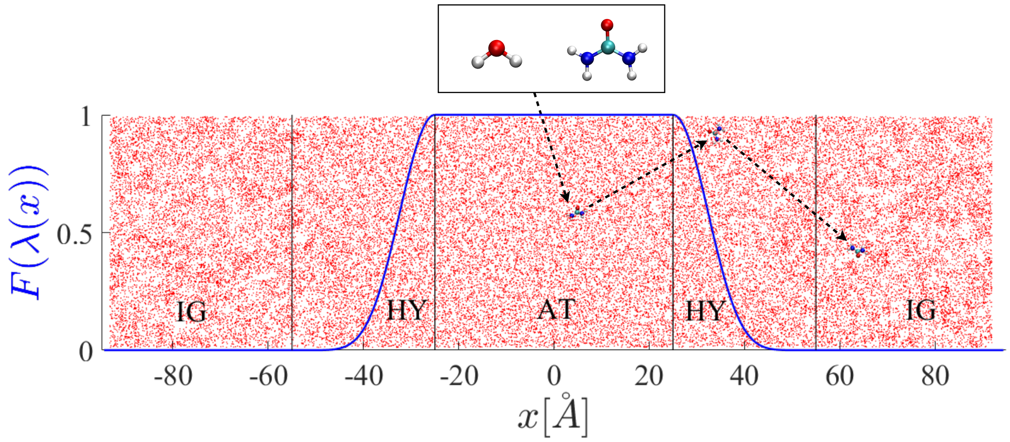


Figure 5.1. Snapshot of a typical `H-AdResS` setup showing the atomistic (AT), hybrid (HY) and ideal gas (IG) regions. The blue curve represents the switching function that smoothly interpolates the intermolecular interactions from fully atomistic to non-interacting. The molecules of interest, water and urea, are initially located in the AT region and then pulled across the simulation box until they reach the IG region.

5.3 Results

A constant chemical potential implies that pulling a water molecule from the AT to the IG region (or vice versa) is a barrier-less process. We illustrate this point with simulations of pure water (see computational details in the SI) carried out by combining umbrella sampling [231] with the SPARTIAN method. We select a water molecule located at the centre of the AT region (schematically indicated in Fig. 5.1) and restrain the X coordinate of its oxygen atom using the harmonic potential V_σ^{res} in Eq. (5.7) with $\kappa = 209.2\,\text{kJmol}^{-1}\text{\AA}^{-2}$. The solute molecule, i. e. the selected water molecule, is moved sequentially by $\Delta x^0_{\text{oxygen}} = 0.2\text{\AA}$ and sampled for 20 ps to construct the biased probability distribution of x_{oxygen}. We perform six uncorrelated simulation runs to calculate the solvation free energy profile. Each profile is shifted in such a way that the average of the solute's free energy in the ideal gas region is equal to zero. Averages and standard deviations (error bars) are reported in Fig. 5.2 (blue points) which show, as expected and within error bars, a flat free energy profile across the simulation box.

The solvation free energy of a water molecule in water, i. e. the work necessary to bring a water molecule from the bulk to the gas phase, precisely amounts to the excess chemical potential of water. In practice, this corresponds to repeating the same umbrella sampling calculation as described before but without applying the FEC

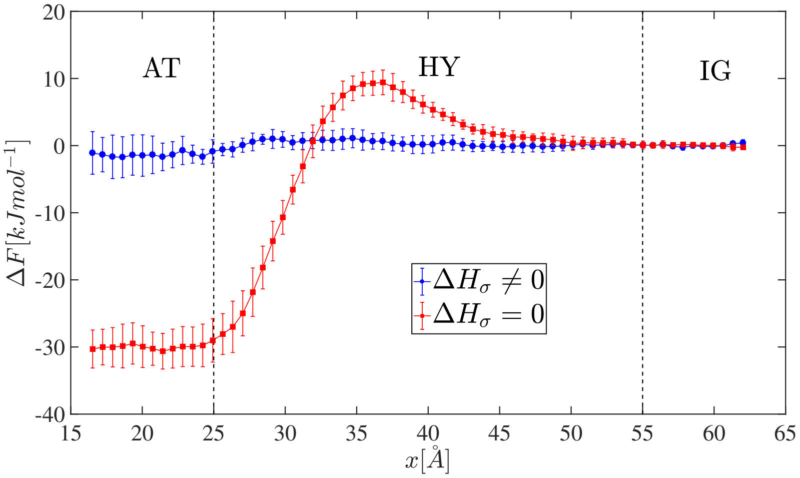


Figure 5.2. Free energy profile of a water molecule transferred between atomistic and ideal gas regions. The blue flat profile corresponds to the case when energy compensations are applied to all molecules in the system and indicates that there is no energy barrier for solvent molecules to freely diffuse between regions, i. e. the chemical potential is constant. The red curve corresponds to the case when there are no compensation forces applied on the *solute* water molecule. The barrier in the HY region is related to the particular choice of the function $\mathcal{F}(\lambda)$. However, given the flat profile observed in both, AT and IG regions, it is possible to estimate a net free energy difference that corresponds to the solvation free energy of water in water, i.e. the water excess chemical potential. The uncertainty in the calculations is indicated by the error bars.

($\Delta H_\sigma = 0$ in Eq. (5.7)) to the *solute* water molecule. Results are also reported in Fig. 5.2 (red points) where a flat energy profile is apparent in both the AT and IG regions and allows one to compute a difference in free energy of -31.03 ± 2.92 kJ/mol [2]. This result agrees within error bars with the excess chemical potential as reported in the literature (-29.52 ± 0.03 kJ/mol [157], -29.18±0.16 kJ/mol [3]). In this way, we have validated the spatially-resolved thermodynamic integration as a method to compute solvation free energies of molecular fluids.

For the second part of this study, we focus on the calculation of the solvation free energy of one urea molecule in water. We simulate urea and water molecules using the OPLS [232, 233] and SPC/E [133–135] force fields, respectively (see the computational

---

[2] The choice of the DSF model for electrostatic interactions implies the use of a correction discussed in the SI.

[3] SPARTIAN result, see the SI for details

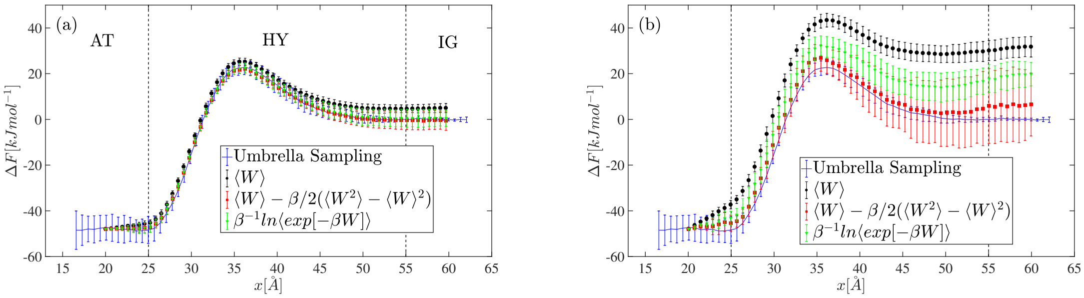


Figure 5.3. Free energy profile of one urea molecule transferred between the atomistic region and the ideal gas with pulling rate (a) $v = 10\AA/ns$ and (b) $v = 50\AA/ns$. The blue curve in each panel is obtained by using the umbrella sampling method. The free profiles calculated by Jarzynski relation using the first, first and second and exponential averages of work are represented by black, red and green curves respectively. The averages and error bars are calculated over 6 sets of 10 independent simulation runs.

details in the SI). First, and as it has been done in the case of pure water, we use SPARTIAN in combination with umbrella sampling. In this case, the carbon atom of the urea molecule is restrained and pulled across the simulation box using the harmonic potential V_σ^{res} in Eq. (5.7) with the same spring constant $\kappa = 209.2\,\text{kJmol}^{-1}\text{\AA}^{-2}$. The resulting free energy profile is shown in Fig. 5.5 (blue points). Similar to the simulation of the water molecule, the free energy profile is flat in both atomistic and ideal gas regions and we use it as the benchmark for the next part of the work.

We now proceed to perform steered molecular dynamics of the solute molecule to bring it from the AT to the IG region and, by accumulating enough statistics for the applied work, we use the JE to compute the solvation free energy. Let us suppose we have the system initially at equilibrium at temperature T in a state A that corresponds to having the solute molecule in the AT region. We apply the restraining potential $V_\sigma^{\text{res}} = \kappa/2(x_{i\in\sigma}(t) - x_{i\in\sigma}^0 - vt)^2$ with v the pulling speed. We use the same value of κ as before which is high enough to neglect, within error bars, the contribution of the restraining potential in the calculation of the free energy. In this way, we calculate the difference in free energy for the unconstrained system [218]. External work W on the system will bring it into a new state B that corresponds to having the solute molecule in the IG region. The difference in free energy ΔF between the states A and B is related to the external work on the system by the thermodynamic inequality $\Delta F \leq W$. The equality $\Delta F = W$ holds for quasi-static processes, which in our case correspond to the umbrella sampling calculations. Alternatively, the JE

$$\langle \mathrm{e}^{-\beta W} \rangle = \mathrm{e}^{-\beta \Delta F}\,, \tag{5.3}$$

with $\beta = 1/k_\mathrm{B}T$, is valid for arbitrarily fast processes. In practice, to make use of the JE it is possible, on the one hand, to express $\langle \mathrm{e}^{-\beta W} \rangle$ in terms of a cumulant expansion [228]:

$$\Delta F = \langle W \rangle - (\beta/2)(\langle W^2 \rangle - \langle W \rangle^2) + \cdots \, , \tag{5.4}$$

with $\langle W^2 \rangle - \langle W \rangle^2$ the mean-squared variation of W. The previous equation is related to the statistical expression $\langle \mathrm{e}^x \rangle \geq \mathrm{e}^{\langle x \rangle}$. On the other hand, it is also possible to directly use the exponential average, i. e.

$$\Delta F = -\beta^{-1} \ln(\langle \mathrm{e}^{-\beta W} \rangle) \, . \tag{5.5}$$

In both cases, it is necessary to compute the applied work W on the system that we obtain as:

$$W = -\kappa v \int_0^\tau dt' \left[x_{i\in\sigma}(t') - x^0_{i\in\sigma} - vt' \right] \, , \tag{5.6}$$

where the initial position of the solute is set at $x^0_{i\in\sigma} = 20\,\text{Å}$ and the simulation total time τ is set such that the final equilibrium position $x^0_{i\in\sigma} + v\tau$ is 60Å. We pull the restrained atom in the solute molecule at two constant speeds: $v = 10\text{Å}\text{ns}^{-1}$ and $v = 50\text{Å}\text{ns}^{-1}$ (the latter being comparable to the thermal velocity of urea at room temperature). To accumulate statistics, we generate 60 uncorrelated trajectories and we group them into 6 blocks of 10 trajectories. The statistical analysis (first and second orders of the cumulants as well as the exponential average) is carried out for each block. In Fig. 5.5 (upper panel), the result of computing $\langle W \rangle$ for a pulling rate of $v = 10\text{Å}\text{ns}^{-1}$ shows a relatively small deviation with respect to the free energy of approximately $4\,\text{kJmol}^{-1}$ (black data points). Nevertheless, the use of the JE requires to either add the contribution of higher order cumulants as in Eq. (5.4) (red data points) or to evaluate the exponential average as in Eq. (5.5) (green data points). In this case, both estimates increase the accuracy with respect to the benchmark free energy curve. We report the solvation free energy of one urea molecule as obtained from the exponential average as $-52.9 \pm 3.6\,\text{kJmol}^{-1}$ in good agreement with the result reported in Ref. [234] of the excess chemical potential of aqueous urea (-55.13 kJmol^{-1}) at a mole fraction of 0.0007657. Concerning the efficiency of the method, the profiles presented in Fig. 5.5 (upper panel) obtained from the JE effectively required to run ten times more molecular dynamics steps than the umbrella sampling result. However, it is also apparent that the JE results are more precise since the error bars are systematically smaller. In the SI we present free energy profiles with reduced size of the statistical sample and we find that to obtain comparable uncertainty in comparison with umbrella sampling, it is required to run only three times more molecular dynamics steps. This result shows that one can confidently use the Jarzynski equality in this context, and also suggests that the (relative) extra computational effort might substantially decreases upon increasing the complexity of the solute.

The amount of non-reversible work increases significantly upon increasing the pulling rate to 50 Å/ns, as presented in the lower panel in Fig. 5.5. The first order cumulant $\langle W \rangle$ (black data points) shows a difference of approximately $30\,\mathrm{kJmol^{-1}}$ with respect to the free energy as computed with the umbrella sampling method (blue points). In this case, the second order cumulant (red points) is more accurate with respect to the umbrella sampling calculation than the logarithmic estimator in Eq. (5.5) (green points). However, the uncertainty in the second order cumulant calculation is large because of the limited sampling [218, 228]. The error in the exponential average is lower, i.e. with smaller fluctuations, but still with a considerable mean deviation of approximately $20\,\mathrm{kJmol^{-1}}$ from the reference value, that evidences the well-documented bias of the estimator [228]. As it has been pointed out in Ref. [218], good statistics in block averaging does not necessarily correlate with better accuracy. At this pulling rate we try the reverse process, namely, to transfer the urea molecule from the ideal gas to the AT region. Results are presented in the SI (Fig. 3) showing a consistent behaviour with the results for the forward process, thus indicating the absence of hysteresis in the simulation setup. Furthermore, work distributions for the forward and reverse processes are presented in the SI (Fig. 4) suggesting that in SPARTIAN forward and reverse processes can be combined in terms of the Crooks fluctuation theorem to also compute solvation free energies.

5.4 Conclusion

In conclusion, we have verified the Jarzynski equality in the context of multiscale simulations of soft matter and used it to compute, rather efficiently and accurately, the absolute solvation free energy of small molecules. In our method, the smooth coupling enforced by the `H-AdResS` method between a fully atomistic model of the solvent and the ideal gas representation allows the simultaneous presence, within the same simulation setup, of the two end states of a thermodynamic integration between a solvated state and a non-interacting, effectively vacuum, state. In addition to umbrella sampling, steered molecular dynamics is used on the solute molecule and its solvation free energy is computed as a function of the work applied on the system via the Jarzynski equality. The most prominent advantage of this computational strategy consists in the possibility of solvating large molecules, e.g. initially inserting them in the ideal gas subregion and pulling them into the AT domain. In this way, the solvation free energy of arbitrarily large and complex molecules –from novel chemical compounds for pharmaceutical use to entire proteins– becomes accessible.

5.5 Supplementary Information

5.5.1 Computational details

We perform all the simulations in this study using the SPC/E [133–135] water model and the OPLS force field for urea [232, 233] implemented in the LAMMPS Molecular Dynamics engine [125] and describe electrostatic interactions with the Damped Shifted Force (DSF) potential [65] with damping coefficient $\alpha = 0.2\,\text{Å}^{-1}$ and cut-off radius $r_c = 12\,\text{Å}$. The `H-AdResS` method is implemented in LAMMPS [61], freely available and ready to use. We start by equilibrating a fully atomistic simulation box for 500 ps in the NPT ensemble with a time step $\delta t = 0.001$ ps. The bonds and angles are constrained using the SHAKE algorithm [235]. The temperature and pressure are enforced at $T = 298$K and $P = 1$ bar using the Nosé-Hoover thermostat and barostat with damping coefficient of 0.1 ps and 1 ps, respectively. In all simulations, the total number of water molecules is 15614 and the dimension of the equilibrated simulation box is $188.6 \times 49.93 \times 49.93$Å. This corresponds to a water number density of approximately 33 water molecules per nm^3. The equilibrated configuration is used as the input for the `H-AdResS` simulation where temperature has been enforced by a Langevin thermostat with the same damping coefficient. Following our previous work, SPARTIAN simulations [93] have been performed to calculate the excess chemical potential with precision comparable to that of the most common and accurate computational methods. In particular, sizes of the atomistic and hybrid region have been chosen as $L_{AT} = 50$Å and $L_{HY} = 30$Å, respectively.

The solvent `H-AdResS` Hamiltonian of the system has the form [93, 230]:

$$H = \mathcal{K} + V^{int} + \sum_{\alpha} \mathcal{F}(\lambda_\alpha) V_\alpha^{AT}\,, \tag{5.7}$$

where $\mathcal{K}$ is the kinetic energy and V^{int} includes all the intramolecular interactions. For a given molecule α with coordinates R_α, the function $\mathcal{F}(\lambda_\alpha)$, which takes values between 0 and 1, specifies its resolution. That is, for $\mathcal{F}(\lambda_\alpha) = 0$ intermolecular interactions are set to zero and the molecules behave as ideal gas (IG) particles and for $\mathcal{F}(\lambda_\alpha) = 1$ the molecules interact fully atomistically (AT). In between these two cases, a hybrid (HY) region is defined where a smooth interpolation between 0 and 1 unambiguously defines the resolution of a molecule based on its position.

The total force acting on an atom i of a molecule α reads:

$$\mathbf{F}_{\alpha i} = \mathbf{F}_{\alpha i}^{int} + \sum_{\beta, \beta \neq \alpha} \left\{ \frac{\mathcal{F}(\lambda_\alpha) + \mathcal{F}(\lambda_\beta)}{2} \mathbf{F}_{\alpha i|\beta}^{AT} \right\} - V_\alpha^{AT} \left. \frac{\partial \mathcal{F}}{\partial \lambda} \right|_{\lambda = \lambda_\alpha} \nabla_{\alpha i} \lambda_\alpha\,, \tag{5.8}$$

where $\mathbf{F}_{\alpha i}^{int}$ contains the intramolecular forces, and the intermolecular forces $\mathbf{F}_{\alpha i|\beta}^{AT}$ are weighted by the average resolution between the molecules α and β.

A spurious *drift* force proportional to the gradient of $\mathcal{F}(\lambda_\alpha)$ pushes molecules in/out of the hybrid region depending on where the Helmholtz free energy is the lowest. Accordingly, the Hamiltonian Eq. 5.7 should be modified to neutralize such a force. The new Hamiltonian H_Δ takes the form:

$$H_\Delta = H - \sum_\alpha \Delta H(\mathcal{F}(\lambda(\mathbf{R}_\alpha)))\,. \tag{5.9}$$

In the particular case of the coupling to an IG representation we have [93, 230]:

$$\left.\frac{d\Delta H(\mathcal{F}(\lambda(\mathbf{R}_\alpha)))}{d\mathcal{F}(\lambda(\mathbf{R}_\alpha))}\right|_{\lambda=\lambda_\alpha} = \langle V_\alpha^{AT}\rangle_{\mathbf{R}_\alpha}\,. \tag{5.10}$$

The total drift force thus becomes:

$$\hat{\mathbf{F}}_\alpha^{dr} = \left(V_\alpha^{AT} - \langle V_\alpha^{AT}\rangle_{\mathbf{R}_\alpha}\right)\nabla\lambda\left(\mathbf{R}_\alpha\right)\,, \tag{5.11}$$

and $\langle\hat{\mathbf{F}}_\alpha^{dr}\rangle = 0$ as required. The associated compensation to the energy guarantees that both, AT and IG regions, attain the same pressure. However, their density might not be the same [49, 50]. To ensure a flat density profile, a thermodynamic force is iteratively computed and applied [61]:

$$\mathbf{F}_{n+1}^{Th} = \mathbf{F}_n^{Th} + \frac{c\nabla\rho_n(x)}{\rho^*}\,. \tag{5.12}$$

with n the iteration step, ρ^* the reference density and ρ the actual density. The parameter c fine-tunes the force correction and has units of energy. The total force acting on the molecules present in the HY region is equal to the sum of the compensation needed to neutralize the drift force plus the thermodynamic force. The corresponding energy compensation can be obtained upon integration.

$$\Delta H(R_b) = -\int_{R_a}^{R_b} dR\left[-\mathcal{V}(\lambda)\nabla\lambda(R) + \mathbf{F}^{Th}(R)\right]\,, \tag{5.13}$$

where $R_a = r_{AT} + d_{HY}$, $R_b = r_{AT}$, with r_{AT} and d_{HY} the linear dimensions of the AT and HY regions, respectively, and $\mathcal{V}(\lambda) \equiv \langle V^{AT}\rangle_{\mathbf{R}}$.

The iterative calculations of the drift and thermodynamic force are simultaneously started. The drift forces is updated every 10 ps. The resolution interval is divided into 1000 bins of size $\Delta\lambda = 0.001$. The *on the fly* density balancing method is applied simultaneously to the drift force correction [61]. In this case, the length of the simulation box is uniformly discretized into slabs of size $\Delta x = 1.0\,\text{Å}$ and the thermodynamic force is updated every 50 ps. We employed values of $c = 4.148$ kJ mol^{-1}Å^{-1}, $\sigma = 3\,\text{Å}$, $l = 6\,\text{Å}$ and $\rho^* = 0.03Å^{-3}$ for smoothing and scaling the thermodynamic force. The updates of drift and thermodynamic forces are continued for 5 ns.

Following the `H-AdResS` parameterization, the umbrella sampling or steered molecular dynamics is performed in canonical ensembles using Langeving thermostat with damping coefficient of 0.1 ps and temperature 298 K. In all simulations, the spring stiffness of the restraining potential is $\kappa_b = 209.2\mathrm{kJmol}^{-1}\mathrm{\AA}^{-2}$. For the umbrella sampling in particular, the biased probability distribution of the restrained atom/molecule position is constructed for each $\Delta x_b = 0.2\,\mathrm{\AA}$ increment along the x-axis. In each increment, the position of the solute is sampled every 5 fs for 30000 fs long.

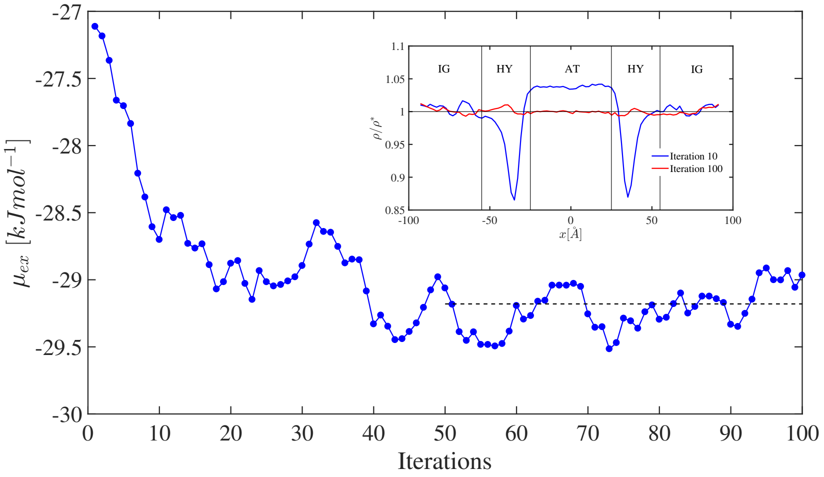


Figure 5.4. Excess chemical potential of SPC/E water molecules plotted against iterations to compute the FEC in the SPARTIAN method. The average and standard deviation of the excess chemical potential during the last 50 iterations are obtained as $\mu_{\mathrm{exc}} = -29.18 \pm 0.16$ kJ/mol (indicated by dashed black line). In the inset the density profiles of the system after 10 and 100 iterations are shown with blue and red lines, respectively.

The evolution of the FEC, i. e. excess chemical potential μ_{exc}, is presented in Fig. 5.4 where fluctuations around a limiting value are apparent after 50 iterations. Once convergence has been achieved, a flat density profile evidences a constant chemical potential throughout the simulation box. Finally, we report $\mu_{\mathrm{exc}} = -29.18 \pm 0.16$ kJ/mol [93] that reasonably agrees with the result $(29.53 \pm 0.03\,\mathrm{kJmol}^{-1})$ obtained by using the Bennet acceptance ratio method [157].

5.5.2 Umbrella sampling

We use the umbrella sampling [231] technique to obtain the free energy difference resulting from transferring the molecule from the AT to the IG region. A harmonic potential is introduced into the Hamiltonian of the system to perform biased sampling. This potential is resolution independent (similar to V^{int}) and constrains the x-component of an atom "i" of the molecule (x_p).

$$V_i^b = \frac{\kappa}{2}\left(x_p - x_i^0\right)^2 . \tag{5.14}$$

The equilibrium position (x_i^0) of the harmonic potential is varied incrementally (Δx_b) along the reaction coordinate, which is set equal to the x-axis. This gives a total number of windows N_w. For every increment, the biased probability distribution of the position $P_i^b(x_b)$ is calculated by sampling N_i number of steps in every window:

$$P_i^u(x_p) = P_i^b(x_p)\mathrm{exp}\left[\beta V_i^b(x_p)\right]\left\langle \mathrm{exp}\left[-\beta V_i^b(x_p)\right]\right\rangle \tag{5.15}$$

where $\beta = 1/k_BT$ and the ensemble average is obtained using the unbiased probability distribution,

$$\begin{aligned}\left\langle \mathrm{exp}\left[-\beta V_i^b(x_p)\right]\right\rangle &= \int dx_p P^u(x_p)\mathrm{exp}\left[-\beta V_i^b(x_p)\right] \\ &= \int dx_p \mathrm{exp}\{-\beta\left[\Delta F(x_p) - V_i^b(x_p)\right]\}\end{aligned} \tag{5.16}$$

Here, $\Delta F(x) = -(1/\beta)\log\left[P^u(x)\right]$ is the free energy of the unbiased system. Using the Weighted Histogram Analysis Method [231, 236] (WHAM), the global unbiased probability distribution (P^u) is calculated as the weighted average of unbiased probability of each window:

$$P^u(x_p) = \sum_i^{N_w} \alpha_i(x_p) P_i^u(x_p) \tag{5.17}$$

The prefactor α_i which is obtained by minimizing the statistical error of P^u reads:

$$\alpha_i(x_p) = \frac{N_i \mathrm{exp}\left[-\beta\left(V_i^b(x_p) - F_i\right)\right]}{\sum_j N_j \mathrm{exp}\left[-\beta\left(V_j^b - F_j\right)\right]} , \tag{5.18}$$

where N_i is the number of steps sampled for window i and F_i is given by Eq. 5.16

$$\exp\left(-\beta F_i\right) = \left\langle \exp\left[-\beta V_i^b(x_p)\right]\right\rangle . \tag{5.19}$$

Equations 5.15 to 5.17 have to be iterated until convergence. In all simulations, we used equal number of samples at each window, i.e. $N_i = N_j$ for all i and j, and the convergence criteria is set to when $\sum_i \left|F_i^{new} - F_i^{old}\right| < 10^{-4}$.

5.5.3 Mean-squared displacement of urea molecule

We compute the mean-squared displacement (MSD) of a single urea molecule in water. From the plot of the MSD in figure 5.5, we obtain the self-diffusion constant $D = 2.8 \times 10^{-3}\mathrm{nm}^2\mathrm{ps}^{-1}$ in good agreement with the reference value reported in [237]

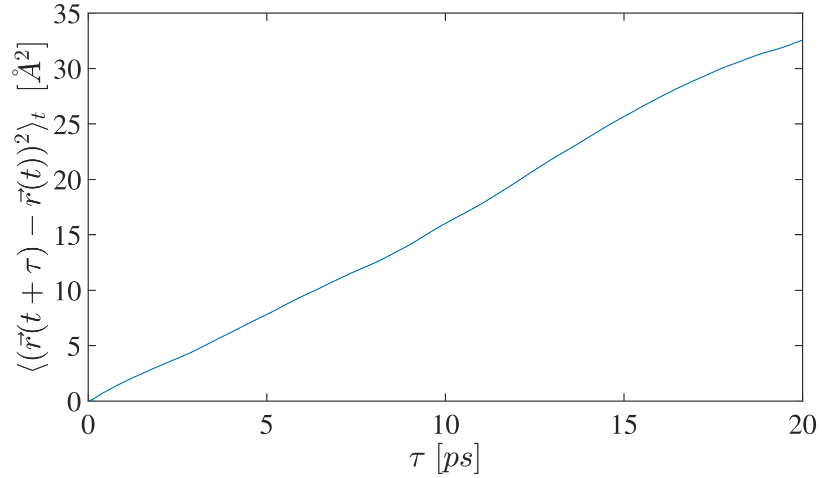


Figure 5.5. Mean-squared displacement of a single urea molecule in 15615 SPC/E water molecules (mole fraction is 6.4×10^{-5}). The self-diffusion constant is obtained $D = 2.8 \times 10^{-3}\mathrm{nm}^2\mathrm{ps}^{-1}$ which agrees well with the reference [237]

5.5.4 Increasing number of trajectories

In this section we show that by doubling the size of the statistical sample the presented results of the solvation free energy as obtained with the JE (pulling rate $v = 50\,\text{Å/ns}$) do not significantly improve in terms of accuracy. This is presented in Figure 5.6 where energy profiles obtained by the average, second order cumulant and logaritmic average of the work, obtained over twelve blocks consisting of ten independent simulation runs, are very similar to the ones presented in Fig. 4(b) of the main manuscript.

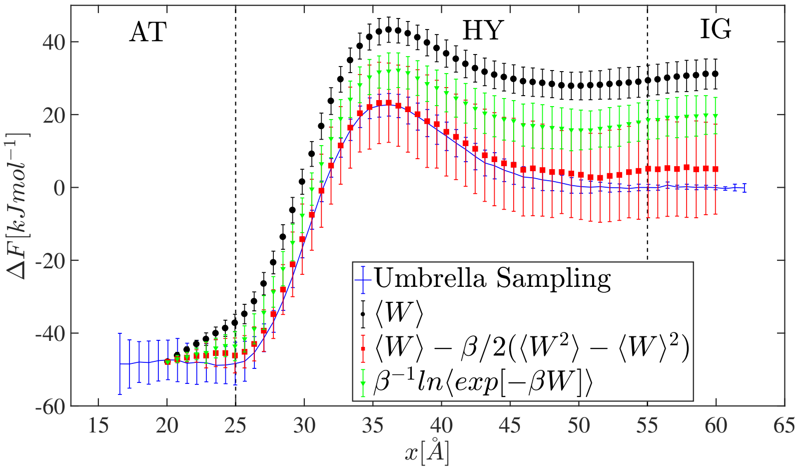


Figure 5.6. Free energy profile of a single urea molecule being transferred between the atomistic and the ideal gas regions with pulling rate $v = 50\AA/ns$. The free profiles calculated by the first and second and exponential averages of work are represented by red, black and green curves respectively. The averages and error bars are calculated over 12 sets of 10 independent simulation runs. The shaded color around the curve indicates the error bars.

5.5.5 Crooks fluctuation theorem

In addition to the JE, the Crooks fluctuation theorem (CFT) also relates the work performed on a system during a nonequilibrium transformation to the difference in free energy between the final and initial states of the transformation. For a system, initially at equilibrium, the CFT has the form [92]:

$$\frac{P_{\mathrm{F}}(+\beta W)}{P_{\mathrm{R}}(-\beta W)} = \mathrm{e}^{-\Delta F + \beta W}\,, \tag{5.20}$$

where W is the work performed on the system resulting from forward (F) and reverse (R) processes and $\beta = 1/k_{\mathrm{B}}T$. The difference in free energy ΔF is obtained by evaluating the ratio of the distributions of forward and reverse work $P_{\mathrm{F}}(+\beta W)/P_{\mathrm{R}}(-\beta W)$.

To make use of the CFT, we perform a number of reverse equal to the number of forward processes. Reverse in this context means to bring a urea molecule from the IG to the AT region. Free energy profiles are presented in Figure 5.7. For this case as well, the best estimate with respect to the umbrella sampling profile is given by the second order cumulant expansion (red points).

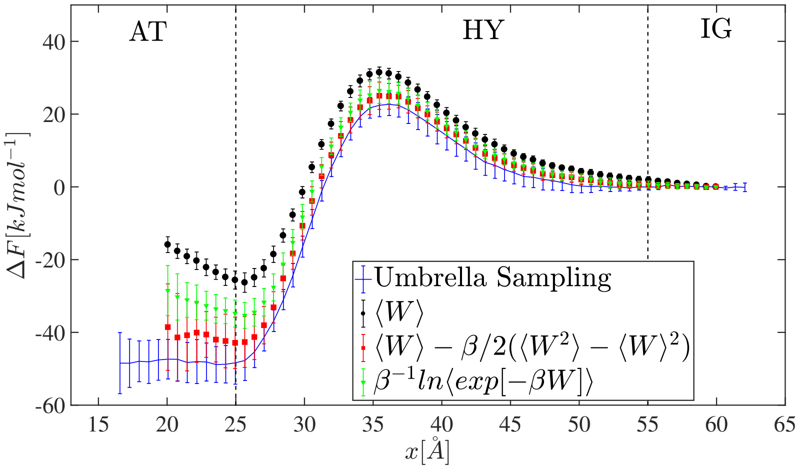


Figure 5.7. Free energy profile of a single urea molecule being transferred from the ideal gas region into atomistic region with pulling rate $v = 50$ Å/ns. The free profiles calculated by the first and second and exponential averages of work are represented by red, black and green curves respectively. The averages and error bars are calculated over 6 sets of 10 independent simulation runs. The shaded color around the curve indicates the error bars.

Finally, we compute the work distributions $P(\pm\beta W)_{\mathrm{F/R}}$ for both reverse and forward processes and present them in Figure 5.8. To estimate the intersection of the distributions, i. e. the point at which the $W = \Delta F$, we fit a gaussian curve and report $\Delta F = 44.75\,\mathrm{kJmol^{-1}}$. It is thus clear that more statistics and data from different pulling rates need to be collected in order to convincingly calculate ΔF from the CFT.

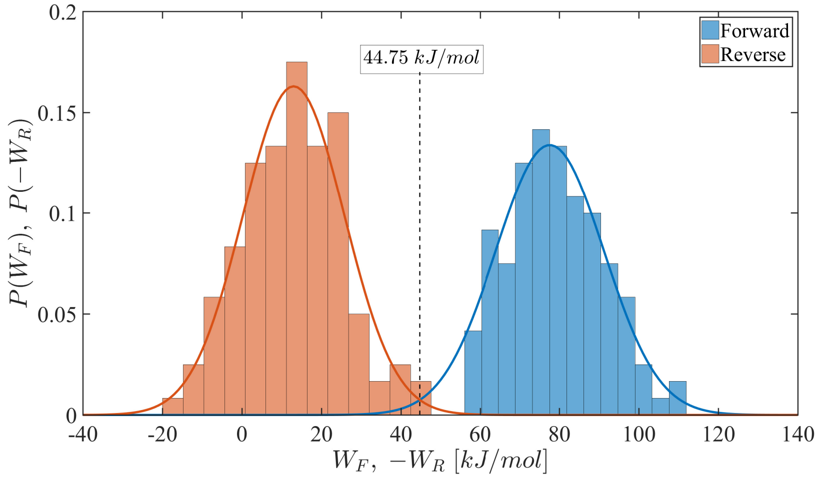


Figure 5.8. Work distributions for urea molecules being pulled forward from atomistic region into ideal gas region (blue bars) and reverse (red bars). The distributions are obtained from 120 trajectories under the pulling rate $v = 50\,\text{Å/ns}$. The solid curves represent the Gaussian curves fitted to each distributions. The Gaussian curves intersect at 44.75kJ mol^{-1}.

5.5.6 Comparing efficiency

We use the free energy profile obtained from umbrella sampling calculations as the reference data for the nonequilibrium experiments. The error bars reported in Figure 4(a) in the main manuscript for the nonequilibrium calculations, in particular the ones obtained with the logarithmic estimator, are systematically smaller than the ones of the reference profile. In order to assess the efficiency of the nonequilibrium calculations, we should compare results with similar precision. In Figure 5.9 we decrease the size of the statistical sample until we observe comparable error bars between the nonequilibrium and the reference profiles. Using the information provided here, we conclude that, at a pulling rate of $v = 10\,\text{Å/ns}$, the nonequilibrium calculation is only three times less efficient than the umbrella sampling calculation.

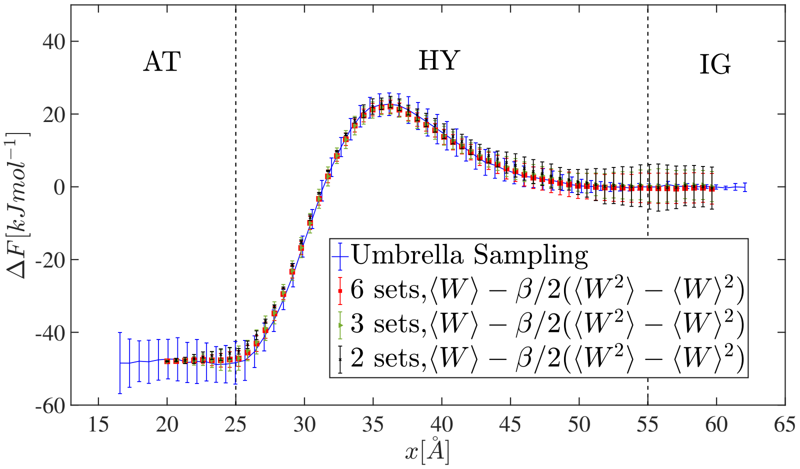


Figure 5.9. Free energy profiles of a single urea molecule being transferred from atomistic region into ideal gas region. The profiles are obtained using umbrella sampling method (blue) and Jarzynski equality when the pulling rate is $v = 10\,\text{Å/ns}$ and the number of sets in block-averaging is 6(red), 3(green) and 2(black). The maximum error bars obtained using umbrella sampling method is comparable to that obtained using Jarzynski equality when the second cumulant is averaged over 2 sets.

5.5.7 Work fluctuations

In theory the JE and CFT results are independent of the pulling rate. In practice, however, the obtained value of the free energy difference depends on the distribution of work values which strongly depends on the pulling rate. In particular for the use of the JE, a few authors have suggested [224, 238, 239] that the optimal pulling rate in terms of the efficiency of the averaging process obeys $\sigma_W \approx k_{\mathrm{B}}T$ with σ_W the standard deviation of the distribution of work. Consistently with the results presented in the main manuscript, Figure 5.10 shows the standard deviation of the work values during the pulling process indicating that the optimal pulling rate for the use of the JE is $v = 10\,\text{Å/ns}$.

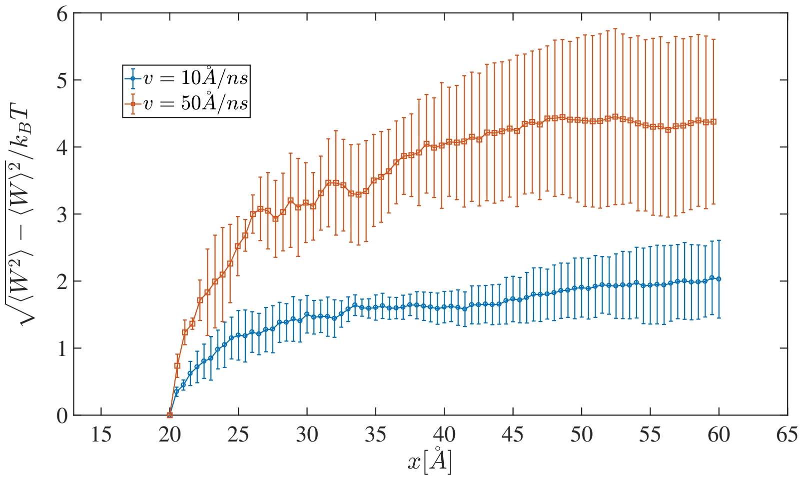


Figure 5.10. Standard deviation of the work during pulling process. The work fluctuations is often used as a measure of applicability of Jarzynski's equality [218]. For the case of slow pulling rate, the ultimate standard deviation is $\approx 2\mathrm{k}_BT$ while this quantity rises to $5\mathrm{k}_BT$ for high pulling rate.

5.5.8 Correction to the chemical potential from the DSF treatment of electrostatic interaction

For a system having N charged particle, the short-ranged electrostatic DSF potential is given by:

$$\begin{aligned} V^{\mathrm{DSF}} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\prime=1}^{N_i(r_c)} q_i q_j \left[\frac{\mathrm{erfc}(\alpha r_{ij})}{r_{ij}} - \frac{\mathrm{erfc}(\alpha r_c)}{r_c}\right. \\ &+ \left.\left(\frac{\mathrm{erfc}(\alpha r_c)}{r_c^2} + \frac{2\alpha}{\pi^{1/2}}\frac{\exp\left(-\alpha^2 r_c^2\right)}{r_c}\right)(r_{ij} - r_c)\right] \\ &+ V_{\mathrm{self}} - V_{\mathrm{excl}}\,, \end{aligned} \tag{5.21}$$

where $N_i(r_c)$, is the number of particles which are located within the sphere of radius r_c and the center of particle i. The decay of the potential is set by the parameter α whose dimension is an inverse length, and $\mathrm{erfc}(r)$ is the complementary error function [82]. Since the net charge of all $N_i(r_c)$ particles inside the truncating sphere is not zero, a correction to the potential is needed to compensate the lack of charge neutrality. This is done by approximating the net charge of the truncating sphere and placing the approximated "imaginary" charge on the surface of the sphere [186, 187]. For the

DSF potential, the electrostatic interaction of particle i with the imaginary charge is expressed as [187]:

$$V^{\text{self}} = -\left(\frac{\text{erfc}(\alpha r_c)}{r_c} + \frac{\alpha}{\sqrt{\pi}} + \frac{\alpha}{\sqrt{\pi}}\text{exp}(-\alpha^2 r_c^2)\right)\sum_{i=1}^{N} q_i^2 \tag{5.22}$$

For the case of rigid structural molecules, the intra-molecular electrostatic interaction has to be excluded. This is done by subtracting the corresponding Coulomb interaction from the DSF potential. Thus, for particles i and j separated by a distance r_{ij} in a rigid molecule, the excluded electrostatic interaction is computed as:

$$V_{ij}^{\text{excl}}(r_{ij}) = \frac{q_i q_j}{r_{ij}}. \tag{5.23}$$

The value of the correction to the chemical potential of SPC/E water and urea molecules using OPLS force field is $-1.07\text{k}_B\text{T}$ and $-4.17\text{k}_B\text{T}$, respectively.

5.6 Acknowledgments

The authors thank Nancy C. Forero-Martinez and Jan Smrek for a critical reading and insightful comments. M.H., K.K. and R.P. gratefully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG). R.C.-H. gratefully acknowledges the Alexander von Humboldt Foundation for financial support. This work has been supported by the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement No. 340906-MOLPROCOMP.

6. Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: revisiting the spatial block analysis method

This chapter is a research article that has been published in journal of Entropy.

Maziar Heidari, Kurt Kremer, Raffaello Potestio, and Robinson Cortes-Huerto
Fluctuations, finite-size effects and the thermodynamic limit in computer simulations: revisiting the spatial block analysis method
Entropy 2018, 20(4), 222
DOI: 10.3390/e20040222
©2018 by the authors. Licensee MDPI, Basel, Switzerland

Abstract

The spatial block analysis (SBA) method has been introduced to efficiently extrapolate thermodynamic quantities from finite-size computer simulations of a large variety of physical systems. In the particular case of simple liquids and liquid mixtures, by subdividing the simulation box into *blocks* of increasing size and calculating volume dependent fluctuations of the number of particles, it is possible to extrapolate the bulk isothermal compressibility and Kirkwood-Buff integrals in the thermodynamic limit. Only by explicitly including finite-size effects, ubiquitous in computer simulations, into the SBA method, the extrapolation to the thermodynamic limit can be achieved. In this review, we discuss two of these finite-size effects in the context of the SBA method due to i) the statistical ensemble and ii) the finite integration domains used in computer simulations. To illustrate the method, we consider prototypical liquids and liquid mixtures described by truncated and shifted Lennard-Jones (TSLJ) potentials. Furthermore, we show some of the most recent developments of the SBA method, in particular, its use to calculate chemical potentials of liquids in a wide range of density/concentration conditions.

6.1 Introduction

In the last decades, computational studies of soft matter have gained ground in the no-man's land between purely theoretical studies and experimental investigations. Arguably, this success is due to the use of statistical mechanics relations between macroscopic thermodynamic properties and microscopic components and interactions of a physical system in the thermodynamic limit (TL) [240, 241]. However, and apart from few examples [242–244], computer simulations are mainly constrained to consider *closed* systems with a finite and usually small number of particles N_0. These limitations introduce spurious finite-size effects, apparent in the simulation results, that in spite of the current computing capabilities are still the subject of intense investigations [245–252].

A meaningful comparison between computer simulations of finite systems and experimental results has been always a difficult task. In principle, it is possible to extrapolate the simulation data to the quantities of interest in the thermodynamic limit by considering systems of increasing size and performing simulations for each of them. The SBA method has been proposed as a more efficient alternative where only one system is examined and then subdivided into *blocks* of different size from which the data are extracted. The method is rather general since it was originally proposed to study the critical behaviour of Ising systems [94, 253] and then extended to study liquids [95, 254–258] and even the elastic constants of model solids [259].

In this paper, we examine the SBA method focusing on the extrapolation of bulk thermodynamic properties of simple liquids. We use prototypical liquids and mixtures described by TSLJ potentials to discuss the original ideas [95, 254] and explore the background [257, 258, 260–263] for the most recent developments [245, 246, 248] of the method. The simple examples presented here, in addition to the results available in the literature [245], suggest that the method is suitable for the calculation of trends in the chemical potential of complex liquids in a wide range of density/concentration conditions.

The paper is organised as follows: In Section 6.2 we introduce the relevant finite-size effects present in standard computer simulations. In Section 6.3 we introduce the finite-size integral equations for liquids and illustrate the procedure to extrapolate thermodynamic quantities. In Section 6.4 we discuss the extension of the block analysis method to liquid mixtures. We conclude the paper in Section 6.5.

6.2 Boundary and ensemble finite-size effects

Statistical mechanics establishes the connection between macroscopic thermodynamic properties and the microscopic components and interactions of a physical system. An interesting example of this relation is provided by the compressibility equation that identifies the density fluctuations of a system in the grand canonical ensemble with the bulk isothermal compressibility κ_T [265]. In the thermodynamic limit (TL), the isothermal compressibility of a homogeneous system is related to the fluctuations of the number of particles via the expression [240]:

$$\chi_T^{\infty} = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle}, \tag{6.1}$$

with $\langle N \rangle$ the average number of particles contained in a volume V of the fluid. The reduced isothermal compressibility $\chi_T^{\infty} = \rho k_{\mathrm{B}} T \kappa_T$ is the ratio between the bulk

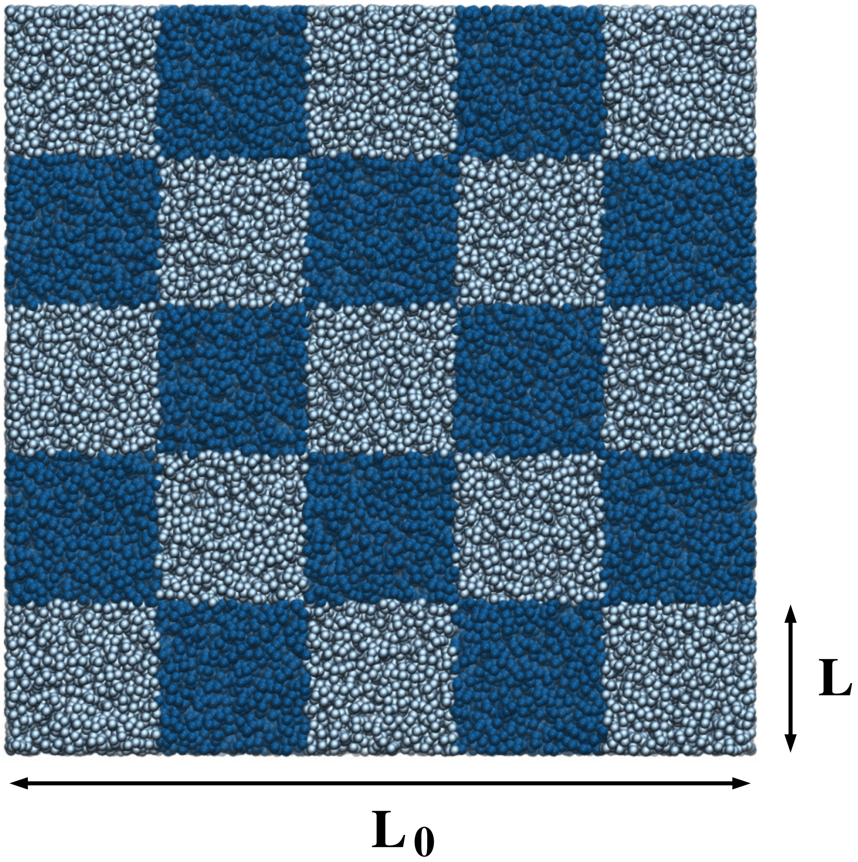


Figure 6.1. Snapshot of the simulation box for a system of particles interacting *via* a TSLJ potential at density $\rho\sigma^3 = 0.1$ and temperature $k_{\mathrm{B}}T = 1.2\epsilon$. In this particular example, a box of linear size L_0 has been subdivided in blocks of linear dimension $L = L_0/5$ as indicated by the different colour shades. The figure has been rendered with VMD [264].

isothermal compressibility of the system, κ_T, and the isothermal compressibility of the ideal gas $(\rho k_{\mathrm{B}}T)^{-1}$ with $\rho = \langle N\rangle/V$.

Various finite-size effects can be included in the block analysis aiming at extrapolating interesting thermodynamic quantities. In practice, let us consider a system of N_0 particles where the simulation box of volume $V_0 = L_0^3$ is divided in subdomains of volume $V = L^3$, as illustrated in Fig. 6.1. By evaluating the fluctuations of the number of particles in these subdomains, it is possible to obtain the distribution $P_{L,L_0}(N)$ of the number of particles, with k-moments given by [262]:

$$\langle N^k\rangle_{L,L_0} = \sum_{N=0}^{N_0} N^k \, P_{L,L_0}(N) \, . \tag{6.2}$$

The second moment of the distribution is related to the reduced isothermal compressibility of the finite system $\chi_T(L, L_0)$ [94, 95, 254, 262]:

$$\chi_T(L, L_0) = \frac{\langle N^2 \rangle_{L,L_0} - \langle N \rangle^2_{L,L_0}}{\langle N \rangle_{L,L_0}} \,. \tag{6.3}$$

The finite-size reduced isothermal compressibility, $\chi_T(L, L_0)$, can be extrapolated to the reduced isothermal compressibility in the TL, χ_T^∞, taking the limits $L, L_0 \to \infty$. Originally [254, 262], by applying periodic boundary conditions (PBCs) to the total linear size L_0 and taking into account volumes such that $L \gg \zeta$ with ζ the correlation length of the system, it has been proposed that the difference between $\chi_T(L, L_0)$ and χ_T^∞ is related to boundary effects associated to the finite-size of the subdomains. This difference takes the form [95, 254]:

$$\chi_T(L, L_0 \to \infty) = \chi_T^\infty + \frac{c}{L} + O\left(\frac{1}{L^2}\right) \,, \tag{6.4}$$

with c a constant. Recently, Eq. (6.4) has been obtained [266] using arguments based on the thermodynamics of small systems [267, 268], underpinning the consistency of the result.

To investigate this expression, we consider a liquid system whose potential energy is described by a 12–6 Lennard-Jones potential truncated, with cutoff radius $r_c/\sigma = 2^{1/6}$, and shifted. The parameters ϵ, σ and m, define the units of energy, length and mass, respectively. All the results are expressed in LJ units with time $\sigma(m/\epsilon)^{1/2}$, temperature ϵ/k_B and pressure ϵ/σ^3. Various system sizes, namely $N_0 = 10^4, 10^5$ and 10^6, are considered, and the density is fixed at $\rho\sigma^3 = 0.864$ thus defining the linear size of the simulation box L_0. The systems are equilibrated at $k_\mathrm{B}T = 1.2\epsilon$, enforced with a Langevin thermostat with damping coefficient $\gamma(\sigma(m/\epsilon)^{1/2}) = 1.0$, for 2×10^6 MD steps using a time step of $\delta t/(\sigma(m/\epsilon)^{1/2}) = 10^{-3}$. Production runs span 10^6 MD steps. All the simulations have been performed with the ESPResSo++ [269] simulation package.

To use the block analysis method we compute the fluctuations of the number of particles. In particular, we choose domains of size $1 < L/\sigma < L_0/\sigma$ to scan *continuously* the fluctuations as a function of domain size. To increase the amount of statistics we use 100 randomly positioned subdomains per simulation frame.

In Fig. 6.2 we report $\chi_T(L, L_0)$ as a function of σ/L. The linear behaviour predicted in Eq. 6.4 is apparent for $L \ll L_0$. There are evident deviations from the linear behaviour which are not included in Eq. (6.4) since this equation has been obtained for a system in the grand canonical ensemble. As a matter of fact, the deviations from linearity are mainly related to the fixed size of the system because when $L \to L_0$, $\chi_T(L_0, L_0) = 0$, that is, the fluctuations of the number of particles for a *closed* system are equal to zero. In principle, the isothermal compressibility in the TL can be extracted by extrapolating a line to the y-axis, i. e. $\sigma/L \to 0$, and determining the y-intercept. This procedure, however, might lead to ambiguous and strongly-size-dependent results as suggested by the same plot.

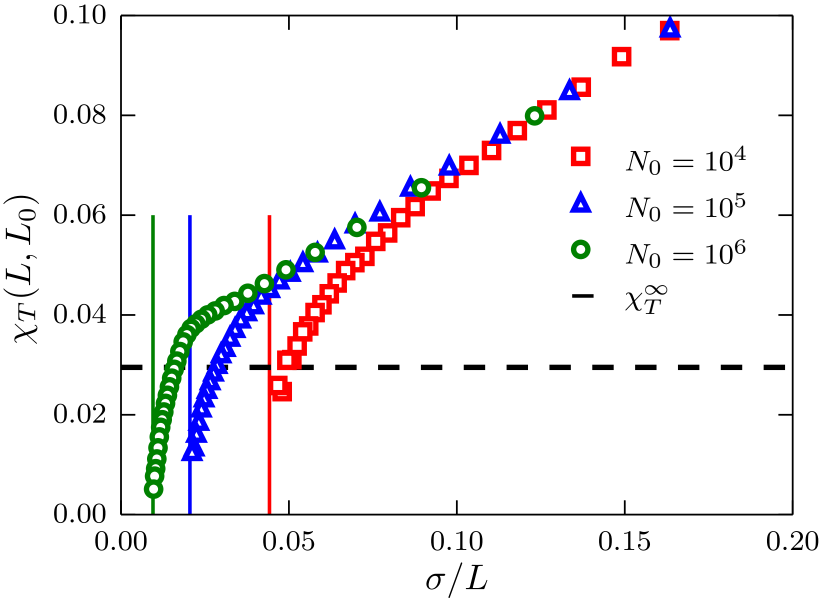


Figure 6.2. Fluctuations of the number of particles $\chi_T(L, L_0)$ as a function of σ/L for systems described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$. Data corresponding to system sizes $N_0 = 10^4, 10^5$ and 10^6 are presented using red squares, blue triangles and green circles, respectively. The vertical lines indicate the limit σ/L_0 at which fluctuations become zero. The black horizontal dashed line indicates the value $\chi_T^\infty = \rho k_{\mathrm{B}} T \kappa_T = 0.0295$ with κ_T the bulk compressibility obtained with the method described in Ref. [245].

From the previous discussion, Eq. 6.4 satisfactorily describes the *boundary* size effects present in a system described in the grand canonical ensemble. However, *ensemble* size effects, i. e. the fact that we are computing quantities defined in the grand canonical ensemble using information obtained from a system in a canonical ensemble, are important even in cases where the size of the system might appear to be enormous ($L_0/\sigma = 105$ for $N_0 = 10^6$ where $\zeta/\sigma \approx 10$).

It is thus clear that the isothermal compressibility of a finite-size system in the TL, i. e. $L, L_0 \to \infty$ with $\rho = N_0/L_0^3$, should equate the bulk isothermal compressibility κ_T. An elegant analysis using probabilistic arguments for the ideal gas case [270] shows that the finite-size reduced isothermal compressibility can be written as:

$$\chi_T(L, L_0) = \chi_T^\infty \left(1 - \left(\frac{L}{L_0}\right)^3\right) . \tag{6.5}$$

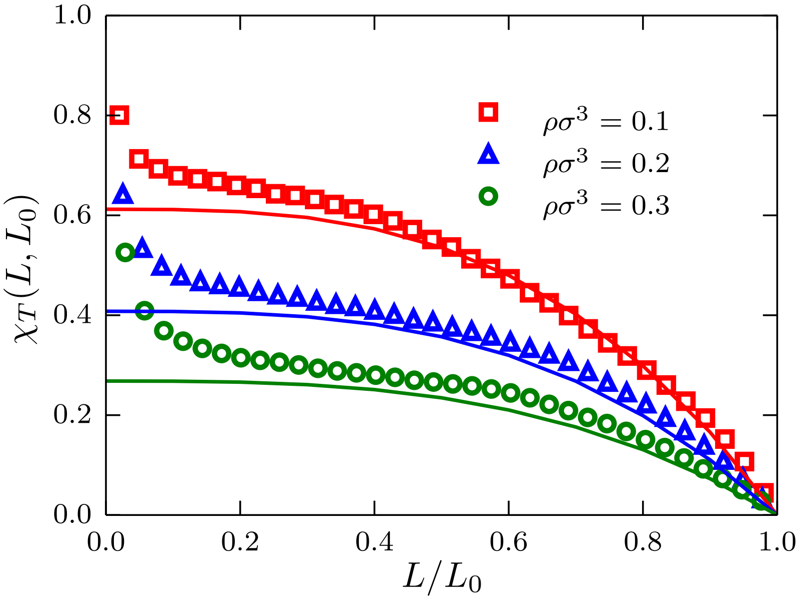


Figure 6.3. Fluctuations of the number of particles $\chi_T(L, L_0)$ as a function of the ratio L/L_0 for systems described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$. Results corresponding to systems of $N_0 = 10^5$ particles with densities $\rho\sigma^3 = 0.1$, 0.2 and 0.3 are presented using red squares, blue triangles and green circles, respectively. The theoretical prediction presented in the text is plotted using the corresponding value for χ_T^∞, obtained as described in Ref. [245], and solid-line curves with the same color code.

In spite of the simplicity of the system chosen in this study, it can not be identified with the ideal gas. However, at very low densities and temperature $k_\mathrm{B}T = 1.2\epsilon$, the system behaves more like a real gas and a meaningful trend could be identified. Therefore, to investigate Eq. 6.5 we consider the density range $\rho\sigma^3 = 0.1\,, \cdots, 1.0$ for systems of size $N_0 = 10^5$ particles. Results are presented in Fig. 6.3 for the cases $\rho\sigma^3 = 0.1$, 0.2 and 0.3. The three data sets follow the theoretical prediction in Eq. (6.5) with deviations from this behaviour for $L \ll L_0$ thus indicating the signature of boundary finite-size effects. As expected, the data presented also suggest that upon increasing density the deviations from the ideal gas behaviour become more evident, as can be seen in the case $\rho\sigma^3 = 0.3$.

This is also seen in Fig. 6.4 where for a system with density $\rho\sigma^3 = 0.864$ the deviations from the ideal gas case are much more evident. As a matter of fact, even for the largest size considered ($N_0 = 10^6$) it is not possible to convincingly reproduce the ideal gas behaviour.

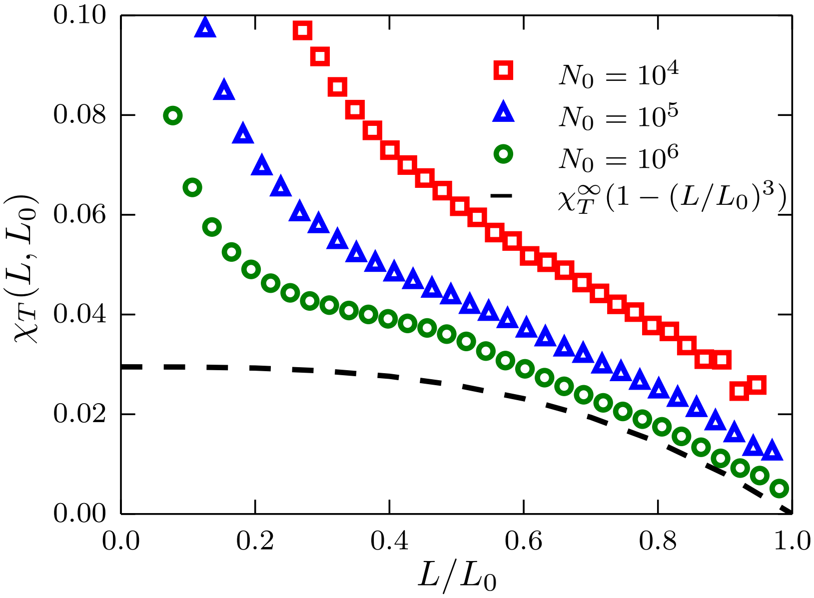


Figure 6.4. Fluctuations of the number of particles $\chi_T(L, L_0)$ as a function of the ratio L/L_0 for systems described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$. Results corresponding to sizes $N_0 = 10^4$, 10^5 and 10^6, with density $\rho\sigma^3 = 0.864$, using red squares, blue triangles and green circles, respectively. The theoretical prediction presented in the text is plotted as the black dashed curve using $\chi_T^\infty = 0.0295$.

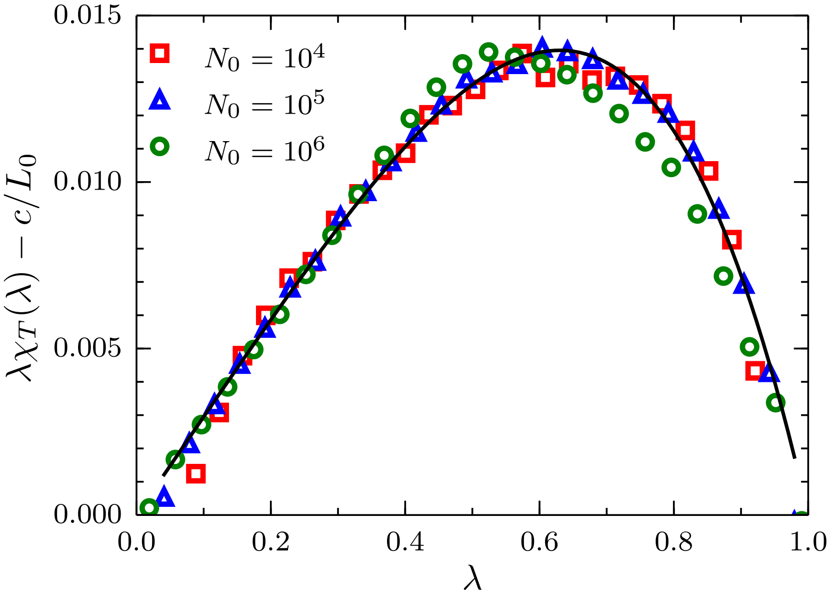


Figure 6.5. Scaled fluctuations of the number of particles $\lambda\chi_T(L, L_0)$, minus c/L_0, versus the ratio $\lambda = L/L_0$ for systems described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$. Results corresponding to sizes $N_0 = 10^4$, 10^5 and 10^6, with density $\rho\sigma^3 = 0.864$, using red squares, blue triangles and green circles, respectively. The theoretical prediction Eq. 6.7 presented in the text is plotted as the black solid curve using $\chi_T^\infty = 0.0295$ and $c = 0.415\sigma$.

Nonetheless, one intuitively could imagine that the following expression:

$$\chi_T(L, L_0) = \chi_T^\infty \left(1 - \left(\frac{L}{L_0}\right)^3\right) + \frac{c}{L} + O\left(\frac{1}{L^2}\right) , \tag{6.6}$$

captures the two finite-size effects, *ensemble* and *boundary* [262]. By neglecting the $O(1/L^2)$ terms, defining $\lambda = L/L_0$ and multiplying everything times λ we obtain:

$$\lambda\chi_T(\lambda) = \lambda\chi_T^\infty \left(1 - \lambda^3\right) + \frac{c}{L_0} . \tag{6.7}$$

Equation 6.7 is more convenient to analyse because in the limit $\lambda \to 0$, provided that $\zeta < L < L_0$, λ^3 is negligible and this expression can be approximated to a linear function in λ with slope χ_T^∞ and y-intercept equal to c/L_0. In particular, we use a simple linear regression in the interval $0.0 < \lambda < 0.3$, with the fluctuations data for $N_0 = 10^5$, to find $\chi_T^\infty = 0.0295(5)$ and $c = 0.415(5)\sigma$. Results of the scaled fluctuations $\lambda\chi_T(\lambda)$ minus c/L_0 are presented in Fig. 6.5 where the intensive character of the constant c becomes clear. By replacing the calculated values χ_T^∞ and c in Eq. (6.7) we obtain the black curve that superimposes on the simulation data in the full range $0 < \lambda < 1$.

In addition to the *explicit* finite-size effects discussed above, there is another type of effects related to the periodicity of the simulation box. This is the case of *implicit* finite-size effects that appear due to anisotropies in the pair correlation function of the system, generated by the use of PBCs [271, 272]. These effects, extremely important for small simulation setups, appear as oscillations in $\lambda\chi_T(\lambda)$ for $\lambda \approx 1$ caused by short range interactions between the system and its nearest neighbour images. However, given the large sizes of the systems considered here, implicit finite-size effects can be safely ignored in the present discussion.

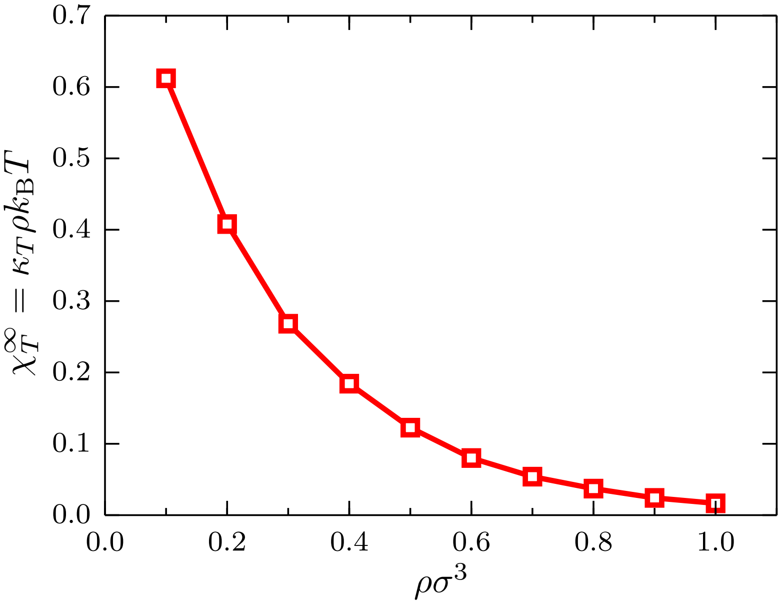


Figure 6.6. Ratio $\chi_T^\infty = \kappa_T/\kappa_T^{IG}$ at $k_{\mathrm{B}}T = 1.2\epsilon$ as a function of the density for systems described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$, with $\kappa_T^{IG} = (\rho k_{\mathrm{B}} T)^{-1}$ the isothermal compressibility of the ideal gas. The red curve is a guide to the eye.

With the trajectories of the system with $N_0 = 10^5$ particles in the density interval $0.1 < \rho\sigma^3 < 1.0$ we compute the scaled fluctuations $\lambda\chi_T(\lambda)$ and determine, as before, the ratio $\chi_T^\infty = \kappa_T/\kappa_T^{IG}$ as a function of the density, with $\kappa_T^{IG} = (\rho k_{\mathrm{B}} T)^{-1}$ the isothermal compressibility of the ideal gas (See Fig. 6.6). As expected for this system at $k_{\mathrm{B}}T = 1.2\epsilon$, a monotonically decreasing behaviour is observed since the system becomes less compressible as the density increases.

The isothermal compressibility as a function of the density allows one to investigate more interesting thermodynamic properties, as it has been recently demonstrated [245, 246]. For example, the isothermal compressibility can be written as:

$$\kappa_T = \frac{1}{\rho^2} \frac{\partial \rho}{\partial \mu}\bigg|_T , \tag{6.8}$$

which can be rearranged, in terms of the chemical potential μ, as:

$$\delta\mu = \int_{\rho_0}^{\rho} \frac{d\rho'}{\rho'^2 \kappa_T} \tag{6.9}$$

with $\delta\mu = \mu - \mu_0$ and μ_0 the chemical potential of the system at the reference density ρ_0. In practice, one usually is interested in the excess chemical potential [1]:

$$\delta\mu^{\mathrm{ex}} = \delta\mu - k_{\mathrm{B}} T \ln \rho, \tag{6.10}$$

obtained by subtracting from $\delta\mu$ the density dependent part of the chemical potential of the ideal gas.

To validate the results obtained using Eq. (6.10) it is necessary to use a different computational method to evaluate μ_0. For that purpose any computational method aiming at calculating chemical potentials could be used. In particular, we use the spatially resolved thermodynamic integration (SPARTIAN) method [93], recently implemented by us. In SPARTIAN, the target system, described with atomistic resolution, is embedded in a reservoir of ideal gas particles. An interface between the two subdomains is defined such that molecules are free to diffuse adapting their resolution on the fly. A uniform density across the simulation box is guaranteed by applying a single-molecule external potential that is identified with the difference in chemical potential between the two resolutions, i. e. the excess chemical potential of the target system. This method has been validated by calculating excess chemical potentials for Lennard-Jones liquids, mixtures, as well as for SPC, SPC/E water and aqueous sodium chloride solutions, all in good agreement with state-of-the-art computational methods.

For the comparison, we consider the same system at the same temperature with densities $\rho\sigma^3$ =0.2, 0.4, 0.6, 0.8 and 1.0. Results for the excess chemical potential as a function of the density are presented in Fig. 6.7 where the value of $\rho_0\sigma = 0.6$ has been used as the reference value. Once $\delta\mu^{ex}$ is rescaled, it becomes clear that the agreement between the two methods is remarkable. This result suggest that the simple calculation of the fluctuations of the number of particles, used in combination with Eq. (6.7), provides us with an efficient and accurate method to compute chemical potential of simple liquids, that can be extended to more complex fluids [245].

---

[1] In this context, the word excess should be replaced with residual. The residual chemical potential is the difference between the chemical potential of the target system and that of an ideal gas at the same density, temperature and composition. We misuse the expression *excess chemical potential* to match the modern literature.

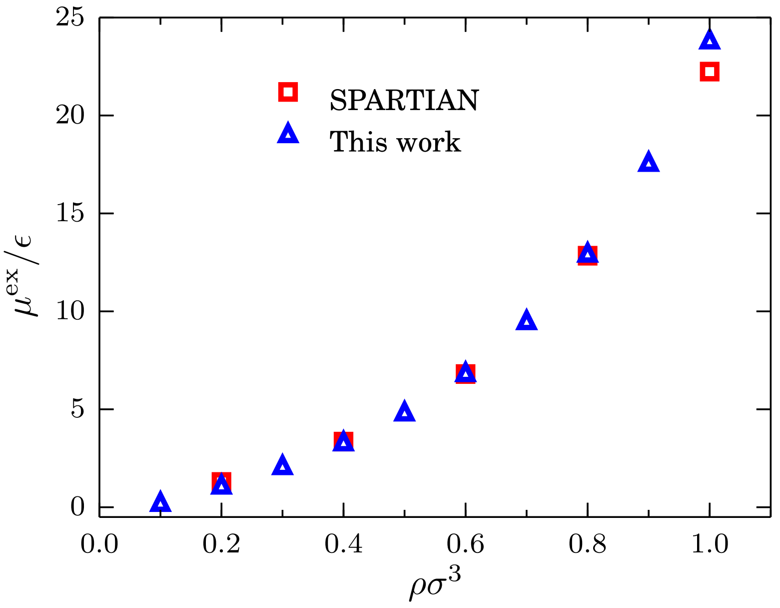


Figure 6.7. Excess chemical potential μ^{ex}/ϵ at $k_{\mathrm{B}}T = 1.2\epsilon$ as a function of the density for systems described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$. Red squares indicate the data obtained with the SPARTIAN method [93] and the blue triangles are the data points obtained with the method outlined in the text.

In this section, Eq. (6.7) has been introduced in a rather intuitive manner. However, the presented results suggest that it encompasses the relevant finite-size effects of the system and allows one to compute bulk thermodynamic quantities. In the following section, we derive Eq. (6.7) more rigorously and explore, using a different example, its range of validity.

6.3 Finite-size Ornstein-Zernike integral equation

Fluctuations of the number of particles are related to the local structure of a liquid. Let us consider a molecular liquid of average density ρ at temperature T in equilibrium with a reservoir of particles, i.e. an *open* system. The fluctuations of the number of molecules are related to the local structure of the liquid via the Ornstein-Zernike integral equation [240, 273]

$$\frac{\Delta^2(N)}{\langle N\rangle} = 1 + \frac{\rho}{V}\int_V\int_V [g^{\circ}(\mathbf{r}_1, \mathbf{r}_2) - 1]\, d\mathbf{r}_1\, d\mathbf{r}_2\,, \tag{6.11}$$

where $\Delta^2(N)/\langle N\rangle$ are the fluctuations of the number of particles, $\Delta^2(N) = \langle N^2\rangle - \langle N\rangle^2$ and $g^{\mathrm{o}}(\mathbf{r}_1, \mathbf{r}_2)$ is the pair correlation function of the open system and $\mathbf{r}_1, \mathbf{r}_2$ the position vectors of a pair of fluid particles. To solve the integral in Eq. (6.11) one assumes that the fluid is homogeneous, isotropic and that the system is in the thermodynamic limit (TL), i.e. $V \to \infty$, $\langle N\rangle \to \infty$ with $\rho = \langle N\rangle/V =$ constant. An infinite, homogeneous and isotropic system is translationally invariant, therefore we rewrite Eq. (6.11) as [240]:

$$\chi_T^{\infty} = \frac{\Delta^2(N)}{\langle N\rangle} = 1 + 4\pi\rho \int_0^{\infty} (g^{\mathrm{o}}(r) - 1)\, r^2\, dr\,, \tag{6.12}$$

with $\chi_T^{\infty} = \rho k_{\mathrm{B}} T \kappa_T$, κ_T being the isothermal compressibility of the bulk system. We have replaced $g^{\mathrm{o}}(\mathbf{r}_1, \mathbf{r}_2)$ with $g^{\mathrm{o}}(r)$ the radial distribution function (RDF) of the open system, with $r = |\mathbf{r}_2 - \mathbf{r}_1|$.

An alternative version of the OZ integral equation for finite systems has been introduced [262]. For a finite system with total volume V_0 with PBCs we have:

$$\chi_T(V, V_0) = \frac{\Delta^2(N; V, V_0)}{\langle N\rangle_{V,V_0}} = 1 + \frac{\rho}{V} \int_V \int_V [g^{\mathrm{c}}(r_{12}) - 1]\, d\mathbf{r}_1\, d\mathbf{r}_2\,, \tag{6.13}$$

where $g^{\mathrm{c}}(r_{12})$, $r_{12} = |\mathbf{r}_2 - \mathbf{r}_1|$, is the pair correlation function of the closed system with total number of particles N_0, and $\Delta^2(N; V, V_0) = \langle N^2\rangle_{V,V_0} - \langle N\rangle^2_{V,V_0}$. The fluctuations of the number of particles thus depend on both subdomain and simulation box volumes.

For a single component fluid of density ρ at temperature T with fixed number of particles N_0 and volume V_0, its RDF can be written in terms of an expansion around N_0 as [260–263, 270] :

$$g^{\mathrm{c}}(r) = g^{\mathrm{o}}(r) - \frac{\chi_T^{\infty}}{N_0}\,. \tag{6.14}$$

As a matter of fact, the expansion includes terms that depend on the partial derivative of $g^{\mathrm{o}}(r)$ with respect to the density. However, we anticipate here that for the present analysis their contribution is negligible [245]. By replacing $g^{\mathrm{c}}(r)$ in the integral on the r.h.s of Eq. (6.13) we obtain:

$$\frac{\rho}{V} \int_V \int_V (g^{\mathrm{c}}(r_{12}) - 1)\, d\mathbf{r}_1\, d\mathbf{r}_2 = I_{V,V} - \frac{V}{V_0}\chi_T^{\infty}\,, \tag{6.15}$$

where

$$I_{V,V} = \frac{\rho}{V} \int_V \int_V (g^{\mathrm{o}}(r_{12}) - 1)\, d\mathbf{r}_1\, d\mathbf{r}_2\,, \tag{6.16}$$

and we use that $\rho = N_0/V_0$.

Next, we include explicitly the second finite-size effect, i.e. the fact that the volume V is finite and embedded into a finite volume V_0 with PBCs. For this we rewrite $I_{V,V}$ as [254]

$$I_{V,V_0-V} = I_{V,V_0} - I_{V,V} \ ,$$

with

$$\begin{aligned} I_{V,V_0} &= \frac{\rho}{V}\int_V\int_{V_0}(g^{\mathrm{o}}(r_{12})-1)\,d\mathbf{r}_1\,d\mathbf{r}_2 \\ I_{V,V_0-V} &= \frac{\rho}{V}\int_V\int_{V_0-V}(g^{\mathrm{o}}(r_{12})-1)\,d\mathbf{r}_1\,d\mathbf{r}_2 \,. \end{aligned}$$

As pointed out by Rovere, Heermann and Binder [254], the two integrals $I_{V,V}$ and I_{V,V_0} are equal when $\mathbf{r}_1$ and $\mathbf{r}_2$ are both within the volume V. When $r_{12} > \zeta$ the integrand $(g^{\mathrm{o}}(r_{12})-1)=0$ and it does not contribute to the integrals. Close to the boundary of the subdomain V, for $r_{12} < \zeta$, and in particular when $\mathbf{r}_1$ lies inside and $\mathbf{r}_2$ outside the volume V, there are contributions missing in $I_{V,V}$ which are present in I_{V,V_0}. Therefore, the difference between the two integrals $I_{V,V_0-V} = I_{V,V_0} - I_{V,V}$, must be proportional to the surface volume ratio of the subdomain V [254], i. e.

$$I_{V,V_0-V} = \frac{c_1}{L} + \left(\frac{c_2}{L}\right)^2 + O\left(\frac{1}{L^3}\right) \ , \tag{6.17}$$

with c_1, c_2 proportionality constants with units of length that, at this point, we assume to be intensive .

To compute I_{V,V_0}, we require that $\zeta < L < L_0$. Since we assume PBCs, the system is translationally invariant. Hence, upon applying the transformation $\mathbf{r}_{12} \to \mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$, the expression:

$$I_{V,V_0} = \rho\int_{V_0}(g^{\mathrm{o}}(r)-1)\,d\mathbf{r} = \chi_T^{\infty} - 1 \tag{6.18}$$

is obtained, where we assume that $g^{\mathrm{o}}(r > \zeta) = 1$ thus ignoring fluctuations of the RDF beyond the volume V. By combining these two results we obtain

$$I_{V,V} = \chi_T^{\infty} - 1 + \frac{c_1}{L} + \left(\frac{c_2}{L}\right)^2 \ , \tag{6.19}$$

and by including this result in Eq. (6.15) we arrive to the following expression:

$$\frac{\rho}{V}\int_V\int_V(g^{\mathrm{c}}(r_{12})-1)\,d\mathbf{r}_1\,d\mathbf{r}_2 = \chi_T^{\infty}\left(1-\left(\frac{L}{L_0}\right)^3\right) - 1 + \frac{c_1}{L} + \left(\frac{c_2}{L}\right)^2 \,. \tag{6.20}$$

Finally this expression becomes:

$$\chi_T(L, L_0) = \chi_T^\infty\left(1 - \left(\frac{L}{L_0}\right)^3\right) + \frac{c_1}{L} + \left(\frac{c_2}{L}\right)^2 , \tag{6.21}$$

and by defining $\lambda = L/L_0$ we write:

$$\lambda\chi_T(\lambda) = \lambda\chi_T^\infty\left(1 - \lambda^3\right) + \frac{c_1}{L_0} + \left(\frac{c_2}{L_0}\right)^2 \frac{1}{\lambda} . \tag{6.22}$$

Eqs (6.22) and (6.7) differ in the $c_2^2/L_0^2\lambda$ term that appear from considering the *boundary* finite size effects. One possible scenario in which this difference might play a role is in the case of simulations near critical conditions where the correlation length of the system tends to infinity.

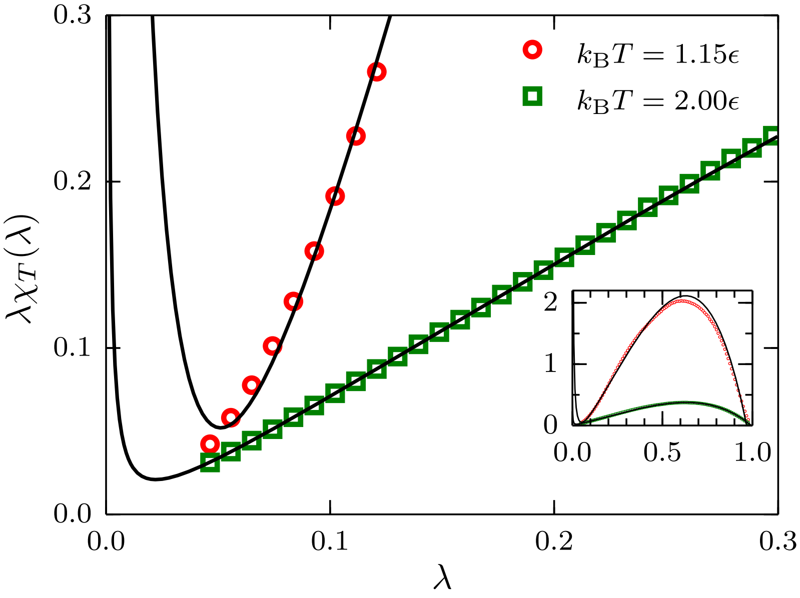


Figure 6.8. Reduced fluctuations as a function of λ for systems described by a TSLJ potential with $r_c/\sigma = 2.5$ with density $\rho\sigma^3 = 0.3$ at temperatures $k_\mathrm{B}T = 2.00\,\epsilon$ and $1.15\,\epsilon$. For the latter case, it is apparent that the contribution proportional to λ^{-1} is not negligible. The inset shows the full range $0 < \lambda < 1$. The black curves are the result of fitting the data to Eq. (6.22).

To test this expression we perform simulations of systems with potential energy described by the truncated, at $r_c/\sigma = 2.5$, and shifted 12–6 Lennard-Jones potential. We consider systems with $N_0 = 24000$ particles, with densities spanning the range

$0.05 < \rho\sigma^3 < 0.70$. Two temperatures were considered, $k_{\mathrm{B}}T = 2.00\,\epsilon$ and $1.15\,\epsilon$. The critical point of this system has been reported at $\rho_c\sigma^3 = 0.319$ and $k_{\mathrm{B}}T_c = 1.086\epsilon$ [274].

We report the reduced fluctuations $\lambda\chi_T(\lambda)$ as a function of λ for $\rho\sigma^3 = 0.3$ in Fig. 6.8. In the case $k_{\mathrm{B}}T = 2.00\,\epsilon$ the effect of the λ^{-1} term in Eq. (6.22) is negligible and a linear approximation in the region $\lambda < 0.3$ seems to be well justified. However, for the case close to the critical point, i. e. $k_{\mathrm{B}}T = 1.15\,\epsilon$, the effect of this term is evident and should be included in the extrapolation to χ_T^∞.

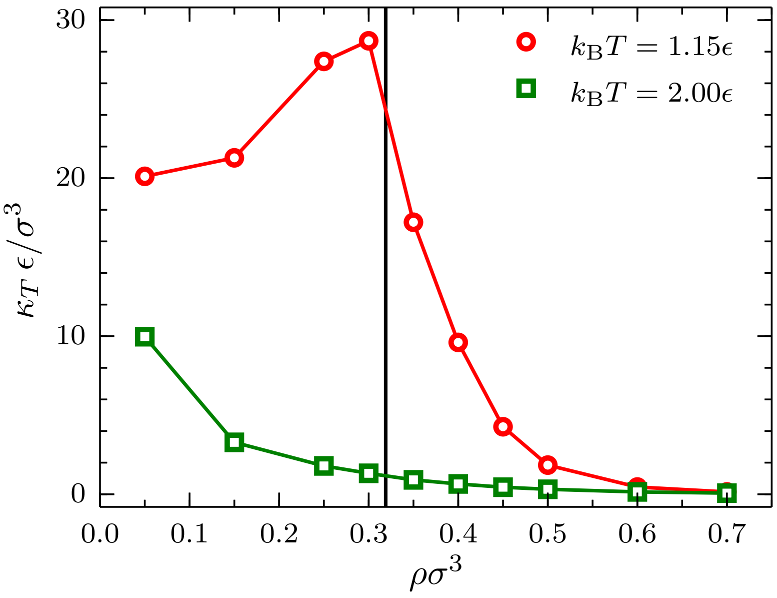


Figure 6.9. Bulk isothermal compressibility κ_T as a function of the density ρ at $k_{\mathrm{B}}T = 1.15\epsilon$ (red circles) and $k_{\mathrm{B}}T = 2.00\epsilon$ (green squares) for systems described by a TSLJ potential with $r_c/\sigma = 2.5$. The vertical black line indicates the location of the critical density $\rho\sigma^3 = 0.319$ [274].

Finally, upon extrapolating to χ_T^∞, an interesting behaviour is observed for the bulk isothermal compressibility κ_T as a function of density (Fig. 6.9). In the case $k_{\mathrm{B}}T = 2.00\,\epsilon$, as expected, a monotonically decreasing behaviour with increasing density is observed. More interestingly, in the case $k_{\mathrm{B}}T = 1.15\,\epsilon$ the monotonically decreasing behaviour is interrupted by a singularity in the isothermal compressibility in the vicinity of the critical density. This cusp in the curve is expected since the isothermal compressibility of a fluid at the critical point is infinite.

The use of finite-size integral equations is general enough to admit generalisations of other systems of interest. In the next section, we describe one of such possible extensions: the study of binary mixtures.

6.4 Mixtures

Kirkwood-Buff (KB) theory [275] is arguably the most successful framework to investigate the properties of liquid mixtures that relates the local structure of a system to density fluctuations in the grand canonical ensemble. These quantities are in turn related to equilibrium thermodynamic quantities such as the compressibility, the partial molar volumes, and the derivatives of the chemical potentials [241]. Formulated more than sixty years ago, KB enjoys renewed interest in computational soft-matter and statistical physics communities [245,246,248–252]. Recent works have shown promising applications related to solvation of biomolecules [276] and potential uses to compute multicomponent diffusion in liquids [277] and to study complex phenomena such as self-assembly of proteins [278] and polymer conformation in complex mixtures [243,279].

For a multicomponent fluid of species i, j in equilibrium at temperature T, the Kirkwood-Buff integral (KBI) is defined as:

$$G^{\mathrm{o}}_{ij} = V\left(\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_i\rangle}\right) = \frac{1}{V}\int_V\int_V [g^{\mathrm{o}}_{ij}(\mathbf{r}_{12}) - 1]\, d\mathbf{r}_1\, d\mathbf{r}_2\,, \tag{6.23}$$

with δ_{ij} the Kronecker delta. The superscript (o) indicates that this definition holds for an open system, i.e. a system in the grand canonical ensemble. In practice, we compute fluctuations of the number of particles in a subdomain of volume V embedded in a reservoir whose size goes to infinity. Thus, $\langle N_i\rangle$ is the average number of i-particles inside V, or $\rho_i = \langle N_i\rangle/V$. $g^{\mathrm{o}}_{ij}(\mathbf{r}_{12})$ is the multicomponent radial distribution function (RDF) of the infinite system, with $\mathbf{r}_{12} = \mathbf{r}_2 - \mathbf{r}_1$.

Let us recall that in computer simulations one considers systems with total fixed number of particles N_0 and volume V_0 with PBCs. In this case we have [272]:

$$G_{ij}(L, L_0) = V\left(\frac{\langle N_i N_j\rangle' - \langle N_i\rangle'\langle N_j\rangle'}{\langle N_i\rangle'\langle N_j\rangle'} - \frac{\delta_{ij}}{\langle N_i\rangle'}\right) = \frac{1}{V}\int_V\int_V [g^{\mathrm{c}}_{ij}(\mathbf{r}_{12}) - 1]\, d\mathbf{r}_1\, d\mathbf{r}_2\,. \tag{6.24}$$

The finite-size KBI $G_{ij}(L, L_0)$ is evaluated by computing fluctuations of the number of particles in finite subdomains of volume V inside a simulation box of volume V_0. The average number of i-particles $\langle N_i\rangle' \equiv \langle N_i\rangle_{V,V_0}$ depends on both subdomain and simulation box volumes. Moreover, the integral on the r.h.s. of Eq. (6.24) should be evaluated for the RDF of the finite system $g^{\mathrm{c}}_{ij}(\mathbf{r}_{12})$ with volume V_0 by using a finite integration domain V.

As has been done for the single component case, we include in this example both, *ensemble* and *boundary*, finite-size effects. For the former the following correction has been suggested [280]:

$$g^{\mathrm{c}}_{ij}(r) = g^{\mathrm{o}}_{ij}(r) - \frac{1}{V_0}\left(\frac{\delta_{ij}}{\rho_i} + G^{\infty}_{ij}\right)\,, \tag{6.25}$$

based on the asymptotic limit $g_{ij}^{\mathrm{c}}(r \gg \zeta) = 1 - (\delta_{ij}/\rho_i + G_{ij}^{\infty})/V_0$ discussed in Ref. [241]. As expected, when the total volume $V_0 \to \infty$ we recover $g_{ij}^{\mathrm{c}}(r) = g_{ij}^{\mathrm{o}}(r)$. By including Eq. (6.25) in the integral on the r.h.s. of Eq. (6.24) and evaluating the finite-size integral as for the single component case, we finally obtain:

$$\lambda G_{ij}(\lambda) = \lambda G_{ij}^{\infty}\left(1 - \lambda^3\right) - \lambda^4 \frac{\delta_{ij}}{\rho_i} + \frac{\alpha_{ij}}{L_0}\,, \tag{6.26}$$

with $\lambda \equiv L/L_0$ and α_{ij} an intensive parameter with units of length. In the limit $L_0 \to \infty$ the following expression is obtained:

$$G_{ij}(L, L_0 \to \infty) = G_{ij}^{\infty} + \frac{\alpha_{ij}}{L}\,, \tag{6.27}$$

that describes the finite-size effects on the KBIs for a system in the grand canonical ensemble. Consistent with this limiting case in Eq. (6.26), Eq. (6.27) has been obtained from the thermodynamics of small systems [281, 282].

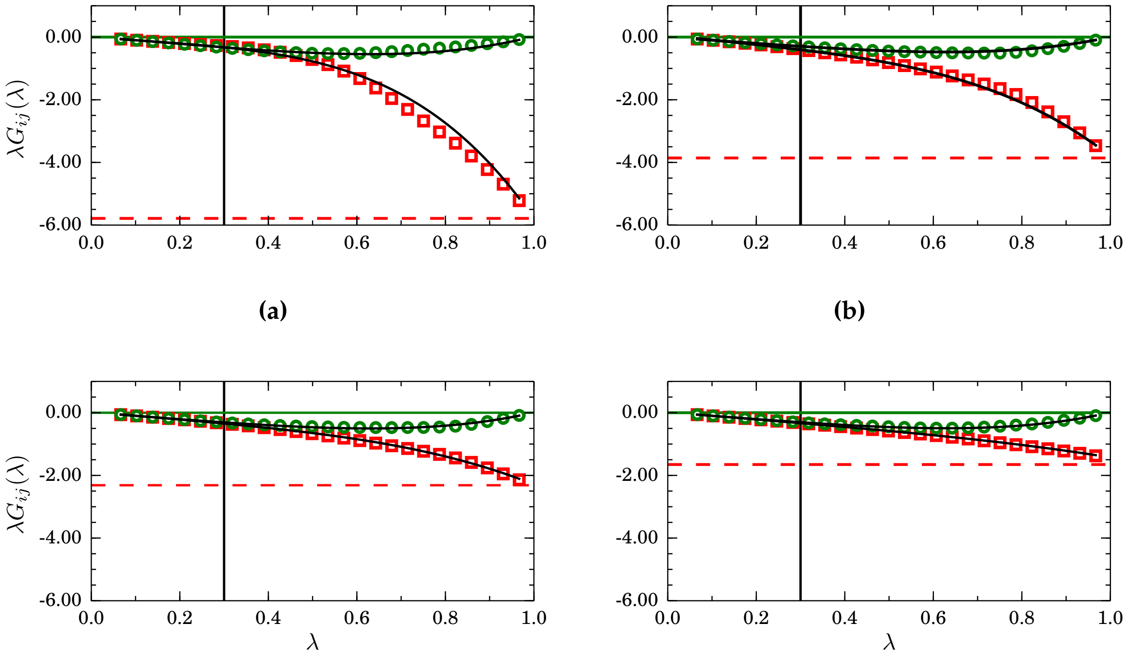
For the investigation of Eq. (6.26), we perform simulations for binary mixtures (A, B) of Lennard-Jones (LJ) fluids. We use a purely repulsive 12-6 LJ potential truncated and shifted with cutoff radius $2^{1/6}\sigma$. The potential parameters are chosen as $\sigma_{AA} = \sigma_{BB} = \sigma_{AB} = \sigma$, and $\epsilon_{AA} = 1.2\epsilon$, $\epsilon_{BB} = 1.0\epsilon$ with $\epsilon_{AB} = (\epsilon_{AA} + \epsilon_{BB})/2 = 1.1\epsilon$. All the results are expressed in LJ units with energy ϵ, length σ, mass $m_A = m_B = m$, time $\sigma(m/\epsilon)^{1/2}$, temperature ϵ/k_{B} and pressure ϵ/σ^3. As before, simulations are carried out using ESPResSo++ [269] with a time step of $\delta t/(\sigma(m/\epsilon)^{1/2}) = 10^{-3}$. Constant temperature $k_{\mathrm{B}}T = 1.2\epsilon$ is enforced through a Langevin thermostat with damping coefficient $\gamma(\sigma(m/\epsilon)^{1/2}) = 1.0$. The size of the system is $N_0 = 23328$ in the range of mole fractions of A-molecules $x_A = 0.1, \cdots, 1.0$. The pressure is fixed at $P\sigma^3/\epsilon = 9.8$ by adjusting the number density of the system at values around $\rho\sigma^3 \approx 0.86$ (or $L_0/\sigma \approx 30$). We perform equilibration runs of $64 \cdot 10^6$ MD steps and production runs of $2 \cdot 10^6$ MD steps. To compute $G_{ij}(\lambda)$, we select 800 frames per trajectory and for each frame identify 1000 randomly positioned subdomains with linear sizes ranging from $2 < L/\sigma < L_0/\sigma$.

In Fig. 6.10, results for finite-size KBIs are presented for four mole fractions, namely (a) $x_A = 0.20$, (b) $x_A = 0.30$, (c) $x_A = 0.50$ and (d) $x_A = 0.80$. Plots of G_{AB} (green circles) tend to zero when $\lambda \to 1$, as indicated by the horizontal green lines. By contrast, $G_{AA} \to 1/\rho_A$ (indicated by horizontal red lines) when $\lambda \to 1$. The region $\lambda < 3$, indicated by vertical black lines, is where simple linear regression is used to find G_{ij}^{∞} and α_{ij}. By replacing such values in Eq. (6.26), we obtained the black curves that, in all cases, superimpose on the simulation data for the full interval $0 < \lambda < 1$.

The bulk KBIs are related to various thermodynamic quantities. For example, the isothermal compressibility is given by [275]

$$\kappa_T = \frac{1 + \rho_A G_{AA} + \rho_B G_{BB} + \rho_A \rho_B (G_{AA} G_{BB} - G_{AB}^2)}{k_{\mathrm{B}}T(\rho_A + \rho_B + \rho_A \rho_B (G_{AA} + G_{BB} - 2G_{AB}))}\,, \tag{6.28}$$

with $\rho_{A,B}$ the number density of the corresponding species.



(a) **(b)** **(c)** **(d)**

Figure 6.10. Scaled finite-size Kirkwood-Buff integrals $\lambda G_{ij}(\lambda)$ as a function of λ for different mole fractions: (a) $x_A = 0.20$, (b) $x_A = 0.30$, (c) $x_A = 0.50$ and (d) $x_A = 0.80$, for mixtures described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$. For clarity, only the cases G_{AA} (red squares) and G_{AB} (green circles) are plotted. In the asymptotic case $\lambda \to 1$, $G_{AB} \to 0$ and $G_{AA} \to 1/\rho_A$, as indicated by the horizontal green and red lines, respectively. The black curves correspond to Eq. (6.26) with G_{ij}^{∞} and α_{ij} obtained from a simple regression analysis in the interval $\lambda < 0.3$.

Results for the isothermal compressibility obtained from the G_{ij}^{∞} values are presented in Fig. 6.11. Single component cases corresponding to systems composed by only type-A and type-B particles are indicated by the horizontal black lines. As expected, the system composed by strongly interacting particles, i. e. the type-A, has a lower compressibility. The behaviour of the isothermal compressibility is nearly ideal since it follows closely the relation $\kappa_T = (1 - x_A)\kappa_T^B + x_A\kappa_T^A$, with $\kappa_T^A\epsilon/\sigma^3 = 0.012(1)$ and $\kappa_T^B\epsilon/\sigma^3 = 0.0281(8)$, as indicated by the solid black line.

Finally, the extrapolated KBIs have been used to compute the derivative of the chemical potential of type-A particles with respect to the number density ρ_A using the expression [275]:

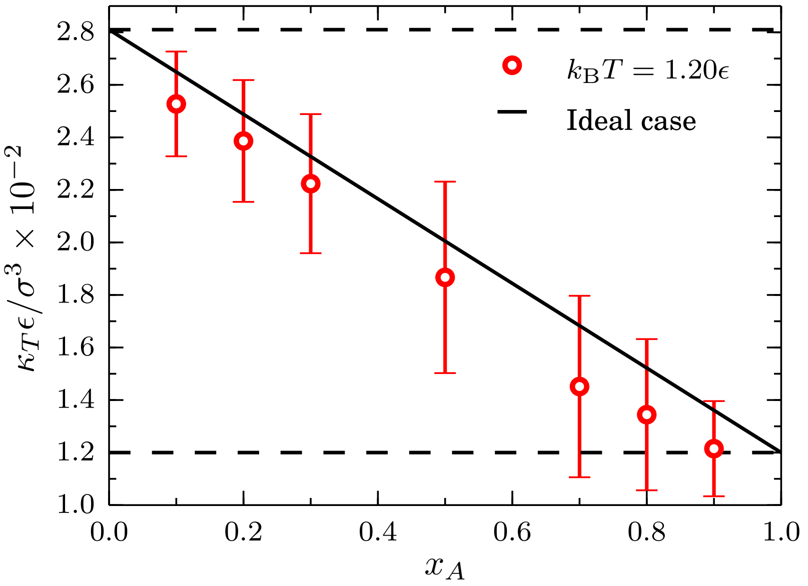


Figure 6.11. Isothermal compressibility at $k_{\mathrm{B}}T = 1.20\epsilon$ and $P\sigma^3/\epsilon = 9.8$ as a function of the mole fraction of type-A particles x_A for mixtures described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$. The horizontal black lines indicate the compressibility for a pure system of type-A particles $\kappa_T^A\epsilon/\sigma^3 = 0.012(1)$ and for a pure system of type-B particles $\kappa_T^B\epsilon/\sigma^3 = 0.0281(8)$. The red line is a guide to the eye. The ideal case corresponds to $\kappa_T = (1 - x_A)\kappa_T^B + x_A\kappa_T^A$.

$$\frac{1}{k_{\mathrm{B}}T}\left(\frac{\partial\mu_A}{\partial\rho_A}\right)_{P,T} = \frac{1}{\rho_A} + \frac{G_{AB} - G_{AA}}{1 + \rho_A(G_{AA} - G_{AB})}, \tag{6.29}$$

that, as it has been done for the single component case, can be integrated to obtain [245]:

$$\delta\mu_A = k_{\mathrm{B}}T\int_{\rho_A^0}^{\rho_A}\left[\frac{1}{\rho'_A} + \frac{G_{AB} - G_{AA}}{1 + \rho'_A(G_{AA} - G_{AB})}\right] d\rho'_A\,. \tag{6.30}$$

This is the chemical potential shifted by a reference chemical potential computed at density ρ_A^0 [243, 279]. By removing the density and concentration terms of the chemical potential of an ideal mixture, the excess chemical potential can be written as:

$$\delta\mu_A^{ex} = \delta\mu_A - k_{\mathrm{B}}T\ln(x_A) - k_{\mathrm{B}}T\ln(\rho_A)\,. \tag{6.31}$$

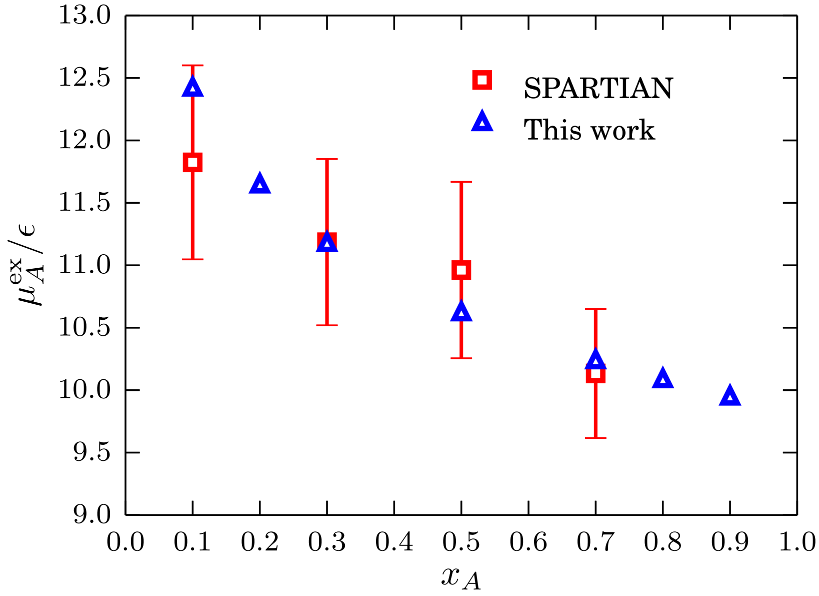


Figure 6.12. Excess chemical potential of type-A particles as a function of the mole fraction x_A for mixtures described by a TSLJ potential with $r_c/\sigma = 2^{1/6}$ at $k_{\rm B}T = 1.2\epsilon$ and $P\sigma^3/\epsilon = 9.8$. Data points obtained with the method in Ref. [93], in particular for $x_A = 0.3$, are used as a reference for the data points obtained with Eqs 6.30 and 6.31.

We compare the results obtained using Eqs 6.30 and 6.31 with the results obtained with the SPARTIAN method [93] and use the excess chemical potential result from $x_A = 0.3$ to find the reference value. We present the results in Fig. 6.12 where a good agreement between the two data sets is apparent. To conclude this section, it has been shown that the block analysis method constitutes a robust strategy to compute chemical potentials of liquids and mixtures in a wide range of density/concentration conditions.

6.5 Summary and Concluding Remarks

In general, a direct comparison between a real system and a finite-size simulation is prevented by the fixed and relatively small number of particles used in the latter. As it has been encoded in the title of the paper, the spatial block analysis method employs a clever combination of finite-size effects, *ensemble* and *boundary*, and density fluctuations to extrapolate bulk isothermal compressibilities and Kirkwood-Buff integrals in the thermodynamic limit.

In this work, we have illustrated with prototypical Lennard-Jones liquids and liquid mixtures the working mechanisms of the method. Upon identifying the relevant finite-size effects and assessing their impact on the simulation results, we have intuitively introduced an analytical expression connecting the fluctuations measured in a small subdomain of the simulation box with the bulk isothermal compressibility for a single component fluid.

Subsequently, the same analytical expression has been rigorously obtained from a finite-size version of the Ornstein-Zernike integral equation. Using a challenging system close to critical point conditions, we have tested the range of validity of the method and obtained results in line with theoretical expectations.

Then, for a multicomponent system, we have applied the same protocol to the Kirkwood-Buff integrals. Using the corresponding analytical expression it is possible to obtain the Kirkwood-Buff integrals in the thermodynamic limit. In both, single and multicomponent systems, the method allows one to compute the chemical potential of a liquid/mixture for a wide range of density/concentration conditions, provided a single reference chemical potential has been determined using a different computational method. These results contribute to establish the spatial block analysis method as a powerful tool to investigate systems where the accurate computation of the chemical potential is of paramount importance.

6.6 Acknowledgement

We thank Debashish Mukherji and Roberto Menichetti for many stimulating discussions. We also thank Claudio Perego and Nancy C. Forero Martinez for a critical reading of the manuscript. MH, KK and RP gratefully acknowledge funding from SFB-TRR146 of the German Research Foundation (DFG). RCH gratefully acknowledges the Alexander von Humboldt Foundation for financial support.

7. Conclusion

7.1 Summary

When I joined the theory group lead by Prof. Kremer in 2015, the Hamiltonian adaptive resolution method had just been recently introduced. At the time, nearly two years after the first `H-AdResS` publication, all the physical basis of the method were fully established. However, there were two major technical aspects whose solution became my project's first goal. The first aspect concerns the treatment of electrostatic interactions within `H-AdResS`. Before my thesis and for practical reasons, the method of choice for the description of electrostatic interactions in adaptive resolution simulations was the reaction field method. However, the strong requirement of a well-defined dielectric constant around every charge centre and the daunting possibility of having to modify the force field to reach the desired accuracy made us consider a more robust method to describe electrostatics. The second aspect lies at the core of the method, namely, concerns the possibility to efficiently calculate on-the-fly the free energy compensations that otherwise required an additional, often cumbersome, thermodynamic integration calculation.

The solution to both issues is the subject of Chapter 2. There, we have introduced the damped shift force (DSF) potential as an accurate treatment of electrostatics in the context of Adaptive Resolution Simulations. Compared with the long-ranged EWALD summation-based scheme, the short-ranged DSF potential has been shown to accurately reproduce the physical (structural and dynamical) properties of many charged systems. Compared with other short-ranged representations such as Reaction Field (RF), the DSF potential does not need to be pre-parametrized based on emergent properties of the system, e.g. the dielectric constant, but contains only two free parameters instead. Another limitation of the RF is that the dielectric constant is calculated under the assumption that the medium is homogeneous, which does not always apply. Thus, we have shown that the DSF method is ideally suited for dual-resolution simulations, and we have demonstrated the validity of its implementation by performing simulations of liquid water and comparing the structural and dynamic properties to a reference, fully atomistic simulation. The accurate treatment of the Coulomb potential in the context of adaptive resolution simulations without the need to re-parametrize the atomistic force field paves the way to the efficient simulation of complex systems in which electrostatic interaction is known to play a primary role, for instance ions in solutions, ionic liquids, and nucleic acids. Finally, we have implemented the `H-AdResS` method, with an on-the-fly calculation of the free energy compensations, in the `LAMMPS` simulation package. Supplemented with the DSF method for Coulomb potential and equipped with a detailed user guide, the new package provides a broad community with an effective instrument to investigate soft and biological matter.

Once these major technical points were addressed, I devoted my research efforts to investigate and exploit the more fundamental analogy between `H-AdResS` and thermodynamic integration methods. In the original `H-AdResS` papers, the free energy compensation has been identified with the Gibbs (Helmholtz) free energy, provided the density (pressure) of the whole system remains uniform throughout the simulation box. By coupling a target atomistic system with a reservoir of thermalized, non-interacting particles (ideal gas) it is possible to extract the excess (or residual) free energy of the target system. This is the subject of Chapter 3, where we have presented a method, spatially resolved thermodynamic integration, or `SPARTIAN`, aiming at computing the chemical potential of liquids and liquid mixtures. First, we have implemented a thermodynamically consistent coupling between the target system and a reservoir of ideal gas particles using the hybrid region as an interface. In this hybrid region, compensating free energies are computed and applied iteratively leading to a uniform density profile for the whole system. Upon convergence of the free energy compensations, the system reaches thermal equilibrium and the chemical potential of both target system and ideal gas reservoir equate. Therefore, the free energy compensations are identified with the difference in chemical potential between the two representations, which is precisely the excess chemical potential of the target system. We demonstrated the efficiency and general applicability of the `SPARTIAN` method by computing the chemical potential of different liquids and mixtures such as pure and multi-component Lennard Jones liquids, pure water, and aqueous solutions of sodium chloride. The computed values of the excess chemical potential for all systems under examination were consistent with the data reported in the literature. For dense liquids and highly concentrated mixtures, the `SPARTIAN` method exceeds standard methods based on test particle insertion algorithms. This is due to the fact that our approach benefits from the larger statistical population of molecules in the hybrid region that helps the convergence of statistical averages.

In chapter 4 we have demonstrated that within the context of the `H-AdResS` framework it is possible to couple all-atom models to their corresponding ideal representations not only for liquids but also for solids, i.e. ideal gas for liquids and ideal Einstein crystal for solids. We have shown that such a coupling does not affect the physical properties of the fully atomistic subdomain. As a matter of fact, we have reported the excellent agreement of structural and dynamical quantities as computed in both `H-AdResS` and fully atomistic simulations. The proposed coupling scheme provides the conceptual and technical foundation for applications with important consequences. Additionally, the computational cost of these simulations is reduced to that of the high-resolution domain since the ideal gas/solid parts require a negligible amount of force calculations. The size of the system can substantially increase thus minimizing finite size effects and enabling the calculation of free energies for systems composed of complex molecules. A particularly useful application of this solid and liquid coupling to ideal representations concerns the calculation of chemical potentials of complex molecular crystals, a topic of increasing relevance for its industrial and pharmaceutical applications.

In chapter 5, we have verified the Jarzynski equality in the context of multiscale simulations of soft matter and used it to compute the absolute solvation free energy of small molecules. In the proposed method, we have used the `SPARTIAN` method to enforce a thermodynamically consistent coupling between an atomistic model of the solvent and its ideal gas representation. We have shown that the smooth coupling between the atomistic and ideal gas representations provides two end states obtained through a thermodynamic integration path, i.e. the state in which the molecule is in the atomistic region (solvated state) or in the non-interacting ideal gas region (unsolvated state). We have used umbrella sampling and steered molecular dynamics on the solute molecule and computed its solvation free energy using both equilibrium statistics and non-equilibrium work-free-energy relations (Jarzynski equality, Crooks theorem). The advantage of this strategy might become more prominent when it is intended to compute the solvation free energy of large molecules. This can be accomplished, for instance by initially inserting them in the ideal gas subregion and pulling them into the atomistic domain. In this way, the solvation free energy of arbitrarily large and complex molecules –from novel chemical compounds for pharmaceutical use to entire proteins– becomes accessible.

Finally, in Chapter 6 we have presented the results of a side project that started as a first approximation to perform simulations in the grand canonical ensemble using the `H-AdResS` method. Following ideas developed from the Sixties until the late Nineties, we have revisited the spatial block analysis method and used it to obtain isothermal compressibilities and Kirkwood-Buff integrals in the thermodynamic limit from relatively small-sized molecular dynamics simulations. By using thermodynamic relations and comparing with our own `SPARTIAN` results, we have used such quantities to calculate chemical potentials in a wide range of density/concentration conditions.

7.2 Outlook

In this thesis, we have presented the Hamiltonian adaptive resolution method in a different light. Instead of focusing on the obvious computational advantage of the method, we have introduced to the community an efficient and versatile method to compute free energy differences for a large variety of systems. The method resembles the thermodynamic integration method without the disadvantages resulting from test particle insertion procedures. More advantages and drawbacks of the method will be apparent only through a thorough application in many different and challenging conditions.

One potential application revolves around the extension of the method to perform molecular dynamics at constant chemical potential or to perform open boundary molecular dynamics (grand canonical) simulations. There are various physical-chemical processes in liquid solutions such as self-assembly, crystallization and growth which are the subject of intense investigation in view of their wide spectrum of applications. However, due to the limited computational resources, we are constrained to study the small time and length scales of such processes thus dramatically impacting the obtained results. In the example of crystal growth, following the crystallization, the solution is depleted and this substantially affects the growth process, because the ensembles of solutions around the growing faces of the crystal are no longer at constant chemical potential. In these situations, one can include a particle exchange algorithm in `H-AdResS` to control the density of the component or equivalently the corresponding chemical potential. Given the advances presented in this thesis, this could be an easy and accessible task.

Another area for future research is to use the `H-AdResS` method to perform non-equilibrium molecular dynamics in the presence of gradients of chemical potential. Recently, the subject of diffusio-osmotic flows in liquid mixtures has received great attention. These phenomena occur in proximity to a fluid-solid interface, when gradients in the local solute concentration are induced. In the region far from the interface, i.e. in a bulk fluid, a diffusion-based mechanism leads to a diffusive flux of both components. However, close to an interface, due to the solute-wall interaction which can adsorb or deplete the solute particles, the solute concentration is perturbed and differs from that in the bulk. As a result, near the interface, the force balance between the solvent and solute is broken and consequently the driving force results in diffusio-osmotic flow [85, 86]. The molecular dynamics simulations of such phenomena for the molecular liquids and dense regimes is challenging because as explained thoroughly in chapter 3, performing particle exchange mechanism so as to change the concentration and the local chemical potential in the fluid is very difficult. However, one can take advantage of coupling fluid mixtures to ideal gas particles in the `H-AdResS` scheme. Then, the desired value of the chemical potential can be easily imposed via a particle exchange mechanism.

SYMBOLS

m	Mass
v	Velocity
t	Time
ps	Pico-second
fs	Femto-second
$\AA$	Angstrom
μ	Micro
q	Charge
$\mathrm{k_B}$	Boltzmann constant
T	Temperature
H	Hamlitonian
$\mathcal{K}$	Kinetic energy
$\Delta\mathrm{H}$	Compensating energy
H_Δ	Compensated Hamlitonian
U	Interacting potential
$\mathrm{V^{AT}}$	Atomistic potential
$\mathrm{V^{CG}}$	Coarse-grained potential
λ	Switching function
F	Helmholtz free energy
G	Gibbs free energy
μ_{ex}	Excess chemical potential
μ_{id}	Ideal chemical potential
$\mathbf{F}_{\mathrm{th}}$	Thermodynamic force
$\mathbf{F}_{\mathrm{dr}}$	Drift force
Q	Canonical partition function
$\mathbf{p}$	Momenta in phase space
$\mathbf{r}$	Coordinates in phase space
W	Work
ρ	Density
ρ^*	Reference density
P	Pressure
ϵ_0	Vacuum permittivity
ϵ_{RF}	Medium permittivity
$\mathrm{C_{vv}}$	Velocity auto-correlation function
σ	Lennard-Jones length scale
ϵ	Lennard-Jones energy scale
χ_{T}	Isothermal compressibility

ABBREVIATIONS

Abbreviation	Meaning
AdResS	Adaptive Resolution Simulation
AT	Atomistic
BAR	Bennett acceptance ratio
CFT	Crooks fluctuation theorem
CG	Coarse-grained
DSF	Damped Shifted Force
FEC	Free energy compensation
KB	Kirkwood-Buff
KBI	Kirkwood-Buff integral
HY	Hybrid
H-AdResS	Hamiltonian Adaptive Resolution Simulation
IEC	Ideal Einstein crystal
IG	Ideal gas
JC	Joung and Cheatham
JE	Jarzynski equality
LJ	Lennard-Jones
MD	Molecular Dynamics
MC	Monte Carlo
MSD	Mean-squared displacement
P3M	Particle-Particle-Particle Mesh
PBC	Periodic boundary condition
RC	Real Crystal
RDF	Radial distribution function
RF	Reaction Field
SBA	Spatial Block Analysis
SFE	Solvation free energy
SPARTIAN	Spatially resolved thermodynamic integration
TI	Thermodynamic integration
TSLJ	Truncated and shifted Lennard-Jones
VACF	Velocity auto-correlation function
WCA	Weeks-Chandler-Andersen

CONTRIBUTIONS

The chapters 2, 3, 4 and 6 have been published as scientific research papers and chapter 5 is a draft that has been submitted at the time of submission of this thesis. These works are carried out in collaborations with colleagues from the Max Planck Institute for Polymer Research, Mainz and the University of California, Davis. In the following, we present in detail the individual contributions.

Chapter 2

The simulation setups were conceived by Maziar Heidari, Robinson Cortes-Huerto, Davide Donadio and Raffaello Potestio. Maziar Heidari implemented `H-AdResS` in the LAMMPS simulation package. Maziar Heidari and Robinson Cortes-Huerto ran simulations and conducted the data analysis. The paper was written by Maziar Heidari, Robinson Cortes-Huerto, Davide Donadio and Raffaello Potestio.

Chapter 3

The original idea was developed by Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. The simulation setups were conceived by Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Maziar Heidari performed all SPARTIAN simulations and analyzed data and Robinson Cortes-Huerto ran TI and Widom insertion simulations. The paper was written by Robinson Cortes-Huerto, and Raffaello Potestio, incorporating critical comments from Kurt Kremer.

Chapter 4

The original idea of concurrent coupling of solids with ideal Einstein crystal was developed by Maziar Heidari, Robinson Cortes-Huerto, Kurt Kremer, and Raffaello Potestio. The simulation setups were conceived by Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Maziar Heidari performed all `H-AdResS` simulations and analyzed data. The paper was written by Maziar Heidari, Robinson Cortes-Huerto, and Raffaello Potestio. Kurt Kremer provided insightful ideas and help.

Chapter 5

The original idea was developed by Raffaello Potestio and Kurt Kremer. The simulation setups were conceived by Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio. Maziar Heidari performed all `H-AdResS` simulations and analyzed the results of work distribution and umbrella sampling method. The paper was written by Maziar Heidari, Robinson Cortes-Huerto, Raffaello Potestio and Kurt Kremer.

Chapter 6

The study was conceived by Robinson Cortes-Huerto. Kurt Kremer, Raffaello Potestio and Robinson Cortes-Huerto planned the computer simulations. Maziar Heidari, Kurt Kremer, Robinson Cortes-Huerto, and Raffaello Potestio discussed the results, helped with their interpretation and contributed to the final manuscript.

REFERENCES

REFERENCES

[1] Physical properties of solids elucidated by zooming in and out of high resolution. `https://www.springer.com/gp/about-springer/media/research-news/all-english-research-news/physical-properties-of-solids-elucidated-by-zooming-in-and-out-of-high-resolution/15785180`, 2018.

[2] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2001.

[3] E. Fermi, J. Pasta, and S. Ulam. Studies of nonlinear problems. *Los Alamos National Laboratory*, Document LA-1940, 1955.

[4] T. Dauxois. Fermi, pasta, ulam and a mysterious lady. *arXiv preprint arXiv:0801.1590*, 2008.

[5] T. P. Weissert. *The genesis of simulation in dynamics: pursuing the Fermi-Pasta-Ulam problem*. Springer Science & Business Media, 2012.

[6] B. K. Shivamoggi. *Nonlinear dynamics and chaotic phenomena: An introduction*, volume 103. Springer, 2014.

[7] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. i. general method. *J. Chem. Phys.*, 31(2):459, 1959.

[8] N. B. Kotadiya, H. Lu, A. Mondal, Y. Ie, D. Andrienko, P. W. M. Blom, and G. A. H. Wetzelaer. Universal strategy for ohmic hole injection into organic semiconductors with high ionization energies. *Nature Materials*, 17(4):329–334, 2018.

[9] P. Kordt, J. J. M. van der Holst, M. Al Helwi, W. Kowalsky, F. May, A. Badinski, C. Lennartz, and D. Andrienko. Modeling of organic light emitting diodes: From molecular to device properties. *Advanced Functional Materials*, 25(13):1955–1971, 2015.

[10] R. Menichetti, K. H. Kanekal, K. Kremer, and T. Bereau. In silico screening of drug-membrane thermodynamics reveals linear relations between bulk partitioning and the potential of mean force. *J. Chem. Phys.*, 147(12):125101, 2017.

[11] R. Potestio, F. Pontiggia, and C. Micheletti. Coarse-grained description of protein internal dynamics: An optimal strategy for decomposing proteins in rigid subunits. *Biophysical Journal*, 96(12):4993 – 5002, 2009.

[12] R. Potestio, C. Micheletti, and H. Orland. Knotted vs. unknotted proteins: Evidence of knot-promoting loops. *PLOS Computational Biology*, 6(7):1–10, 07 2010.

[13] K. Kremer and G. S. Grest. Dynamics of entangled linear polymer melts: a moleculardynamics simulation. *J. Chem. Phys.*, 92(8):5057–5086, 1990.

[14] H. Hsu and K. Kremer. Static and dynamic properties of large polymer melts in equilibrium. *J. Chem. Phys.*, 144(15):154907, 2016.

[15] K. Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136(15):150901, 2012.

[16] R. O. Jones. Density functional theory: Its origins, rise to prominence, and future. *Rev. Mod. Phys.*, 87:897–923, Aug 2015.

[17] M. E. Tuckerman. *Statistical mechanics: theory and molecular simulation.* Oxford University Press, 2010.

[18] A. Chaimovich and M. S. Shell. Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.*, 134(9):094112, 2011.

[19] J. F. Rudzinski and W. G. Noid. Coarse-graining entropy, forces, and structures. *J. Chem. Phys.*, 135(21):214101, 2011.

[20] L. D. Landau and E. M. Lifshitz, editors. *Fluid mechanics, Course of Theoretical Physics,.* Pergamon Press, second edition, 1987.

[21] L. D. Landau, L. P. Pitaevskii, A. M. Kosevich, and E. M. Lifshitz, editors. *Theory of Elasticity, Course of Theoretical Physics,.* Butterworth-Heinemann, third edition, 2012.

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

[23] H. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, 1980.

[24] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, Oct 1980.

[25] N. Shuichi. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.

[26] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126(1):014101, 2007.

[27] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, Mar 1985.

[28] B. Dünweg and W. Paul. Brownian dynamics simulations without GAUSSIAN random numbers. *International Journal of Modern Physics C*, 02(03):817–827, 1991.

[29] V. Molinero and E. B. Moore. Water modeled as an intermediate element between carbon and silicon. *J. Phys. Chem. B*, 113(13):4008–4016, 2009. PMID: 18956896.

[30] C. Globisch, V. Krishnamani, M. Deserno, and C. Peter. Optimization of an elastic network augmented coarse grained model to study ccmv capsid deformation. *PLoS ONE*, 8(4):e60582, 04 2013.

[31] A. C. Fogarty, R. Potestio, and K. Kremer. A multi-resolution model to capture both global fluctuations of an enzyme and molecular recognition in the ligand-binding site. *PROTEINS*, 84:1902–1913, 2016.

[32] S. J. Marrink, A. H. de Vries, and A. E. Mark. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B*, 108(2):750–760, 2004.

[33] B. J. Reynwar, G. Illya, V. A. Harmandaris, M. M. Müller, K. Kremer, and M. Deserno. Aggregation and vesiculation of membrane proteins by curvature-mediated interactions. *Nature*, 447:461 EP –, May 2007.

[34] R. Potestio, F. Pontiggia, and C. Micheletti. Coarse-grained description of protein internal dynamics: An optimal strategy for decomposing proteins in rigid subunits. *Biophys. J.*, 96(12):4993 – 5002, 2009.

[35] M. Heidari, M. Mehrbod, M. R. Ejtehadi, and M. R. K. Mofrad. Cooperation within von willebrand factors enhances adsorption mechanism. *J. Royal Society Interface*, 12(109):20150334, Aug 2015.

[36] G. A. Voth. *Coarse-Graining of Condensed Phase and Biomolecular Systems*. CRC Press, 2008.

[37] F. Ercolessi and J. B. Adams. Interatomic potentials from first-principles calculations: The force-matching method. *EPL (Europhysics Letters)*, 26(8):583, 1994.

[38] S. Izvekov and G. A. Voth. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B*, 109(7):2469–2473, 02 2005.

[39] R. L. McGreevy and L. Pusztai. Reverse Monte Carlo simulation: A new technique for the determination of disordered structures. *Molecular Simulation*, 1(6):359–367, 1988.

[40] A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E*, 52:3730–3737, Oct 1995.

[41] M. S. Shell. *Thermodynamics and statistical mechanics: an integrated approach*. Cambridge University Press, 2015.

[42] P. Carloni, U. Rothlisberger, and M. Parrinello. The role and perspective of ab initio molecular dynamics in the study of biological systems. *Accounts of Chemical Research*, 35(6):455–464, 06 2002.

[43] H. M. Senn and W. Thiel. QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, 48(7):1198–1229, 2009.

[44] B. Wang and D. G. Truhlar. Combined quantum mechanical and molecular mechanical methods for calculating potential energy surfaces: Tuned and balanced redistributed-charge algorithm. *J. Chem. Theory Comput.*, 6(2):359–369, 02 2010.

[45] F. H. Wallrapp and V. Guallar. Mixed quantum mechanics and molecular mechanics methods: Looking inside proteins. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(2):315–322, 2011.

[46] W. Shi and E. J. Maginn. Atomistic simulation of the absorption of carbon dioxide and water in the ionic liquid 1-n-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ([hmim][tf2n]. *J. Phys. Chem. B*, 112(7):2045–2055, 2008.

[47] M. W. van der Kamp and A. J. Mulholland. Combined quantum mechanics/molecular mechanics (qm/mm) methods in computational enzymology. *Biochemistry*, 52(16):2708–2728, 04 2013.

[48] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution molecular-dynamics simulation: Changing the degrees of freedom on the fly. *J. Chem. Phys.*, 123(22):224106–14, 2005.

[49] R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio. Hamiltonian adaptive resolution simulation for molecular liquids. *Phys. Rev. Lett.*, 110:108301, Mar 2013.

[50] R. Potestio, P. Español, R. Delgado-Buscalioni, R. Everaers, K. Kremer, and D. Donadio. Monte Carlo adaptive resolution simulation of multicomponent molecular liquids. *Phys. Rev. Lett.*, 111:060601, Aug 2013.

[51] 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. *Phys. Rev. Lett.*, 108(17):170602, APR 27 2012.

[52] A. C. Fogarty, R. Potestio, and K. Kremer. Adaptive resolution simulation of a biomolecule and its hydration shell: Structural and dynamical properties. *J. Chem. Phys.*, 142(19):195101, 2015.

[53] K. Kreis, R. Potestio, K. Kremer, and A. C. Fogarty. Adaptive resolution simulations with self-adjusting high-resolution regions. *J. Chem. Theory Comput.*, 12(8):4067–4081, 2016. PMID: 27384753.

[54] S. Fritsch, C. Junghans, and K. Kremer. Structure formation of toluene around C60: Implementation of the adaptive resolution scheme (AdResS) into GROMACS. *J. Chem. Theory Comput.*, 8(2):398–403, FEB 2012.

[55] M. Praprotnik, L. Delle Site, and K. Kremer. Multiscale simulation of soft matter: From scale bridging to adaptive resolution. *Ann. Rev. Phys. Chem.*, 59(1):545–571, 2008.

[56] M. Praprotnik, L. Delle Site, and K. Kremer. A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation. *J. Chem. Phys.*, 126:134902, 2007.

[57] M. Praprotnik, L. Delle Site, and K. Kremer. Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids. *Phys. Rev. E*, 73:066701, 2006.

[58] R. Fiorentini, K. Kremer, R. Potestio, and A. C. Fogarty. Using force-based adaptive resolution simulations to calculate solvation free energies of amino acid sidechain analogues. *J. Chem. Phys.*, 146, 2017.

[59] J. Zavadlav, R. Podgornik, M. N. Melo, S. J. Marrink, and M. Praprotnik. Adaptive resolution simulation of an atomistic DNA molecule in MARTINI salt solution. *Eur. Phys. J. Spec. Top.*, 225(8):1595–1607, Oct 2016.

[60] J. Zavadlav, R. Podgornik, and M. Praprotnik. Order and interactions in DNA arrays: Multiscale molecular dynamics simulation. *Scientific Reports*, 7(1):4775, 2017.

[61] M. Heidari, R. Cortes-Huerto, D. Donadio, and R. Potestio. Accurate and general treatment of electrostatic interaction in Hamiltonian adaptive resolution simulations. *Eur. Phys. J.: Special Topics*, 225(8):1505–1526, 2016.

[62] P. Español, R. Delgado-Buscalioni, R. Everaers, R. Potestio, D. Donadio, and K. Kremer. Statistical mechanics of Hamiltonian adaptive resolution simulations. *J. Chem. Phys.*, 142(6):064115, 2015.

[63] K. Kreis. *Advanced adaptive resolution methods for molecular simulation*. PhD thesis, Johannes Gutenberg-Universität Mainz, 2017.

[64] J. G. Kirkwood. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 3(5):300, 1935.

[65] C. J. Fennell and J. D. Gezelter. Is the Ewald summation still necessary? pairwise alternatives to the accepted standard for long-range electrostatics. *J. Chem. Phys.*, 124(23), 2006.

[66] M. Born. Volumen und hydratationswärme der ionen. *Zeitschrift für Physik*, 1(1):45–48, Feb 1920.

[67] P. J. Steinbach and B. R. Brooks. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.*, 15(7):667–683, 1994.

[68] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 64:253–287, 1921.

[69] L. Onsager. Electric moments of molecules in liquids. *J. Am. Chem. Soc*, 58(8):1486–1493, 1936.

[70] J. A. Barker and R. O. Watts. Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.*, 26:789 – 792, 1973.

[71] D. van der Spoel, P. J. van Maaren, and H. J. C. Berendsen. A systematic study of water models for molecular simulation: derivation of water models optimized for use with a reaction field. *J. Chem. Phys.*, 108:10220, 1998.

[72] J. W. Perram, Petersen H. G., and De Leeuw S. W. An algorithm for the simulation of condensed matter which grows as the 3/2 power of the number of particles. *Mol. Phys.*, 65(4):875–893, 1988.

[73] R. W. Hockney and J. W Eastwood. *Computer simulation using particles*. CRC Press, 1988.

[74] T. Darden, D. York, and L. Pedersen. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys.*, 98(12):10089–10092, 1993.

[75] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh Ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.

[76] Y.-J. Rhee, J. W. Halley, Joseph Hautman, and A. Rahman. Ewald methods in molecular dynamics for systems of finite extent in one of three dimensions. *Phys. Rev. B*, 40:36–42, Jul 1989.

[77] D. E. Parry. The electrostatic potential in the surface region of an ionic crystal. *Surface Science*, 49(2):433 – 440, 1975.

[78] I.-C. Yeh and M. L. Berkowitz. Ewald summation for systems with slab geometry. *J. Chem. Phys.*, 111(7):3155–3162, 1999.

[79] W. Weber, P. H. Hünenberger, and J. A. McCammon. Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: influence of artificial periodicity on peptide conformation. *J. Phys. Chem. B*, 104(15):3668–3675, 2000.

[80] J. S. Hub, B. L. de Groot, H. Grubmüller, and G. Groenhof. Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J. Chem. Theory Comput.*, 10(1):381–390, 2014. PMID: 26579917.

[81] D. Wolf. Reconstruction of NaCl surfaces from a dipolar solution to the Madelung problem. *Phys. Rev. Lett.*, 68:3315, 1992.

[82] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht. Exact method for the simulation of Coulombic systems by spherically truncated, pairwise r-1 summation. *J. Chem. Phys.*, 110(17):8254–8282, 1999.

[83] M. Schrader, P. Virnau, and K. Binder. Simulation of vapor-liquid coexistence in finite volumes: A method to compute the surface free energy of droplets. *Phys. Rev. E*, 79:061104, Jun 2009.

[84] N. E. R. Zimmermann, B. Vorselaars, D. Quigley, and B. Peters. Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates. *J. Am. Chem. Soc.*, 137(41):13352–13361, 2015. PMID: 26371630.

[85] H. Yoshida, S. Marbach, and L. Bocquet. Osmotic and diffusio-osmotic flow generation at high solute concentration. ii. molecular dynamics simulations. *J. Chem. Phys.*, 146(19):194702, 2017.

[86] Y. Liu, R. Ganti, and D. Frenkel. Pressure gradients fail to predict diffusio-osmosis. *J. Phys. Condens. Matter*, 30(20):205002, 2018.

[87] C. Perego, F. Giberti, and M. Parrinello. Chemical potential calculations in dense liquids using metadynamics. *Eur. Phys. J.: Special Topics*, 225(8):1621–1628, 2016.

[88] S. G. Moore and D. R. Wheeler. Chemical potential perturbation: A method to predict chemical potentials in periodic molecular simulations. *J. Chem. Phys.*, 134(11):114514, 2011.

[89] C. Perego, O. Valsson, and M. Parrinello. Chemical potential calculations in non-homogeneous liquids. *J. Chem. Phys.*, 149(7):072305, 2018.

[90] O. Valsson and M. Parrinello. Variational approach to enhanced sampling and free energy calculations. *Phys. Rev. Lett.*, 113:090601, Aug 2014.

[91] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690, 4 1997.

[92] G. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60(3):2721–2726, 1999.

[93] M. Heidari, K. Kremer, R. Cortes-Huerto, and R. Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. *J. Chem. Theory. Comput.*, 14(7):3409–3417, 2018.

[94] K. Binder. Finite size scaling analysis of ising model block distribution functions. *Z. Phys. B*, 43(2):119–140, 1981.

[95] M. Rovere, D. W. Hermann, and K. Binder. Block density distribution function analysis of two-dimensional Lennard-Jones fluids. *EPL*, 6(7):585, 1988.

[96] G. S. Grest and K. Kremer. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A*, 33:3628–3631, May 1986.

[97] K. Kremer, G. S. Grest, and I. Carmesin. Crossover from rouse to reptation dynamics: A molecular-dynamics simulation. *Phys. Rev. Lett.*, 61:566–569, Aug 1988.

[98] L. Yelash, M. Müller, W. Paul, and K. Binder. How well can coarse-grained models of real polymers describe their structure? the case of polybutadiene. *J. Chem. Theory Comput.*, 2(3):588–597, MAY 2006.

[99] T. Spyriouni, C. Tzoumanekas, D. Theodorou, F. Müller-Plathe, and G. Milano. Coarse-grained and reverse-mapped united-atom simulations of long-chain atactic polystyrene melts: Structure, thermodynamic properties, chain conformation, and entanglements. *Macromolecules*, 40(10):3876–3885, 2007.

[100] J.A. McCammon and M. Karplus. Internal motions of antibody molecules. *Nature*, 268(5622):765–766, Aug 1977.

[101] M. Karplus and J.A. McCammon. Protein structural fluctuations during a period of 100 ps. *Nature*, 277(5697):578–578, Feb 1979.

[102] P. Raiteri, A. Laio, F. L. Gervasio, C. Micheletti, and M. Parrinello. Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Phys. Chem. B*, 110(8):3533–3539, Mar 2006.

[103] H. Lou and R. I. Cukier. Molecular dynamics of apo-adenylate kinase: A distance replica exchange method for the free energy of conformational fluctuations. *J. Phys. Chem. B*, 110(25):12796–12808, Jun 2006.

[104] K. Arora and C. L. Brooks. Large-scale allosteric conformational transitions of adenylate kinase appear to involve a population-shift mechanism. *Proc. Natl. Acad. Sci. U. S. A.*, 104(47):18496–18501, Nov 2007.

[105] F. Pontiggia, A. Zen, and C. Micheletti. Small and large scale conformational changes of adenylate kinase: a molecular dynamics study of the subdomain motion and mechanics. *Biophys J*, 95(12):5901–5912, Dec 2008.

[106] M. M. Tirion and D. ben Avraham. Normal mode analysis of g-actin. *J. Mol. Biol.*, 230:186–195, 1993.

[107] M. M. Tirion. Large amplitude elastic motions in proteins from a single–parameter, atomic analysis. *Phys. Rev. Lett.*, 77:1905–1908, 1996.

[108] I. Bahar, A. R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. *Folding and Design*, 2:173–181, 1997.

[109] C. Micheletti, P. Carloni, and A. Maritan. Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models. *Proteins*, 55(3):635–645, May 2004.

[110] R. Potestio, F. Pontiggia, and C. Micheletti. Coarse-grained description of proteins' internal dynamics: an optimal strategy for decomposing proteins in rigid subunits. *Biophys J*, 96:4993–5002, 2009.

[111] K. Kremer. *Computer simulations in soft matter science*, volume 53 of *SUSSP proceedings*. IOP Publishing Ltd., 2000.

[112] K. Kremer and F. Müller-Plathe. Multiscale problems in polymer science: Simulation approaches. *MRS Bulletin*, 26(3):205–210, MAR 2001.

[113] N. F. A. van der Vegt, C. Peter, and K. Kremer. *Structure-Based Coarse- and Fine-Graining in Soft Matter Simulations*. CRC Press - Taylor and Francis Group, 2009.

[114] C. Hijón, E. Vanden-Eijnden, R. Delgado-Buscalioni, and P. Español. Mori-Zwanzig formalism as a practical computational tool. *Farad. Discuss.*, 144(1):301–22; discussion 323–45, 467–81, January 2010.

[115] W.G. Noid. *Systematic methods for structurally consistent coarse-grained models*, volume 924 of *Methods in Molecular Biology*. Humana Press, 2013.

[116] W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.*, 139(9):090901, 2013.

[117] A. Agarwal, H. Wang, C. Schütte, and L. Delle Site. Chemical potential of liquids and mixtures via adaptive resolution simulation. *J. Chem. Phys.*, 141(3):034102, 2014.

[118] K. Kreis, D. Donadio, K. Kremer, and R. Potestio. A unified framework for force-based and energy-based adaptive resolution simulations. *Europhys. Lett.*, 108(3):30007, 2014.

[119] B. P. Jr. Lambeth, C. Junghans, K. Kremer, C. Clementi, and L. Delle Site. Communication: On the locality of hydrogen bond networks at hydrophobic interfaces. *J. Chem. Phys.*, 133(22):221101, 2010.

[120] S. Bevc, C. Junghans, K Kremer, and M. Praprotnik. Adaptive resolution simulation of salt solutions. *New J. Phys.*, 15(10):105007, 2013.

[121] I. Fukuda and H. Nakamura. Non-Ewald methods: theory and applications to molecular systems. *Biophys. Rev.*, 4:161, 2012.

[122] M. A. Kastenholz, , and P. H. Hünenberger. Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods. *J. Phys. Chem. B*, 108(2):774–788, 2004.

[123] D. A. C. Beck, R. S. Armen, and V. Daggett. Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides. *Biochemistry*, 44(2):609–616, 2005.

[124] R. D. Lins and U. Röthlisberger. Influence of long-range electrostatic treatments on the folding of the n-terminal h4 histone tail peptide. *J. Chem. Theory Comput.*, 2(2):246–250, 2006.

[125] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.*, 117(1):1 – 19, 1995.

[126] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids.* Clarendon Press, Oxford, 1987.

[127] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *J. Comp. Chem.*, 26(16):1701 – 1718, 2005.

[128] S. Fritsch, R. Potestio, D. Donadio, and K. Kremer. Nuclear quantum effects in water: A multiscale study. *J. Chem. Theory Comput.*, 10(2):816–824, 2014. PMID: 26580055.

[129] F. Bresme, A. Lervik, D. Bedeaux, and S. Kjelstrup. Water polarization under thermal gradients. *Phys. Rev. Lett.*, 101:020602, 2008.

[130] J. R. Errington and P. G. Debenedetti. Relationship between structural order and the anomalies of liquid water. *Nature*, 409:318 – 321, 2001.

[131] J. Kohanoff. Phonon spectra from short non-thermally equilibrated molecular dynamics simulations. *Comput. Mat. Sci.*, 2(2):221 – 232, 1994.

[132] T. Youngs. dlputils: Calculate properties from molecular dynamics trajectories, 2016. https://www.projectaten.com/dlputils.

[133] 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.

[134] L. X. Dang and B. M. Pettitt. Simple intramolecular model potentials for water. *J. Phys. Chem.*, 91(12):3349–3354, 1987.

[135] Y. Wu, H. L. Tepper, and G. A. Voth. Flexible simple point-charge water model with improved liquid-state properties. *J. Chem. Phys.*, 124(2):–, 2006.

[136] J. D. Weeks, D. Chandler, and H. C. Andersen. Role of repulsive forces in determining equilibrium structure of simple liquids. *J. Chem. Phys.*, 54(12):5237 – 5247, 1971.

[137] E. Dubou-Dijon and D. Laage. Characterization of the local structure in liquid water by various order parameters. *J. Phys. Chem. B*, 119(26):8406–8418, 2015. PMID: 26054933.

[138] C. Avendaño and A. Gil-Villegas. Monte Carlo simulations of primitive models for ionic systems using the wolf method. *Mol. Phys.*, 104(9):1475–1486, 2006.

[139] T. G. Desai. Molecular dynamics study of screening at ionic surfaces. *J. Chem. Phys.*, 127(15), 2007.

[140] Y. Nagata and S. Mukamel. Vibrational sum-frequency generation spectroscopy at the water/lipid interface: Molecular dynamics simulation study. *J. Am. Chem. Soc.*, 132(18):6434–6442, 2010.

[141] E. E. Gdoutos, R. Agrawal, and H. D. Espinosa. Comparison of the Ewald and Wolf methods for modeling electrostatic interactions in nanowires. *International Journal for Numerical Methods in Engineering*, 84(13):1541–1551, 2010.

[142] Wei Shi, , and Edward J. Maginn*. Atomistic simulation of the absorption of carbon dioxide and water in the ionic liquid 1-n-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ([hmim][tf2n]. *The Journal of Physical Chemistry B*, 112(7):2045–2055, 2008.

[143] K. Kreis, A. C. Fogarty, K. Kremer, and R. Potestio. Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations. *Eur. Phys. J. Special Topics*, 224(12):2289–2304, 2015.

[144] S. Bellissima, M. Neumann, E. Guarini, U. Bafile, and F. Barocchi. Time dependence of the velocity autocorrelation function of a fluid: An eigenmode analysis of dynamical processes. *Phys. Rev. E*, 92:042166, Oct 2015.

[145] G. Job and F. Herrmann. Chemical potential - a quantity in search of recognition. *Eur. J. Phys.*, 27(2):353, 2006.

[146] R. Baierlein. The elusive chemical potential. *Am. J. Phys.*, 69(4):423–434, 2001.

[147] K. Voïtchovsky, D. Giofre, J. J. Segura, F. Stellacci, and M. Ceriotti. Thermally-nucleated self-assembly of water and alcohol into stable structures at hydrophobic interfaces. *Nat. Comm.*, 7:13064, 2016.

[148] P. Raiteri and J. D. Gale. Water is the key to nonclassical nucleation of amorphous calcium carbonate. *J. Am. Chem. Soc.*, 132(49):17623–17634, 2010.

[149] R. Demichelis, P. Raiteri, J. D. Gale, D. Quigley, and D. Gebauer. Stable prenucleation mineral clusters are liquid-like ionic polymers. *Nat. Commun.*, 2(590):590, 2011.

[150] M. De La Pierre, P. Raiteri, A. G. Stack, and J. D. Gale. Uncovering the atomistic mechanism for calcite step growth. *Angew. Chem.*, 129:8584, 2017.

[151] N. E. R. Zimmermann, B. Vorselaars, D. Quigley, and B. Peters. Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates. *J. Am. Chem. Soc.*, 137(41):13352–13361, 2015. PMID: 26371630.

[152] M. Ferrario, G. Ciccotti, E. Spohr, T. Cartailler, and P. Turq. Solubility of KF in water by molecular dynamics using the Kirkwood integration method. *J. Chem. Phys.*, 117(10):4947–4953, 2002.

[153] J. L. Aragones, E. Sanz, and C. Vega. Solubility of NaCl in water by molecular simulation revisited. *J. Chem. Phys.*, 136(24):244508, 2012.

[154] A. L. Benavides, J. L. Aragones, and C. Vega. Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route. *J. Chem. Phys.*, 144(12):124504, 2016.

[155] J. R. Espinosa, J. M. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. G. Debenedetti, and A. Z. Panagiotopoulos. On the calculation of solubilities via direct coexistence simulations: Investigation of NaCl aqueous solutions and Lennard-Jones binary mixtures. *J. Chem. Phys.*, 145(15):154111, 2016.

[156] Z. Mester and A. Z. Panagiotopoulos. Temperature-dependent solubilities and mean ionic activity coefficients of alkali halides in water from molecular dynamics simulations. *J. Chem. Phys.*, 143(4):044505, 2015.

[157] Z. Mester and A. Z. Panagiotopoulos. Mean ionic activity coefficients in aqueous NaCl solutions from molecular dynamics simulations. *J. Chem. Phys.*, 142(4):044507, 2015.

[158] F. Moučka, I. Nezbeda, and W. R. Smith. Molecular simulation of aqueous electrolytes: Water chemical potential results and gibbs-duhem equation consistency tests. *J. Chem. Phys.*, 139(12):124505, 2013.

[159] K. B. Daly, J. B. Benziger, P. G. Debenedetti, and A. Z. Panagiotopoulos. Massively parallel chemical potential calculation on graphics processing units. *Comp. Phys. Comm.*, 183(10):2054 – 2062, 2012.

[160] D. A. Kofke and P. T. Cummings. Quantitative comparison and optimization of methods for evaluating the chemical potential by molecular simulation. *Mol. Phys.*, 92(6):973–996, 1997.

[161] I. Nezbeda and J. Kolafa. A new version of the insertion particle method for determining the chemical potential by Monte Carlo simulation. *Mol. Sim.*, 5(6):391–403, 1991.

[162] A. M. Ferrenberg and R. H. Swendsen. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.*, 61:2635–2638, Dec 1988.

[163] A. M. Ferrenberg and R. H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63:1195–1198, Sep 1989.

[164] A. M. Ferrenberg, D. P. Landau, and R. H. Swendsen. Statistical errors in histogram reweighting. *Phys. Rev. E*, 51:5092–5100, May 1995.

[165] B. Widom. Some topics in the theory of fluids. *J. Chem. Phys.*, 39(11):2808–2812, 1963.

[166] C. H. Bennett. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.*, 22(2):245 – 268, 1976.

[167] M. R. Shirts, E. Bair, G. Hooker, and V. S. Pande. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.*, 91:140601, Oct 2003.

[168] M. R. Shirts and V. S. Pande. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *J. Chem. Phys.*, 122(14):144107, 2005.

[169] H. Paliwal and M. R. Shirts. A benchmark test set for alchemical free energy transformations and its use to quantify error in common free energy methods. *J. Chem. Theory Comput.*, 7(12):4115–4134, 12 2011.

[170] R. W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *J. Chem. Phys.*, 22(8):1420–1426, 1954.

[171] D. Frenkel and A. J. C. Ladd. New Monte Carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres. *J. Chem. Phys.*, 81(7):3188–3193, 1984.

[172] M. Rossi, P. Gasparotto, and M. Ceriotti. Anharmonic and quantum fluctuations in molecular crystals: A first-principles study of the stability of paracetamol. *Phys. Rev. Lett.*, 117:115702, Sep 2016.

[173] Z. Sun, X. Wang, and J. Song. Extensive assessment of various computational methods for aspartate's pka shift. *J. Chem. Inf. Model*, 57(7):1621–1639, 07 2017.

[174] M. Grünwald and C. Dellago. Ideal gas pressure bath: a method for applying hydrostatic pressure in the computer simulation of nanoparticles. *Mol. Phys.*, 104(22-24):3709–3715, 2006.

[175] J. G. Powles, B. Holtz, and W. A. B. Evans. New method for determining the chemical potential for condensed matter at high density. *J. Chem. Phys.*, 101(9):7804–7810, 1994.

[176] S. G. Moore and D. R. Wheeler. Chemical potential perturbation: Extension of the method to lattice sum treatment of intermolecular potentials. *J. Chem. Phys.*, 136(16):164503, 2012.

[177] R.L. Rowley, T.D. Shupe, and M.W. Schuck. A direct method for determination of chemical potential from osmotic molecular dynamics simulations. *Fluid Phase Equilib.*, 104:159 – 171, 1995.

[178] I. S. Joung and T. E. Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B*, 112(30):9020–9041, 2008.

[179] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.*, 222(6):529 – 539, 1994.

[180] M. Heidari, R. Cortes-Huerto, K. Kremer, and R. Potestio. Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations. In Press.

[181] W. Kob and H. C. Andersen. Scaling behavior in the β-relaxation regime of a supercooled Lennard-Jones mixture. *Phys. Rev. Lett.*, 73:1376–1379, Sep 1994.

[182] J. Quintana and A. D. J. Haymet. The chemical-potential of water - molecular-dynamics computer-simulation of the cf and spc models. *Chem. Phys. Lett.*, 189(3):273–277, Feb 1992.

[183] J. Sauter and A. Grafmüller. Predicting the chemical potential and osmotic pressure of polysaccharide solutions by molecular simulations. *J. Chem. Theory Comput.*, 12(9):4375–4384, 2016.

[184] A. Ben-Naim and Y. Marcus. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.*, 81:2016–2027, 1984.

[185] K. Z. Takahashi. Truncation effects of shift function methods in bulk water systems. *Entropy*, 15(8):3249–3264, 2013.

[186] D. Zahn, B. Schilling, and S. M. Kast. Enhancement of the Wolf Damped Coulomb potential: Static, dynamic, and dielectric properties of liquid water from molecular simulation. *J. Phys. Chem. B*, 106(41):10725–10732, 2002.

[187] G. Fanourgakis. An extension of wolf's method for the treatment of electrostatic interactions: Application to liquid water and aqueous solutions. *J. Phys. Chem. B*, 119(5):1974–1985, 2015.

[188] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, Jun 1982.

[189] D. Frenkel and J. P. Hansen. Understanding liquids: a computer game? *Physics World*, 9(4):35, 1996.

[190] W. F. van Gunsteren and A. E. Mark. Validation of molecular dynamics simulation. *J. Chem. Phys.*, 108(15):6109–6116, 1998.

[191] W. G. Hoover. 50 Years of Computer Simulation – a Personal View. *ArXiv e-prints*, December 2008.

[192] A. Mulero, editor. *Theory and Simulation of Hard-Sphere Fluids and Related Systems*, volume 753 of *Lecture Notes in Physics*. Springer, Berlin, Heidelberg, 2008.

[193] R. E. Caflisch, G. Ceder, K. Kremer, T. Pollock, M. Scheffler, and E. G. Wang, editors. *Focus on Novel Materials Discovery*. Focus on collections. New J. Phys., 2014.

[194] M. Karplus. Molecular dynamics simulations of biomolecules. *Accounts of Chemical Research*, 35(6):321–323, 2002. PMID: 12069615.

[195] A. Prez, F. J. Luque, and M. Orozco. Frontiers in molecular dynamics simulations of DNA. *Accounts of Chemical Research*, 45(2):196–205, 2012. PMID: 21830782.

[196] P. Ballone. Modeling potential energy surfaces: From first-principle approaches to empirical force fields. *Entropy*, 16(1):322–349, 2014.

[197] U. R. Pedersen, F. Hummel, G. Kresse, G. Kahl, and C. Dellago. Computing gibbs free energy differences by interface pinning. *Phys. Rev. B*, 88:094101, Sep 2013.

[198] P. Español, R. Delgado-Buscalioni, R. Everaers, R. Potestio, D. Donadio, and K. Kremer. Statistical Mechanics of Hamiltonian Adaptive Resolution Simulations. *ArXiv e-prints*, July 2014.

[199] M. Heidari, K. Kremer, R. Cortes-Huerto, and R. Potestio. Spatially resolved thermodynamic integration: An efficient method to compute chemical potentials of dense fluids. Submitted.

[200] K. Kreis, A. C. Fogarty, K. Kremer, and R. Potestio. Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations. *Eur. Phys. J. Special Topics*, 224:2289–2304, 2015.

[201] F. Pavia and W. A. Curtin. Parallel algorithm for multiscale atomistic/continuum simulations using lammps. *Modelling and Simulation in Materials Science and Engineering*, 23(5):055002, 2015.

[202] R. E. Rudd and J. Q. Broughton. Concurrent coupling of length scales in solid state systems. *Phys. Status Solidi B-Basic Res.*, 217(1):251–291, JAN 2000.

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

[204] G. Csanyi, T. A., M. C. Payne, and A. De Vita. “Learn on the fly”: A hybrid classical and Quantum-Mechanical molecular dynamics simulation. *Phys. Rev. Lett.*, 93(17):175503, 2004.

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

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

[207] J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel. Finite-size corrections to the free energies of crystalline solids. *J. Chem. Phys.*, 112(12):5339–5342, 2000.

[208] M. A. van der Hoef. Free energy of the Lennard-Jones solid. *J. Chem. Phys.*, 113(18):8142–8148, 2000.

[209] C. Vega and E. G. Noya. Revisiting the Frenkel-Ladd method to compute the free energy of solids: The Einstein molecule approach. *J. Chem. Phys.*, 127(15):154113, 2007.

[210] C. Bustamante, J. Liphardt, and F. Ritort. The nonequilibrium thermodynamics of small systems. *Phys. Today*, 58(7):43–48, 5 2005.

[211] C. Jarzynski. Nonequilibrium work relations: foundations and applications. *Eur. Phys. J. B*, 64(3-4):331–340, 8 2008.

[212] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys. Rev. E*, 56:5018, 11 1997.

[213] G. Hummer and A. Szabo. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci.*, 98(7):3658–3661, 1 2001.

[214] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante. Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski’s equality. *Science*, 296(5574):1832–1835, 2 2002.

[215] G. Hummer and A. Szabo. Free energy profiles from single-molecule pulling experiments. *Proc. Natl. Acad. Sci.*, 107(50):21441–21446, 2010.

[216] S. Raman, T. Utzig, T. Baimpos, B. R. Shrestha, and M. Valtiner. Deciphering the scaling of single-molecule interactions using Jarzynskis equality. *Nat. Comm.*, 5(1):5539, 2014.

[217] M. Ø. Jensen, S. Park, E. Tajkhorshid, and K. Schulten. Energetics of glycerol conduction through aquaglyceroporin glpf. *Proc. Natl. Acad. Sci.*, 99(10):6731–6736, 2 2002.

[218] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. Free energy calculation from steered molecular dynamics simulations using Jarzynskis equality. *J. Chem. Phys.*, 119(6):3559–3566, 2003.

[219] M. Goette and H. Grubmüller. Accuracy and convergence of free energy differences calculated from nonequilibrium switching processes. *J. Comput. Chem.*, 30(3):447–456, 2009.

[220] D. Frenkel. *Lecture Notes on: Free-Energy Calculations*, pages 85–117. Springer Netherlands, Dordrecht, 1991.

[221] R. Roth, Y. Harano, and M. Kinoshita. Morphometric approach to the solvation free energy of complex molecules. *Phys. Rev. Lett.*, 97:078101, Aug 2006.

[222] A. M. A. West, R. Elber, and D. Shalloway. Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide. *J. Chem. Phys.*, 126(14):145104, 2007.

[223] N. Hansen and W. F. van Gunsteren. Practical aspects of Free-Energy calculations: A review. *J. Chem. Theory. Comput.*, 10(7):2632–47, 2014.

[224] F. M. Ytreberg, R. H. Swendsen, and D. M. Zuckerman. Comparison of free energy methods for molecular systems. *J. Chem. Phys.*, 125(18), 6 2006.

[225] L. Yang, A. Ahmed, and S. I. Sandler. Comparison of two simulation methods to compute solvation free energies and partition coefficients. *J. Comput. Chem.*, 34(4):284–293, 2013.

[226] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. Mobley. Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the freesolv database. *bioRxiv*, 2017.

[227] L. Li, T. Totton, and D. Frenkel. Computational methodology for solubility prediction: Application to the sparingly soluble solutes. *J. Chem. Phys.*, 146(21):214110, 2017.

[228] J. Gore, F. Ritort, and C. Bustamante. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl. Acad. Sci.*, 100(22):12564–12569, 3 2003.

[229] C. Dellago and G. Hummer. Computing equilibrium free energies using Non-Equilibrium molecular dynamics. *Entropy*, 16(1):41–61, 2014.

[230] M. Heidari, R. Cortes-Huerto, K. Kremer, and R. Potestio. Concurrent coupling of realistic and ideal models of liquids and solids in Hamiltonian adaptive resolution simulations. *Eur. Phys. J. E*, 41(5):64, May 2018.

[231] G.M. Torrie and J.P. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.*, 23(2):187 – 199, 1977.

[232] W. L. Jorgensen and J. Tirado-Rives. *J. Am. Chem. Soc.*, 110(6):1657–1666, 1988.

[233] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.

[234] H. Kokubo, J. Rösgen, D. W. Bolen, and B. M. Pettitt. Molecular basis of the apparent near ideality of urea solutions. *Biophys. J.*, 93(10):3392–3407, 7 2007.

[235] 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. *J. Comput. Phys.*, 23(3):327 – 341, 1977.

[236] C. Bartels. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chem. Phys. Lett.*, 331(5):446 – 454, 2000.

[237] L. J. Smith, H. J. C. Berendsen, and W. F. van Gunsteren. Computer simulation of ureawater mixtures: a test of force field parameters for use in biomolecular simulation. *J. Phys. Chem. B*, 108(3):1065–1071, 2004.

[238] G. Hummer. Fast-growth thermodynamic integration: Error and efficiency analysis. *J. Chem. Phys.*, 114(17):7330–7337, 1 2001.

[239] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61(3):2361–2366, 2000.

[240] J.-P. Hansen and I. R. McDonald. *Theory of Simple Liquids*. Academic Press, third edition, 2006.

[241] A. Ben-Naim. *Molecular Theory of Solutions*. Oxford University Press, 2006.

[242] D. J. Adams. Chemical potential of hard-sphere fluids by Monte Carlo methods. *Mol. Phys.*, 28(5):1241–1252, 1974.

[243] D. Mukherji and K. Kremer. Coil-Globule-Coil transition of PNIPAm in aqueous methanol: Coupling All-Atom simulations to Semi-Grand canonical Coarse-Grained reservoir. *Macromolecules*, 46:9158, 2013.

[244] H. Wang, C. Hartmann, C. Schütte, and L. Delle Site. Grand-Canonical-like Molecular-Dynamics simulations by using an Adaptive-Resolution technique. *Phys. Rev. X*, 3:011018, 2013.

[245] M. Heidari, K. Kremer, R. Potestio, and R. Cortes-Huerto. Finite-size integral equations in the theory of liquids and the thermodynamic limit in computer simulations. 2018.

[246] A. A. Galata, S. D. Anogiannakis, and D. N. Theodorou. Thermodynamic analysis of Lennard-Jones binary mixtures using Kirkwood-Buff theory. *Fluid Phase Equilibria*, 2017.

[247] L. Delle Site, G. Ciccotti, and C. Hartmann. Partitioning a macroscopic system into independent subsystems. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(8):083201, 2017.

[248] B. A. Strom, J. M. Simon, S. K. Schnell, S. Kjelstrup, J. He, and D. Bedeaux. Size and shape effects on the thermodynamic properties of nanoscale volumes of water. *Phys. Chem. Chem. Phys.*, 19:9016–9027, 2017.

[249] N. Dawass, P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, J. M. Simon, and T. J. H. Vlugt. Finite-size effects of Kirkwood-Buff integrals from molecular simulations. *Molecular Simulation*, 0(0):1–14, 2017.

[250] J. Milzetti, D. Nayar, and N. F. A. van der Vegt. Convergence of Kirkwood-Buff integrals of ideal and nonideal aqueous solutions using molecular dynamics simulations. *J. Phys. Chem. B*, 0(0):null, 0. PMID: 29342355.

[251] D. M. Rogers. Extension of Kirkwood-Buff theory to the canonical ensemble. *J. Chem. Phys.*, 148(5):054102, 2018.

[252] N. Dawass, P. Krüger, J. M. Simon, and T. J. H. Vlugt. Kirkwood–buff integrals of finite systems: shape effects. *Mol. Phys.*, 0(0):1–8, 2018.

[253] K. Binder. Critical properties from Monte Carlo coarse graining and renormalization. *Phys. Rev. Lett.*, 47:693–696, Aug 1981.

[254] M. Rovere, D. W. Heermann, and K. Binder. The gas-liquid transition of the two-dimensional Lennard-Jones fluid. *J. Phys.: Condens. Matter*, 2(33):7009, 1990.

[255] M. Rovere, P. Nielaba, and K. Binder. Simulation studies of gas-liquid transitions in two dimensions via a subsystem-block-density distribution analysis. *Z. Phys. B*, 90(2):215–228, 1993.

[256] H. Weber, D. Marx, and K. Binder. Melting transition in two dimensions: A finite-size scaling analysis of bond-orientational order in hard disks. *Phys. Rev. B*, 51:14636–14651, May 1995.

[257] F. L. Román, J. A. White, and S. Velasco. Block analysis method in off-lattice fluids. *EPL*, 42(4):371, 1998.

[258] J. J. Salacuse. Particle fluctuations within sub-regions of an n-particle, three-dimensional fluid: Finite-size effects and compressibility. *Physica A*, 387(13):3073 – 3083, 2008.

[259] S. Sengupta, P. Nielaba, M. Rao, and K. Binder. Elastic constants from microscopic strain fluctuations. *Phys. Rev. E*, 61:1072–1080, Feb 2000.

[260] J. L. Lebowitz and J. K. Percus. Long-range correlations in a closed system with applications to nonuniform fluids. *Phys. Rev.*, 122:1675–1691, Jun 1961.

[261] J. J. Salacuse, A. R. Denton, and P. A. Egelstaff. Finite-size effects in molecular dynamics simulations: Static structure factor and compressibility. i. theoretical method. *Phys. Rev. E*, 53:2382–2389, Mar 1996.

[262] F. L. Román, J. A. White, and S. Velasco. Fluctuations in an equilibrium hard-disk fluid: Explicit size effects. *J. Chem. Phys.*, 107:4635, 1997.

[263] D. Villamaina and E. Trizac. Thinking outside the box: fluctuations and finite size effects. *Eur. J. Phys.*, 35(3):035011, 2014.

[264] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[265] J. S. Rowlinson. The equation of state of dense systems. *Reports on Progress in Physics*, 28(1):169, 1965.

[266] S. K. Schnell, T. J.H. Vlugt, J.-M. Simon, D. Bedeaux, and S. Kjelstrup. Thermodynamics of a small system in a μt reservoir. *Chem. Phys. Lett.*, 504(46):199 – 201, 2011.

[267] J. L. Lebowitz and J. K. Percus. Thermodynamic properties of small systems. *Phys. Rev.*, 124:1673–1681, Dec 1961.

[268] T. L. Hill. *Thermodynamics of Small Systems.* Dover, 1963.

[269] J. D. Halverson, T. Brandes, O. Lenz, A. Arnold, S. Bevc, V. Starchenko, K. Kremer, T. Stuehn, and D. Reith. Espresso++: A modern multiscale simulation package for soft matter systems. *Comp. Phys. Comm.*, 184(4):1129 – 1149, 2013.

[270] F. L. Román, A. González, J. A. White, and S. Velasco. Fluctuations in the number of particles of the ideal gas: A simple example of explicit finite-size effects. *Am. J. Phys.*, 67:1149, 1999.

[271] F. L. Román, J. A. White, A. González, and S. Velasco. Fluctuations in a small hard-disk system: Implicit finite size effects. *J. Chem. Phys.*, 110:9821, 1999.

[272] F. L. Román, J. A. White, A. González, and S. Velasco. *Theory and Simulation of Hard-Sphere Fluids and Related Systems*, chapter Ensemble Effects in Small Systems, pages 343–381. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[273] L. S. Ornstein and F. Zernike. Accidental deviations of density and opalescence at the critical point of a single substance. *Proc. Akad. Sci. (Amsterdam)*, 17:793, 1914.

[274] M. Thol, G. Rutkai, R. Span, J. Vrabec, and R. Lustig. Equation of state for the Lennard-Jones truncated and shifted model fluid. *International Journal of Thermophysics*, 36(1):25–43, 2015.

[275] J. G. Kirkwood and F. P. Buff. The statistical mechanical theory of solutions. i. *J. Chem. Phys.*, 19(6):774–777, 1951.

[276] V. Pierce, M. Kang, M. Aburi, S. Weerasinghe, and P. E. Smith. Recent applications of Kirkwood-Buff theory to biological systems. *Cell Biochem. Biophys.*, 50(1):1–22, 2008.

[277] S. Kjelstrup, S. K. Schnell, T. J. H. Vlugt, J.-M. Simon, A. Bardow, D. Bedeaux, and T. Trinh. Bridging scales with thermodynamics: from nano to macro. *Adv. Nat. Sci.: Nanosci. Nanotechnol.*, 5(2):023002, 2014.

[278] A. Ben-Naim. Theoretical aspects of self-assembly of proteins: A Kirkwood-Buff-theory approach. *J. Chem. Phys.*, 138(22):224906, 2013.

[279] D. Mukherji, C. M. Marques, T. Stuehn, and K. Kremer. Depleted depletion drives polymer swelling in poor solvent mixtures. *Nature Communications*, 8(1):1374, 2017.

[280] R. Cortes-Huerto, K. Kremer, and R. Potestio. Communication: Kirkwood-Buff integrals in the thermodynamic limit from small-sized molecular dynamics simulations. *J. Chem. Phys.*, 145(14):141103, 2016.

[281] S. K. Schnell, X. Liu, M. Simon J, A. Bardow, D. Bedeaux, T. J. H. Vlugt, and S. Kjelstrup. Calculating thermodynamic properties from fluctuations at small scales. *J. Phys. Chem. B*, 115:10911, 2011.

[282] P. Ganguly and N. F. A. van der Vegt. Convergence of sampling Kirkwood-Buff integrals of aqueous solutions with molecular dynamics simulations. *J. Chem. Theory Comput.*, 9(3):1347–1355, 2013.