School of Molecular and Life Sciences

Improving Simulations of Aqueous Systems through Experimental Bias

Tiong Tze Ling

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

August 2020

Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Date: August 18, 2020

Acknowledgements

The author is grateful to Professor Andrew Rohl, Professor Julian Gale, and Associate Professor Paolo Raiteri for their guidance, programming advice, and critical reading of the manuscript. The author acknowledges Cray Inc. for the financial support through the Australian Technology Network (ATN) Industry Doctoral Training Centre (IDTC) program. This work was supported by the resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

Abstract

An atomistic understanding of aqueous solutions is important in many scientific fields such as chemistry and biotechnology. The *ab initio* molecular dynamics (AIMD) simulation of liquid water, a condensed phase system, is typically necessary to model reactive aqueous chemistry that involves the breaking and forming of bonds. However, density functional theory (DFT) that is commonly used for such simulations produces overstructured and slowly diffusing liquid water at ambient temperature. In order to improve these properties, the Boltzmann inversion directed simulation (BIDS) method was developed in this work. This method is able to derive a corrective bias to the system potential, either from a force field or *ab initio* model. The bias acts as an empirical correction that enables routine-level AIMD simulation of DFT water at ambient temperature to achieve comparable liquid structure to experiment. This is accompanied by some improvement to the self-diffusion of water molecules. Furthermore, the negligible increment in computational cost from implementing an empirical correction has no impact on the consideration of the size and time scales that can be explored with AIMD simulations.

Chapter 1 provides an introduction to the significance of water and its anomalous properties, which spurs its interest in various fields of science, including computational chemistry. Some experimentally measured bulk properties used in evaluating the accuracy of water models are presented. These properties are also used to evaluate the bias potential derived from the BIDS method. The types of models that are used to describe water are also discussed.

Chapter 2 explains the fundamental quantum mechanical theories underlying the models used in the AIMD simulation of water, including long-range dispersion-corrected DFT which will be investigated in this work. The variety of empirical models, or force fields, that are parameterised to reproduce the bulk properties of water are also reviewed. The force fields are ideal test cases for investigating the effects of the BIDS biases due to their lower computational cost. The relationship between the pair correlation function, a bulk static property, and atomistic interactions in terms of statistical fluid mechanics is provided. Accordingly, methods that derive pair potential using information from the pair correlation function, which the BIDS method draws from, are examined.

Section 3.1 details the derivation of the bias potential from the pair correlation function using the BIDS method for use in molecular dynamics (MD) programs.

The quality of the selected pair correlation function, the oxygen-oxygen (O-O) radial distribution function (RDF), was optimised by smooth splining and through the use of an appropriate bin size. This reduces the artefacts due to the statistics from simulations that would otherwise be translated to the bias. Details are also given for the post-processing of the bias by extrapolation and padding for use in tabulated form to be read into the MD programs.

Chapter 4 investigates and develops the BIDS methodology using water force fields, instead of the more computationally expensive DFT. However, the statistics produced by the short time scale that is typical of DFT simulations are emulated, in addition to the statistics from the long time scale achieved by force field simulations. The O-O RDF was converged to the target across the iterative scheme and the corresponding changes in the bulk properties; the self-diffusion coefficient and static dielectric constant, were evaluated. The static dielectric constant generally did not respond to the effects of the bias in contrast to the self-diffusion coefficient which generally improved. Provided smooth splining was undertaken, the biases obtained from short time scale simulations were confirmed to converge across iterations to that obtained from long time scale simulations. A cutoff at 7 Å for the bias was selected to avoid the region where the long-range dispersion correction in the later DFT simulations is well-defined, without significantly affecting the quality of the bias. Moreover, the bias can be optimised through a scaling coefficient at each step in the iteration to prevent overcorrection to the O-O RDF.

Chapter 5 investigates the efficacy of the BIDS method by transforming between the variants of known water force fields, *i.e.* rigid and flexible, three-site and four-site. The biases obtained between the rigid and flexible variants were found to be similar. They can be well approximated in the short range with a parameterised Halgren Buffered 45-15 form. The other pair correlation functions, oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H) RDFs, which were not directly biased, were found to have a better fit to that of the target force field. However, the molecular orientation of water molecules is shown to affect that goodness of fit. The corresponding changes in the self-diffusion coefficient transport property due to the changes in the liquid structure from the bias potential were also observed. This can be explained through the hydrogen bond network. For example, a diminished liquid structure indicates a weakened hydrogen bond network resulting in an increase of the self-diffusion.

Chapter 6 compares the results from improving water models using a BIDS bias derived from experimental data to that of similar methods deriving a pair potential using information from the pair correlation function. When the BIDS

method was applied to long-range dispersion-corrected DFT water, the O-O RDF well matched the experimental data, along with improvements to the other pair correlation functions. The self-diffusion coefficient obtained is similar to that in literature using a purely repulsive bias to reproduce the experimental O-O RDF. This gives an idea for the degree of improvement achievable in the self-diffusion of DFT water through a potential bias that targets the liquid structure. In force field simulations, the BIDS method retains the atomistic description of the system in contrast to methods that optimised coarse-grained (CG) potentials to the O-O RDF. The coarse-graining led to a speed up of the self-diffusion coefficient by a factor of 3–5 times, resulting in a value significantly greater than the experiment. Furthermore, valuable insight into the shortcomings of the Lennard-Jones (LJ) potential form can also be derived from the resulting net potential between the BIDS bias and the system potential.

Chapter 7 concludes with a summary of the findings in this thesis. A recommendation was given to explore the optimisation of the bias potential to multiple thermodynamic states in order to improve its transferability. This could be beneficial for the application of the BIDS bias in the AIMD simulation of DFT water for a range of temperatures. Furthermore, the bias can be constrained to satisfy the relationship between the pair potential and the pair correlation function found in statistical fluid mechanics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Water is an abundant resource as it occupies 70–75 % of the surface of the Earth.[1,2] However, the Earth contains only 0.07 % water by mass or 0.4 % by volume, which is smaller in volume than numerous icy bodies in the Solar System.[3] The relative amount of the liquid water of the Earth in comparison to that of the icy moons Europa and Titan of Jupiter and Saturn, respectively, is illustrated in Figure 1.1. The mixing of fluids from deep sea hydrothermal vents with the abundant salt water has been suggested to produced conditions conducive to abiotic synthesis.[4] This underscores the role of water as the fundamental matrix that facilitates the functions of life. So much so that the search for life by National Aeronautics and Space Administration (NASA) is predicated on the search for water.[1] Physiologically, water makes up 70–90 % of the bodies of animals and vegetables.[2] In the human body, water acts as a solvent for transport of nutrients, removal of excretory products, as well as being a major thermal regulator. It is also the driving force for some biological structure, *i.e.* the aversion of hydrocarbons to water is the major stabilising factor for membranes.[1] Arguably no single factor is so responsible for the climate and geological evolution of Earth as is water.[2] Small wonder then that water is deliberately chosen as the reference fluid to define many properties, including specific gravity, heat capacity, *etc.* and to obtain fixed points in thermometry.[2,5] Moreover, many important chemical reactions occur in aqueous solutions, where reacting species are in constant motion and readily interact, in which water is the dissolving medium, or solvent. In the chemical industry, water is employed in many processes for purification, separation, and physical processing.

Water has one of the simplest molecular structures, yet the collective effect of interacting water molecules gives rise to anomalous properties.[7] Water has the highest heat capacity of any liquid, except ammonia, which provides a stabilising effect on temperature. Most solids expand when melting but water expands when

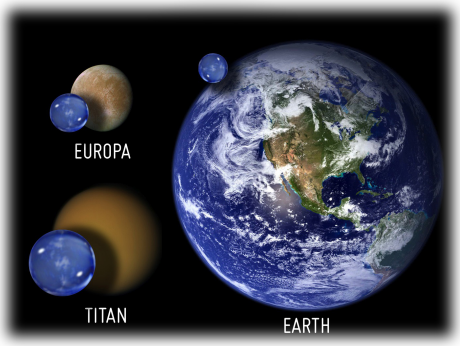


Figure 1.1: Water spheres representing the relative amount of the liquid water volume of the Earth in comparison to that of Europa and Titan. Adapted from the Planetary Habitability Laboratory managed by the University of Puerto Rico at Arecibo with the collaboration of NASA under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA). The data for the water of Earth can be obtained from Shiklomanov's estimate of global water distribution.[6]

freezing. In fact, the maximum density of water occurs above its freezing point at a temperature of 4 °C at atmospheric pressure. Compared to values expected in similar compounds, *e.g.* H_2S, water has a boiling point almost 200 °C higher and a melting point 100 °C higher.[1] Other properties, such as the critical temperature and pressure, heats of vaporisation and fusion, as well as the association constant are unusually high for such a simple composition, whereas the boiling point elevation and freezing-point depression constants are unusually low.[2] The emergence of these anomalies can be attributed to the structural behaviour of water. The water molecules constantly rearrange into an ever-changing tetrahedral network by breaking and forming hydrogen bonds through reorientation of water molecules within an experimentally estimated range of 1–6 ps.[8–11]

Water excels all other solvents in its ability to dissolve a large variety of substances and in its capacity to cause electrolytic dissociation. This is due to its polarity, high dielectric constant (relative permittivity) and small molecular size. The aforementioned rearrangement also occurs when the water hydration pattern adapts to the presence of a solute. As water influences the structure and dynamics of solvates, its structure and dynamics are in turn, affected. These interactions become increasingly complex depending on the solute, from ion pairs to ion clusters and macromolecules such as proteins. These structural fluctuations, in tandem with the large dipole moments of water molecules, high relative permittivity and water-mediated ion pairing give rise to aqueous reactivity. Certain pericyclic reactions, such as Diels-Alder cycloadditions and Claisen rearrangements of hydrophobic compounds, have been noted by Narayan et al.[12] to be accelerated in aqueous suspension.

1.1 Experimental Measurements of Water at Ambient Conditions

Water is at once familiar and ubiquitous yet atypical as a liquid and is still not completely understood. It is then not surprising that water has garnered so much interest to become one of the most studied liquids by experimental and computational chemists alike. Of key interest is the structure of liquid water, which describes the local water molecule packing and more importantly, provides insight into the thermodynamic and kinetic anomalies of water.[13–16] The structural order in water manifests itself as characteristic correlations in space that can be measured in radiation scattering experiments, mostly X-ray[17,18] and neutron.[19] The radial distribution function (RDF) can be determined using the scattered intensity, a long time-averaged scattering profile of the differential scattering cross section per atom,[20] from these measurements. The oxygen-oxygen (O-O) RDF describes the structure of liquid water as the variation in local number density with displacement from a reference oxygen atom and with respect to the average number density of oxygen. Unfortunately, there has been no uniform agreement on the height and profile of the O-O RDF. This is due to the experimental uncertainties, *e.g.* statistical noise and scattering vector range, as well as the required approximations in the techniques, *e.g.* Compton scattering and self-scattering, for extracting the RDF. The resulting O-O RDFs from four X-ray diffraction data sets in liquid water are compared in Figure 1.2. However, there is consensus on the persistence of the tetrahedral ice-like order in the liquid phase from interpretation of the O-O RDFs. Figure 1.3 is provided as reference for a typical O-O RDF. The coordination number of water, determined from the area under the first peak (1st coordination shell) up to the first minimum $\sim$3.4 Å, is between 4 and 5, which suggests tetrahedral coordination. If σ is defined as the O-O distance of 2.76 Å in ice Ih, this is further reinforced by the position of the second peak (2nd coordination shell) at $\sim 1.63\sigma$ that corresponds with the angle between two bonds of $\sim$109.5° in a tetrahedral geometry (Figure 1.4). Fortunately, Soper has attempted to find a consistent RDF using existing experimental data combined with data analysis and computer simulation tools.[21]

The self-diffusion coefficient of liquid water is an important transport property that describes the translational motions of water molecules in the absence of a chemical potential gradient. Early experimental mainstays in determining the coefficient were isotropic tracer techniques, such as the capillary method[24] and the diaphragm cell method.[25] Due to isotopes having minor differences in masses,

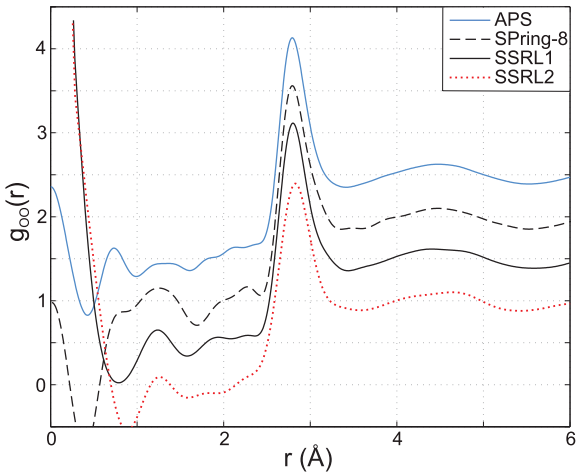


Figure 1.2: Comparison of the O-O RDF derived from four X-ray datasets: advanced photon source (APS), non-Compton discriminating SPring-8, Stanford synchroton radiation lightsource beamlines 7-2 (SSRL1) and 10-2 (SSRL2).[22] The first peak is broader and blunter in the SSRL2 dataset. For clarity, the O-O RDFs are shifted vertically by 0.5 from each other. Reprinted from L. B. Skinner et al. "Benchmark Oxygen-Oxygen Pair-Distribution Function of Ambient Water from X-Ray Diffraction Measurements with a Wide Q-Range". In: *The Journal of Chemical Physics* 138.7 (2013), p. 074506, with the permission of AIP Publishing.

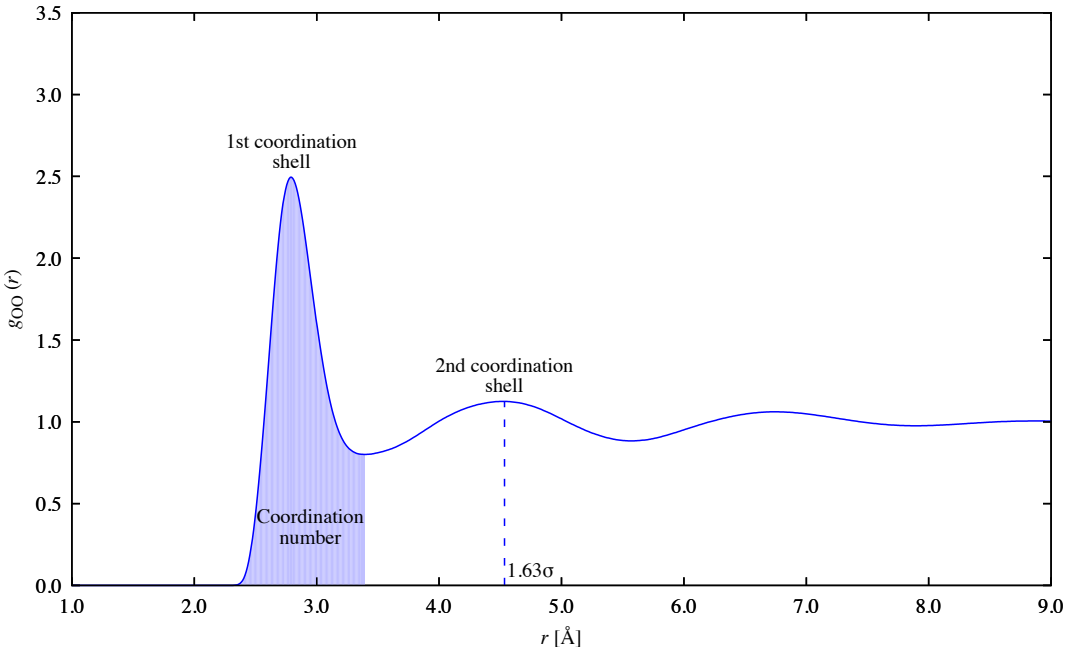


Figure 1.3: Typical O-O RDF of liquid water, where σ is the O-O distance in ice Ih.

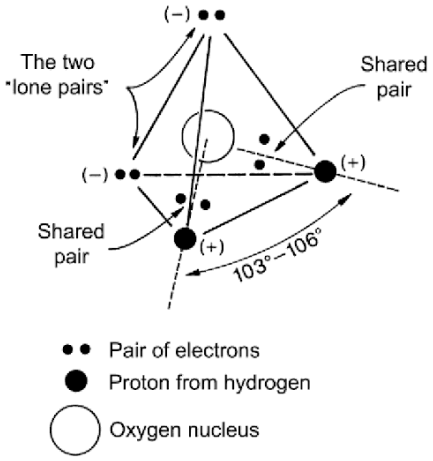


Figure 1.4: Tetrahedral charge structure of a water molecule.[23] Reprinted from M. Kirkham. "Chapter 3 - Structure and Properties of Water". In: *Principles of Soil and Plant Water Relations (Second Edition)*. Ed. by M. Kirkham. Second Edition. Boston: Academic Press, 2014, pp. 27 –40, with permission from Elsevier.

there are slight deviations from the molecular motion in a uniform molecular environment. Longsworth[26] has indicated a small but definite influence of concentration when deuterium oxide was used as the tracer in water. More recently, the nuclear magnetic resonance (NMR) spin-echo technique has increasingly become the method of choice. NMR spectroscopy requires small sample volumes, allows for a wide range in pressure and temperature control, as well as being comparatively fast when taking measurements. Furthermore, there is no special handling required for radioactive samples and it provides measurements that are not influenced by undesirable isotope effects. Using pulsed magnetic field gradient NMR, Holz & Weingärtner[27] obtained $2.31 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$, in good agreement to within error limits $\leq 1\,\%$ for the self-diffusion data of water at 25 °C, with that of Mills[28] from the cell diaphragm method. The self-diffusion coefficient ($2.299 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$) of the latter is taken as the most reliable value for water at this temperature.

Another often investigated property is the temperature-dependent static (zero-frequency) dielectric constant (relative permittivity) of liquid water that correlates with its ability to dissolve salts. Fernández *et al.*[29] compiled and evaluated high-quality data of the dielectric constant of water for a temperature range between 273 and 373 K at ambient pressure. Using a combined data set, they calculated a value of 78.405 ± 0.009 at a nominal temperature of 298.15 K. The data were obtained by either capacitance measurements[30–32] or resonance techniques.[33] The capacitance measurements are made in the audio frequency range and require water samples to have low ionic concentrations to prevent electrode polarisation. The high intrinsic conductivity of water at the upper temperature

range can affect the measurement, which is sensitive to conductance. The resonance techniques involve an electromagnetic resonator connected to a source of electromagnetic radiation. Modes of oscillation in the radio frequency and microwave ranges are used for the dielectric measurements of water.

1.2 Role of Water in Simulations of Aqueous Systems

So important are the interactions provided by water that computational chemists have spent significant resources to simulate the aqueous environment. In 1986, Levitt spent half a million dollars simulating an amount of water that would scarcely wet the point of a pin.[34] It was to pay for around two weeks of processing time on a then state-of-the-art supercomputer to model how explicit water molecules affect the structure and dynamics of a relatively small protein, bovine pancreatic trypsin inhibitor (BPTI). The simulation demonstrated that the previous *in vacuo* model was a poor predictor of the real world behaviour of BPTI and thus established the necessity to model biological molecules in their native environment surrounded by water molecules. This established the importance of including water in simulations, as the interactions between the object of interest and the surrounding water are necessary to model reality.

Simulations have played a significant role in gaining a better understanding of water as a liquid and as a solvent in numerous applications. Density functional theory (DFT) has been used to describe water structure close to certain metal surfaces, which is important in electrochemistry, catalysis and the study of corrosion.[35] The lattice Boltzmann method[36] or dissipative particle dynamics (DPD)[37] can give qualitative insight into hydrodynamic effects despite being coarse-grained. Empirical force fields fitted to quantum mechanical data have been used to show the dynamic behaviour of prenucleation clusters in an aqueous calcium carbonate system.[38] The prediction of energetics in aqueous systems from simulation is another area of interest in organic and pharmaceutical chemistry. Molecular dynamics (MD) simulations have suggested that water strongly affects the binding conformation of an organic molecule at the calcite crystal surface.[39] Accurate calculation of the solvation free energies is required in determining fundamental properties such as solubility and partitioning for preliminary drug design screening.[40] The need to achieve so-called "chemical accuracy" for hydration free energies required in rational drug design is one motivation of studies for more accurate, yet computationally facile, water models.

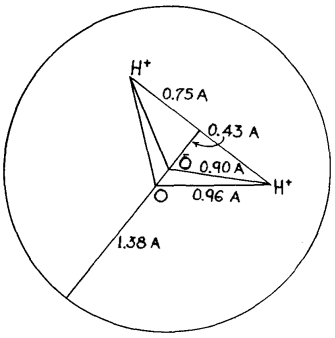


Figure 1.5: The 1933 Bernal & Fowler water molecule model[41] showed remarkable insight into the structure of the water molecule. $\bar{O}$ is the proposed position of the negative charge of the oxygen atom, which was later found to be correct. Reprinted from J. D. Bernal and R. H. Fowler. "A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions". In: *The Journal of Chemical Physics* 1.8 (1933), pp. 515–548, with the permission of AIP Publishing.

1.3 Modelling Water at Ambient Conditions

Any water model must be able to accurately describe the three "types" of water in an aqueous environment; isolated water, bulk liquid water, and water interacting with a solute. However, sufficiently describing the physics of these interactions to achieve a common water model capable of producing the breadth of experimental thermodynamic properties has proven to be a challenge. The water-water interaction and the hydrogen bonding networks in the liquid phase are sensitive to various model descriptions and parameters. This has led to the development of numerous water models with varying degrees of complexity. Unfortunately, the accuracy of the model does not necessarily correlate with its complexity.

In the empirical models of water, force fields are parameterised to fit one or more thermodynamic or other physical properties, including RDF, density, heat of vaporisation and dipole moment. The development of an empirical water model dates back to the seminal paper by Bernal and Fowler in 1933, which reproduced the main features of an early X-ray diffraction pattern of water.[41] Based on the argument of a near-tetrahedral arrangement of the charges in a water molecule with two bonds and two lone pairs, it was proposed that a single water molecule formed a similar arrangement in its coordination with four neighbours. This was likened to the structure of silicates. Although it was understood that liquid water was a disordered structure, the silica analogy allowed the use of Bragg's law to interpret the X-ray diffraction pattern of water. The resultant Bernal-Fowler (BF) water model was essentially a disordered quartz-like structural model. In fact, the model of the water molecule itself (Figure 1.5) is mimicked in some of the modern molecular models, *e.g.* the geometry of the TIP4P model.[42]

Modern empirical water models are based on the widely accepted view that the hydrogen bond is the result of competing interplay between the long-range attractive Coulomb and short-range repulsive energies.[43] The repulsion between molecules is mostly described by a Lennard-Jones (LJ) potential centred on the oxygen. Incorporated in this potential is the dispersion energy that along with classical electrostatic interactions describe the attraction between molecules. The empirical water models can generally be classified based on: (i) rigid or flexible geometry, (ii) number of sites, and (iii) polarisation effects. Several of the currently used models were developed in the 1980s. The SPC model was first introduced in 1981[44] followed by the TIP3P model published in 1983.[45] These three-site models have rigid bonds and angles with point charges on the oxygen and hydrogen atoms to describe polarity. Extensions have since been developed to provide a more comprehensive physical description of water, *e.g.* SPC/E accounts for the averaged polarisation effect. The inclusion of polarisation effects, *i.e.* the electron density redistribution due to the electric field of other molecules,[46] better reproduces the molecular dipole and by extension, the hydrogen-bonding network and the interaction energies. This also allows for a simultaneous description of both liquid and vapour phase electrostatics due to the induction of higher dipoles in the condensed phase over that of the gas phase. Additionally, fictitious charge sites can be located in the plane of the molecular frame, *e.g.* TIP4P. Great effort has gone into improving empirical water models and yet they are unable to simultaneously reproduce all the physical properties and lack transferability. Their continued development will likely remain active moving forward. A comprehensive review of water models up to 2002 can be found in the works by Guillot and Guissani[47,48] as well as the more recent Vega *et al.*[49] with a focus on rigid, non-polarisable water models. Any empirical models investigated as part of this study will be covered in more detail in a later chapter.

In contrast with the increasingly realistic water models, alternative water models that are very simple have also emerged and are capable of predicting certain features of the properties of water. The Nezbeda model describes water as a fused-hard-sphere body, *i.e.* a spherical segment.[50] It uses an attractive square-well potential to model hydrogen-bonding interactions and a hard sphere potential for all other interactions. It is able to quantitatively describe the P-V-T phase behaviour of water while being computationally efficient. However, it is not suitable for application in MD due to the discontinuous interaction potential. The Mercedes-Benz model is a planar LJ disk with three orientation-dependent hydrogen-bonding arms.[51] It can qualitatively describe a number of anomalous properties of water. A three-dimensional variant has been developed, but the

orientation-based hydrogen-bonding potential makes it difficult to implement.[52] The mW model coarse-grains water to a tetrahedrally coordinated monoatomic system with only short-ranged interactions,[53] similar to silicon. It exploits a reduced Stillinger-Weber potential for silicon,[54] where the energy scale, density scale and the nonbond angular-dependent term enforcing the tetrahedral topology were tuned for water. The model has demonstrated remarkable success in reproducing the energetics, density and structure of liquid water, its anomalies and phase transitions at an extremely low computational cost.

A more rigorous water description can be obtained through a first principles or *ab initio* approach. *Ab initio* methods use quantum mechanics (QM) to describe atoms and explicitly enable breaking and forming of chemical bonds, cooperative polarisation effects, proton transfer, *etc.* This quantum mechanical description can be provided by wavefunction theory that can be improved systematically or DFT that heavily depends on the approximation of its density functional. The use of high-level wavefunction theory for liquid water has been challenging and is mostly limited to Møller-Plesset second-order perturbation theory (MP2). In fact, successful *ab initio* molecular dynamics (AIMD) simulations of liquid water with MP2 have required the employment of stochastic methods[55,56] or fragment-based methods.[57,58] It is so computationally demanding that only a few applications in other condensed phase systems have been reported.[59–61] Nonetheless, MP2 is one of the best methods for the WATER27 benchmark due to a good description of hydrogen bonding in water clusters and the inclusion of non-empirical long-range dispersion interactions.[62] Introduced by Bryantsev *et al.*[63] the WATER27 benchmark consists of ten neutral structures of $(H_2O)_{n=2-8}$, four isomers of $(H_2O)_{20}$, five protonated water clusters $(H_3O)^+(H_2O)_{n=1-3,6}$, seven hydrated clusters $OH^-(H_2O)_{n=1-6}$ and one hydroxonium-hydroxide zwitterion $(H_3O)^+(H_2O)_6OH^-$. All the geometries were optimised using B3LYP/6-311++G (2d,2p). For the binding energies of water clusters, they used the MP2/complete basis set energies corrected with higher order correlations at the CCSD(T)/aug-cc-pVDZ level on all isomers except $(H_2O)_{20}$. Furthermore, MP2 has achieved results for water clusters comparable with the higher-level coupled cluster theory.[64–67]

The bulk of *ab initio* simulations of liquids and solutions, such as water, have employed DFT. It strikes a good balance between accuracy and computational cost. DFT has been applied to study water in various states. Common density functionals usually give satisfactory binding energies for the water dimer[68–71] and the common form of ice.[72,73] AIMD investigations of liquid water have been found to reproduce the radial positions of the main features in the liquid struc-

ture that agree with experiment.[74–76] However, DFT does encounter difficulties in describing water systems. Commonly used density functionals lead to an over-structuring of water and at room temperature, water behaves as a glass with a lower self-diffusion coefficient than observed experimentally.[77,78] DFT water also has melting point well above the correct point.[79,80] In order to provide a more satisfactory density functional approximation, the accurate description of the long-range dispersion is important. Its addition softens liquid structure and increases diffusivity. Moreover, the deficiencies of DFT water structure and dynamics are reflected in the excess proton diffusion. Common density functionals underestimate the proton transport barrier due to the artificial stabilisation of delocalised states induced by the self-interaction error (SIE). The inclusion of Hartree-Fock (HF) exchange via a hybrid of wavefunction theory and DFT can partially correct this problem. Proton transport via the Grotthuss mechanism and acid-base reactions are thus difficult to describe by pure DFT.[81,82]

The quality of the *ab initio* water is dependent on the level of theory and basis set for a description of the potential energy surface. It also depends on a good description of the long-range dispersion and the fundamental nature of the hydrogen bond. Due to the light mass of the hydrogen atoms, a quantum mechanical treatment of its nuclear dynamics is likely required. However, the atomic nuclei are treated as classical particles in most AIMD simulations of water. Such an approximation is insufficient as hydrogen atoms in liquid water deviate significantly from classical behaviour at ambient temperature.[83–85] Furthermore, the neglect of nuclear quantum effects (NQE) also contribute to the previously mentioned over-structuring of DFT water.[86–88] This is supported by experimental studies of D_2O that show softening of the liquid structure.[89] A crude approximation of the NQE can be achieved by a simple 30 K increase to the simulated temperature, which was investigated by DiStasio *et al.*[90] This treatment is suggested by the lowest-order perturbative expansion in $\hbar$ of the free energy, where the momentum distribution remains Gaussian but with an increased temperature for each particle.[91] Although the NQE approximation leads to better agreement in the O-O RDF, it does not reproduce the width and intensity of the O-H RDF. The proper inclusion of NQE can be performed via the Feynman discretised path integral scheme. With recent algorithmic developments,[92] the additional cost is said to be mitigated. Empirical models can give insight into other important physics that need to be captured in *ab initio* water. The parameterised many-body potential model, MB-Pol, established that accurate description of the three-body energy expansions and the many-body polarisation effects are required to simultaneously reproduce various thermodynamic properties of water.[93,94]

1.4 Improving Water Models using a Boltzmann Inverted Bias

Computing power has advanced leaps and bounds since Levitt's simulation of BPTI. The same BPTI in an aqueous environment can currently be modeled on a personal computer in a couple of days. Computational chemists are however constantly running up against the computational limit as more complex systems and high-level theories are pursued to achieve an increasingly realistic description at the atomistic scale. The current gold-standard in quantum chemistry is CCSD(T), which scales prohibitively at the seventh power with the size of the system. It has been applied to water clusters up to the 26-mers.[95] Moreover, the previously unrelenting march of Moore's law is stagnating.[96] Without a breakthrough from conventional computers to quantum computers, the computational processing power for the performance metrics required to ensure statistical convergence of such demanding systems are elusive. Hence, there is a desire to eke out more from the current computational resources that are available. One way is to directly apply an auxiliary coupled "knowledge field", which contains information from experimental data, as an empirical correction to an approximate and computationally viable model.[97] By virtue of being empirical, there should be no additional computational costs. We turn to the unique correspondence that exists between a RDF and its pair potential, as per Henderson's theorem,[98] in this study. It reasons there necessarily exists an effective bias from the difference between the potentials of two RDFs that maps one to the other. Therefore, the "knowledge field" can be found from the experimental scattering data in order for simulated water to achieve comparable results, at least structurally, with experiment.

Johnson & March[99] noted that any approximate theory which relates the structure factor of a liquid to the pair potential can be inverted to give an empirical potential provided that the relationship in the theory between the liquid structure and the potential is sufficiently transparent.[100] Here, we use the structure of water from the previously mentioned Soper[21] that was refined against a set of scattering data from both X-ray and neutron experiments. These approximate theories, such as hypernetted-chain (HNC)[101] and Percus-Yevick (PY),[102] provide the closures necessary to complete the Ornstein-Zernike (OZ) integral equation approach.[103] Madden & Rice[100] inverted structure factors using various closures to produce meaningful estimations of the pair potentials. The structure factors used were the Fourier transformed experimental X-ray diffraction data of

liquid argon, sodium and potassium. They emphasised the requirement of high quality diffraction data with consideration of low angle scattering, form factors, normalisation and correction for multiple scattering. This was followed by Zerah & Hansen[104] who reproduced the shape of a known LJ potential with reasonable success from the calculated RDF with a thermodynamically self-consistent mixed integral equation that interpolates continuously between the HNC and the soft-core mean spherical approximation (SMSA) closures. However, the well depth was found to be underestimated by 15%. The application of inversion schemes using the OZ integral equation has been extended to other interesting systems. Rajagopalan & Srinivasa Rao[105] presented the predictor-corrector inversion scheme for the interactions in charged colloidal particles. A more recent example using the predictor-corrector inversion scheme is the study of the density-dependence of the interaction potential between colloidal particles and an oil-water interface.[106]

The inverse treatment can also be applied to correlation functions that obey a Boltzmann distribution, which Reith *et al.*[107] demonstrated as capable of successfully deriving coarse-grained (CG) interaction potentials mapped to atomistic models using polyisoprene solution and melt RDFs. The procedure is performed iteratively and is hence called iterative Boltzmann inversion (IBI). It has been used to reduce the degrees of freedom (DoFs) for various polymers to increase the length- and time-scales allowed in their simulations.[108–110] This is primarily due to the short-chain CG potentials being transferable to long-chains. In the case of an alternating copolymer, Agrawal *et al.*[111] proposed that simultaneous, rather than sequential, optimisation of the pair potentials is arguably better for such heterogeneous systems. However, the potentials produced are heavily dependent on the given thermodynamic state. Moore *et al.*[112] addressed this by developing a multi-state extension that finds a less-state dependent potential within the region of three overlapping phase spaces at various thermodynamic states. The improved method was demonstrated as being able to reproduce the known potential of a monoatomic LJ fluid. They further demonstrated the efficacy of this multistate iterative Boltzmann inversion (MS IBI) using propane and *n*-dodecance.

The IBI method is evidently a powerful tool used to find effective potentials from RDFs in the coarse-graining of polymers and has even been applied to water. It was found that the four-water model is the optimal number of water molecules mapped to a single bead.[5] Subsequently, a CG water force field was developed for the four-water bead model using MS IBI.[113] The effective potential derived from Boltzmann inversion is empirical, much like a LJ potential in empirical models or the C_6 based dispersion correction in *ab initio* models and can thus be applied routinely with only minimal additional computational cost. We therefore

investigate the use of a Boltzmann inverted bias in improving the structure and dynamics of simulated liquid water at ambient temperature. The performance will be evaluated based on improvements to the correlation functions, diffusion coefficient and dielectric constant. We start by using empirical water models to compare the resulting bias and its effects given statistics from different time length simulations, before applying it to our intended *ab initio* water model.

References

[1] W. Stillwell. "Chapter 3 - Water and the Hydrophobic Effect". In: *An Introduction to Biological Membranes*. Ed. by W. Stillwell. San Diego: Elsevier, 2013, pp. 29 –41. DOI: 10.1016/B978-0-444-52153-8.00003-9.

[2] W. C. Fernelius. "An Ammonia World". In: *Journal of Chemical Education* 8.1 (1931), p. 55. DOI: 10.1021/ed008p55.

[3] V. Kotwicki. "Water in the Universe". In: *Hydrological Sciences Journal* 36.1 (1991), pp. 49–66. DOI: 10.1080/02626669109492484.

[4] F. L. Sousa et al. "Early Bioenergetic Evolution". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 368.1622 (2013). DOI: 10.1098/rstb.2013.0088.

[5] K. R. Hadley and C. McCabe. "On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties". In: *The Journal of Physical Chemistry B* 114.13 (2010). PMID: 20230012, pp. 4590–4599. DOI: 10.1021/jp911894a.

[6] P. H. Gleick. *Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford University Press, 1993.

[7] A. Nilsson and L. G. M. Pettersson. "The Structural Origin of Anomalous Properties of Liquid Water". In: *Nature Communications* 6.8998 (2015).

[8] F. N. Keutsch et al. "Hydrogen Bond Breaking Dynamics of the Water Trimer in the Translational and Librational Band Region of Liquid Water". In: *Journal of the American Chemical Society* 123.25 (2001). PMID: 11414826, pp. 5938–5941. DOI: 10.1021/ja003683r.

[9] M. D. Fayer. *Ultrafast Infrared Vibrational Spectroscopy*. EBL-Schweitzer. CRC Press, 2013.

[10] J. B. Asbury et al. "Dynamics of Water Probed with Vibrational Echo Correlation Spectroscopy". In: *The Journal of Chemical Physics* 121.24 (2004), pp. 12431–12446. DOI: 10.1063/1.1818107.

[11] J. B. Asbury et al. "Water Dynamics: Vibrational Echo Correlation Spectroscopy and Comparison to Molecular Dynamics Simulations". In: *The Journal of Physical Chemistry A* 108.7 (2004), pp. 1107–1119. DOI: 10.1021/jp036266k.

[12] S. Narayan et al. ""On Water": Unique Reactivity of Organic Compounds in Aqueous Suspension". In: *Angewandte Chemie International Edition* 44.21 (2005), pp. 3275–3279. DOI: 10.1002/anie.200462883.

[13] J. R. Errington and P. G. Debenedetti. "Relationship between Structural Order and the Anomalies of Liquid Water". In: *Nature* 409 (Jan. 2001), p. 318.

[14] Z. Yan et al. "Structure of the First- and Second-Neighbor Shells of Simulated Water: Quantitative Relation to Translational and Orientational Order". In: *Phys. Rev. E* 76 (5 2007), p. 051201. DOI: 10.1103/PhysRevE.76.051201.

[15] R. M. Lynden-Bell and P. G. Debenedetti. "Computational Investigation of Order, Structure, and Dynamics in Modified Water Models". In: *The Journal of Physical Chemistry B* 109.14 (2005). PMID: 16851733, pp. 6527–6534. DOI: 10.1021/jp0458553.

[16] T. Head-Gordon and S. W. Rick. "Consequences of Chain Networks on Thermodynamic, Dielectric and Structural Properties for Liquid Water". In: *Phys. Chem. Chem. Phys.* 9 (1 2007), pp. 83–91. DOI: 10.1039/B614742A.

[17] J. M. Sorenson et al. "What can X-ray Scattering Tell Us about the Radial Distribution Functions of Water?" In: *The Journal of Chemical Physics* 113.20 (2000), pp. 9149–9161. DOI: 10.1063/1.1319615.

[18] G. Hura et al. "A High-Quality X-ray Scattering Experiment on Liquid Water at Ambient Conditions". In: *The Journal of Chemical Physics* 113.20 (2000), pp. 9140–9148. DOI: 10.1063/1.1319614.

[19] A. Soper. "The Radial Distribution Functions of Water and Ice from 220 to 673 K and at Pressures up to 400 MPa". In: *Chemical Physics* 258.2 (2000), pp. 121 –137. DOI: 10.1016/S0301-0104(00)00179-8.

[20] "P. A. Egelstaff: An Introduction to the Liquid State, Second Edition, Vol. 7, Aus der Reihe: Oxford Series on Neutron Scattering in Condensed Matter. Clarendon Press, Oxford, ISBN 0-19-851012-8, 1992". In: *Berichte der Bunsengesellschaft für physikalische Chemie* 96.12 (1992), pp. 1898–1898. DOI: 10.1002/bbpc.19920961238.

[21] A. K. Soper. "The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure?" In: *ISRN Physical Chemistry* 2013.279463 (2013), p. 67. DOI: 10.1155/2013/279463.

[22] L. B. Skinner et al. "Benchmark Oxygen-Oxygen Pair-Distribution Function of Ambient Water from X-Ray Diffraction Measurements with a Wide Q-Range". In: *The Journal of Chemical Physics* 138.7 (2013), p. 074506. DOI: 10.1063/1.4790861.

[23] M. Kirkham. "Chapter 3 - Structure and Properties of Water". In: *Principles of Soil and Plant Water Relations (Second Edition)*. Ed. by M. Kirkham. Second Edition. Boston: Academic Press, 2014, pp. 27 –40. DOI: 10.1016/B978-0-12-420022-7.00003-3.

[24] T. Graham. "XVII. On the Molecular Mobility of Gases". In: *Philosophical Transactions of the Royal Society of London* 153 (1863), pp. 385–405. DOI: 10.1098/rstl.1863.0017.

[25] J. H. Northrop and M. L. Anson. "A Method for the Determination of Diffusion Constants and the Calculation of the Radius and Weight of the Hemoglobin Molecule". In: *The Journal of General Physiology* 12.4 (1929), pp. 543–554. DOI: 10.1085/jgp.12.4.543.

[26] L. G. Longsworth. "Temperature Dependence of Diffusion in Aqueous Solutions". In: *The Journal of Physical Chemistry* 58.9 (1954), pp. 770–773. DOI: 10.1021/j150519a017.

[27] M. Holz and H. Weingärtner. "Calibration in Accurate Spin-Echo Self-Diffusion Measurements using ^{1}H and Less-Common Nuclei". In: *Journal of Magnetic Resonance (1969)* 92.1 (1991), pp. 115 –125. DOI: 10.1016/0022-2364(91)90252-O.

[28] R. Mills. "Self-Diffusion in Normal and Heavy Water in the Range 1-45°". In: *The Journal of Physical Chemistry* 77.5 (1973), pp. 685–688. DOI: 10.1021/j100624a025.

[29] D. P. Fernández, A. R. H. Goodwin, and J. M. H. L. Sengers. "Measurements of the Relative Permittivity of Liquid Water at Frequencies in the Range of 0.1 to 10 kHz and at Temperatures between 273.1 and 373.2 K at Ambient Pressure". In: *International Journal of Thermophysics* 16.4 (1995), pp. 929–955. DOI: 10.1007/BF02093474.

[30] L. A. Dunn and R. H. Stokes. "Pressure and Temperature Dependence of the Electrical Permittivities of Formamide and Water". In: *Trans. Faraday Soc.* 65 (0 1969), pp. 2906–2912. DOI: 10.1039/TF9696502906.

[31] G. A. Vidulich, D. F. Evans, and R. L. Kay. "The Dielectric Constant of Water and Heavy Water between 0 and 40°". In: *The Journal of Physical Chemistry* 71.3 (1967), pp. 656–662. DOI: 10.1021/j100862a028.

[32] R. Deul and E. U. Franck. "The Static Dielectric Constant of the Water-Benzene Mixture System to 400 °C and 2800 bar". In: *Berichte der Bunsengesellschaft für physikalische Chemie* 95.8 (1991), pp. 847–853. DOI: 10.1002/bbpc.19910950801.

[33] B. B. Owen et al. "The Dielectric Constant of Water as a Function of Temperature and Pressure". In: *The Journal of Physical Chemistry* 65.11 (1961), pp. 2065–2070. DOI: 10.1021/j100828a035.

[34] M. Gerstein and M. Levitt. "Simulating Water and the Molecules of Life". In: *Scientific American* 279.5 (1998), pp. 100–105.

[35] D. Sebastiani and L. Delle Site. "Adsorption of Water Molecules on Flat and Stepped Nickel Surfaces from First Principles". In: *Journal of Chemical Theory and Computation* 1.1 (2005). PMID: 26641118, pp. 78–82. DOI: 10.1021/ct049955o.

[36] B. Dünweg and A. J. C. Ladd. "Lattice Boltzmann Simulations of Soft Matter Systems". In: *Advanced Computer Simulation Approaches for Soft Matter Sciences III*. Ed. by C. Holm and K. Kremer. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 89–166. DOI: 10.1007/978-3-540-87706-6_2.

[37] P. J. Hoogerbrugge and J. M. V. A. Koelman. "Simulating Microscopic Hydrodynamic Phenomena with Dissipative Particle Dynamics". In: *EPL (Europhysics Letters)* 19.3 (1992), p. 155.

[38] R. Demichelis et al. "Stable Prenucleation Mineral Clusters are Liquid-Like Ionic Polymers". In: *Nature Communications* 2.590 (Dec. 2011).

[39] H. Nada. "Importance of Water in the Control of Calcite Crystal Growth by Organic Molecules". In: *Polymer Journal* 47.84 (Oct. 2014).

[40] J. P. M. Jämbeck and A. P. Lyubartsev. "Update to the General Amber Force Field for Small Solutes with an Emphasis on Free Energies of Hydration". In: *The Journal of Physical Chemistry B* 118.14 (2014). PMID: 24684585, pp. 3793–3804. DOI: 10.1021/jp4111234.

[41] J. D. Bernal and R. H. Fowler. "A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions". In: *The Journal of Chemical Physics* 1.8 (1933), pp. 515–548. DOI: 10.1063/1.1749327.

[42] J. L. Finney. "Bernal and the Structure of Water". In: *Journal of Physics: Conference Series* 57.1 (2007), p. 40.

[43] M. S. Gordon and J. H. Jensen. "Understanding the Hydrogen Bond Using Quantum Chemistry". In: *Accounts of Chemical Research* 29.11 (1996), pp. 536–543. DOI: 10.1021/ar9600594.

[44] H. J. C. Berendsen et al. "Interaction Models for Water in Relation to Protein Hydration". In: *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*. Ed. by B. Pullman. Dordrecht: Springer Netherlands, 1981, pp. 331–342. DOI: 10.1007/978-94-015-7658-1_21.

[45] W. L. Jorgensen et al. "Comparison of Simple Potential Functions for Simulating Liquid Water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: 10.1063/1.445869.

[46] S. W. Rick and S. J. Stuart. "Potentials and Algorithms for Incorporating Polarizability in Computer Simulations". In: *Reviews in Computational Chemistry*. Wiley-Blackwell, 2003. Chap. 3, pp. 89–146. DOI: 10.1002/0471433519.ch3.

[47] B. Guillot. "A Reappraisal of What We Have Learnt During Three Decades of Computer Simulations on Water". In: *Journal of Molecular Liquids* 101.1 (2002). Molecular Liquids. Water at the New Millenium, pp. 219 –260. DOI: 10.1016/S0167-7322(02)00094-6.

[48] B. Guillot and Y. Guissani. "How to Build a Better Pair Potential for Water". In: *The Journal of Chemical Physics* 114.15 (2001), pp. 6720–6733. DOI: 10.1063/1.1356002.

[49] C. Vega et al. "What Ice can Teach us about Water Interactions: A Critical Comparison of the Performance of Different Water Models". In: *Faraday Discuss.* 141 (0 2009), pp. 251–276. DOI: 10.1039/B805531A.

[50] I. Nezbeda and L. Vlček. "Thermophysical Properties of Fluids: From Realistic to Simple Models and their Applications". In: *International Journal of Thermophysics* 25.4 (2004), pp. 1037–1049. DOI: 10.1023/B:IJOT.0000038498.47700.3f.

[51] K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill. "A Simple Model of Water and the Hydrophobic Effect". In: *Journal of the American Chemical Society* 120.13 (1998), pp. 3166–3175. DOI: 10.1021/ja973029k.

[52] A. Bizjak, V. Vlachy, and T. Urbic. "Phase Diagram of the Lennard-Jones System of Particles from the Cell Model and Thermodynamic Perturbation Theory". In: *Acta Chimica Slovenica*. Vol. 56. Nov. 2008.

[53] V. Molinero and E. B. Moore. "Water Modeled As an Intermediate Element between Carbon and Silicon". In: *The Journal of Physical Chemistry B* 113.13 (2009). PMID: 18956896, pp. 4008–4016. DOI: 10.1021/jp805227c.

[54] F. H. Stillinger and T. A. Weber. "Computer Simulation of Local Order in Condensed Phases of Silicon". In: *Phys. Rev. B* 31 (8 1985), pp. 5262–5271. DOI: 10.1103/PhysRevB.31.5262.

[55] M. Del Ben et al. "Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory". In: *The Journal of Physical Chemistry Letters* 4.21 (2013), pp. 3753–3759. DOI: 10.1021/jz401931f.

[56] M. Del Ben, J. Hutter, and J. VandeVondele. "Probing the Structural and Dynamical Properties of Liquid Water with Models including Non-local Electron Correlation". In: *The Journal of Chemical Physics* 143.5 (2015), p. 054506. DOI: 10.1063/1.4927325.

[57] J. Liu, X. He, and J. Z. H. Zhang. "Structure of Liquid Water - A Dynamical Mixture of Tetrahedral and 'Ring-and-Chain' Like Structures". In: *Phys. Chem. Chem. Phys.* 19 (19 2017), pp. 11931–11936. DOI: 10.1039/C7CP00667E.

[58] S. Y. Willow et al. "Ab Initio Molecular Dynamics of Liquid Water using Embedded-Fragment Second-Order Many-Body Perturbation Theory Towards its Accurate Property Prediction". In: *Scientific Reports* 5.14358 (Sept. 2015).

[59] A. Grüneis, M. Marsman, and G. Kresse. "Second-Order Møller-Plesset Perturbation Theory Applied to Extended Systems. II. Structural and Energetic Properties". In: *The Journal of Chemical Physics* 133.7 (2010), p. 074107. DOI: 10.1063/1.3466765.

[60] A. Erba et al. "A Post-Hartree-Fock Study of Pressure-Induced Phase Transitions in Solid Nitrogen: The Case of the α, γ, and ϵ Low-Pressure Phases". In: *The Journal of Chemical Physics* 134.7 (2011), p. 074502. DOI: 10.1063/1.3553206.

[61] F. Göltl et al. "Van der Waals Interactions between Hydrocarbon Molecules and Zeolites: Periodic Calculations at Different Levels of Theory, from Density Functional Theory to the Random Phase Approximation and Møller-Plesset Perturbation Theory". In: *The Journal of Chemical Physics* 137.11 (2012), p. 114111. DOI: 10.1063/1.4750979.

[62] L. Goerigk and S. Grimme. "A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions". In: *Phys. Chem. Chem. Phys.* 13 (14 2011), pp. 6670–6688. DOI: 10.1039/C0CP02984J.

[63] V. S. Bryantsev et al. "Evaluation of B3LYP, X3LYP, and M06-Class Density Functionals for Predicting the Binding Energies of Neutral, Protonated, and Deprotonated Water Clusters". In: *Journal of Chemical Theory and Computation* 5.4 (2009). PMID: 26609610, pp. 1016–1026. DOI: 10.1021/ct800549f.

[64] S. S. Xantheas, C. J. Burnham, and R. J. Harrison. "Development of Transferable Interaction Models for Water. II. Accurate Energetics of the First Few Water Clusters from First Principles". In: *The Journal of Chemical Physics* 116.4 (2002), pp. 1493–1499. DOI: 10.1063/1.1423941.

[65] S. S. Xantheas and E. Aprà. "The Binding Energies of the D_{2d} and S_4 Water Octamer Isomers: High-Level Electronic Structure and Empirical Potential Results". In: *The Journal of Chemical Physics* 120.2 (2004), pp. 823–828. DOI: 10.1063/1.1626624.

[66] I. Shin et al. "Structure and Spectral Features of $H^+(H_2O)_7$: Eigen versus Zundel Forms". In: *The Journal of Chemical Physics* 125.23 (2006), p. 234305. DOI: 10.1063/1.2404659.

[67] E. E. Dahlke et al. "Assessment of the Accuracy of Density Functionals for Prediction of Relative Energies and Geometries of Low-Lying Isomers of Water Hexamers". In: *The Journal of Physical Chemistry A* 112.17 (2008). PMID: 18393474, pp. 3976–3984. DOI: 10.1021/jp077376k.

[68] F. Sim et al. "Gaussian Density Functional Calculations on Hydrogen-Bonded Systems". In: *Journal of the American Chemical Society* 114.11 (1992), pp. 4391–4400. DOI: 10.1021/ja00037a055.

[69] K. Laasonen, F. Csajka, and M. Parrinello. "Water Dimer Properties in the Gradient-Corrected Density Functional Theory". In: *Chemical Physics Letters* 194.3 (1992), pp. 172 –174. DOI: 10.1016/0009-2614(92)85529-J.

[70] K. Laasonen et al. "Structures of Small Water Clusters using Gradient-Corrected Density Functional Theory". In: *Chemical Physics Letters* 207.2 (1993), pp. 208 –213. DOI: 10.1016/0009-2614(93)87016-V.

[71] B. Santra, A. Michaelides, and M. Scheffler. "On the Accuracy of Density-Functional Theory Exchange-Correlation Functionals for H Bonds in Small Water Clusters: Benchmarks Approaching the Complete Basis Set Limit". In: *The Journal of Chemical Physics* 127.18 (2007), p. 184104. DOI: 10.1063/1.2790009.

[72] D. R. Hamann. "H_2O Hydrogen Bonding in Density-Functional Theory". In: *Phys. Rev. B* 55 (16 1997), R10157–R10160. DOI: 10.1103/PhysRevB.55.R10157.

[73] P. J. Feibelman. "Lattice Match in Density Functional Calculations: Ice Ih*vs.*β-AgI". In: *Phys. Chem. Chem. Phys.* 10 (32 2008), pp. 4688–4691. DOI: 10.1039/B808482N.

[74] K. Laasonen et al. ""Ab Initio" Liquid Water". In: *The Journal of Chemical Physics* 99.11 (1993), pp. 9080–9089. DOI: 10.1063/1.465574.

[75] M. Sprik, J. Hutter, and M. Parrinello. "Ab Initio Molecular Dynamics Simulation of Liquid Water: Comparison of Three Gradient-Corrected Density Functionals". In: *The Journal of Chemical Physics* 105.3 (1996), pp. 1142–1152. DOI: 10.1063/1.471957.

[76] P. L. Silvestrelli and M. Parrinello. "Structural, Electronic, and Bonding Properties of Liquid Water from First Principles". In: *The Journal of Chemical Physics* 111.8 (1999), pp. 3572–3580. DOI: 10.1063/1.479638.

[77] Y. S. Tse, C. Knight, and G. A. Voth. "An Analysis of Hydrated Proton Diffusion in *ab initio* Molecular Dynamics". In: *The Journal of Chemical Physics* 142.1 (2015), p. 014104. DOI: 10.1063/1.4905077.

[78] J. C. Grossman et al. "Towards an Assessment of the Accuracy of Density Functional Theory for First Principles Simulations of Water". In: *The Journal of Chemical Physics* 120.1 (2004), pp. 300–311. DOI: 10.1063/1.1630560.

[79] S. Yoo and S. S. Xantheas. "Communication: The Effect of Dispersion Corrections on the Melting Temperature of Liquid Water". In: *The Journal of Chemical Physics* 134.12 (2011), p. 121105. DOI: 10.1063/1.3573375.

[80] S. Yoo, X. C. Zeng, and S. S. Xantheas. "On the Phase Diagram of Water with Density Functional Theory Potentials: The Melting Temperature of Ice Ih with the Perdew-Burke-Ernzerhof and Becke-Lee-Yang-Parr Functionals". In: *The Journal of Chemical Physics* 130.22 (2009), p. 221102. DOI: 10.1063/1.3153871.

[81] M. G. Medvedev et al. "Density Functional Theory is Straying from the Path Toward the Exact Functional". In: *Science* 355.6320 (2017), pp. 49–52. DOI: 10.1126/science.aah5975.

[82] M. J. Gillan, D. Alfè, and A. Michaelides. "Perspective: How Good is DFT for Water?" In: *The Journal of Chemical Physics* 144.13 (2016), p. 130901. DOI: 10.1063/1.4944633.

[83] B. Chen et al. "Hydrogen Bonding in Water". In: *Phys. Rev. Lett.* 91 (21 2003), p. 215503. DOI: 10.1103/PhysRevLett.91.215503.

[84] J. A. Morrone and R. Car. "Nuclear Quantum Effects in Water". In: *Phys. Rev. Lett.* 101 (1 2008), p. 017801. DOI: 10.1103/PhysRevLett.101.017801.

[85] M. Ceriotti et al. "Nuclear Quantum Effects and Hydrogen Bond Fluctuations in Water". In: *Proceedings of the National Academy of Sciences* 110.39 (2013), pp. 15591–15596. DOI: 10.1073/pnas.1308560110.

[86] M. Ceriotti and D. E. Manolopoulos. "Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei". In: *Phys. Rev. Lett.* 109 (10 2012), p. 100604. DOI: 10.1103/PhysRevLett.109.100604.

[87] F. Paesani, S. Iuchi, and G. A. Voth. "Quantum Effects in Liquid Water from an Ab Initio-Based Polarizable Force Field". In: *The Journal of Chemical Physics* 127.7 (2007), p. 074506. DOI: 10.1063/1.2759484.

[88] G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas. "A Quantitative Account of Quantum Effects in Liquid Water". In: *The Journal of Chemical Physics* 125.14 (2006), p. 141102. DOI: 10.1063/1.2358137.

[89] A. K. Soper and C. J. Benmore. "Quantum Differences between Heavy and Light Water". In: *Phys. Rev. Lett.* 101 (6 2008), p. 065502. DOI: 10.1103/PhysRevLett.101.065502.

[90] R. A. DiStasio et al. "The Individual and Collective Effects of Exact Exchange and Dispersion Interactions on the Ab Initio Structure of Liquid Water". In: *The Journal of Chemical Physics* 141.8 (2014), p. 084502. DOI: 10.1063/1.4893377.

[91] L. D. Landau and E. M. Lifshitz. *Statistical Physics*. Vol. 5. Course of Theoretical Physics. Pergamon Press, 1969.

[92] M. Ceriotti, J. More, and D. E. Manolopoulos. "i-PI: A Python Interface for Ab Initio Path Integral Molecular Dynamics Simulations". In: *Computer Physics Communications* 185.3 (2014), pp. 1019 –1026. DOI: 10.1016/j.cpc.2013.10.027.

[93] G. R. Medders et al. "On the Representation of Many-Body Interactions in Water". In: *The Journal of Chemical Physics* 143.10 (2015), p. 104102. DOI: 10.1063/1.4930194.

[94] G. R. Medders, V. Babin, and F. Paesani. "Development of a "First-Principles" Water Potential with Flexible Monomers. III. Liquid Phase Properties". In: *Journal of Chemical Theory and Computation* 10.8 (2014). PMID: 26588266, pp. 2906–2910. DOI: 10.1021/ct5004115.

[95] J. Friedrich et al. "Water 26-mers Drawn from Bulk Simulations: Benchmark Binding Energies for Unprecedentedly Large Water Clusters and Assessment of the Electrostatically Embedded Three-Body and Pairwise Additive Approximations". In: *The Journal of Physical Chemistry Letters* 5.4 (2014). PMID: 26270834, pp. 666–670. DOI: 10.1021/jz500079e.

[96] H. N. Khan, D. A. Hounshell, and E. R. H. Fuchs. "Science and Research Policy at the End of Moore's Law". In: *Nature Electronics* 1.1 (Jan. 2018), pp. 14–21.

[97] A. D. White and G. A. Voth. "Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data". In: *Journal of Chemical Theory and Computation* 10.8 (2014). PMID: 26588273, pp. 3023–3030. DOI: 10.1021/ct500320c.

[98] R. L. Henderson. "A Uniqueness Theorem for Fluid Pair Correlation Functions". In: *Physics Letters A* 49.3 (1974), pp. 197 –198. DOI: 10.1016/0375-9601(74)90847-0.

[99] M. D. Johnson and N. H. March. "Long-Range Oscillatory Interaction between Ions in Liquid Metals". In: *Physics Letters* 3.7 (1963), pp. 313 –314. DOI: 10.1016/0031-9163(63)90170-7.

[100] W. G. Madden and S. A. Rice. "The Mean Spherical Approximation and Effective Pair Potentials in Liquids". In: *The Journal of Chemical Physics* 72.7 (1980), pp. 4208–4215. DOI: 10.1063/1.439651.

[101] T. Morita. "Theory of Classical Fluids: Hyper-Netted Chain Approximation, I: Formulation for a One-Component System". In: *Progress of Theoretical Physics* 20.6 (1958), pp. 920–938. DOI: 10.1143/PTP.20.920.

[102] J. K. Percus and G. J. Yevick. "Analysis of Classical Statistical Mechanics by Means of Collective Coordinates". In: *Phys. Rev.* 110 (1 Apr. 1958), pp. 1–13. DOI: 10.1103/PhysRev.110.1.

[103] L. S. Ornstein and F. Zernike. "Accidental Deviations of Density and Opalescence at the Critical Point of a Single Substance". In: *Proc. Acad. Sci. Amsterdam.* Vol. 17. 2. 1914, pp. 793–806.

[104] G. Zerah and J. Hansen. "Self-Consistent Integral Equations for Fluid Pair Distribution Functions: Another Attempt". In: *The Journal of Chemical Physics* 84.4 (1986), pp. 2336–2343. DOI: 10.1063/1.450397.

[105] R. Rajagopalan and K. Srinivasa Rao. "Interaction Forces in Charged Colloids: Inversion of Static Structure Factors". In: *Phys. Rev. E* 55 (4 Apr. 1997), pp. 4423–4432. DOI: 10.1103/PhysRevE.55.4423.

[106] L. Parolini et al. "Interaction between Colloidal Particles on an Oil-Water Interface in Dilute and Dense Phases". In: *Journal of Physics: Condensed Matter* 27.19 (2015), p. 194119.

[107] D. Reith, M. Pütz, and F. Müller-Plathe. "Deriving Effective Mesoscale Potentials from Atomistic Simulations". In: *Journal of Computational Chemistry* 24.13 (2003), pp. 1624–1636. DOI: 10.1002/jcc.10307.

[108] F. Müller-Plathe. "Coarse-Graining in Polymer Simulation: From the Atomistic to the Mesoscopic Scale and Back". In: *ChemPhysChem* 3.9 (2002), pp. 754–769. DOI: 10.1002/1439-7641(20020916)3:9<754::AID-CPHC754>3.0.CO;2-U.

[109] G. Milano and F. Müller-Plathe. "Mapping Atomistic Simulations to Mesoscopic Models: A Systematic Coarse-Graining Procedure for Vinyl Polymer Chains". In: *The Journal of Physical Chemistry B* 109.39 (2005). PMID: 16853395, pp. 18609–18619. DOI: 10.1021/jp0523571.

[110] G. Maurel et al. "Multiscale Modeling Approach toward the Prediction of Viscoelastic Properties of Polymers". In: *Journal of Chemical Theory and Computation* 8.11 (2012). PMID: 26605616, pp. 4570–4579. DOI: 10.1021/ct300582y.

[111] V. Agrawal, G. Arya, and J. Oswald. "Simultaneous Iterative Boltzmann Inversion for Coarse-Graining of Polyurea". In: *Macromolecules* 47.10 (2014), pp. 3378–3389. DOI: 10.1021/ma500320n.

[112] T. C. Moore, C. R. Iacovella, and C. McCabe. "Derivation of Coarse-Grained Potentials via Multistate Iterative Boltzmann Inversion". In: *The Journal of Chemical Physics* 140.22, 224104 (2014). DOI: 10.1063/1.4880555.

[113] T. C. Moore, C. R. Iacovella, and C. McCabe. "Development of a Coarse-Grained Water Forcefield via Multistate Iterative Boltzmann Inversion". In: *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*. Ed. by R. Q Snurr, C. S. Adjiman, and D. A. Kofke. Singapore: Springer Singapore, 2016, pp. 37–52. DOI: 10.1007/978-981-10-1128-3_3.

Chapter 2

Theory behind Water Modelling

In this chapter, first principles theories used in water modelling are discussed, primarily density functional theory (DFT) due to its prominence in simulating liquid water. Wavefunction theory is covered briefly for the understanding of hybrid functionals in DFT. This is followed by a discussion of water force fields that provide a fast empirical approach to water modelling. Next the integral equation theory from the statistical mechanics of liquid is covered, which establishes some analytical relationships between bulk liquid properties and the potential between its constituents. The chapter is closed with the iterative Boltzmann inversion (IBI) method used to derive potentials between coarse-grained (CG) beads for a given pair correlation. This method has also been applied to coarse-grained liquid water modelled by force fields.

2.1 Quantum Mechanics

Molecules are defined as being composed of atoms, each consisting of a number of electrons orbiting a positively charged nucleus. Presented with a set of nuclei and electrons, computational chemistry is capable, within the constraints of available computational power, of calculating useful quantities such as energies, properties (dipole moment, polarisability, nuclear magnetic resonance (NMR) coupling constants, *etc.*), rates of transformation from one molecule to another, time dependence of molecular structure in addition to the geometrical arrangements of stable molecules.[114]

Although the prime objective of computational chemistry is to achieve quantitative results relevant to chemical problems, there exists an undeniable interaction with the explanation of chemical phenomena using theoretical physics. Discoveries of new theories may prompt computational investigations while computational

results may provide insight to refine theory. Quantum theory is testament to this reciprocal relationship. It is a double-edged sword having breathtaking generality but with governing equations that are intractable for all but the most ideal of systems.[115]

The advancement of computational performance (floating-point operations per second of processors, size of memory, *etc.*) complemented by the decline of computational cost (code that takes best advantage of the computational environment, *etc.*) has granted greater accessibility to quantum models. We note that although machine learning has achieved groundbreaking results without the implementation of a physical basis,[116–118] successful models are in principle based on quantum mechanics (QM). The resultant quantum solutions can be used to calibrate the more intuitive empirical method, or force fields. The features that differentiate QM from force fields include the explicit representation of electrons in calculations and the derivation of properties that depend upon electronic distribution.[119] While the ability to investigate chemical reactions where bonds are broken and formed are inherent in QM, this is not the case for standard force fields. However, reactive force fields that include connection-dependent terms allow changes in atom connectivity and thereby, model such chemical reactions.[120]

2.1.1 Schrödinger Equation

The translation of Newton's second law in classical mechanics into the domain of quantum mechanics takes the form of the time-dependent Schrödinger equation:

$$\hat{H}\Psi(r,t) = i\hbar\frac{\partial\Psi(r,t)}{\partial t} \tag{2.1}$$

This deceptively simple equation, which can be traced back to Schrödinger's landmark contributions,[121] stands as a cornerstone of quantum mechanics. The Hamiltonian operator $\hat{H}$ consists of the total kinetic and potential energy for all particles in a system and $\hbar$ is the reduced Planck constant. The particle motion through space and time can be described mathematically in the form of a wavefunction, Ψ. The spatial and time dependencies of the wavefunction are denoted by r and t respectively.

Hereafter, consider the case where the Hamiltonian operator is independent of time. The wavefunction becomes a product of its spatial and time terms. Thus, the time-independent Schrödinger equation becomes:

$$\hat{H}\Psi(r) = E\Psi(r) \tag{2.2}$$

This is an eigenvalue equation in which the Hamiltonian operator acts on a wavefunction and gives a result proportional to that wavefunction. Hence, the wavefunction is a stationary state and the proportionality constant is the scalar system energy E of the state.

2.1.2 Hamiltonian

As suggested in Eq. (2.2), the operator that returns the system energy when acting on the wavefunction is named the Hamiltonian. This operator can be divided into two contributions; namely the kinetic and potential energy. The total kinetic energy can be split into the motion of the electrons and nuclei. The components of the potential energy, which take the form of Coulomb's law, are the attraction of electrons to the nuclei, interelectronic repulsion and internuclear repulsion. Thus, the Hamiltonian can be expanded as:

$$\begin{aligned}\hat{H} = &-\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 - \sum_a \frac{\hbar^2}{2m_a}\nabla_a^2 \\ &-\sum_i\sum_a k_e\frac{e^2 Z_a}{r_{ia}} + \sum_{i<j} k_e\frac{e^2}{r_{ij}} + \sum_{a<b} k_e\frac{e^2 Z_a Z_b}{r_{ab}}\end{aligned} \tag{2.3}$$

i and j run over the set of electrons with mass m_e of $9.105\,93 \times 10^{-31}$ kg whereas a and b run over the set of nuclei. m is the mass of the particles in the system. The electrostatic constant is represented by $k_e = 1/(4\pi\epsilon_0)$, where ϵ_0 is the vacuum permittivity. e is the electron charge which is -1.60219×10^{-19} C. Z is the atomic number, and r is the distance between particles. In Cartesian coordinates, the Laplacian operator ∇^2 takes the form:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.4}$$

The wavefunction is thus a function of $3n$ coordinates, where n is the number of particles, i.e. the electrons and nuclei. Note that factors such as the presence of an external electric or magnetic field, the presence of significant spin-orbit coupling in heavy elements, relativistic effects particularly for $Z > 35$ and transition metals, *etc.* contribute to other terms into the Hamiltonian.[115]

2.1.3 Born-Oppenheimer Approximation

The motions of particles in a many-particle molecular system are correlated, as suggested by the pairwise attraction and repulsion terms contained in the Hamiltonian in Eq. (2.3). Consequently, this interdependency complicates the expres-

sion of the wavefunction. By invoking the Born-Oppenheimer approximation, the Schrödinger equation can be approximated by decoupling of the wavefunction into its electronic and nuclear components:

$$\Psi = \Psi_e \Psi_n \tag{2.5}$$

The nuclear component of this electronic wavefunction is treated classically and depends solely on the nuclei positions. Such an approximation is grounded in the fact that the rest mass of the lightest nucleus, the proton, is 1836 times heavier than the rest mass of electron[122] and the nuclear velocities are minute in comparison to those of the electrons. Thus, the electronic relaxation with respect to the nuclear motion is essentially instantaneous and solving the Schrödinger equation for fixed nuclear positions gives the electronic energy of the electronic state. By taking this electronic Schrödinger equation over all possible nuclear positions, the potential energy surface (PES) can be determined. Valuable information can be drawn from the critical points on the PES, as the minima and the saddle points correspond to the equilibrium and the transition state geometries respectively. A complete PES thus provides all the possible chemical structures and pathways interconnecting them.

2.1.4 Born Interpretation

The Born interpretation postulates that the product of the wavefunction and its complex conjugate gives the probability density. Consequently, the integral over a region of space gives the probability of finding that particle within that region of space:

$$\int \Psi^* \Psi d\tau = 1 \tag{2.6}$$

where $d\tau$ denotes integration over all space. Given that the integral must be unity as the particle must be somewhere, wavefunctions that satisfy this condition are normalised. Additionally, a general requirement for different solutions to the Schrödinger equation is that they are orthogonal for stationary states:

$$\int \Psi_i^* \Psi_j d\tau = 0 \ (i \neq j) \tag{2.7}$$

Wavefunctions that satisfy both the normalisation and the orthogonality conditions simultaneously are referred to as being orthonormal, which can be written as:

$$\int \Psi_i^* \Psi_j d\tau = \delta_{ij} \tag{2.8}$$

where the Kronecker delta, δ_{ij}, is used to express both conditions and takes the value 1 if the variables are equal ($i = j$), and 0 otherwise ($i \neq j$).

2.1.5 Pauli Exclusion Principle

Elementary particles of half-integral spin that fall under the distribution of Fermi-Dirac statistics,[123] including electrons, are referred to as Fermions. Electrons thus obey the Pauli exclusion principle that states that any two electrons in a system cannot be defined by the same set of quantum numbers. As electrons are indistinguishable, the electron density distribution remains the same despite the exchange of any pair of electrons. Thus, the wavefunction, Ψ, of a multi-Fermionic system assumes an antisymmetric form that can be expressed as a Slater determinant of spin orbitals χ:

$$\Psi(x_1, x_2, \ldots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \cdots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \cdots & \chi_N(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \cdots & \chi_N(x_N) \end{vmatrix} \tag{2.9}$$

where the coefficient $\frac{1}{\sqrt{N!}}$ acts as a normalisation factor to the wavefunction.

The antisymmetric expression is consistent with the characteristics of a Fermionic system. The exchange of any two electrons is reflected in the exchange of any two rows resulting in an identical wavefunction except for a sign change. Given two identical rows that correspond to two electrons occupying the same spin orbital, the wavefunction vanishes.

2.1.6 Hartree-Fock Method

The variational principle states that the calculated energy from an approximation to the true wavefunction is greater than or equal to the true energy of the ground state. The better wavefunction therefore generates the lower energy. Essentially, solving the Schrödinger equation becomes a minimisation procedure to find the lowest energy eigenstate of the system. The Hartree-Fock (HF) method is such a means to solve the electronic Schrödinger equation by iteratively refining the orbitals in the wavefunction in a self-consistent field (SCF) procedure.

Given a symmetric energy expression, the variational principle holds for a Fermionic system where the N-body wavefunction can be described by a single Slater determinant composed of one spin-orbital per electron. The wavefunction can thus be expanded in terms of a linear combination of a finite series of atomic

orbitals. Lagrange multipliers can be employed to maintain the orthonormality condition while relativistic effects are often neglected. The HF equation can be written as:

$$\hat{F}\Psi = E\Psi \tag{2.10}$$

where $\hat{F}$ is the Fock operator consisting of the core Hamiltonian, Coulomb term and exchange term. The core Hamiltonian corresponds to the motion of a single electron in the field of the nuclei. The Coulomb term denotes the mean field approximation and describes the interaction of a single electron with the average charge distribution of the other electrons. The exchange term arises from the antisymmetry requirement of the wavefunction. Unfortunately, the HF equation is difficult to solve as it leads to near-intractable integro-differential equations. The Roothaan-Hall (RH) approach[124,125] to the HF equation introduces basis sets in place of the atomic orbitals that can similarly be solved in a self-consistent manner. In contrast to HF, the RH equation has a matrix form that is suitable for solving numerically.

2.1.7 Density Functional Theory

In the fifty years since its formalisation, DFT has proven to be a contender to wavefunction theory as a reliable research tool in many fields that employ quantum mechanics,[126,127] although from the personal reflection of Kohn,[128] one can certainly appreciate the struggle of DFT in its early stages to be recognised among quantum theorists.

The breakthrough of a formal exact theory that maps electron density to electronic energy can be traced back to the landmark work by Hohenberg & Kohn in 1964.[129] The first Hohenberg-Kohn (HK) theorem states that the ground state density determines the external potential energy to within a trivial additive constant. The implications of this statement is that for a given density, there will be a single external potential. Thus, the energy is a unique functional of the density. In consequence, the energy and energy-derivative properties of the quantum mechanical system can be determined from the ground state density. In accordance with the second HK theorem, the variational principle holds as the energy functional returns the ground state energy if and only if provided with the ground state density.

In the following year, a practical scheme for its implementation was devised when Kohn & Sham[130] derived its self-consistent form that includes exchange and correlation effects, albeit approximately. Within the Kohn-Sham (KS) framework, the energy functional of an N-electron system of density $\rho(r)$ subject to

potential $v(r)$ is written as:

$$E[\rho(r)] = \int v(r)\rho(r)dr + \frac{1}{2}\iint \frac{\rho(r)\rho(r')}{|r-r'|}drdr' + T[\rho(r)] + E_{\text{xc}}[\rho(r)] \quad (2.11)$$

The first term is the influence of the external potential $v(r)$ on the electrons, which is normally the Coulomb potential exerted by the nuclei. The second term is the classical Coulomb energy between two charge densities. The third term is the kinetic energy of the KS reference system of noninteracting electrons. The fourth term constitutes the remaining energy, which by definition, is the exchange-correlation energy.

The exchange-correlation energy can be treated either as a single expression or its separate components, exchange and correlation. Exchange is the manifestation of the Pauli exclusion principle, which states that the wavefunction for Fermions must be antisymmetric with respect to the interchange of electrons. Correlation describes the probability distribution due to the dependency of each electron's motion on other electrons in the system. The correlation energy term also contains the contribution that accounts for the difference between the kinetic energy of a noninteracting electron gas and the real system. Both contribute to the tendency of electrons to avoid each other and lead to the formation of the exchange-correlation hole. This hole is a region about each electron where the probability of finding another electron is close to zero and in accordance with the sum rule, integrates over all space to a deficit of exactly one electron. The rules governing this hole must be obeyed to sufficiently describe the effect of exchange-correlation.

The following one-electron KS equation can be solved self-consistently with an initial guess of the density of the system:

$$\left\{ v(r) + \int \frac{\rho(r')}{|r-r'|}dr' - \frac{\hbar^2}{2m_e}\nabla_i^2 + V_{\text{xc}}(r) \right\} \psi_i(r_i) = \epsilon_i \psi_i(r_i) \quad (2.12)$$

where the density of the system is defined as the sum of the square moduli of a set of one-electron orthonormal orbitals:

$$\rho(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 \quad (2.13)$$

The expressions of the equations thus far are exact for the electronic ground state. However, the form of the exchange-correlation functional is not explicitly known and cannot be approached systematically. Therefore, the quality of DFT depends

critically on the form of the exchange-correlation functional.

A major source of error in the exchange-correlation functional, due to its approximate nature, is the self-interaction error (SIE). The exchange and Coulomb self-energies are, in general, not exactly cancelled. This self-interaction problem is due to the spurious interaction of an electron in the mean-field generated by all electrons inclusive of itself. The use of functionals without correcting for the SIE produces erroneous behaviour at the dissociation limit,[131] and contravenes Koopmans' theorem.[132,133] Moreover, the energy barriers for chemical reactions are underestimated,[134] while the intermolecular interactions for charge transfer complexes are overestimated.[135]

It is tempting to delegate all errors to the inadequacy of the exchange-correlation approximation, but such a notion would be remiss. A problem in the DFT approach itself is the inherent error when calculating the band gap, a property of the excited state, caused by the discontinuity in the exchange-correlation potential when an electron is added. This was shown by Godby, Schlüter & Sham[136] by reproducing the bulk of the error in the silicon band gap even with an exact exchange-correlation potential.

2.1.7.1 Local Density Approximation

The most basic method, as proposed by Kohn & Sham,[130] to obtain the exchange-correlation contribution is the local density approximation (LDA). This approximation sets the exchange-correlation energy at each position r to that of a homogeneous electron gas of density equal to that found at r. The exchange-correlation energy can thus be expressed as the integral of the exchange-correlation energy per unit volume in space:

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho(r)] = \int \rho(r)\epsilon_{\mathrm{xc}}[\rho(r)]dr \tag{2.14}$$

where ϵ_{xc} is the exchange-correlation energy per electron of the homogeneous electron gas. From quantum Monte Carlo simulations,[137] the exchange-correlation energies for a homogeneous electron gas are accurately known.

The LDA performs well for systems with slowly varying densities, most likely because it satisfies the sum rule, and has been widely used in solid-state physics.[138] However, the LDA exhibits limitations in describing systems with localised electrons, including ionic solids, which leads to systematic overestimation of binding energies and underestimation of bond lengths.[139]

2.1.7.2 Generalised Gradient Approximation

The behaviour of the LDA can be improved by also considering the gradient of the density in addition to the value of the density at each position r in space. This generalised gradient approximation (GGA)[140] takes an extended expression of the LDA exchange-correlation energy:

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[\rho(r)] = \int \rho(r)\epsilon_{\mathrm{xc}}[\rho(r), \nabla\rho(r)]dr \tag{2.15}$$

where $\nabla\rho$ describes the gradient correction that improves the shape of the exchange-correlation hole. Additionally, a comprehensive GGA should observe other physical boundaries including correct exchange-correlation energy in the homogeneous electron limit, coordinate scaling relations and global bounds on the exchange-correlation energy.[141] Despite the expansion of the exchange-correlation functional with a gradient term, it does not universally yield a systematic improvement over LDA and in some cases can produce greater deviation from experimental values.[142] For condensed systems, GGA is usually preferred over LDA and was used in the first DFT simulations of liquid water.[82] GGA gives satisfactory binding energies for the water dimer[68–71] and the common form of ice[72,73] as well as a reasonable structure for liquid water.[74–76]

GGAs come in a variety of forms. One popular form is the Becke exchange with Lee-Yang-Parr correlation (BLYP) functional. Its constituent exchange and correlation components are constructed separately. In 1988, the Becke exchange (B88) functional,[143] which gives comparable exchange energies to HF, was developed:

$$E_{\mathrm{x}} = E_{\mathrm{x}}^{\mathrm{LDA}} - \beta\sum_{\sigma}\int \rho_{\sigma}^{4/3}\frac{x_{\sigma}^{2}}{(1+6\beta x_{\sigma}\sinh^{-1}x_{\sigma})}dr \tag{2.16}$$

where σ is the electron spin and x_σ is the dimensionless number given by $|\nabla\rho_\sigma|/\rho_\sigma^{4/3}$. An attractive feature of this functional is the expression of its semiempirical attribute in a single constant β. Through a least-squares fit to the HF exact exchange energies of the six noble gas atoms, the constant β is set at 0.0042 a.u. By construction, it recovers the correct $-1/r$ exchange potential decay at the asymptotic limit.

In the same year, the Lee-Yang-Parr correlation (LYP) functional[144] was devised. The form taken is that of a second-order gradient expansion functional,

which for a closed shell can be written as:

$$E_{\mathrm{c}} = -a \int \frac{1}{1 + d\rho^{-1/3}} \Bigg\{ \rho + b\rho^{-2/3} \Bigg[C_{\mathrm{F}} \rho^{5/3} - 2\tau_{\mathrm{W}} + \left(\frac{\tau_{\mathrm{W}}}{9} + \frac{\nabla^2 \rho}{18} \right) \Bigg] e^{-c\rho^{-1/3}} \Bigg\} dr \tag{2.17}$$

where τ_{W} is the local Weizsäcker kinetic energy density and C_{F} is the coefficient as defined in the Thomas-Fermi kinetic energy density model. It is clear, given the form of the functional, that it is not an *ad hoc* correction to the correlation contribution in the LDA, but a calculation of the correlation energy *in toto*. The parameters a, b, c and d, which are 0.04918 a.u., 0.132 a.u., 0.2533 a.u. and 0.349 a.u., respectively, illustrates its semiempirical character. They were obtained from fitting to the HF orbital for the helium atom, based on the proposition that the second-order density matrix including correlation may be approximated by the HF second-order density matrix times a correlation factor. Notably, the functional vanishes for one-electron systems and hence, provides an exact cancellation of the one-electron SIE.

2.1.7.3 Meta-Generalized Gradient Approximation

A natural expansion to the GGA is to include the Laplacian, which is the second derivative, of the density. However, numerically stable calculations of the Laplacian proved to be a challenge. A more stable alternative is to use the kinetic energy density. These higher functionals, termed meta-GGA, can be developed for exchange, correlation or both. The first case of such a functional, by Becke & Roussel[145] for exchange, is dependent on both the Laplacian and the kinetic energy density. The general form taken is expressed here, showing all possible dependencies:

$$E_{\mathrm{xc}}^{\mathrm{mGGA}}[\rho(r)] = \int \rho(r) \epsilon_{\mathrm{xc}}[\rho(r), \nabla\rho(r), \nabla^2\rho(r), \tau(r)] dr \tag{2.18}$$

2.1.7.4 Hybrid Functionals

The problem of self-interaction continues to plague DFT. As HF exactly cancels the SIE, it follows that mixing some exact exchange contribution into exchange-correlation functionals should reduce some portion of this error. In fact, it seems promising to simply sum the exact exchange energy with a correlation energy from DFT. Based on the electron density, the density functionals do not simulate

multi-configuration mixing and the short-range correlation hole generated does not balance the long-range attributes of the exact exchange hole.[146] This produces poor left-right correlation and consequently, fails to achieve the expected improvements.

Drawing from the adiabatic connection model,[147] Becke[148] proposed a more promising mixing scheme based on the formula:

$$E_{\mathrm{xc}} = \int_0^1 U_{xc}^{\lambda} d\lambda \tag{2.19}$$

where λ is an inter-electronic coupling-strength parameter that varies the exchange-correlation potential $U_{\mathrm{xc}}^{\lambda}$. This parameter constrains the KS system between the noninteracting reference system at a value of 0 with the fully interacting real system at a value of 1, bridged by a continuum of partially interacting systems. The same density is shared across the values of the coupling strength.

Thus, the foundation was laid for the construction of hybrid functionals, such as the Becke exchange with Lee-Yang-Parr correlation adiabatic connection model with three parameter (B3LYP) functional[149] and the Perdew-Burke-Ernzerhof exchange correlation adiabatic connection model with zero parameter (PBE0) functional,[150] which perform better than their pure density functional components. Unlike the former that is tuned via three parameters fitted to experimental data of atomisation energies, ionisation potentials, proton affinities and total atomic energies of the ten first-row elements, the latter is designed with a single coefficient determined based on the convergence of fourth-order perturbation theory.[151] This predefined coefficient of 1/4 in conjunction with the implementation of the parameter-free Perdew-Burke-Ernzerhof exchange correlation (PBE) functional[152] results in a model without any adjustable parameters:

$$E_{\mathrm{xc}}^{\mathrm{PBE0}} = E_{\mathrm{xc}}^{\mathrm{PBE}} + \frac{1}{4}(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{PBE}}) \tag{2.20}$$

For most properties, *e.g.* geometry, atomisation energy and dipole moment, hybrid functionals typically perform better than GGA.[82,153] In fact, B3LYP is considered the most widely used functional for DFT calculations. However, hybrid functionals are approximately an order of magnitude more expensive in terms of computational cost compared to GGA when employed in periodic boundary conditions, such as in liquid water simulations.[82]

2.1.8 London Dispersion Forces

The exchange-correlation functionals described so far are approximations using a semilocal electron density, its gradient, its second derivative or the local kinetic energy density. Nonetheless, there exist nonlocal contributions, such as London dispersion forces that arise from the instantaneous dipole-induced dipole moments from density fluctuations due to correlated motion of electrons, which results in an attraction that is always present. Unfortunately, semilocal density functionals cannot produce the correct asymptotic $-C_6/R^6$ behaviour of the dispersion interaction energy, where R is the interatomic distance. Their effect is important in crystal packing despite being weaker than covalent and ionic bonding. Thus, dispersion must inevitably be accounted for in order to give accurate structural prediction of molecular crystals and their associated properties, such as solubility, morphology and stability.[154] Similarly for liquid water, GGAs do not provide a completely satisfactory description[78,155–158] due to their failure to correctly account for dispersion. Moreover, earlier successes of GGAs were discovered to be from the fortuitous cancelation of errors[78,159,160] and the energy differences predicted between extended and compact structures of some water systems are qualitatively incorrect[161–163].

Dispersion is dominant at long-range whereas semilocal density functionals sufficiently describe the short-range region, but they do not overlap seamlessly in the medium range so that assimilation of the two is difficult. Still, many methods that successfully incorporate the dispersion interaction into density functionals have appeared. Given the myriad of dispersion correction methods available, it is only appropriate that they are identified according to how dispersion is implemented, as demonstrated by Grimme[164]: C_6 based corrections, parameterised functionals, nonlocal van der Waals (vdW) corrections and effective one-electron potential corrections.

2.1.9 Empirical Addition of Dispersion Corrections

A dispersion energy term with a $-C_6/R^6$ dependence is appended to the energy of the system. This C_6 based correction within the DFT framework has been referred as the dispersion-corrected density functional theory (DFT-D).

$$E_{\text{disp}}^{\text{DFT-D}} = -\sum_{AB}^{N_{\text{at}}} \sum_{n=6,8,\ldots} s_n \frac{C_n^{AB}}{R_{AB}^n} f_{\text{damp}}(R_{AB}) \tag{2.21}$$

Here, a general form is presented for the dispersion energy term that adds a pairwise attractive energy for each atom pair AB with an internuclear distance R_{AB} in the system. The order of the dispersion energy series is denoted by n. C_n^{AB} is the n-th order atom pairwise dispersion coefficient as defined by the DFT-D variant. The scaling factor s_n is in place to adjust the atom pairwise addition accordingly. In order to avoid near singularities for small internuclear distances and ensure overlapping effects of correlation at intermediate distances are counted only once, the damping function f_{damp} modulates the effective onset of the correction.

2.1.9.1 Grimme

The DFT-D series developed by Grimme, consisting of DFT-D1,[165] DFT-D2,[166] and DFT-D3,[167] has a numerically simple and robust design. It is recognised that although the lack of both dependency and effect on the electronic structure pose some limitations on accuracy, this compromise is necessary to maintain low numerical complexity.

The first two versions in the series remain at the 6th order dispersion, where the scaling factor s_6 is dependent on the chosen density functional. The pairwise dispersion coefficients C_6 of the two versions employ different combination rules for the atomic C_6 coefficients. The atomic coefficients for DFT-D1 are based on the work of Wu & Yang,[168] whereas those for DFT-D2 are computed based on atomic ionisation potentials and static dipole polarisabilities. Both versions use a Fermi-type function for damping, with the steepness parameter d reduced from 23 to 20 going from DFT-D1 to DFT-D2:

$$f_{\text{damp}}(R_{AB}) = \frac{1}{1 + \exp\left[-d(R_{AB}/R_{AB}^0 - 1)\right]} \tag{2.22}$$

where R_{AB}^0 is the sum of the atomic vdW radii. This function is used as it decays sufficiently fast to zero at small R so that the dispersion correction becomes negligible below the vdW distances.

The more recent DFT-D3 made significant improvements to its predecessors. The resultant performance is better due to an overhaul of the dispersion coefficient as a geometric-dependent function based on the fractional coordination number $\hat{f}$. This emulates a change in the dispersion with the hybridisation state of the atom in the molecule via the change in the electronic structure from bond formation.

The aforementioned function took on a Gaussian-distance weighted average form:

$$C_6^{AB}(\hat{f}^A, \hat{f}^B) = \frac{\sum_i^{N_A} \sum_j^{N_B} C_{6,\mathrm{ref}}^{AB}(\hat{f}_i^A, \hat{f}_j^B) e^{-k_3\left[(\hat{f}^A - \hat{f}_i^A)^2 + (\hat{f}^B - \hat{f}_i^B)^2\right]}}{\sum_i^{N_A} \sum_j^{N_B} e^{-k_3\left[(\hat{f}^A - \hat{f}_i^A)^2 + (\hat{f}^B - \hat{f}_i^B)^2\right]}} \tag{2.23}$$

where $C_{6,\mathrm{ref}}^{AB}$ is precomputed using time-dependent density functional theory (TD-DFT) with an equation based on the Casimir-Polder integral.[169] N_A and N_B are the number of reference systems for atoms A and B respectively. $\hat{f}_i^A$ and $\hat{f}_j^B$ are coordination numbers for the atom pair AB for the two reference systems i and j, while $\hat{f}^A$ and $\hat{f}^B$ are those for the system of interest. The parameter k_3 was designated as 4 to produce smooth potential curves and plateaus close to integer $\hat{f}$ values.

The scaling factor s_6 in Eq. (2.21) normally takes on a value of unity resulting in the exact asymptotic limit. In density functionals where some of the long range dispersion energy is built-in, s_6 can be a proper fraction. Although higher order dispersion coefficients can be determined via the recursion relations,[170] the dispersion terms of the order greater than 8 exhibit numeric instability without justifiable improvements. Hence, n is truncated after 8, where C_8^{AB} needs to be scaled by s_8 depending on the chosen density functional as it interferes appreciably with the short-range electron correlation.

It is worth noting that a nonadditive three-body energy can be implemented using a damped Axilrod-Teller-Muto (ATM) formula,[171] where its dispersion coefficient C_9^{ABC} is simplified to a geometric mean of the two-body coefficients. The effect of including this three-body term increases the mean absolute deviation (MAD) by approximately 0.02 kcal/mol for the S22 benchmark[172] set for most density functionals. This may be due to the inherent overestimation of the three-body contributions by density functionals in overlapping density regions.[173] Although negligible for small systems, the three-body effects may be substantial and should be given due consideration in the case of larger systems.

The dependency of the overall performance of the DFT-D model on the choice of damping function f_{damp} was found to be overemphasised, when a wide range of systems and density functional were considered.[168,174] Hence, emphasis is instead given to one that was convenient for higher dispersion orders. A simple variant of the damping function by Chai and Head-Gordon[175] was prescribed:

$$f_{\mathrm{damp}}(R_{AB}) = \frac{1}{1 + 6(R_{AB}/(s_{r,n} R_{AB}^0))^{-\alpha_n}} \tag{2.24}$$

which can easily be applied to any dispersion order through the scaling factor $s_{r,n}$ of the atom pairwise cutoff radii R^0_{AB}.[176] The value $s_{r,6}$ is important for adapting to the chosen density functional and determined by a least-square fit to reference energies across multiple benchmark sets, whereas $s_{r,8}$ is fixed to unity. The cutoff radii R_0^{AB} is defined as the distance where the first-order interaction energy between atoms A and B equals a cutoff energy. The steepness parameters α_n are adjusted to attain a dispersion correction less than 1% of the maximum magnitude of the dispersion energy for typical covalent bond distances. Here, α_6 is set at 14 followed by α_8 at 16.

An issue with damping to zero, as per the previous damping functions in the DFT-D series, is that it leads to repulsive dispersion in the short-range region if the parameters are not chosen carefully. Instead of damping to zero, Becke & Johnson[177–179] proposed the use of rational damping to finite values for short interatomic distances. This was physically more justified based on the multipole-expansion of the dispersion energy that was found to converge on a constant finite value as R_{AB} approaches zero.[180] The rational damping was incorporated into a revised DFT-D3 so that the dispersion energy is given by:

$$E_{\text{disp}}^{\text{DFT-D3(BJ)}} = -\frac{1}{2}\sum_{A\neq B}\left[s_6\frac{C_6^{AB}}{R_{AB}^6 + (a_1 R_{AB}^0 + a_2)^6} + s_8\frac{C_8^{AB}}{R_{AB}^8 + (a_1 R_{AB}^0 + a_2)^8}\right] \tag{2.25}$$

where the cutoff radius R^0_{AB} is defined as $\sqrt{C_8^{AB}/C_6^{AB}}$. This reduces to the geometric mean of atomic expectation values, as the higher ranked C_8 values can be computed recursively from C_6 values. The scaling factor s_6 is set to unity for GGA and hybrid functionals to ensure the correct long-range asymptotic behaviour, whereas s_8 is adapted for the repulsive short- to medium-range behaviour of the exchange-correlation functional.

The revised DFT-D3 with Becke-Johnson (BJ) damping does not require a pair-specific cut-off radii and unlike the zero damping, it does not provide artificial repulsive interatomic forces at short distances. In fact, BJ damping performs better than zero damping for thermochemical problems that are sensitive to medium-range correlation effects. Although BJ damping does provide small correlation energies for spatially close atom pairs, there is no significant "over-correlation" using common density functionals for many chemically relevant processes, based on the GMTKN30[181] molecular energy database.

2.1.9.2 Tkatchenko-Scheffler

Similar to the previously mentioned DFT-D3, the Tkatchenko-Scheffler (TS) scheme[182] accounts for hybridisation effects for atoms in molecules. It does so using the relative polarisability for an atom in a molecule that is calculated from the electron density. The analytical gradients or atomic forces are, however, not readily available for performing structural optimisation.

From the London formula,[183] the atom pairwise dispersion coefficient can be derived in a form that depends on homonuclear parameters:

$$C_6^{AB} = \frac{2C_6^{AA}C_6^{BB}}{\frac{\alpha_B^0}{\alpha_A^0}C_6^{AA} + \frac{\alpha_A^0}{\alpha_B^0}C_6^{BB}} \tag{2.26}$$

where α^0 is the static atomic polarisability and C_6^{AA} and C_6^{BB} are the dispersion coefficients for the homonuclear dimers. These parameters can be drawn from the free-atom reference data,[184] which are computed using self-interaction corrected TDDFT. As per the DFT-D series so far, barring the revised DFT-D3 with BJ damping, a simple Fermi-type damping function[168] is used here:

$$f_{\text{damp}}(R_{AB}) = \frac{1}{1 + \exp[-d(R_{AB}/(s_r R_{AB}^0) - 1)]} \tag{2.27}$$

where the pairwise vdW radii R_{AB}^0 is the sum of its atomic vdW radii that are defined as half the distance where the repulsion caused by the Pauli repulsion principle balances the attraction due to London dispersion forces.[185] A proportionality relationship between the atomic volume and the atomic vdW radii is once again employed to allow the calculation of the effective atom-in-a-molecule value from the free-atom value. Deferring to Grimme's use of a similar type damping function in DFT-D2,[166] the steepness d is satisfactorily optimal at 20. The scaling factor s_r, as taken from the S22 database of Jurecka *et al.*,[176] enables adaptation to the specific density functional.

A linear correlation exists between atomic volume and polarisability,[186] so that a proportionality relationship can be established between atomic volume and polarisability for the atom-in-a-molecule relative to the free-atom. The atomic volume is in turn defined by the Hirshfeld partitioning of electron density.[187] Likewise, a proportionality holds between the atomic volume and the dispersion coefficients, where the proportionality constant is assumed to be unity. This allows Eq. (2.26) to be redefined in terms of the effective dispersion coefficients for an atom in a molecule. The only caveat is that the assumption does not stand for the smallest H_2 molecule, deviating by 44% for the effective dispersion

coefficient. The scheme is usually limited to the C_6 term but has the flexibility for generalisation to higher-order dispersion. The scaling factor s_6 in Eq. (2.21) is set to 1/2.

In contrast to a free atom, that same atom in a molecule will experience a dynamic electric field from other surrounding atoms. This electrostatic screening induces an electrodynamic response, such as polarisation and depolarisation effects, as well as anisotropy in molecular polarisability. The electrodynamic response, which becomes more pronounced with increasing system size, is accounted for in the TS scheme via a discretised form of the classical electrodynamics self-consistent screening (SCS) equation:

$$\alpha_A^{SCS}(i\omega) = \alpha_A^{TS}(i\omega) + \alpha_A^{TS}(i\omega) \sum_{A \neq B}^{N} \tau_{AB} \alpha_A^{SCS}(i\omega) \tag{2.28}$$

where α_A^{TS} is the TS-vdW effective atomic polarisability and τ_{AB} is the dipole interaction tensor. The solution for every frequency of the electric field yields the frequency-dependent atomic polarisability tensors $\alpha_A^{SCS}(i\omega)$. A new set of parameters can be derived from this SCS result to be used as per the original scheme, *e.g.* using $\alpha_A^{SCS}(i\omega)$ in the Casimir-Polder integral to obtain the dispersion coefficient C_{6AA}^{SCS}.

The TS-vdW + SCS method has seen further improvement by including the fully nonadditive many-body dispersion effects up to the Nth order. This is accomplished using the coupled fluctuating dipole moment, in which atoms are treated as coupled isotropic three-dimensional quantum harmonic oscillators. The many-body interaction energy is then calculated as the difference between the zero-point energies of the coupled and uncoupled quantum harmonic oscillators, which is coupled to density functionals with a Coulomb potential. The many-body effects are expected to be more important for large systems and have been demonstrated, along with SCS, to improve the mean absolute relative error (MARE) of the S22 benchmark set from 9.2% to 5.4%.

2.1.9.3 Exchange-hole Dipole Moment

Becke & Johnson developed the exchange-hole dipole moment (XDM) model based on the idea that the instantaneous dipole moments arise due to the spherical asymmetries in the exchange hole.[188,189] The resulting nonzero dipole moment of the exchange hole, whose shape depends on the associated electron's position, in one system induces an instantaneous dipole moment in another nonoverlapping system a large separation away. The occurrences of this instantaneous dipole

moment thereby give rise to the dispersion force.

Applying the second-order perturbation of the dispersion interaction[190] to the model, the dispersion coefficients can be obtained in terms of the position-dependent dipole moment of an electron and its accompanying exchange hole. The form taken is:

$$C_6^{AB} = \frac{\alpha_A \alpha_B \langle M_l^2 \rangle_A \langle M_l^2 \rangle_B}{\langle M_l^2 \rangle_A \alpha_B + \langle M_l^2 \rangle_B \alpha_A} \tag{2.29}$$

where α is the atomic polarisability and $\langle M_l^2 \rangle$ is the expectation value of the squared l-th multipole moment integral. This equation can be rewritten in terms where the nonoverlapping systems refer to molecules. Derivations have also been expanded to the higher-order C_8 and C_{10} dispersion coefficients.[191] The third-order perturbation of dispersion interaction can correspondingly be used to derive the isotropic three-body dispersion coefficients; a procedure which can be repeated to find the many-body dispersion to any order.[192] Since the dispersion term within the context of this model is not scaled, s_n in Eq. (2.21) is unity.

A separation-based damping function[193] employing the sum of the vdW radii of an atom pair has been proposed with the use of the model. This sum of the vdW radii can be well defined through the behaviour of higher-order dispersion terms with respect to internuclear separation between the atom pair. As the separation decreases from a large distance, the dispersion energies of the higher-order terms diverge faster than those of the lower-order terms leading to intersections. Thus, there exists a critical separation where the dispersion terms are approximately equal. Assuming a linear relationship between the sum of effective vdW radii of an atom pair with the critical separation, the damping function can be expressed as:

$$f_{\text{damp}}(R_{AB}) = \frac{R_{AB}}{R_{AB} + a_1 R_{c,AB} + a_2} \tag{2.30}$$

where the parameters of the aforementioned linear relationship a_1 and a_2 are set to be 0.83 and 1.55 Å, respectively, obtained from the root-mean-square percent error minimisation of binding energies of 45 intermolecular complexes.

2.1.10 Density Based Dispersion Functionals

2.1.10.1 Parameterised Density Functionals

In the absence of dispersion corrections, some semilocal or hybrid functionals are capable of emulating dispersion effects up to the medium-range. These include pure density functionals that provide remarkable results from the fortuitous cancellation of errors or the fortuitous balance of HF and DFT, rather than correctly

describing nonlocal dispersion, leading to good results. The X3LYP functional significantly improves the accuracy of hybrid GGA methods for rare-gas dimers[194] and the water dimer[195], but fails qualitatively for base stacking[196]. In contrast, the binding energies of π-stacked complexes from the Becke half-and-half (BH&H) functional agree well with those from MP2 and CCSD(T).[197]

There are also functionals that incorporate dispersion interaction by design. These highly parametrised density functionals are usually empirically fitted to benchmark data sets, which include systems dominated by dispersion interactions and π-π stacking. A prominent example are the Minnesota-type functionals. In fact, the M06-2X functional is probably the most accurate functional without an explicit dispersion correction, as it produces good results for the S22 set and stacked aromatic structures.[198] However, it contains many highly parameterised terms in a power series expansion that leads to numerical instability resulting in artificial vdW minima and noisy potential energy curves.[199] On top of that, the more recent M11 and MN12 show that simply fitting for dispersion interactions without an appropriate functional form inadequately described the adsorption of aromatic structures on graphene.[200] Furthermore, two M11-types were reported to underbind the helium, argon, coronene and naphthalene dimers.[201,202] The caveat of parameterised density functionals is that the medium-range dispersion energy asymptotically approaches zero. This prevents application in extended systems, such as solids and biomolecules, where long-range contributions are important.

2.1.10.2 Nonlocal Correction

In the van der Waals density functional (vdW-DF),[203] the dispersion energy for any system can be calculated in a non-empirical way based on the electron density. The total exchange correlation energy E_{xc} of the system is simply the addition of the nonlocal term $E_{\mathrm{c}}^{\mathrm{nl}}$ for the dispersion energy to the standard exchange E_{x} and correlation E_{c} components to describe the short range:

$$E_{\mathrm{xc}} = E_{\mathrm{x}} + E_{\mathrm{c}} + E_{\mathrm{c}}^{\mathrm{nl}} \tag{2.31}$$

The short-range components are typically described by LDA or semilocal GGA. It was found that repulsive short-range GGA components best complemented the vdW-DF, which can be further improved through empirical optimisation.[204] In essence, vdW-DF provides an advantage in seamlessly accounting for the dispersion effects and its charge dependence via the nonlocal term. The

nonlocal correlation energy takes the form:

$$E_{\mathrm{c}}^{\mathrm{nl}} = \frac{1}{2} \int \int \rho(r)\phi(r,r')\rho(r')drdr' \tag{2.32}$$

where ρ is the electron density. The electron coordinates are denoted by r and r'. The variant of vdW-DF is determined by the choice of the nonlocal correlation kernel ϕ, which are based on the local approximations to the frequency-dependent averaged dipole polarisability. vdW-DF2[205] goes beyond the local approximations for the kernel giving a more consistent description of interaction energies and intermolecular distances in complex geometries. However, the kernels are constructed to give $1/r^6$ in the long-range limit, which is usually not the correct form at low dimensionality and in metallic systems.

Due to the complexity of the self-consistent approach, the nonlocal term was usually computed using an *a posteriori* energy correction scheme or restricting the relaxation of the structure. This often only has minor effects on energies but impedes the computation of forces during geometry optimisation. To remedy this, an efficient linear-scaling SCF implementation of the nonlocal term has been developed.[206]

The earlier versions of vdW-DF were derived for non-overlapping densities and required empirical damping functions to ensure finite values for short interfragment distances. In contrast, the modern versions of the nonlocal term are undamped and provide contributions to the short-range correlation energy affecting the thermochemistry. It is currently not known whether double counting effects of correlation at short range is present in modern vdW-DF.[164]

2.1.11 Effective One-Electron Potential

The dispersion energy is inherently a many-electron correlation effect. However, the polarisabilities of molecules are represented well by adding local atom-like quantities as in DFT-D.[164] The dispersion can thus be described by an effective atom-centred one-electron potential (1ePOT), which can change the electronic charge density. This approach was first used by von Lilienfeld *et al.*[207–209] using an optimised dispersion-corrected atom-centred potential (DCACP). It has been used to model attractive long-range vdW forces for argon-benzene and graphite-graphite complexes, as well as calculate interaction energies for polyaromatic hydrocarbon molecules[210,211] and the adsorption of argon on graphite[212].

Although its empirical nature means computational viability, the number of free parameters required is equal to the number of different elements in the sys-

tem. So far, only a small fraction of the periodic table has managed to be covered. Since the atomic parameters are fixed for each element, the changes of dispersion coefficients with the hybridisation or oxidisation state of atoms can not be reflected. Another caveat is that the 1ePOT method does not show correct asymptotic R^{-6} behaviour and the exponential decay with interatomic distances is too rapid. This results in graphene sheets being underbound by about 20 % of the interlayer binding energy.[213]

2.1.12 Performance of Density Functional Theory for Water

In the case of water, the DFT approach has been successfully used ever since the simulations of liquid water pioneered by Parrinello and Car *et al.*[74–76,214,215] Not only has DFT yielded satisfactory descriptions for liquid water, clusters and ice structures, it has been applied to general aqueous systems, including acid-base systems,[216–218] confined water,[219–221] and water adsorbed on surfaces.[222–224]

Amongst the various DFT approaches, GGAs serve as the foundation of widely used density functionals in investigating the *ab initio* molecular dynamics (AIMD) simulations of liquid water. However, the description of liquid water by pure GGAs are insufficient and require high pressures to maintain the experimental equilibrium density. These approximations are affected by the so-called SIE. This contributes to the spurious delocalisation of protons and consequently strengthens the hydrogen bonds in liquid water. The stronger hydrogen bonds hinders the degree of cross-shell penetration leading to liquid water that is over-structured and under-diffusive. At a specified density and temperature, Gillan *et al.*[82] reasoned that the structure and diffusivity largely corresponds with the strength of the hydrogen bonds from the choice of GGA. In a paper by Mattsson and Mattsson,[225] the binding energy of the water dimer was calculated for semi-local functionals, including GGAs, as it describes the strength of the hydrogen bonding. The emerging trend of the GGA functionals: RPBE, BLYP, PBE, based on those binding energies were in the same order as the increasing structural ordering in liquid water at 300 K and the leftward shift of the position of the first peak.

The degree at which BLYP liquid water is over-structured can be observed in the oxygen-oxygen (O-O) radial distribution function (RDF). The first peak and first depth are somewhat exaggerated but fortunately, the radial positions of the main features are in agreement with experiment. Up to half of this over-structuring can be attributed to ignoring nuclear quantum effects (NQE) as suggested by DFT simulations,[84,226] with the remainder due to the functional errors

including the lack of nonlocal vdW interactions. With the over-structuring comes a corresponding self-diffusion that is lower, usually by a factor of ∼3, than experiment. Additionally, the peak of the oxygen-hydrogen (O-H) RDF at ∼1.75 Å is useful to probe the hydrogen bond between the donor O and the receptor H. For BLYP-based simulations, the height of this hydrogen bond peak was reported to be overestimated by ∼30 %.[78,84,105,227,228]

The over-structuring observed in the O-O RDF from liquid water simulations using PBE is more severe than BLYP, consistent with the trend seen in dimer binding energies. Compared to Soper's experimental first peak height of 2.49 and first minimum depth of 0.80,[21] these features are substantially exaggerated in PBE-based simulations. The resulting features are typically ∼3.4 and ∼0.4 respectively.[229] As expected from the greater over-structuring, the intimately linked self-diffusion is underestimated by a larger factor of ∼10. The more ordered peaks suggests the lack of frequency of the water molecules crossing from the second shell into the first shell. The O-H RDF is correspondingly over-structured, where the height of the hydrogen bond peak at ∼1.75 Å is overestimated by ∼50 %.[156,229]

Despite the shortcomings of PBE, revisions made to the functional have yielded more satisfactory properties for liquid water. Namely revPBE and RPBE achieved more diminished structures and higher diffusivities than BLYP and PBE.[229,230] The addition of a fraction of exact exchange into the density functional such as for the hybrid PBE0 can also soften the structure of the liquid water, although this effect is weak.[90,231,232] This is due to the mitigation of delocalisation error or SIE, which can be seen in the improvement of binding energies.[233] Guidon *et al.*[234,235] reported very little change in the liquid structure when the fraction of exact exchange in PBE0 was increased from one fourth to unity. However, this has been attributed to the hardening from the basis-set errors in the auxiliary density matrix technique used.[56] Consequently, a larger basis set is required to reduce the severity of basis-set errors. The arguable improvements in liquid structure and diffusivity are in part due to the lack of nonlocal correlations, which is mainly a two-body interaction in water systems,[236] in the hybrids and similarly, in the pure GGAs.

The significance of nonlocal correlations is further corroborated when the addition of dispersion changes the properties of liquid water for most functionals in the right direction, such as diminished liquid structure and increased diffusivity. These changes also include an increased equilibrium density, which is a property that is underestimated in pure GGAs due to their inaccurate pressure-volume relations for liquid water. Fortunately, the poor prediction in routinely calculated pressure is straightforward to detect in constant-volume simulations. Although

dispersion-inclusive approximations demonstrate changes in the right direction generally across the board, the extent of which is as important do depend on the pairing of functional and dispersion correction.[237] According to Gillan *et al.*[82] by using the BLYP functional corrected by various dispersions, the equilibrium density increased by between $100\,\mathrm{kg\,m^{-3}}$ and $250\,\mathrm{kg\,m^{-3}}$.[55,238–242] As opposed to the underestimation using uncorrected BLYP, this is an overestimation of the property by ∼5 % and is consistent with the tendency to compress water by BLYP corrected for one- and two-body errors using Gaussian approximation potentials. As expected for the liquid structure of BLYP when corrected for dispersion, the first maximum and the first minimum are usually lowered by ∼0.30 and raised by ∼0.20 respectively. In effect, the first minimum is improved significantly to ∼0.78 that is closer to the 0.84 value for experiment. The diffusion is likewise improved considerably through an increase by a factor of 2–3.

On the other hand, the performance of PBE can vary widely with the dispersion method. This can be observed in the degree of softening in the liquid structure. The various PBE-D[55,227,229,239,243] using Grimme's pairwise semiempirical corrections and PBE-DCACP[242] produce a first maximum of 3.2–3.4 and a first minimum of 0.4–0.5, which indicates a lack of structural softening. The PBE-TS improves slightly over the previously mentioned functionals with a modest softening that results in a first maximum and a first minimum of 2.99 and 0.54 respectively.[90] The softening is much stronger for vdW-DF, typically reaching a first maximum of ∼2.6 and a first minimum of ∼0.8.[227,231,232,244,245] The overall liquid structure unfortunately shows large disagreements with experiment for the original vdW-DF or DRSLL, which combines the revPBE exchange with the nonlocal vdW correlation. The second coordination shell is significantly disrupted and an anomalous bump manifest at 3.8 Å.[244] This can mostly be alleviated using PBE-DRSLL, where the exchange component incorporates PBE instead of revPBE. Although PBE-DRSLL experienced an increase in equilibrium density, it is overestimated by ∼13 %. As the almost exact Gaussian approximation potential correction for one- and two-body errors does not alter the pressure significantly in compressed water systems,[246] this overestimation appears to be a beyond two-body effect.

Compared to hybrids, the meta-GGA level has not been as widely used in condensed phase simulations of water. This can be attributed to the poor results produced in early studies[160,247] and the prevalent delocalisation error. Fortunately, recent meta-GGA functionals cost a fraction of hybrids at a comparable accuracy for non-covalent interactions. More importantly, the addition of physics, namely the nuclear quantum effects, bring meta-GGAs in line with the cluster and liquid

state properties observed in experiment. Ruiz Pestana *et al.*[248] have investigated two of the more successful meta-GGAs corrected for dispersion, B97M-rV[249] and M06-L[250] with D3 correction, which demonstrated good results in dimer binding energies, barrier heights and thermochemistry. Although the resulting first peak in their liquid structure were at a similar height to revPBE-D3, they were located to the right of the experimental first peak. The other identifiable features in B97M-rV are observed to have shifted to greater radial positions and exhibited a shallower first minimum. This diminished structure beyond the first peak was also evident in M06-L-D3 that failed to reproduce the second peak, indicating a loss of tetrahedral structure. The dipole moments were found to be $\sim$0.2 D lower than experiment, which showed that the lack of structure beyond the first peak can be attributed to weaker hydrogen bonds. As for the self-diffusion coefficient, B97M-rV achieved a value in a comparable bracket with experiment but overestimated at $2.9 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$. In contrast, M06-L-D3 performed poorly with a suppressed value of $0.3 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ for diffusivity.

One promising meta-GGA is the strongly constrained and appropriately normed (SCAN) density functional.[251] Where qualitative assessment of the gas-phase water hexamers and ice phases were difficult using other functionals, including with long-range dispersion corrections, SCAN was able to make a quantitative prediction.[252] This was achieved as the functional satisfied all seventeen known exact constraints on semi-local exchange-correlation functionals. When used to investigate structural, electronic and dynamic properties of liquid water, SCAN produced results in agreement with experiment.[253] By simply elevating the temperature by 30 K to emulate NQE and without correcting for dispersion, the functional was able to closely reproduce the disordered and densely-packed water structure, albeit with a slightly higher and left-shifted first peak. This is due to the ability of SCAN to describe the vdW interactions at intermediate length-scales, which resulted in non-directional attraction that drew surrounding water molecules into the interstitial spaces between the hydrogen-bonded waters. The inherent vdW description also led to the correct density ordering between liquid water and ice Ih, *i.e* that ice floats on liquid water, which have is not predicted by almost all other density functionals. Furthermore, SCAN has been found to produce weak directional hydrogen bond strength that contributed to improved electronic properties of water, such as a dipole moment of 2.97 D in good agreement with experiment. The weaker hydrogen bond also resulted in water diffusing fast at $1.9 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$, within good accuracy from the experimental diffusivity.

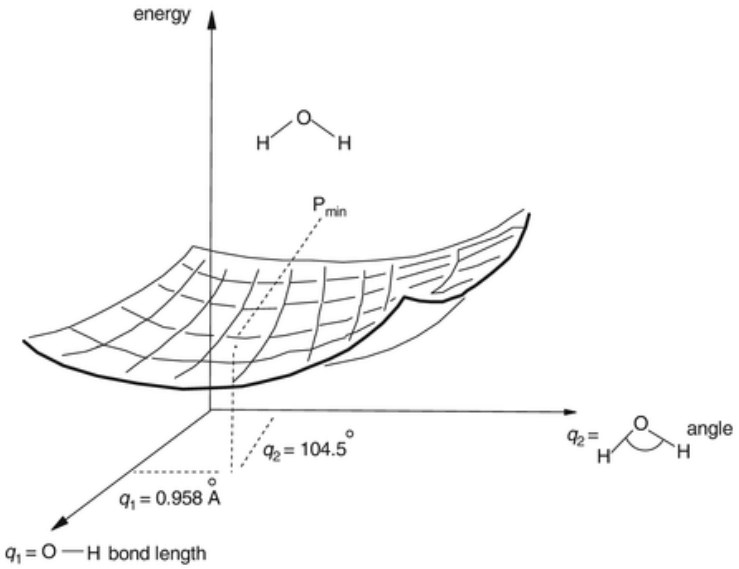


Figure 2.1: The PES of a water molecule with C_{2v} symmetry.[256] The point P_{min} corresponds to the minimum-energy geometry, *i.e.* equilibrium geometry, for the water molecule. Reprinted by permission from Springer Nature Customer Service Centre GmbH: E. G. Lewars. "The Concept of the Potential Energy Surface". In: *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Cham: Springer International Publishing, 2016, pp. 9–49.

2.2 Molecular Mechanics

The potential energy surface, commonly abbreviated as PES, describes the relationship between the energy of a molecule and its geometry. A core part of chemistry is arguably in the study of the stationary points on the PES, which correspond to stable chemical species, and the pathways connecting them. It is then not surprising that in computational chemistry the PES is a central concept and is generated by determining the molecular energy for discrete points on the surface. Water is a triatomic molecule with two bond lengths and a bond angle. Assuming that the bond lengths are the same, *i.e.* C_{2v} symmetry, the structure can be defined in terms of two geometric parameters, the O-H bond length and the H-O-H bond angle. The resultant PES of the water molecule is reproduced in Figure 2.1, where the minimum denotes the equilibrium geometry of the water molecule. Some of the earliest attempts to determine the PES for the water molecule made use of experimental spectroscopic data.[254,255] Molecular mechanics (MM) enables fast exploration of the PES of large systems by modelling the atoms and their bonds in molecules as a collection of masses held together by springs. MM ignores electrons as such and cannot explicitly describe electronic properties.

MM expresses the energy of a molecule as a sum of various energy interaction

terms that describe its resistance toward changes from some equilibrium geometry, *e.g.* bond stretching, angle bending, torsion and nonbonded interactions. The energy terms contain empirical values that are parameterised to the structure and to reproduce some thermodynamic properties. Consequently, MM is both fast and accurate in predicting geometries and relative energies for the kinds of molecules for which it has been parameterised, usually at the expense of transferability. The energy expression, E_{FF}, and the parameters in it constitute a force field,[114] which can be minimised as a function of the atom coordinates in order to locate a PES minima:

$$\begin{aligned} E_{\mathrm{FF}} = &\sum E_{\mathrm{stretch}} + \sum E_{\mathrm{bend}} + \sum E_{\mathrm{torsion}} \\ &+ \sum E_{\mathrm{vdw}} + \sum E_{\mathrm{electrostatic}} + \sum E_{\mathrm{cross}} \end{aligned} \tag{2.33}$$

E_{stretch} and E_{bend} are the energy contributions from the stretching of the bond between two atoms and from bending the angle formed by three atoms, respectively. The simple harmonic form is typically sufficient for these two terms in applications close to the equilibrium geometry. E_{torsion} is the torsional energy for rotation around a bond in a four atom sequence. The energy function must be periodic in the angle of rotation, which can be described as a Fourier series. Naturally, the torsional term is not applicable for triatomic molecules such as water. These contributions make up the bonded interactions.

The coupling between the above intramolecular terms is described by E_{cross}. Consider the equilibrium geometry of the water molecule with an O-H distance of 0.958 Å and an angle of 104.5°. When the angle is compressed to 90°, the lowest energy bond length is 0.968 Å. Likewise, widening the angle leads to a shorter bond length. Coupled stretch-bend interactions can capture such effects, but are not commonly used in water force fields.

E_{vdw} and $E_{\mathrm{electrostatic}}$ encompass the nonbonded interactions. The former is the energy contribution from the van der Waals forces describing repulsion and attraction between atoms. The latter is due to the distribution of electrons creating positive and negative regions. This electronic property is accounted for by assigning partial charges, q, to the atoms. In most water force fields, the Coulombic interaction between all intermolecular pairs of charges along with a potential energy curve between oxygen atoms determine the nonbonded interaction energy:

$$E_{\mathrm{nonbonded}} = \sum_i \sum_{j>i} \left(k_e \frac{q_i q_j}{r_{ij}} + V_{ij}^{\mathrm{LJ}} \right) \tag{2.34}$$

The electrostatic constant is represented by $k_e = 1/(4\pi\epsilon_0)$, where ϵ_0 is the vac-

Table 2.1: Water force fields mentioned in this work.

(Date)	Force field	Type	Sites
(1933)	BF[41]	Rigid	4
(1981)	TIPS[257]	Rigid	3
(1981)	SPC[44]	Rigid	3
(1983)	TIP3P[45]	Rigid	3
(1985)	SPC/F[258]	Flexible	3
(1987)	SPC/E[259]	Rigid	3
(1989)	SPCP[260]	Polarisable	3
(1994)	TIP4P/FQ[261]	Polarisable	4
(1997)	SPC/F2[262]	Flexible	3
(1999)	TIP3P/Fs[263]	Flexible	3
(2000)	TIP5P[264]	Rigid	5
(2005)	TIP4P/2005[265]	Rigid	4
(2006)	SPC/Fw[266]	Flexible	3

uum permittivity. The potential energy curve between two oxygen atoms can be approximated by the Lennard-Jones (LJ) 12-6 potential that is made up of a twelfth-power repulsive term and a sixth-power attractive term. It is written as a function of the distance r_{ij} between the atoms:

$$V_{ij}^{\mathrm{LJ}} = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.35}$$

where ϵ_{ij} is the depth of the energy well and σ_{ij} is the finite distance where there is zero interaction energy. The (negative) derivative of the potential gives the force between the atoms.

2.2.1 Performance of Water Force Fields

Water force fields, such as the ones mentioned in this work (Table 2.1), are able to predict a combination of physical properties of water successfully, but none have been wholly satisfactory and are usually limited within a range of state points. In fact, none of them are able to reproduce accurately all the properties from the triple point to the critical point.[267] When the phase diagrams for various water models were determined, their performance was noted to be quite different.[265,268,269] Most water force fields also fail to reproduce quantitatively both the temperature of maximum density[270–273] and the liquid-vapour coexistence curve.[274–277] SPC/E, that incorporates polarisation effects in a mean-field manner, represents quite well the coexistence curve.[278] This suggests the importance of polarisation effects in reproducing bulk and vapour properties over a broad range of densities and temperatures. However, its addition does not necessarily

improve the coexistence curve, as shown in the studies of polarisable water force fields. In some cases, such as SPCP,[277,279] the agreement with the experimental coexistence curve may worsen. Notable properties that are typically reported in literature to assess the performance of different water force fields include the atom-atom pair distribution function, the self-diffusion coefficient and the dielectric constant.

The most distinctive property of water is arguably its microscopic structure and it is essential for water force fields to be able to reproduce the structural data from radiation scattering experiments. In the review by Guillot,[47] marks were assigned to a number of empirical and *ab initio* water models by comparing to the structural data from the X-ray scattering of Sorensen *et al.*[18] and the neutron data of Bellissent-Funel et al.[280] Early water force fields in the 1980s and 1990s were however parameterised using inaccurate structural data that were due to experimental uncertainties and the data treatment. The Bernal-Fowler (BF) model was based on early X-ray diffraction data and does not reproduce the second peak at $\sim$4.5 Å in the oxygen-oxygen pair correlation function, which is a signature of the hydrogen-bond network. Improvements in X-ray and neutron scattering techniques, *e.g.* isotropic substitution, have since reconciled their structural data.[281]

Although the interaction potentials in water are not usually parametrised to reproduce the self-diffusion coefficient, $D_s = 2.299 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$, it is one of the more important transport coefficients of water in molecular dynamics (MD) studies. As a measure of the mobility of water molecules, the self-diffusion coefficient is an indicator of the effect of hydrogen bonds on the translational motions. In force field simulations, the flexibility of the molecule has an appreciable effect on the self-diffusion.[282] It has been reported that the flexible variants of the rigid force field, SPC, have lower self-diffusion coefficients, which was attributed to the larger individual average dipole moments.[283,284] In addition to the rotational motion of water, the quantum effects of the hydrogen atoms were found to appreciably influence the self-diffusion coefficient at ambient temperature.[262,285] In order to account for the quantum effects in force fields, the Feynman-Hibbs path integral approach can be implemented.[286] Unfortunately, the translational motion of water is also affected by the finite size effects arising from the hydrodynamic interactions between periodic images. Consequently, this transport property of water linearly depends on the inverse of the effective cell length. Calculations of the diffusion coefficient are hence not comparable for different cell sizes without correction. The overall effect of the system-size dependence for a cubic simulation

box is a decrease in the diffusion coefficient, following the scaling:

$$D_{s,\mathrm{PBC}} = D_{s,0} - \frac{k_\mathrm{B}T\xi}{6\pi\eta L} \tag{2.36}$$

where η is the liquid viscosity and L is the cubic cell length. The expression was derived from the use of the Oseen tensor to describe the hydrodynamic interactions between particles and the Ewald summation to account for periodicity.[287,288] The value of ξ is approximately 2.837, whose dominant contribution comes from the background force density that ensures the overall conservation of momentum.

The relative permittivity or dielectric constant, $\epsilon_r = 78.4$, of liquid water is a measure of the orientational correlations throughout the liquid, from which the electric dipole moments can be deduced. A good water model should be able to reproduce simultaneously the aforementioned value for the dielectric constant and an average dipole moment around 3.0 Debye units, when polarisation effects are included. However, most water models with a polarisation term tend to overestimate the dielectric constant. In contrast, some non-polarisable force fields have been observed to be quite successful, *e.g.* TIP5P, SPC/F, TIP4P/FQ, but this is most likely due to the fortuitous cancellation of errors as none of them were parameterised to reproduce the dielectric constant.[289] Unfortunately, reproducing these two properties do not go hand in hand, as shown by TIP3P and SPC/E which have very different dielectric constants despite sharing the same average dipole moment. Their different H-O-H angles were shown to be responsible for this.[290] When reproducing the dielectric constant in simulations, its value has demonstrated a measurable dependence on how the long-range interactions are taken into account, *e.g.* Ewald summation, reaction field.[291] Since the reaction field method is more uncertain and tends to give a smaller value,[292] investigations of water models using the Ewald summation are more comparable. We also note that the dielectric constant is slow to converge near ambient conditions.[293]

2.2.2 Investigated Water Force Fields

Simulations of water were pioneered by Watts & Barker[294] and Rahman & Stillinger.[295] In empirical water simulations, the choice of water force field to model the interaction between water molecules is important.[296–299] The water force fields investigated here are the popular non-polarisable variants from the simple point charge (SPC) group and the transferable intermolecular potentials with n interaction sites (TIPnP), *i.e.* SPC and SPC/Fw for the former and TIP3P, TIP3P/Fs and TIP4P/2005 for the latter. Their parameters are detailed in Table 2.2.

Table 2.2: The parameters of selected LJ-based water force fields. q_M is the charge assigned to the additional fictitious site for TIP4P/2005, located 0.1546 Å from the oxygen atom at the bisector of the H-O-H angle. A long-range vdW tail correction is applied to SPC and SPC/Fw.

Model	$k_{OH}/2$ [eV·Å^{-2}]	R^0_{OH} [Å]	$k_{\angle HOH}/2$ [eV·rad^{-2}]	$\theta^0_{\angle HOH}$ [°]	ϵ_{OO} [$\times 10^{-3}$eV]	σ_{OO} [Å]	q_O [e]	q_M [e]
SPC	–	1.0000	–	109.47	6.7390	3.1655	-0.8200	0
SPC/Fw	22.9648	1.0120	1.6457	113.24	6.7390	3.1655	-0.8200	0
TIP3P	–	0.9527	–	104.52	6.5960	3.1506	-0.8340	0
TIP3P/Fs	22.9648	0.9600	1.4763	104.50	6.6000	3.1506	-0.8340	0
TIP4P/2005	–	0.9572	–	104.52	8.0314	3.1589	0	-1.1128

Despite being among the first, SPC and TIP3P remain as standard water force fields.[47] This can be attributed to the development of many potentials for electrolyte solutions and organic molecules in conjunction with them. Both of those simple force fields are based on the minimalist transferable intermolecular potentials (TIPS) form[257] with three point charges located on the atomic nuclei. A LJ potential acts between the water oxygen atoms with a spherical truncation at a given O...O distance. This cutoff was 9.0 Å for 216 water molecules using SPC and 7.5 Å for 125 water molecules using TIP3P. The parameters of SPC were chosen to reproduce some gas phase properties, *e.g.* dipole moment and dimer interaction energy, or recover some liquid state properties at ambient conditions, *e.g.* heat of vapourisation and density. TIP3P was parameterised to reproduce the density of water at room temperature and pressure as well as the vapourisation enthalpy at room temperature. Most results, *e.g.* density, energy and structure, from fixed-charge water force fields show little variation for system sizes above ∼125 molecules and cutoffs beyond ∼8.0 Å.[300,301]

Unlike TIP3P, which has a geometry close to the experimental geometry of the water molecule, SPC has a greater O-H bond length and the ideal tetrahedral angle for the H-O-H bond angle. Their geometries are rigid preventing bond stretching and angle bending. Although many bulk water properties under ambient conditions are well reproduced by SPC and TIP3P, the self-diffusion coefficient and the static dielectric constant are poorly described. The introduction of flexibility allowing internal changes to the conformation of the water molecule attempts to improve this behaviour.[266] Their flexible variants are SPC/Fw and TIP3P/Fs respectively. The molecular geometry of TIP3P/Fs remained similar to its rigid counterpart, whereas SPC/Fw geometry was optimised to fit the experimental bulk diffusion and dielectric constants.

A four-site variant of the TIPnP proposed in 2005 is perhaps the most satisfactory of the rigid non-polarisable water force fields and has been used to discuss their limitations.[302] This TIP4P/2005 force field has similar geometries to TIP3P, but different charge distributions. It has the negative charge moved to a fictious

site 0.15 Å from the oxygen atom along the bisector of the H-O-H bond angle that improves the quadrupole moment and liquid structure, which is similar to the BF model.[45] The additional fictive site mean that ten distances are required to evaluate the interactions between pairs instead of the nine for a three-site model.[45] TIP4P/2005 has been parameterised to reproduce the room pressure isobar densities and some properties related to the density and stability of the ice polymorphs. Despite the lack of parametrisation against the vapourisation enthalpy at room temperature, it is able to reproduce this value well when a polarisation correction is applied.

2.3 Integral Equation Theory

The integral equation theory is a statistical mechanics approach to predicting the bulk properties of classical liquids from interactions between its constituent particles. The most prominent theory in the theoretical studies of homogeneous liquids is the Ornstein-Zernike (OZ) equation. Its formalism is exact but requires closure relations to provide an approximate analytical relationship between the pair correlation function and the pair potential. For LJ liquids, the closure relation scheme proposed by Zerah & Hansen[104] yields good results for the structural and thermodynamical properties. In statistical mechanics, the inversion of the OZ equation has been used to derive potentials from some of the first scattering experimental data.

2.3.1 Ornstein-Zernike Integral

Ornstein and Zernike initially proposed the integral equation theory as an approach to describe the density fluctuations in classical fluids near the critical point.[103] They supposed the molecules to be spherical and rigid particles in a homogeneous and isotropic distribution. The structure is then spherically symmetric as for monoatomic liquids and hence, dependent only on the separation of the particles. The OZ equation can be derived using functional calculus from the partition function and defines the direct correlation function $c(r)$ as:

$$h(r) = c(r) + \rho \int c(r')h(|r - r'|)dr' \tag{2.37}$$

where r is the separation between particles and ρ is the number density. The total correlation function $h(r)$ relates to the fluid structure via:

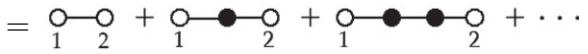
$$h(r) = g(r) - 1 \tag{2.38}$$

where $g(r)$ is the pair distribution function. The second term in Eq. (2.37), the indirect correlation, takes the form of a convoluted integral. The Fourier transform of this convolution is simply the pointwise product of Fourier transforms. Thus, the Fourier transform of the OZ equation defined in k-space is routinely used:

$$\hat{h}(k) = \hat{c}(k) + \rho\hat{c}(k)\hat{h}(k) \tag{2.39}$$

The OZ equation can also be derived using diagrammatic methods of which an introductory account is given by Hansen & McDonald.[303] Not only does a diagram representation (called bridge or elementary diagrams) enable the use of graph theoretical methods, it lends itself to illustrate the physical interpretation of the OZ equation:

$h(1,2) =$ [all bridge diagrams consisting of two terminal white circles labelled 1 and 2, black circles and bonds] (2.40)



The bridge diagrams in Eq. (2.40) are reprinted from J. Hansen and I. McDonald. "Chapter 3 - Static Properties of Liquids: Thermodynamics and Structure". In: *Theory of Simple Liquids (Fourth Edition)*. Ed. by J. P. Hansen and I. R. McDonald. Fourth Edition. Oxford: Academic Press, 2013, pp. 61 –104, with permission from Elsevier. It describes the total correlation between a pair of particles 1 and 2, represented by $h(1,2)$, which is due in part to their direct correlation between 1 and 2 or $c(1,2)$, illustrated by the first diagrammatic term, and also the indirect correlation propagated by increasing numbers of intermediate particles, illustrated by the subsequent diagrammatic terms in the series expansion.[303] The indirect correlation can be written as $\rho \int c(1,3)h(3,2)d3$ in the formalism using particle indices $i = 1, 2, 3, \ldots$ which represent the position R_i, such that $(1,2) \equiv (R_{1,2}) = (|R_1 - R_2|)$ and $d3 \equiv dR_3$.

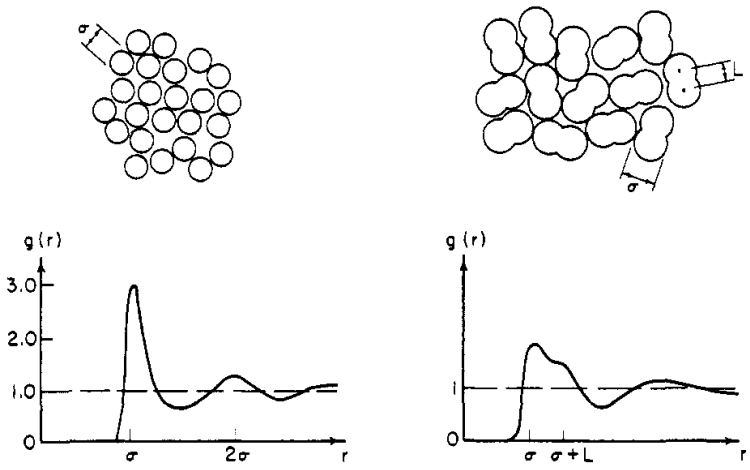


Figure 2.2: Radial distribution functions and regions of a dense fluid for an atomic fluid (left) and a homonuclear diatomic fluid (right).[304] Reprinted with permission from D. Chandler and D. M. Richardson. "Theory of Orientational Pair Correlations in Molecular Fluids". In: *The Journal of Physical Chemistry* 87.12 (1983), pp. 2060–2064, copyright American Chemical Society.

2.3.2 Reference Interaction Site Model

Molecules characteristically have explicit features that do not conform to a spherical shape. Figure 2.2 illustrates that as opposed to an atomic packing, the coupling of the intramolecular length to the intermolecular length in diatomic molecules produces a shoulder or bump in the RDF after the first peak and broader subsequent extrema, *i.e.* produces a more complicated packing. The constraint in configurational space due to molecular bonds leads to angular dependence in the correlation functions. Generalisation of the OZ equation in Eq. (2.37) is thus necessary to capture the molecular orientations, which can be written in the formalism using particle indices as:

$$h(1,2) = c(1,2) + \frac{\rho}{\Omega} \int c(1,3)h(3,2)d3 \tag{2.41}$$

where $\Omega \equiv \int d\Omega_i$, the unweighted integral over the angles for a molecule, depends on the linearity of the molecular geometry so that:

$$\Omega = \begin{cases} \iint d(\cos\theta_i)d\phi_i = 4\pi & \text{, if linear} \\ \iiint d(\cos\theta_i)d\phi_i d\chi_i = 8\pi^2 & \text{, if non-linear} \end{cases} \tag{2.42}$$

Since the molecular OZ equation depends on position R_i and orientation Ω_i, the particle index is introduced as $i \equiv (R_i, \Omega_i)$. The position conveniently corresponds with the molecular centre of mass or a coordinate of high symmetry in the molecule. The orientation is expressed either in polar angles (θ_i, ϕ_i) for linear

molecules or in Euler angles $(\theta_i, \phi_i, \chi_i)$ for non-linear molecules. Extending the formalism to Eq. (2.38), its molecular analogue can simply be written as:

$$h(1,2) = g(1,2) - 1 \tag{2.43}$$

Based on the argument that an interaction potential can be sufficiently described by the site-site correlation function, the molecular OZ equation is approximated in terms of the separation between sites. This site-site OZ equation or reference interaction site model (RISM) contains the intramolecular correlation function ω, so that its Fourier transform is given by:

$$\hat{h} = \hat{\omega}\hat{c}\hat{\omega} + \hat{\omega}\rho\hat{c}\hat{h} \tag{2.44}$$

2.3.3 Closure Relations

Since the functions $c(r)$ and $h(r)$ are unknowns, another relation between those two is necessary to complete a closed set of equations. A formally exact form of the required closure relation is:

$$g(r) = \exp\left[-\frac{1}{k_\mathrm{B}T}V(r) + h(r) - c(r) + B(r)\right] \tag{2.45}$$

where k_B is the Boltzmann constant and T is the absolute temperature. A potential $V(r)$ acts through the closure relation so that the pair distribution $g(r)$ can be approximated. An inversion must then hold true such that given a pair distribution, an effective potential can be extracted. $B(r)$ is the so-called "bridge" function that can be expressed as a sum over the infinite set of bridge diagrams.[104] The simplification of Eq. (2.45) to tractable forms leads to well-known closures, such as the hypernetted-chain (HNC) and Percus-Yevick (PY) approximations.

When $B(r) \equiv 0$, the closure relation becomes the HNC approximation that is well suited to long-range potentials and in particular Coulombic systems. The exact closure is reduced to:

$$g(r) = \exp\left[-\beta v(r)\right] \exp\left[h(r) - c(r)\right] \tag{2.46}$$

Upon linearisation of the second exponential in the HNC equation, the closure relation becomes the PY approximation. It is best applied to describe the correlations between atoms with strong repulsive short-range forces, *i.e.* the hard sphere model. The exact closure is further simplified to:

$$g(r) = \exp\left[-\beta v(r)\right][1 + h(r) - c(r)] \tag{2.47}$$

The interpolation between these two approximations results in the soft-core mean spherical approximation (SMSA), which is adapted for interatomic potentials with a soft core repulsion and an attractive potential tail.

2.4 Iterative Boltzmann Inversion

The large number of degrees of freedom (DoFs) is a limiting factor in the investigations of fully atomistic models. Naturally, a direct approach to this problem is mapping the atomistic model to the mesoscopic scale to remove the superfluous DoFs. This CG model maintains the relevant DoFs required to retain the chemical identity of the parent atomistic model. Coarse-graining has been applied extensively in polymer systems.

The IBI method[107] introduces an automatic optimisation scheme to extract effective potentials in a standardised framework. Soper[305] proposed to invert the RDF for one-component simple liquid systems by taking a simple Boltzmann inverse of the pair correlation function. This is exact only in the limit of infinitely dilute systems.

2.4.1 Single-State

In a canonical ensemble, the correlation functions that depend on a single coordinate, q, obey the Boltzmann distribution. Therefore, a simple relation between a structural property, $P(q)$, and potential energy, $U(q)$, can be derived:

$$P(q) = Z^{-1}\mathrm{e}^{-U(q)/k_\mathrm{B}T} \tag{2.48}$$

where k_B is the Boltzmann constant and T is the absolute temperature. The correlation functions include pair $g(r)$, bending angle $a(\alpha)$ and dihedral angle $b(\beta)$ distributions. Z is the partition function that becomes an additive constant for the inverse function of Eq. (2.48). Thus, the Boltzmann inversion of a pair distribution g gives the potential of mean force (PMF) V:

$$V(r) = -k_\mathrm{B}T \ln g(r) \tag{2.49}$$

where r is the radial distance between two particles. Eq. (2.49) can be generalised to the other correlation functions in a straightforward manner. The corresponding

force between the particles is simply obtained from its derivative:

$$F(r) = -\nabla V(r) \tag{2.50}$$

A pair interaction potential V generated from the Boltzmann inversion of a given target RDF g^* does not directly reproduce that RDF when the potential is used in a MD simulation. This is due to the higher-order correlation effects. It does, however, provide a reasonable initial guess:

$$V_0(r) = -k_\mathrm{B}T \ln g^*(r) \tag{2.51}$$

that can be refined in an iterative sequence until convergence to the target RDF is reached within some set tolerance, through a self-consistent equation:

$$V_{i+1}(r) = V_i(r) - \alpha k_\mathrm{B}T \ln \left[\frac{g^*(r)}{g_\mathrm{i}(r)}\right] \tag{2.52}$$

where subscript i is the iteration number and $*$ denotes the target system. Likewise, the equation can be generalised to optimise for other structural properties of interest. Since each iteration in this IBI method is prone to overcorrection, α scales the intensity of the corrective potential, where smaller values are used for dense and/or crystalline states while unity is used for the amorphous state.[306]

Despite the ability of IBI to achieve the target RDF, the corresponding potential is non-unique. Using different cutoffs for the trial potentials, Reith *et al.*[107] generated distinct potentials that are capable of reproducing essentially the same RDF within the set tolerance. They further investigated the numerical stability of the IBI method by using a known potential that was then padded with trailing zeros. Nonzero artefacts were observed in the resultant potential as a systematic error in the domain of the trailing zeros. When zooming into the scale of the mesh, unphysical fluctuations are prominent in the intermediate to long-range distances. Thus, smoothing of the trial potentials and the trial RDFs was recommended to reduce statistical noise.

The constant volume IBI simulation results in a potential that has a positive pressure for the CG system, which is contrary to the ambient conditions of the initial atomistic system. When coarse-grained, the virial pressures of TIP3P, SPC and SPC/E increased from 1 bar, 0.82 bar, 0.76 bar to as high as 8536 bar, 8994 bar and 9886 bar respectively. The internal structure is lost when water molecules are substituted with beads and with it, the related interactions contributing to pressure. This is well known for CG simulations.[307,308] Pressure correction can be applied as a post-optimisation process to amend the compressibility without

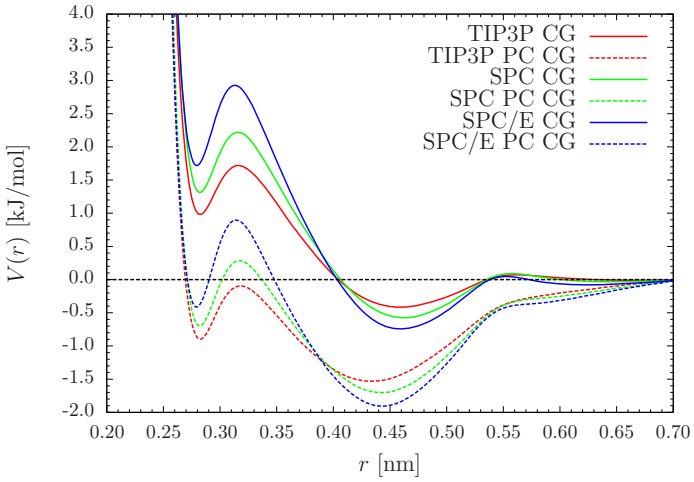


Figure 2.3: Effective one-water bead CG potentials of water force fields: TIP3P, SPC and SPC/E without pressure correction (solid) and with pressure correction (dashed).[309] H. Wang, C. Junghans, and K. Kremer. "Comparative Atomistic and Coarse-Grained Study of Water: What do we Lose by Coarse-Graining?" In: *The European Physical Journal E* 28.2 (2009), pp. 221–229 is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

severely affecting the agreement of the RDF. Reith *et al.*[107] proposed a linear function:

$$\Delta V(r) = A\left(1 - \frac{r}{r_{\text{cutoff}}}\right) \tag{2.53}$$

where $A = -0.1\,k_BT$. The function is A at $r = 0$ and 0 at $r = r_{\text{cutoff}}$. Using the corrected potential as the initial guess, it is then re-optimised until the pressure reflects that of the parent atomistic system. This shifts the tail of the potential downward providing the missing long-range attraction.

In a study by Wang *et al.*[309] of the tradeoff from coarse-graining rigid three-site water force fields using the IBI method, A was instead estimated based on the virial expression of pressure P_i of the i-th step. In order for the corrected potential to match the correct pressure, A_i must satisfy:

$$-\left[\frac{2\pi N\rho}{3r_{\text{cut}}}\int_0^{r_{\text{cut}}} r^3 g_i(r)dr\right] A_i \approx (P_i - P_{\text{target}})V \tag{2.54}$$

By representing each water molecule as a CG bead, a speed-up in computational time up to the order of 50 was obtained. The CG potentials, including those with pressure correction, generated for the water force fields TIP3P, SPC and SPC/E are shown in Figure 2.3. They noted difficulty in simultaneously adjusting for the pressure and the compressibility through the isotropic two-body potentials between beads. In fact, the pressure correction had increased the isothermal

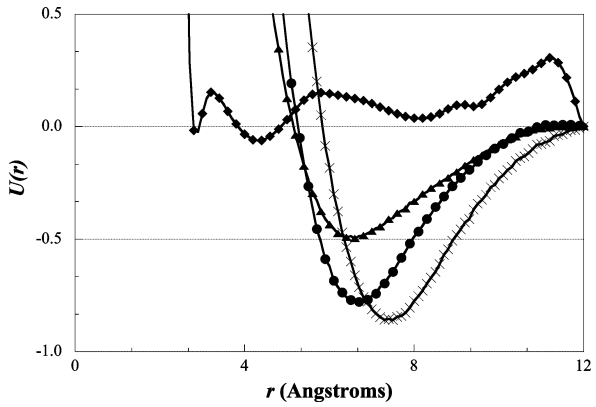


Figure 2.4: Interaction potential between one-water beads (rhombus), four-water beads (triangle), six-water beads (circle) and eight-water beads (cross).[5] Reprinted with permission from K. R. Hadley and C. McCabe. "On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties". In: *The Journal of Physical Chemistry B* 114.13 (2010). PMID: 20230012, pp. 4590–4599, copyright American Chemical Society.

compressibility by an order of magnitude. One significant detrimental effect of the CG water force fields was the increase by an order of magnitude to the diffusion coefficients compared to their all-atom analogue and the experimental value. In order to overcome the shortcomings of coarse-graining, the AdRes method[310] that allows free exchange of molecules and their representation was proposed.

In an effort to further extend CG water to greater temporal and spatial scales, Hadley & McCabe[5] investigated the optimal number of water molecules per bead that retained the structural and solvation properties. This is performed using a K-means algorithm to determine the clustering of one to nine water molecules. Figure 2.4 shows the difference between the potentials of the multiple water molecules per bead models derived using the IBI method. The location of the energy minimum was noted to shift to a larger distance as the size of the bead increases. The four-water bead model was found to have the optimal balance between computational efficiency and its ability to reproduce the density of pure water as well as to solvate the 1-pentanol solute.

2.4.2 Multi-State Extension

The IBI method is limited to optimising a potential for a single property of interest. The derived potential is heavily state-dependent, *e.g.* density, temperature and composition, in terms of transferability. As each distribution is also dependent on every other potential in the system, potentials cannot be iterated independently to reproduce a set of properties.

In order to improve transferability of the potential derived from IBI, other

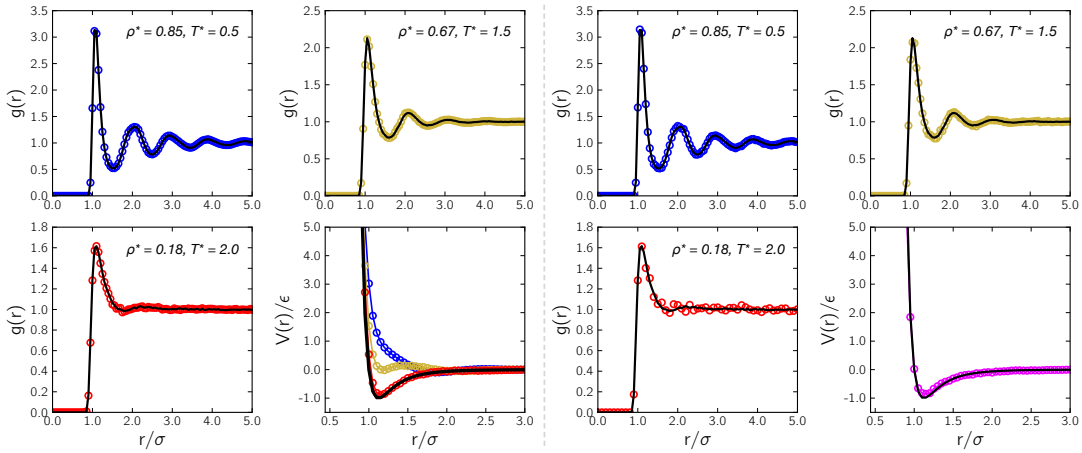


Figure 2.5: Given three states for a LJ system, varying from a gaseous phase to a dense liquid, three separate potentials were derived for each state using single-state IBI (left) and a single potential was derived from multiple states using MS IBI (right).[112] Reprinted from T. C. Moore, C. R. Iacovella, and C. McCabe. "Derivation of Coarse-Grained Potentials via Multistate Iterative Boltzmann Inversion". In: *The Journal of Chemical Physics* 140.22, 224104 (2014), with the permission of AIP publishing.

thermodynamic states of a system can be considered. The multi-state extension was constructed based on two key assumptions that enable the incorporation of other states into the single-state IBI. Firstly, there exists a region in the potential phase space that returns the target RDF for each thermodynamic state. Secondly, the "true" underlying potential can be found within the overlap between these regions. It stands to reason that a state-independent potential can be derived through constraining the sampling of the phase space to the overlap. This is achieved in multistate iterative Boltzmann inversion (MS IBI) by updating a single potential to simultaneously match distinct target RDFs from sufficiently different states. However, large disparities between states may not give rise to the desired overlap. A comparison between the results of single-state IBI and MS IBI for a LJ system, varying from a gaseous phase to a dense liquid, is shown in Figure 2.5.

The single potential starts off as the Boltzmann inversions of the target RDFs, averaged over the N states:

$$V_{\mathrm{s},0}(r) = -\frac{1}{N}\sum_{\mathrm{s}} k_{\mathrm{B}} T_{\mathrm{s}} \ln g^*_{\mathrm{s}}(r) \tag{2.55}$$

where s denotes the state. In Moore *et al.*[112] a set of three states were selected to test the performance of the method. The same potential is applied to each

thermodynamic state and is iteratively updated until a set tolerance is reached:

$$V_{\mathrm{s},i+1}(r) = V_{\mathrm{s},i}(r) - \frac{1}{N}\sum_{\mathrm{s}} \alpha_{\mathrm{s}}(r) k_{\mathrm{B}} T_{\mathrm{s}} \ln\left[\frac{g^{*}_{\mathrm{s}}(r)}{g_{\mathrm{s},i}(r)}\right] \tag{2.56}$$

where $\alpha_{\mathrm{s}}(r)$ is a linear function used to adjust the relative weight exerted by each state so that emphasis can be given to the subset of states of most interest. The function decays linearly from some intensity at no separation to zero at the interaction cutoff distance. The short-range interactions become more prominent in contrast to the long-range interactions, which is reasonable given that short-range interactions may induce long-range correlations. MS IBI further provides a systematic framework to tune the potential to reflect the bulk system properties of interest through the relative weights, as in the case of the tilt angle and nematic order of an *n*-dodecane monolayer.[112] However, approximately 50 iterations were required for a standard potential optimisation to be well converged.

Moore *et al.*[113] later applied MS IBI to develop a CG potential for a 4:1 mapped CG water model based on the TIP3P force field. Rather than using multiple thermodynamics states, they used bulk canonical ensemble (NVT) and bulk isothermal-isobaric ensemble (NPT) to account for the density-pressure relationship. This allowed the resulting potential to reproduce the structure and density of water. The inclusion of the droplet NVT state further constrained the potential to reproduce the interfacial properties. Additionally, when compared to the MARTINI CG water model[311] that is known to spontaneously crystallise at physiological conditions, the MS IBI potential did not demonstrate such tendency.

2.5 Experiment Directed Simulation

Iterative methods, such as IBI, have been applied to build molecular force fields consistent with experiment data.[312–314] Alternatively, we can bias an approximate molecular simulation to match experimental data.[315–319] One prominent example is the use of harmonic constraints in refining NMR structures.[320–322] Unfortunately, the constraint may also alter the dynamics of the system. An improvement to the harmonic constraint, called the restrained-ensemble, was introduced to reduce the undesired alterations.[317,323] Compared with the iterative method, this approach reduces wall-clock time through parallel MD simulations of a number of replicas in the presence of a bias potential. However, the harmonic bias and restrained-ensemble add terms that are quadratic in the biased quantity. Using

a maximum entropy argument, a linear bias can be derived that changes the statistical ensemble the least.[323] This approach requires a single-replica and has the same properties as an infinite-replica restrained-ensemble.[318] These approaches all require a coupling constant to set the strength of the bias.

The experiment directed simulation (EDS)[97] method brings together the harmonic constraint for adjustable parameters while avoiding undesirable alterations and the linear bias for minimal perturbation to the statistical ensemble. It requires a single-replica simulation and can bias multiple correlated or uncorrelated variables simultaneously. Similarly, this approach requires a coupling constant, α, defined such that some expected value $f(r)$ is equal to the experimental value, $\bar{f}$, under the potential $U'(r,\alpha) = U(r) + (\alpha/\bar{f}) \cdot f(r)$. Instead of optimising the potential point-by-point, only the single variable α is optimised for the appropriate strength of the bias. The expected value $f(r)$ is described by the probability distribution function (Eq. (2.48)) so that its functional derivative with respect to the potential energy $U(r')$ can be written as:

$$\frac{\partial}{\partial U(r')} f(r) = \frac{\partial}{\partial U(r')} \frac{\int dr f(r) \mathrm{e}^{-\beta U(r)}}{Z} \tag{2.57}$$

Via the definition of the coupling constant, the functional derivative becomes:

$$\frac{\partial (\langle f(r) \rangle - \bar{f})^2}{\partial \alpha} = -\frac{2\beta}{\bar{f}} \langle f(r) - \bar{f} \rangle \operatorname{Var}(f(r)) \tag{2.58}$$

where a gradient-based method can be used to optimise the coupling constant that can be extended to any number of collective variables, as hypothesised by Pitera & Chodera.[318]

When applied to match the RDF $g(r)$ from more accurate simulations or scattering experiments, an integral transformation is used to derive the coordination number collective variable from the RDF:

$$\langle N(r_0) \rangle = \rho \int_0^R dr [1 - u(r - r_0)] 4\pi r^2 g(r) \tag{2.59}$$

Since matching the coordination number is similar to matching the zeroth moment or the area under the first peak, the peak means, widths and skews are not guaranteed to be reproduced. Thus, multiple higher moments of the RDF must also be matched. White & Voth[97] investigated the matching of multiple moments of a RDF for LJ particles as shown in Figure 2.6, which demonstrated its successful application for highly coupled biases.

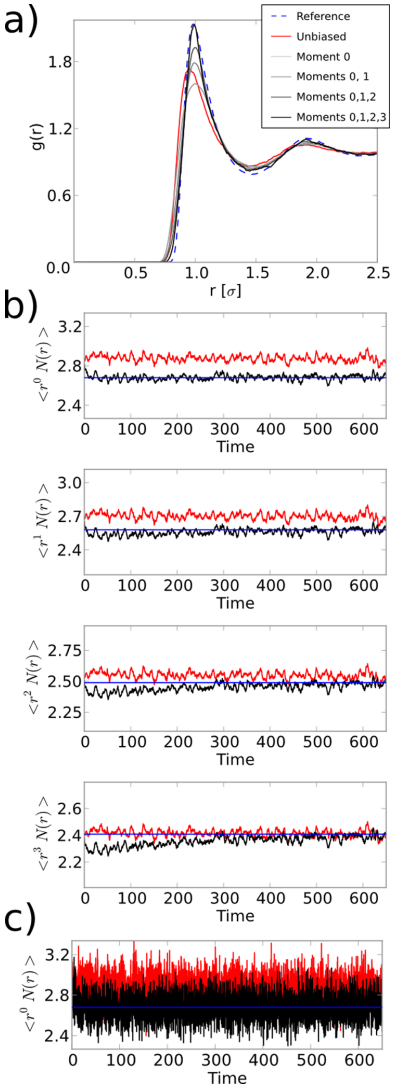


Figure 2.6: RDF for a LJ simulation under different numbers of biases. (a) shows that matching up to the third moment provides good agreement with the reference RDF. (b) shows the convergence of the biased moments over time. (c) shows the unsmoothed zeroth moment and bias over time.[97] Reprinted with permission from A. D. White and G. A. Voth. "Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data". In: *Journal of Chemical Theory and Computation* 10.8 (2014). PMID: 26588273, pp. 3023–3030, copyright American Chemical Society.

References

[5] K. R. Hadley and C. McCabe. "On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties". In: *The Journal of Physical Chemistry B* 114.13 (2010). PMID: 20230012, pp. 4590–4599. DOI: 10.1021/jp911894a.

[18] G. Hura et al. "A High-Quality X-ray Scattering Experiment on Liquid Water at Ambient Conditions". In: *The Journal of Chemical Physics* 113.20 (2000), pp. 9140–9148. DOI: 10.1063/1.1319614.

[21] A. K. Soper. "The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure?" In: *ISRN Physical Chemistry* 2013.279463 (2013), p. 67. DOI: 10.1155/2013/279463.

[41] J. D. Bernal and R. H. Fowler. "A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions". In: *The Journal of Chemical Physics* 1.8 (1933), pp. 515–548. DOI: 10.1063/1.1749327.

[44] H. J. C. Berendsen et al. "Interaction Models for Water in Relation to Protein Hydration". In: *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*. Ed. by B. Pullman. Dordrecht: Springer Netherlands, 1981, pp. 331–342. DOI: 10.1007/978-94-015-7658-1_21.

[45] W. L. Jorgensen et al. "Comparison of Simple Potential Functions for Simulating Liquid Water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: 10.1063/1.445869.

[47] B. Guillot. "A Reappraisal of What We Have Learnt During Three Decades of Computer Simulations on Water". In: *Journal of Molecular Liquids* 101.1 (2002). Molecular Liquids. Water at the New Millenium, pp. 219 –260. DOI: 10.1016/S0167-7322(02)00094-6.

[55] M. Del Ben et al. "Bulk Liquid Water at Ambient Temperature and Pressure from MP2 Theory". In: *The Journal of Physical Chemistry Letters* 4.21 (2013), pp. 3753–3759. DOI: 10.1021/jz401931f.

[56] M. Del Ben, J. Hutter, and J. VandeVondele. "Probing the Structural and Dynamical Properties of Liquid Water with Models including Non-local Electron Correlation". In: *The Journal of Chemical Physics* 143.5 (2015), p. 054506. DOI: 10.1063/1.4927325.

[68] F. Sim et al. "Gaussian Density Functional Calculations on Hydrogen-Bonded Systems". In: *Journal of the American Chemical Society* 114.11 (1992), pp. 4391–4400. DOI: 10.1021/ja00037a055.

[69] K. Laasonen, F. Csajka, and M. Parrinello. "Water Dimer Properties in the Gradient-Corrected Density Functional Theory". In: *Chemical Physics Letters* 194.3 (1992), pp. 172 –174. DOI: 10.1016/0009-2614(92)85529-J.

[70] K. Laasonen et al. "Structures of Small Water Clusters using Gradient-Corrected Density Functional Theory". In: *Chemical Physics Letters* 207.2 (1993), pp. 208 –213. DOI: 10.1016/0009-2614(93)87016-V.

[71] B. Santra, A. Michaelides, and M. Scheffler. "On the Accuracy of Density-Functional Theory Exchange-Correlation Functionals for H Bonds in Small Water Clusters: Benchmarks Approaching the Complete Basis Set Limit". In: *The Journal of Chemical Physics* 127.18 (2007), p. 184104. DOI: 10.1063/1.2790009.

[72] D. R. Hamann. "H_2O Hydrogen Bonding in Density-Functional Theory". In: *Phys. Rev. B* 55 (16 1997), R10157–R10160. DOI: 10.1103/PhysRevB.55.R10157.

[73] P. J. Feibelman. "Lattice Match in Density Functional Calculations: Ice Ih*vs.*β-AgI". In: *Phys. Chem. Chem. Phys.* 10 (32 2008), pp. 4688–4691. DOI: 10.1039/B808482N.

[74] K. Laasonen et al. ""Ab Initio" Liquid Water". In: *The Journal of Chemical Physics* 99.11 (1993), pp. 9080–9089. DOI: 10.1063/1.465574.

[75] M. Sprik, J. Hutter, and M. Parrinello. "Ab Initio Molecular Dynamics Simulation of Liquid Water: Comparison of Three Gradient-Corrected Density Functionals". In: *The Journal of Chemical Physics* 105.3 (1996), pp. 1142–1152. DOI: 10.1063/1.471957.

[76] P. L. Silvestrelli and M. Parrinello. "Structural, Electronic, and Bonding Properties of Liquid Water from First Principles". In: *The Journal of Chemical Physics* 111.8 (1999), pp. 3572–3580. DOI: 10.1063/1.479638.

[78] J. C. Grossman et al. "Towards an Assessment of the Accuracy of Density Functional Theory for First Principles Simulations of Water". In: *The Journal of Chemical Physics* 120.1 (2004), pp. 300–311. DOI: 10.1063/1.1630560.

[82] M. J. Gillan, D. Alfè, and A. Michaelides. "Perspective: How Good is DFT for Water?" In: *The Journal of Chemical Physics* 144.13 (2016), p. 130901. DOI: 10.1063/1.4944633.

[84] J. A. Morrone and R. Car. "Nuclear Quantum Effects in Water". In: *Phys. Rev. Lett.* 101 (1 2008), p. 017801. DOI: 10.1103/PhysRevLett.101.017801.

[90] R. A. DiStasio et al. "The Individual and Collective Effects of Exact Exchange and Dispersion Interactions on the Ab Initio Structure of Liquid Water". In: *The Journal of Chemical Physics* 141.8 (2014), p. 084502. DOI: 10.1063/1.4893377.

[97] A. D. White and G. A. Voth. "Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data". In: *Journal of Chemical Theory and Computation* 10.8 (2014). PMID: 26588273, pp. 3023–3030. DOI: 10.1021/ct500320c.

[103] L. S. Ornstein and F. Zernike. "Accidental Deviations of Density and Opalescence at the Critical Point of a Single Substance". In: *Proc. Acad. Sci. Amsterdam.* Vol. 17. 2. 1914, pp. 793–806.

[104] G. Zerah and J. Hansen. "Self-Consistent Integral Equations for Fluid Pair Distribution Functions: Another Attempt". In: *The Journal of Chemical Physics* 84.4 (1986), pp. 2336–2343. DOI: 10.1063/1.450397.

[105] R. Rajagopalan and K. Srinivasa Rao. "Interaction Forces in Charged Colloids: Inversion of Static Structure Factors". In: *Phys. Rev. E* 55 (4 Apr. 1997), pp. 4423–4432. DOI: 10.1103/PhysRevE.55.4423.

[107] D. Reith, M. Pütz, and F. Müller-Plathe. "Deriving Effective Mesoscale Potentials from Atomistic Simulations". In: *Journal of Computational Chemistry* 24.13 (2003), pp. 1624–1636. DOI: 10.1002/jcc.10307.

[112] T. C. Moore, C. R. Iacovella, and C. McCabe. "Derivation of Coarse-Grained Potentials via Multistate Iterative Boltzmann Inversion". In: *The Journal of Chemical Physics* 140.22, 224104 (2014). DOI: 10.1063/1.4880555.

[113] T. C. Moore, C. R. Iacovella, and C. McCabe. "Development of a Coarse-Grained Water Forcefield via Multistate Iterative Boltzmann Inversion". In: *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*. Ed. by R. Q Snurr, C. S. Adjiman, and D. A. Kofke. Singapore: Springer Singapore, 2016, pp. 37–52. DOI: 10.1007/978-981-10-1128-3_3.

[114] F. Jensen. *Introduction to Computational Chemistry*. John Wiley & Sons, 1999.

[115] C. J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. 2nd. John Wiley & Sons, 2004.

[116] M. Rupp et al. "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning". In: *Phys. Rev. Lett.* 108 (5 2012), p. 058301. DOI: 10.1103/PhysRevLett.108.058301.

[117] K. Hansen et al. "Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies". In: *Journal of Chemical Theory and Computation* 9.8 (2013). PMID: 26584096, pp. 3404–3419. DOI: 10.1021/ct400195d.

[118] G. Montavon et al. "Machine Learning of Molecular Electronic Properties in Chemical Compound Space". In: *New Journal of Physics* 15.9 (2013), p. 095003. DOI: 10.1088/1367-2630/15/9/095003.

[119] A. R. Leach. *Molecular Modelling: Principles and Applications*. 2nd. Harlow: Prentice Hall, 2001. 744 pp.

[120] T. P. Senftle et al. "The ReaxFF Reactive Force-Field: Development, Applications and Future Directions". In: *Npj Computational Materials* 2 (Mar. 2016), p. 15011.

[121] E. Scrhödinger. *Collected Papers on Wave Mechanics*. 3rd. AMS Chelsea Publishing, 1982. 208 pp.

[122] H. Sommer, H. A. Thomas, and J. A. Hipple. "Values of μ_p, F, and $\frac{M_p}{m_e}$ using the Omegatron". In: *Phys. Rev.* 80 (3 Nov. 1950), pp. 487–487. DOI: 10.1103/PhysRev.80.487.

[123] E. Fermi. "On the Quantization of the Monoatomic Ideal Gas". In: *Rendiconti dell'Accademia Nazionale dei Lincei* (1926).

[124] C. C. J. Roothaan. "New Developments in Molecular Orbital Theory". In: *Rev. Mod. Phys.* 23 (2 1951), pp. 69–89. DOI: 10.1103/RevModPhys.23.69.

[125] G. G. Hall. "The Molecular Orbital Theory of Chemical Valency VIII. A Method of Calculating Ionization Potentials". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 205.1083 (1951), pp. 541–552. DOI: 10.1098/rspa.1951.0048.

[126] W. Kohn, A. D. Becke, and R. G. Parr. "Density Functional Theory of Electronic Structure". In: *The Journal of Physical Chemistry* 100.31 (1996), pp. 12974–12980. DOI: 10.1021/jp9606691.

[127] S. Kümmel and L. Kronik. “Orbital-Dependent Density Functionals: Theory and Applications”. In: *Rev. Mod. Phys.* 80 (1 Jan. 2008), pp. 3–60. DOI: 10.1103/RevModPhys.80.3.

[128] W. Kohn and C. D. Sherrill. “Editorial: Reflections on Fifty Years of Density Functional Theory”. In: *The Journal of Chemical Physics* 140.18, 18A201 (2014), pp. –. DOI: 10.1063/1.4870815.

[129] P. Hohenberg and W. Kohn. “Inhomogeneous Electron Gas”. In: *Phys. Rev.* 136 (3B Nov. 1964), B864–B871. DOI: 10.1103/PhysRev.136.B864.

[130] W. Kohn and L. J. Sham. “Self-Consistent Equations including Exchange and Correlation Effects”. In: *Phys. Rev.* 140 (4A Nov. 1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133.

[131] T. Bally and G. N. Sastry. “Incorrect Dissociation Behavior of Radical Ions in Density Functional Calculations”. In: *The Journal of Physical Chemistry A* 101.43 (1997), pp. 7923–7925. DOI: 10.1021/jp972378y.

[132] T. Koopmans. “ber die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms”. In: *Physica* 1.1 (1934), pp. 104 –113. DOI: https://doi.org/10.1016/S0031-8914(34)90011-2.

[133] R. M. Dreizler and E. K. U. Gross. “Explicit Functionals II: The Local Density Approximation and Beyond”. In: *Density Functional Theory: An Approach to the Quantum Many-Body Problem*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990, pp. 173–244. DOI: 10.1007/978-3-642-86105-5_7.

[134] B. G. Johnson et al. “A Density Functional Study of the Simplest Hydrogen Abstraction Reaction. Effect of Self-Interaction Correction”. In: *Chemical Physics Letters* 221.1-2 (1994), pp. 100 –108. DOI: 10.1016/0009-2614(94)87024-1.

[135] E. Ruiz, D. R. Salahub, and A. Vela. “Charge-Transfer Complexes: Stringent Tests for Widely Used Density Functionals”. In: *The Journal of Physical Chemistry* 100.30 (1996), pp. 12265–12276. DOI: 10.1021/jp9533077.

[136] R. W. Godby, M. Schlüter, and L. J. Sham. “Accurate Exchange-Correlation Potential for Silicon and Its Discontinuity on Addition of an Electron”. In: *Phys. Rev. Lett.* 56 (22 June 1986), pp. 2415–2418. DOI: 10.1103/PhysRevLett.56.2415.

[137] D. M. Ceperley and B. J. Alder. “Ground State of the Electron Gas by a Stochastic Method”. In: *Phys. Rev. Lett.* 45 (7 Aug. 1980), pp. 566–569. DOI: 10.1103/PhysRevLett.45.566.

[138] J. P. Perdew and S. Kurth. “Local and Semi-Local Density Functional Approximations for Exchange and Correlation: Why do they Work, and do they Work Best at Zero Temperature?” English. In: *Strongly Coupled Coulomb Systems*. Ed. by G. J. Kalman, J. M. Rommel, and K. Blagoev. Springer US, 2002, pp. 281–285. DOI: 10.1007/0-306-47086-1_46.

[139] R. A. Evarestov. “Semiempirical LCAO Methods for Molecules and Periodic Systems”. English. In: *Quantum Chemistry of Solids*. Vol. 153. Springer Series in Solid-State Sciences. Springer Berlin Heidelberg, 2012, pp. 207–249. DOI: 10.1007/978-3-642-30356-2_6.

[140] D. C. Langreth and M. J. Mehl. "Beyond the Local-Density Approximation in Calculations of Ground-State Electronic Properties". In: *Phys. Rev. B* 28 (4 Aug. 1983), pp. 1809–1834. DOI: 10.1103/PhysRevB.28.1809.

[141] P. L. Taylor and O. Heinonen. "Density Functional Theory". In: *A Quantum Approach to Condensed Matter Physics*. Cambridge Books Online. Cambridge University Press, 2002, pp. 182–209.

[142] C. Filippi, D. J. Singh, and C. J. Umrigar. "All-Electron Local-Density and Generalized-Gradient Calculations of the Structural Properties of Semiconductors". In: *Phys. Rev. B* 50 (20 Nov. 1994), pp. 14947–14951. DOI: 10.1103/PhysRevB.50.14947.

[143] A. D. Becke. "Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior". In: *Phys. Rev. A* 38 (6 Sept. 1988), pp. 3098–3100. DOI: 10.1103/PhysRevA.38.3098.

[144] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density". In: *Phys. Rev. B* 37 (2 Jan. 1988), pp. 785–789. DOI: 10.1103/PhysRevB.37.785.

[145] A. D. Becke and M. R. Roussel. "Exchange holes in inhomogeneous systems: A coordinate-space model". In: *Phys. Rev. A* 39 (8 1989), pp. 3761–3767. DOI: 10.1103/PhysRevA.39.3761.

[146] A. D. Becke. "Exchange-Correlation Approximations in Density-Functional Theory". In: *Modern Electronic Structure Theory, Part II*. World Scientific Publishing Company, 1995. Chap. 4, pp. 1022–1046. DOI: 10.1142/9789812832115_0004.

[147] J. Harris. "Adiabatic-Connection Approach to Kohn-Sham Theory". In: *Phys. Rev. A* 29 (4 Apr. 1984), pp. 1648–1659. DOI: 10.1103/PhysRevA.29.1648.

[148] A. D. Becke. "Density-functional Thermochemistry. III. The Role of Exact Exchange". In: *The Journal of Chemical Physics* 98.7 (1993), pp. 5648–5652. DOI: 10.1063/1.464913.

[149] A. D. Becke. "A New Mixing of Hartree-Fock and Local Density-Functional Theories". In: *The Journal of Chemical Physics* 98.2 (1993), pp. 1372–1377. DOI: 10.1063/1.464304.

[150] C. Adamo and V. Barone. "Toward Reliable Density Functional Methods without Adjustable Parameters: The PBE0 Model". In: *The Journal of Chemical Physics* 110.13 (1999), pp. 6158–6170. DOI: 10.1063/1.478522.

[151] J. P. Perdew, M. Ernzerhof, and K. Burke. "Rationale for Mixing Exact Exchange with Density Functional Approximations". In: *The Journal of Chemical Physics* 105.22 (1996), pp. 9982–9985. DOI: 10.1063/1.472933.

[152] J. P. Perdew, K. Burke, and M. Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Phys. Rev. Lett.* 77 (18 Oct. 1996), pp. 3865–3868. DOI: 10.1103/PhysRevLett.77.3865.

[153] S. F. Sousa, P. A. Fernandes, and M. J. Ramos. "General Performance of Density Functionals". In: *The Journal of Physical Chemistry A* 111.42 (2007). PMID: 17718548, pp. 10439–10452. DOI: 10.1021/jp0734474.

[154] J. Bernstein. *Polymorphism in Molecular Crystals*. International Union of Crystallography Monographs on Crystallography. Oxford University Press, 2007.

[155] D. Asthagiri, L. R. Pratt, and J. D. Kress. "Free Energy of Liquid Water on the Basis of Quasichemical Theory and ab initio Molecular Dynamics". In: *Phys. Rev. E* 68 (4 2003), p. 041505. DOI: 10.1103/PhysRevE.68.041505.

[156] E. Schwegler et al. "Towards an Assessment of the Accuracy of Density Functional Theory for First Principles Simulations of Water. II". In: *The Journal of Chemical Physics* 121.11 (2004), pp. 5400–5409. DOI: 10.1063/1.1782074.

[157] M. V. Fernndez-Serra and E. Artacho. "Network Equilibration and First-Principles Liquid Water". In: *The Journal of Chemical Physics* 121.22 (2004), pp. 11136–11144. DOI: 10.1063/1.1813431.

[158] P. H.-L. Sit and N. Marzari. "Static and Dynamical Properties of Heavy Water at Ambient Conditions from First-Principles Molecular Dynamics". In: *The Journal of Chemical Physics* 122.20 (2005), p. 204510. DOI: 10.1063/1.1908913.

[159] I.-F. W. Kuo et al. "Liquid Water from First Principles: Investigation of Different Sampling Approaches". In: *The Journal of Physical Chemistry B* 108.34 (2004), pp. 12990–12998. DOI: 10.1021/jp047788i.

[160] J. VandeVondele et al. "The Influence of Temperature and Density Functional Models in *ab initio* Molecular Dynamics Simulation of Liquid Water". In: *The Journal of Chemical Physics* 122.1 (2005), p. 014515. DOI: 10.1063/1.1828433.

[161] B. Santra et al. "On the Accuracy of Density-Functional Theory Exchange-Correlation Functionals for H Bonds in Small Water Clusters. II. The Water Hexamer and van der Waals Interactions". In: *The Journal of Chemical Physics* 129.19 (2008), p. 194111. DOI: 10.1063/1.3012573.

[162] G. C. Shields and K. N. Kirschner. "The Limitations of Certain Density Functionals in Modeling Neutral Water Clusters". In: *Synthesis and Reactivity in Inorganic, Metal-Organic, and Nano-Metal Chemistry* 38.1 (2008), pp. 32–39. DOI: 10.1080/15533170701853918.

[163] B. Santra et al. "Hydrogen Bonds and van der Waals Forces in Ice at Ambient and High Pressures". In: *Phys. Rev. Lett.* 107 (18 2011), p. 185701. DOI: 10.1103/PhysRevLett.107.185701.

[164] S. Grimme. "Density Functional Theory with London Dispersion Corrections". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.2 (2011), pp. 211–228. DOI: 10.1002/wcms.30.

[165] S. Grimme. "Accurate Description of van der Waals Complexes by Density Functional Theory including Empirical Corrections". In: *Journal of Computational Chemistry* 25.12 (2004), pp. 1463–1473. DOI: 10.1002/jcc.20078.

[166] S. Grimme. "Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction". In: *Journal of Computational Chemistry* 27.15 (2006), pp. 1787–1799. DOI: 10.1002/jcc.20495.

[167] S. Grimme et al. “A Consistent and Accurate ab initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu”. In: *The Journal of Chemical Physics* 132.15, 154104 (2010), pp. –. DOI: 10.1063/1.3382344.

[168] Q. Wu and W. Yang. “Empirical Correction to Density Functional Theory for van der Waals Interactions”. In: *The Journal of Chemical Physics* 116.2 (2002), pp. 515–524. DOI: 10.1063/1.1424928.

[169] H. B. G. Casimir and D. Polder. “The Influence of Retardation on the London-van der Waals Forces”. In: *Phys. Rev.* 73 (4 Feb. 1948), pp. 360–372. DOI: 10.1103/PhysRev.73.360.

[170] G. Starkschall and R. G. Gordon. “Calculation of Coefficients in the Power Series Expansion of the Long-Range Dispersion Force between Atoms”. In: *The Journal of Chemical Physics* 56.6 (1972), pp. 2801–2806. DOI: 10.1063/1.1677610.

[171] B. M. Axilrod and E. Teller. “Interaction of the van der Waals Type Between Three Atoms”. In: *The Journal of Chemical Physics* 11.6 (1943), pp. 299–300. DOI: 10.1063/1.1723844.

[172] P. Jurečka et al. “Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs”. In: *Phys. Chem. Chem. Phys.* 8 (17 2006), pp. 1985–1993. DOI: 10.1039/B600027D.

[173] A. Tkatchenko and O. A. von Lilienfeld. “Popular Kohn-Sham Density Functionals Strongly Overestimate Many-Body Interactions in van der Waals Systems”. In: *Phys. Rev. B* 78 (4 July 2008), p. 045116. DOI: 10.1103/PhysRevB.78.045116.

[174] Y. Liu and W. A. I. Goddard. “A Universal Damping Function for Empirical Dispersion Correction on Density Functional Theory”. In: *MATERIALS TRANSACTIONS* 50.7 (2009), pp. 1664–1670. DOI: 10.2320/matertrans.MF200911.

[175] J. D. Chai and M. Head-Gordon. “Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections”. In: *Phys. Chem. Chem. Phys.* 10 (44 2008), pp. 6615–6620. DOI: 10.1039/B810189B.

[176] P. Jurečka et al. “Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes compared with ab initio Quantum Mechanics Calculations”. In: *Journal of Computational Chemistry* 28.2 (2007), pp. 555–569. DOI: 10.1002/jcc.20570.

[177] A. D. . Becke and E. R. Johnson. “Exchange-Hole Dipole Moment and the Dispersion Interaction”. In: *The Journal of Chemical Physics* 122.15 (2005), p. 154104. DOI: 10.1063/1.1884601.

[178] A. D. Becke and E. R. Johnson. “A Density-Functional Model of the Dispersion Interaction”. In: *The Journal of Chemical Physics* 123.15 (2005), p. 154101. DOI: 10.1063/1.2065267.

[179] E. R. Johnson and A. D. Becke. “A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections”. In: *The Journal of Chemical Physics* 124.17 (2006), p. 174104. DOI: 10.1063/1.2190220.

[180] A. Koide. "A new expansion for dispersion forces and its application". In: *Journal of Physics B: Atomic and Molecular Physics* 9.18 (1976), pp. 3173–3183. DOI: 10.1088/0022-3700/9/18/009.

[181] L. Goerigk and S. Grimme. "Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals?Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions". In: *Journal of Chemical Theory and Computation* 7.2 (2011). PMID: 26596152, pp. 291–309. DOI: 10.1021/ct100466k.

[182] A. Tkatchenko and M. Scheffler. "Accurate Molecular van der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data". In: *Phys. Rev. Lett.* 102 (7 Feb. 2009), p. 073005. DOI: 10.1103/PhysRevLett.102.073005.

[183] K. T. Tang. "Dynamic Polarizabilities and van der Waals Coefficients". In: *Phys. Rev.* 177 (1 Jan. 1969), pp. 108–114. DOI: 10.1103/PhysRev.177.108.

[184] X. Chu and A. Dalgarno. "Linear Response Time-Dependent Density Functional Theory for van der Waals Coefficients". In: *The Journal of Chemical Physics* 121.9 (2004), pp. 4083–4088. DOI: 10.1063/1.1779576.

[185] A. Bondi. "van der Waals Volumes and Radii". In: *The Journal of Physical Chemistry* 68.3 (1964), pp. 441–451. DOI: 10.1021/j100785a001.

[186] T. Brinck, J. S. Murray, and P. Politzer. "Polarizability and Volume". In: *The Journal of Chemical Physics* 98.5 (1993), pp. 4305–4306. DOI: 10.1063/1.465038.

[187] F. D. Proft et al. "Hirshfeld Partitioning of the Electron Density: Atomic Dipoles and their Relation with Functional Group Properties". In: *Journal of Computational Chemistry* 24.4 (2003), pp. 463–470. DOI: 10.1002/jcc.10241.

[188] A. D. Becke and E. R. Johnson. "Exchange-Hole Dipole Moment and the Dispersion Interaction". In: *The Journal of Chemical Physics* 122.15, 154104 (2005). DOI: 10.1063/1.1884601.

[189] A. D. Becke and E. R. Johnson. "Exchange-Hole Dipole Moment and the Dispersion Interaction Revisited". In: *The Journal of Chemical Physics* 127.15, 154108 (2007). DOI: 10.1063/1.2795701.

[190] A. Dalgarno and W. D. Davison. "The Calculation of van der Waals Interactions". In: *Advances in Atomic and Molecular Physics*. Ed. by D. R. Bates and I. Estermann. Vol. 2. Advances in Atomic and Molecular Physics. Academic Press, 1966, pp. 1 –32. DOI: 10.1016/S0065-2199(08)60216-X.

[191] A. D. Becke and E. R. Johnson. "Exchange-Hole Dipole Moment and the Dispersion Interaction: High-Order Dispersion Coefficients". In: *The Journal of Chemical Physics* 124.1, 014104 (2006). DOI: 10.1063/1.2139668.

[192] A. Otero-de-la Roza and E. R. Johnson. "Many-Body Dispersion Interactions from the Exchange-Hole Dipole Moment Model". In: *The Journal of Chemical Physics* 138.5, 054103 (2013). DOI: 10.1063/1.4789421.

[193] E. R. Johnson and A. D. Becke. "A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections". In: *The Journal of Chemical Physics* 124.17, 174104 (2006). DOI: 10.1063/1.2190220.

[194] X. Xu and W. A. Goddard. "The X3LYP Extended Density Functional for Accurate Descriptions of Nonbond Interactions, Spin States, and Thermochemical Properties". In: *Proceedings of the National Academy of Sciences* 101.9 (2004), pp. 2673–2677. DOI: 10.1073/pnas.0308730100.

[195] X. Xu and W. A. Goddard. "Bonding Properties of the Water Dimer: A Comparative Study of Density Functional Theories". In: *The Journal of Physical Chemistry A* 108.12 (2004), pp. 2305–2313. DOI: 10.1021/jp035869t.

[196] J. Černy and P. Hobza. "The X3LYP Extended Density Functional Accurately Describes H-Bonding but Fails Completely for Stacking". In: *Phys. Chem. Chem. Phys.* 7 (8 2005), pp. 1624–1626. DOI: 10.1039/B502769C.

[197] M. P. Waller et al. "Hybrid Density Functional Theory for π-Stacking Interactions: Application to Benzenes, Pyridines, and DNA Bases". In: *Journal of Computational Chemistry* 27.4 (2006), pp. 491–504. DOI: 10.1002/jcc.20363.

[198] Y. Zhao and D. G. Truhlar. "A Prototype for Graphene Material Simulation: Structures and Interaction Potentials of Coronene Dimers". In: *The Journal of Physical Chemistry C* 112.11 (2008), pp. 4061–4067. DOI: 10.1021/jp710918f.

[199] E. G. Hohenstein, S. T. Chill, and C. D. Sherrill. "Assessment of the Performance of the M05-2X and M06-2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules". In: *Journal of Chemical Theory and Computation* 4.12 (2008). PMID: 26620472, pp. 1996–2000. DOI: 10.1021/ct800308k.

[200] W. Wang et al. "Benchmark Calculations of the Adsorption of Aromatic Molecules on Graphene". In: *Journal of Computational Chemistry* 36.23 (2015), pp. 1763–1771. DOI: 10.1002/jcc.23994.

[201] N. Mardirossian and M. Head-Gordon. "ωB97X-V: A 10-Parameter, Range-Separated Hybrid, Generalized Gradient Approximation Density Functional with Nonlocal Correlation, Designed by a Survival-of-the-Fittest Strategy". In: *Phys. Chem. Chem. Phys.* 16 (21 2014), pp. 9904–9924. DOI: 10.1039/C3CP54374A.

[202] G. A. DiLabio and A. Otero-de-la Roza. "Noncovalent Interactions in Density Functional Theory". In: *Reviews in Computational Chemistry*. Wiley-Blackwell, 2016. Chap. 1, pp. 1–97. DOI: 10.1002/9781119148739.ch1.

[203] M. Dion et al. "Van der Waals Density Functional for General Geometries". In: *Phys. Rev. Lett.* 92 (24 2004), p. 246401. DOI: 10.1103/PhysRevLett.92.246401.

[204] É. D. Murray, K. Lee, and D. C. Langreth. "Investigation of Exchange Energy Density Functional Accuracy for Interacting Molecules". In: *Journal of Chemical Theory and Computation* 5.10 (2009). PMID: 26631788, pp. 2754–2762. DOI: 10.1021/ct900365q.

[205] K. Lee et al. "Higher-accuracy van der Waals Density Functional". In: *Phys. Rev. B* 82 (8 2010), p. 081101. DOI: 10.1103/PhysRevB.82.081101.

[206] A. Gulans, M. J. Puska, and R. M. Nieminen. "Linear-Scaling Self-Consistent Implementation of the van der Waals Density Functional". In: *Phys. Rev. B* 79 (20 2009), p. 201105. DOI: 10.1103/PhysRevB.79.201105.

[207] O. A. von Lilienfeld et al. “Optimization of Effective Atom Centered Potentials for London Dispersion Forces in Density Functional Theory”. In: *Phys. Rev. Lett.* 93 (15 2004), p. 153004. DOI: 10.1103/PhysRevLett.93.153004.

[208] O. A. von Lilienfeld et al. “Performance of Optimized Atom-Centered Potentials for Weakly Bonded Systems using Density Functional Theory”. In: *Phys. Rev. B* 71 (19 2005), p. 195119. DOI: 10.1103/PhysRevB.71.195119.

[209] O. A. von Lilienfeld et al. “Variational Optimization of Effective Atom Centered Potentials for Molecular Properties”. In: *The Journal of Chemical Physics* 122.1 (2005), p. 014113. DOI: 10.1063/1.1829051.

[210] O. A. von Lilienfeld and D. Andrienko. “Coarse-Grained Interaction Potentials for Polyaromatic Hydrocarbons”. In: *The Journal of Chemical Physics* 124.5 (2006), p. 054307. DOI: 10.1063/1.2162543.

[211] G. Cinacchi. “Comment on “Coarse-Grained Interaction Potentials for Polyaromatic Hydrocarbons” [J. Chem. Phys. 124, 054307 (2006)]”. In: *The Journal of Chemical Physics* 125.5 (2006), p. 057101. DOI: 10.1063/1.2234368.

[212] A. Tkatchenko and O. A. von Lilienfeld. “Adsorption of Ar on Graphite using London Dispersion Forces Corrected Kohn-Sham Density Functional Theory”. In: *Phys. Rev. B* 73 (15 2006), p. 153406. DOI: 10.1103/PhysRevB.73.153406.

[213] E. Tapavicza et al. “Weakly Bonded Complexes of Aliphatic and Aromatic Carbon Compounds Described with Dispersion Corrected Density Functional Theory”. In: *Journal of Chemical Theory and Computation* 3.5 (2007). PMID: 26627613, pp. 1673–1679. DOI: 10.1021/ct700049s.

[214] M. E. Tuckerman et al. “Ab initio Simulations of Water and Water Ions”. In: *Journal of Physics: Condensed Matter* 6.23A (1994), A93.

[215] P. L. Silvestrelli, M. Bernasconi, and M. Parrinello. “Ab initio Infrared Spectrum of Liquid Water”. In: *Chemical Physics Letters* 277.5 (1997), pp. 478 –482. DOI: 10.1016/S0009-2614(97)00930-5.

[216] M. Krack, A. Gambirasio, and M. Parrinello. “Ab initio X-Ray Scattering of Liquid Water”. In: *The Journal of Chemical Physics* 117.20 (2002), pp. 9409–9412. DOI: 10.1063/1.1517040.

[217] K. Laasonen and M. L. Klein. “Ab Initio Molecular Dynamics Study of Hydrochloric Acid in Water”. In: *Journal of the American Chemical Society* 116.25 (1994), pp. 11620–11621. DOI: 10.1021/ja00104a073.

[218] A. A. Hassanali et al. “Aqueous Solutions: State of the Art in ab initio Molecular Dynamics”. In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 372.2011 (2014). DOI: 10.1098/rsta.2012.0482.

[219] F. Coudert, R. Vuilleumier, and A. Boutin. “Dipole Moment, Hydrogen Bonding and IR Spectrum of Confined Water”. In: *ChemPhysChem* 7.12 (2006), pp. 2464–2467. DOI: 10.1002/cphc.200600561.

[220] G. Cicero et al. “Water Confined in Nanotubes and between Graphene Sheets: A First Principle Study”. In: *Journal of the American Chemical Society* 130.6 (2008). PMID: 18211065, pp. 1871–1878. DOI: 10.1021/ja074418+.

[221] D. Muoz-Santiburcio, C. Wittekindt, and D. Marx. “Nanoconfinement Effects on Hydrated Excess Protons in Layered Materials”. In: *Nature Communications* 4 (Aug. 2013), p. 2349.

[222] P. J. D. Lindan, N. M. Harrison, and M. J. Gillan. “Mixed Dissociative and Molecular Adsorption of Water on the Rutile (110) Surface”. In: *Phys. Rev. Lett.* 80 (4 1998), pp. 762–765. DOI: 10.1103/PhysRevLett.80.762.

[223] J. Carrasco, A. Hodgson, and A. Michaelides. “A Molecular Perspective of Water at Metal Interfaces”. In: *Nature Materials* 11 (July 2012), p. 667.

[224] B. C. Wood et al. “Hydrogen-Bond Dynamics of Water at the Interface with InP/-GaP(001) and the Implications for Photoelectrochemistry”. In: *Journal of the American Chemical Society* 135.42 (2013). PMID: 24053479, pp. 15774–15783. DOI: 10.1021/ja403850s.

[225] A. E. Mattsson and T. R. Mattsson. “AM05 Density Functional Applied to the Water Molecule, Dimer, and Bulk Liquid”. In: *Journal of Chemical Theory and Computation* 5.4 (2009). PMID: 26609597, pp. 887–894. DOI: 10.1021/ct8004968.

[226] E. A. Cobar et al. “Examination of the Hydrogen-Bonding Networks in Small Water Clusters (n = 2?5, 13, 17) using Absolutely Localized Molecular Orbital Energy Decomposition Analysis”. In: *Phys. Chem. Chem. Phys.* 14 (44 2012), pp. 15328–15339. DOI: 10.1039/C2CP42522J.

[227] A. Bankura et al. “Structure, Dynamics, and Spectral Diffusion of Water from First-Principles Molecular Dynamics”. In: *The Journal of Physical Chemistry C* 118.50 (2014), pp. 29401–29411. DOI: 10.1021/jp506120t.

[228] H. Lee and M. E. Tuckerman. “Structure of Liquid Water at Ambient Temperature from *ab initio* Molecular Dynamics Performed in the Complete Basis Set Limit”. In: *The Journal of Chemical Physics* 125.15 (2006), p. 154507. DOI: 10.1063/1.2354158.

[229] K. Forster-Tonigold and A. Groβ. “Dispersion Corrected RPBE Studies of Liquid Water”. In: *The Journal of Chemical Physics* 141.6 (2014), p. 064501. DOI: 10.1063/1.4892400.

[230] T. D. Kühne, M. Krack, and M. Parrinello. “Static and Dynamical Properties of Liquid Water from First Principles by a Novel Car?Parrinello-like Approach”. In: *Journal of Chemical Theory and Computation* 5.2 (2009). PMID: 26610101, pp. 235–241. DOI: 10.1021/ct800417q.

[231] C. Zhang, L. Spanu, and G. Galli. “Entropy of Liquid Water from *Ab Initio* Molecular Dynamics”. In: *The Journal of Physical Chemistry B* 115.48 (2011). PMID: 21961845, pp. 14190–14195. DOI: 10.1021/jp204981y.

[232] C. Zhang et al. “Structural and Vibrational Properties of Liquid Water from van der Waals Density Functionals”. In: *Journal of Chemical Theory and Computation* 7.10 (2011). PMID: 26598149, pp. 3054–3061. DOI: 10.1021/ct200329e.

[233] M. J. Gillan et al. "Assessing the Accuracy of Quantum Monte Carlo and Density Functional Theory for Energetics of Small Water Clusters". In: *The Journal of Chemical Physics* 136.24 (2012), p. 244105. DOI: 10.1063/1.4730035.

[234] M. Guidon et al. "*Ab Initio* Molecular Dynamics using Hybrid Density Functionals". In: *The Journal of Chemical Physics* 128.21 (2008), p. 214104. DOI: 10.1063/1.2931945.

[235] M. Guidon, J. Hutter, and J. VandeVondele. "Auxiliary Density Matrix Methods for Hartree-Fock Exchange Calculations". In: *Journal of Chemical Theory and Computation* 6.8 (2010). PMID: 26613491, pp. 2348–2364. DOI: 10.1021/ct1002225.

[236] O. Anatole von Lilienfeld and A. Tkatchenko. "Two- and Three-Body Interatomic Dispersion Energy Contributions to Binding in Molecules and Solids". In: *The Journal of Chemical Physics* 132.23 (2010), p. 234109. DOI: 10.1063/1.3432765.

[237] I.-C. Lin et al. "Importance of van der Waals Interactions in Liquid Water". In: *The Journal of Physical Chemistry B* 113.4 (2009). PMID: 19123911, pp. 1127–1131. DOI: 10.1021/jp806376e.

[238] D. Alfè et al. "Communication: Energy Benchmarking with Quantum Monte Carlo for Water Nano-Droplets and Bulk Liquid Water". In: *The Journal of Chemical Physics* 138.22 (2013), p. 221102. DOI: 10.1063/1.4810882.

[239] J. Schmidt et al. "Isobaric-Isothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near-Ambient Conditions". In: *The Journal of Physical Chemistry B* 113.35 (2009). PMID: 19663399, pp. 11959–11964. DOI: 10.1021/jp901990u.

[240] M. D. Baer et al. "Re-Examining the Properties of the Aqueous Vapor?Liquid Interface using Dispersion Corrected Density Functional Theory". In: *The Journal of Chemical Physics* 135.12 (2011), p. 124712. DOI: 10.1063/1.3633239.

[241] Z. Ma, Y. Zhang, and M. E. Tuckerman. "*Ab Initio* Molecular Dynamics Study of Water at Constant Pressure using Converged Basis Sets and Empirical Dispersion Corrections". In: *The Journal of Chemical Physics* 137.4 (2012), p. 044506. DOI: 10.1063/1.4736712.

[242] I.-C. Lin et al. "Structure and Dynamics of Liquid Water from *ab Initio* Molecular Dynamics-Comparison of BLYP, PBE, and revPBE Density Functionals with and without van der Waals Corrections". In: *Journal of Chemical Theory and Computation* 8.10 (2012). PMID: 26593030, pp. 3902–3910. DOI: 10.1021/ct3001848.

[243] A. P. Gaiduk, F. Gygi, and G. Galli. "Density and Compressibility of Liquid Water and Ice from First-Principles Simulations with Hybrid Functionals". In: *The Journal of Physical Chemistry Letters* 6.15 (2015). PMID: 26267178, pp. 2902–2908. DOI: 10.1021/acs.jpclett.5b00901.

[244] J. Wang et al. "Density, Structure, and Dynamics of Water: The Effect of van der Waals Interactions". In: *The Journal of Chemical Physics* 134.2 (2011), p. 024516. DOI: 10.1063/1.3521268.

[245] F. Corsetti et al. "Room Temperature Compressibility and Diffusivity of Liquid Water from First Principles". In: *The Journal of Chemical Physics* 139.19 (2013), p. 194502. DOI: 10.1063/1.4832141.

[246] D. Alfè et al. “Analyzing the Errors of DFT Approximations for Compressed Water Systems”. In: *The Journal of Chemical Physics* 141.1 (2014), p. 014104. DOI: 10.1063/1.4885440.

[247] T. Todorova et al. “Molecular Dynamics Simulation of Liquid Water:? Hybrid Density Functionals”. In: *The Journal of Physical Chemistry B* 110.8 (2006), pp. 3685–3691. DOI: 10.1021/jp055127v.

[248] L. Ruiz Pestana et al. “*Ab Initio* Molecular Dynamics Simulations of Liquid Water using High Quality Meta-GGA Functionals”. In: *Chem. Sci.* 8 (5 2017), pp. 3554–3565. DOI: 10.1039/C6SC04711D.

[249] N. Mardirossian and M. Head-Gordon. “Mapping the Genome of Meta-Generalized Gradient Approximation Density Functionals: The Search for B97M-V”. In: *The Journal of Chemical Physics* 142.7 (2015), p. 074111. DOI: 10.1063/1.4907719.

[250] Y. Zhao and D. G. Truhlar. “A New Local Density Functional for Main-Group Thermochemistry, Transition Metal Bonding, Thermochemical Kinetics, and Noncovalent Interactions”. In: *The Journal of Chemical Physics* 125.19 (2006), p. 194101. DOI: 10.1063/1.2370993.

[251] J. Sun, A. Ruzsinszky, and J. P. Perdew. “Strongly Constrained and Appropriately Normed Semilocal Density Functional”. In: *Phys. Rev. Lett.* 115 (3 2015), p. 036402. DOI: 10.1103/PhysRevLett.115.036402.

[252] J. Sun et al. “Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional”. In: *Nature* 8 (Sept. 2016), p. 831.

[253] M. Chen et al. “*Ab Initio* Theory and Modeling of Water”. In: *Proceedings of the National Academy of Sciences* 114.41 (2017), pp. 10846–10851. DOI: 10.1073/pnas.1712499114.

[254] O. L. Polyansky, P. Jensen, and J. Tennyson. “A Spectroscopically Determined Potential Energy Surface for the Ground State of H_2^16O: A New Level of Accuracy”. In: *The Journal of Chemical Physics* 101.9 (1994), pp. 7651–7657. DOI: 10.1063/1.468258.

[255] P. Jensen. “The Potential Energy Surface for the Electronic Ground State of the Water Molecule Determined from Experimental Data using a Variational Approach”. In: *Journal of Molecular Spectroscopy* 133.2 (1989), pp. 438 –460. DOI: 10.1016/0022-2852(89)90203-8.

[256] E. G. Lewars. “The Concept of the Potential Energy Surface”. In: *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Cham: Springer International Publishing, 2016, pp. 9–49. DOI: 10.1007/978-3-319-30916-3_2.

[257] W. L. Jorgensen. “Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water”. In: *Journal of the American Chemical Society* 103.2 (1981), pp. 335–340. DOI: 10.1021/ja00392a016.

[258] K. Toukan and A. Rahman. “Molecular-Dynamics Study of Atomic Motions in Water”. In: *Phys. Rev. B* 31 (5 1985), pp. 2643–2648. DOI: 10.1103/PhysRevB.31.2643.

[259] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. "The Missing Term in Effective Pair Potentials". In: *The Journal of Physical Chemistry* 91.24 (1987), pp. 6269–6271. DOI: 10.1021/j100308a038.

[260] P. Ahlström et al. "A Molecular Dynamics Study of Polarizable Water". In: *Molecular Physics* 68.3 (1989), pp. 563–581. DOI: 10.1080/00268978900102361.

[261] S. W. Rick, S. J. Stuart, and B. J. Berne. "Dynamical Fluctuating Charge Force Fields: Application to Liquid Water". In: *The Journal of Chemical Physics* 101.7 (1994), pp. 6141–6156. DOI: 10.1063/1.468398.

[262] J. Lobaugh and G. A. Voth. "A Quantum Model for Water: Equilibrium and Dynamical Properties". In: *The Journal of Chemical Physics* 106.6 (1997), pp. 2400–2410. DOI: 10.1063/1.473151.

[263] U. W. Schmitt and G. A. Voth. "The Computer Simulation of Proton Transport in Water". In: *The Journal of Chemical Physics* 111.20 (1999), pp. 9361–9381. DOI: 10.1063/1.480032.

[264] M. W. Mahoney and W. L. Jorgensen. "A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions". In: *The Journal of Chemical Physics* 112.20 (2000), pp. 8910–8922. DOI: 10.1063/1.481505.

[265] J. L. F. Abascal and C. Vega. "A General Purpose Model for the Condensed Phases of Water: TIP4P/2005". In: *The Journal of Chemical Physics* 123.23 (2005), p. 234505. DOI: 10.1063/1.2121687.

[266] Y. J. Wu, H. L. Tepper, and G. A. Voth. "Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties". In: *The Journal of Chemical Physics* 124.2, 024503 (2006). DOI: 10.1063/1.2136877.

[267] A. Wallqvist and R. D. Mountain. "Molecular Models of Water: Derivation and Description". In: *Reviews in Computational Chemistry*. Wiley-Blackwell, 2007, pp. 183–247. DOI: 10.1002/9780470125908.ch4.

[268] E. Sanz et al. "Phase Diagram of Water from Computer Simulation". In: *Phys. Rev. Lett.* 92 (25 2004), p. 255701. DOI: 10.1103/PhysRevLett.92.255701.

[269] J. L. F. Abascal et al. "A Potential Model for the Study of Ices and Amorphous Water: TIP4P/Ice". In: *The Journal of Chemical Physics* 122.23 (2005), p. 234511. DOI: 10.1063/1.1931662.

[270] S. R. Billeter, P. M. King, and W. F. van Gunsteren. "Can the Density Maximum of Water be Found by Computer Simulation?" In: *The Journal of Chemical Physics* 100.9 (1994), pp. 6692–6699. DOI: 10.1063/1.467029.

[271] L. A. Báez and P. Clancy. "Existence of a Density Maximum in Extended Simple Point Charge Water". In: *The Journal of Chemical Physics* 101.11 (1994), pp. 9837–9840. DOI: 10.1063/1.467949.

[272] A. Wallqvist and P.-O. Åstrand. "Liquid Densities and Structural Properties of Molecular Models of Water". In: *The Journal of Chemical Physics* 102.16 (1995), pp. 6559–6565. DOI: 10.1063/1.469370.

[273] W. L. Jorgensen and C. Jenson. "Temperature Dependence of TIP3P, SPC, and TIP4P Water from NPT Monte Carlo Simulations: Seeking Temperatures of Maximum Density". In: *Journal of Computational Chemistry* 19.10 (1998), pp. 1179–1186. DOI: 10.1002/(SICI)1096-987X(19980730)19:10<1179::AID-JCC6>3.0.CO;2-J.

[274] Y. Guissani and B. Guillot. "A Computer Simulation Study of the Liquid-Vapor Coexistence Curve of Water". In: *The Journal of Chemical Physics* 98.10 (1993), pp. 8221–8235. DOI: 10.1063/1.464527.

[275] J. Alejandre, D. J. Tildesley, and G. A. Chapela. "Molecular Dynamics Simulation of the Orthobaric Densities and Surface Tension of Water". In: *The Journal of Chemical Physics* 102.11 (1995), pp. 4574–4583. DOI: 10.1063/1.469505.

[276] M. J. Vlot, J. Huinink, and J. P. van der Eerden. "Free Energy Calculations on Systems of Rigid Molecules: An Application to the TIP4P Model of H_2O". In: *The Journal of Chemical Physics* 110.1 (1999), pp. 55–61. DOI: 10.1063/1.478084.

[277] J. J. de Pablo et al. "Molecular Simulation of Water along the Liquid-Vapor Coexistence Curve from 25 °C to the Critical Point". In: *The Journal of Chemical Physics* 93.10 (1990), pp. 7355–7359. DOI: 10.1063/1.459409.

[278] V. P. Sokhan and D. Tildesley. "The Free Surface of Water: Molecular Orientation, Surface Potential and Nonlinear Susceptibility". In: *Molecular Physics* 92.4 (1997), pp. 625–640. DOI: 10.1080/002689797169916.

[279] K. Kiyohara, K. E. Gubbins, and A. Z. Panagiotopoulos. "Phase Coexistence Properties of Polarizable Water Models". In: *Molecular Physics* 94.5 (1998), pp. 803–808. DOI: 10.1080/002689798167638.

[280] M.-C. Bellissent-Funel et al. "The Structure of Supercritical Heavy Water as Studied by Neutron Diffraction". In: *The Journal of Chemical Physics* 107.8 (1997), pp. 2942–2949. DOI: 10.1063/1.475155.

[281] Y. S. Badyal et al. "Electron Distribution in Water". In: *The Journal of Chemical Physics* 112.21 (2000), pp. 9206–9208. DOI: 10.1063/1.481541.

[282] J. Anderson, J. J. Ullo, and S. Yip. "Molecular Dynamics Simulation of Dielectric Properties of Water". In: *The Journal of Chemical Physics* 87.3 (1987), pp. 1726–1732. DOI: 10.1063/1.453239.

[283] A. Wallqvist and O. Teleman. "Properties of Flexible Water Models". In: *Molecular Physics* 74.3 (1991), pp. 515–533. DOI: 10.1080/00268979100102391.

[284] G. Raabe and R. J. Sadus. "Molecular Dynamics Simulation of the Effect of Bond Flexibility on the Transport Properties of Water". In: *The Journal of Chemical Physics* 137.10 (2012), p. 104512. DOI: 10.1063/1.4749382.

[285] B. Guillot and Y. Guissani. "Quantum Effects in Simulated Water by the Feynman-Hibbs Approach". In: *The Journal of Chemical Physics* 108.24 (1998), pp. 10162–10174. DOI: 10.1063/1.476475.

[286] R. Feynman and A. Hibbs. *Quantum Mechanics and Path Integrals*. International Series in Pure and Applied Physics. McGraw-Hill, 1965.

[287] B. Dünweg and K. Kremer. "Molecular Dynamics Simulation of a Polymer Chain in Solution". In: *The Journal of Chemical Physics* 99.9 (1993), pp. 6983–6997. DOI: 10.1063/1.465445.

[288] I. Yeh and G. Hummer. "System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions". In: *The Journal of Physical Chemistry B* 108.40 (2004), pp. 15873–15879. DOI: 10.1021/jp0477147.

[289] D.-M. Duh, D. N. Perera, and A. D. J. Haymet. "Structure and Properties of the CF1 Central Force Model of Water: Integral Equation Theory". In: *The Journal of Chemical Physics* 102.9 (1995), pp. 3736–3746. DOI: 10.1063/1.468556.

[290] P. Höchtl et al. "Rationalization of the Dielectric Properties of Common Three-Site Water Models in Terms of their Force Field Parameters". In: *The Journal of Chemical Physics* 109.12 (1998), pp. 4927–4937. DOI: 10.1063/1.477104.

[291] N. J. English. "Molecular Dynamics Simulations of Liquid Water using Various Long-Range Electrostatics Techniques". In: *Molecular Physics* 103.14 (2005), pp. 1945–1960. DOI: 10.1080/00268970500105003.

[292] 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". In: *The Journal of Chemical Physics* 108.24 (1998), pp. 10220–10230. DOI: 10.1063/1.476482.

[293] R. D. Mountain and D. Thirumalai. "Ergodic Measures for the Simulation of Dialectric Properties of Water". In: *Computer Physics Communications* 62.2 (1991), pp. 352 – 359. DOI: 10.1016/0010-4655(91)90106-U.

[294] J. Barker and R. Watts. "Structure of Water; A Monte Carlo Calculation". In: *Chemical Physics Letters* 3.3 (1969), pp. 144 –145. DOI: 10.1016/0009-2614(69)80119-3.

[295] A. Rahman and F. H. Stillinger. "Molecular Dynamics Study of Liquid Water". In: *The Journal of Chemical Physics* 55.7 (1971), pp. 3336–3359. DOI: 10.1063/1.1676585.

[296] R. M. Lynden-Bell, J. C. Rasaiah, and J. P. Noworyta. "Using Simulation to Study Solvation in Water". In: *Pure and Applied Chemistry* 73.11 (2001), pp. 1721–1731. DOI: 10.1351/pac200173111721.

[297] K. M. Åberg et al. "Determination of Solvation Free Energies by Adaptive Expanded Ensemble Molecular Dynamics". In: *The Journal of Chemical Physics* 120.8 (2004), pp. 3770–3776. DOI: 10.1063/1.1642601.

[298] M. Ferrario et al. "Solubility of KF in Water by Molecular Dynamics using the Kirkwood Integration Method". In: *The Journal of Chemical Physics* 117.10 (2002), pp. 4947–4953. DOI: 10.1063/1.1498820.

[299] D. Paschek. "Temperature Dependence of the Hydrophobic Hydration and Interaction of Simple Solutes: An Examination of Five Popular Water Models". In: *The Journal of Chemical Physics* 120.14 (2004), pp. 6674–6690. DOI: 10.1063/1.1652015.

[300] C. Pangali, M. Rao, and B. J. Berne. "A Monte Carlo Study of Structural and Thermodynamic Properties of Water: Dependence on the System Size and on the Boundary Conditions". In: *Molecular Physics* 40.3 (1980), pp. 661–680. DOI: 10.1080/00268978000101781.

[301] W. L. Jorgensen and J. D. Madura. "Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water". In: *Molecular Physics* 56.6 (1985), pp. 1381–1392. DOI: 10.1080/00268978500103111.

[302] C. Vega and J. L. F. Abascal. "Simulating Water with Rigid Non-Polarizable Models: A General Perspective". In: *Phys. Chem. Chem. Phys.* 13 (44 2011), pp. 19663–19688. DOI: 10.1039/C1CP22168J.

[303] J. Hansen and I. McDonald. "Chapter 3 - Static Properties of Liquids: Thermodynamics and Structure". In: *Theory of Simple Liquids (Fourth Edition)*. Ed. by J. P. Hansen and I. R. McDonald. Fourth Edition. Oxford: Academic Press, 2013, pp. 61 –104. DOI: 10.1016/B978-0-12-387032-2.00003-9.

[304] D. Chandler and D. M. Richardson. "Theory of Orientational Pair Correlations in Molecular Fluids". In: *The Journal of Physical Chemistry* 87.12 (1983), pp. 2060–2064. DOI: 10.1021/j100235a008.

[305] A. K. Soper. "Empirical Potential Monte Carlo Simulation of Fluid Structure". In: *Chemical Physics* 202.2 (1996), pp. 295 –306. DOI: 10.1016/0301-0104(95)00357-6.

[306] K. R. Hadley and C. McCabe. "A Coarse-Grained Model for Amorphous and Crystalline Fatty Acids". In: *The Journal of Chemical Physics* 132.13, 134505 (2010). DOI: 10.1063/1.3360146.

[307] M. E. Johnson, T. Head-Gordon, and A. A. Louis. "Representability Problems for Coarse-Grained Water Potentials". In: *The Journal of Chemical Physics* 126.14 (2007), p. 144509. DOI: 10.1063/1.2715953.

[308] V. A. Harmandaris et al. "Hierarchical Modeling of Polystyrene: From Atomistic to Coarse-Grained Simulations". In: *Macromolecules* 39.19 (2006), pp. 6708–6719. DOI: 10.1021/ma0606399.

[309] H. Wang, C. Junghans, and K. Kremer. "Comparative Atomistic and Coarse-Grained Study of Water: What do we Lose by Coarse-Graining?" In: *The European Physical Journal E* 28.2 (2009), pp. 221–229. DOI: 10.1140/epje/i2008-10413-5.

[310] M. Praprotnik et al. "Adaptive Resolution Simulation of Liquid Water". In: *Journal of Physics: Condensed Matter* 19.29 (2007), p. 292201.

[311] S. J. Marrink et al. "The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations". In: *The Journal of Physical Chemistry B* 111.27 (2007). PMID: 17569554, pp. 7812–7824. DOI: 10.1021/jp071097f.

[312] L. Wang, J. Chen, and T. Van Voorhis. "Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data". In: *Journal of Chemical Theory and Computation* 9.1 (2013). PMID: 26589047, pp. 452–460. DOI: 10.1021/ct300826t.

[313] R. W. Pastor and A. D. MacKerell. "Development of the CHARMM Force Field for Lipids". In: *The Journal of Physical Chemistry Letters* 2.13 (2011). PMID: 21760975, pp. 1526–1532. DOI: 10.1021/jz200167q.

[314] V. Hornak et al. “Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters”. In: *Proteins: Structure, Function, and Bioinformatics* 65.3 (), pp. 712–725. DOI: 10.1002/prot.21123.

[315] R. B. Best et al. “Simultaneous Determination of Protein Structure and Dynamics”. In: *Nature* 433 (Jan. 2005), pp. 128–132.

[316] R. B. Best and M. Vendruscolo. “Determination of Protein Structures Consistent with NMR Order Parameters”. In: *Journal of the American Chemical Society* 126.26 (2004). PMID: 15225030, pp. 8090–8091. DOI: 10.1021/ja0396955.

[317] S. M. Islam et al. “Structural Refinement from Restrained-Ensemble Simulations Based on EPR/DEER Data: Application to T4 Lysozyme”. In: *The Journal of Physical Chemistry B* 117.17 (2013). PMID: 23510103, pp. 4740–4754. DOI: 10.1021/jp311723a.

[318] J. W. Pitera and J. D. Chodera. “On the Use of Experimental Observations to Bias Simulated Ensembles”. In: *Journal of Chemical Theory and Computation* 8.10 (2012). PMID: 26592995, pp. 3445–3451. DOI: 10.1021/ct300112v.

[319] B. Różycki, Y. C. Kim, and G. Hummer. “SAXS Ensemble Refinement of ESCRT-III CHMP3 Conformational Transitions”. In: *Structure* 19.1 (2011), pp. 109 –116. DOI: 10.1016/j.str.2010.10.006.

[320] A. T. Brünger and M. Nilges. “Computational Challenges for Macromolecular Structure Determination by X-Ray Crystallography and Solution NMR Spectroscopy”. In: *Quarterly Reviews of Biophysics* 26.1 (1993), 49?125. DOI: 10.1017/S0033583500003966.

[321] J. Dolenc et al. “Methods of NMR Structure Refinement: Molecular Dynamics Simulations Improve the Agreement with Measured NMR Data of a C-Terminal Peptide of GCN4-p1”. In: *Journal of Biomolecular NMR* 47.3 (2010), pp. 221–235. DOI: 10.1007/s10858-010-9425-9.

[322] J. Iwahara, C. D. Schwieters, and G. M. Clore. “Ensemble Approach for NMR Structure Refinement against 1H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule”. In: *Journal of the American Chemical Society* 126.18 (2004). PMID: 15125681, pp. 5879–5896. DOI: 10.1021/ja031580d.

[323] B. Roux and J. Weare. “On the Statistical Equivalence of Restrained-Ensemble Simulations with the Maximum Entropy Method”. In: *The Journal of Chemical Physics* 138.8 (2013), p. 084107. DOI: 10.1063/1.4792208.

Chapter 3

Boltzmann Inversion Directed Simulation

This chapter establishes the Boltzmann inversion directed simulation (BIDS) method that is investigated in this thesis. Where BIDS is first introduced, the methods that it draws from are acknowledged with their similarities and differences elucidated. This is followed by an overview of the process flow for the application of BIDS. Here, we highlight the fitting of smooth splines to the oxygen-oxygen (O-O) correlation functions of liquid water to improve the quality of the resulting bias and the processing of the calculated bias to a form that is suitable for application in molecular dynamics (MD) simulation software. For the former, the performance of the smooth splining algorithm for the O-O radial distribution function (RDF) of various water force fields is discussed using the errors between the resultant smooth splines and their respective raw data. This is followed by optimisation of the O-O RDF bin size for the smooth splining algorithm to further improve its performance. For the latter, the extrapolation at the short-range region of the calculated bias that ensured a rapid transition to a force of zero with respect to a decreasing radial displacement is described.

3.1 Introduction

The iterative Boltzmann inversion (IBI) approach[107] is used to determine the bias V to the system potential U within the experiment directed simulation (EDS) framework.[97] The EDS method is based on biasing a molecular simulation with an approximately correct potential energy to match some experimental observable, rather than the parameterisation of novel force fields. In the EDS framework, the system potential U is introduced to an auxiliary "knowledge field" V:

$$U_{i+1} = U_0 + V_{i+1} \tag{3.1}$$

This "knowledge field" draws from the use of harmonic constraints in nuclear magnetic resonance (NMR) structure refinement[320–322] and the linear bias method for minimum change to the statistical ensemble.[318] White *et al.*[324] has also demonstrated its application in the biased *ab initio* molecular dynamics (AIMD) simulation of density functional theory (DFT) water, which will be used as an *ab initio* benchmark in this work. Instead of the potential of mean force (PMF) as per the IBI method, the initial system potential U_0 corresponds with an unbiased Hamiltonian that produces an ensemble of configurations that well approximate the target system, either in classical or *ab initio* MD simulations. Most of the underlying physics captured in the initial system are hence conserved in the EDS framework. The "knowledge field" or bias V can subsequently be updated in a straightforward manner according to the IBI iterative sequence:

$$V_{i+1}(r) = V_i(r) - \alpha k_{\mathrm{B}} T \ln \left[\frac{g^*(r)}{g_{\mathrm{i}}(r)} \right] f_{cut}(r) \tag{3.2}$$

We will refer to this method as the BIDS. The Boltzmann inversion of a site-site correlation function $g(r)$ allows for the many-body contributions to be accounted in the pairwise interaction in an average way. Table 3.1 shows the comparison between BIDS and its constituent methods. We note that whereas BIDS explores the point-wise variations in the bias potential, EDS interrogates the coefficient coupled to the bias potential with a form largely determined by a mollified unit-step function.

The post-optimisation of the bias via the pressure correction[107,309] was not pursued, due to its computational cost. A linear correction was reported by Reith *et al.*[107] to take 10 iteration cycles. The more satisfactory pressure correction via the virial expression was applied at each step of the iteration and after the RDF convergence criteria was met.[309] Moreover, the problem of the significantly deviating virial pressure is related to the loss of internal degrees of freedom through coarse-graining, whereas the water models remained atomistic in our study. In the case where the pressure correction was applied, the coarse-grained (CG) water model would lose agreement with the isothermal compressibility of the atomistic parent model.

In Eq. (3.2), we apply a cutoff function, f_{cut}, to the Boltzmann-inverted bias since information outside the maximum range of the correlation function data or half the dimension of the simulation box are likely artefacts of the empirical simulation. Any strong correlations beyond that separation suggest that the sim-

Table 3.1: Comparison between BIDS and its constituent methods, IBI and EDS. Note that iterative schemes are performed sequentially, whereas schemes that use replicas are performed in parallel.

	BIDS	IBI	EDS
Boltzmann inversion	Yes	Yes	No
Information used	RDF	RDF	Coordination number and its moments
Scheme	Iterative	Iterative	Replicas
Convergence criteria	RDF	RDF	Coupling constant
Initial system	Approximate model	Target PMF or approximate potential	Approximate model
Potential form	Point-wise	Point-wise	Function
Derived potential	Bias	Novel	Bias
System scale	Atomistic	Coarse-grained	Atomistic
AIMD application	Yes	No	Yes

ulation box is too small. Moreover, the non-bonded interactions between atoms must asymptotically approach zero at large distances. Steinbach & Brooks[325] mentioned that there have been three basic approaches to cut off the potential: discontinuous truncation at a cutoff distance, smooth switching to zero over an interval, and continuous shifting of the potential at all distances such that its value and derivative are zero at the cutoff. Here, the smooth switching function is most suitable as it gives a zero derivative at the cutoff. The function chosen here takes the form of a generalised Mei-Davenport-Fernando (MDF) taper[326] that decays the energy and forces smoothly between an inner cutoff r_m and outer cutoff r_cut:

$$f_\mathrm{cut}(r) = \begin{cases} 1 & \text{, if } r < r_\mathrm{m} \\ (1-\phi)(1-x)^3(1+3x+6x^2)+\phi & \text{, if } r_\mathrm{m} \leq r < r_\mathrm{cut} \\ \phi & \text{, if } r \geq r_\mathrm{cut} \end{cases} \tag{3.3a}$$

$$x = (r - r_\mathrm{m})/(r_\mathrm{cut} - r_\mathrm{m}) \tag{3.3b}$$

where ϕ is a unit interval[327] or closed interval $[1, 0]$ that allows the bias to arrive at some relative amplitude. The bias is effectively truncated at the outer cutoff when ϕ is set to zero. The position of the outer cutoff becomes important to maintain smoothness at that cutoff when ϕ is not zero. A position that corresponds with a root, *i.e.* x-axis intercept, of the bias should ensure an appropriate smoothness. The position of the inner cutoff is usually located slightly after the second peak of the O-O RDF. This ensures that damping does not occur in the region of the first peak that describes the tetrahedral coordination of water and

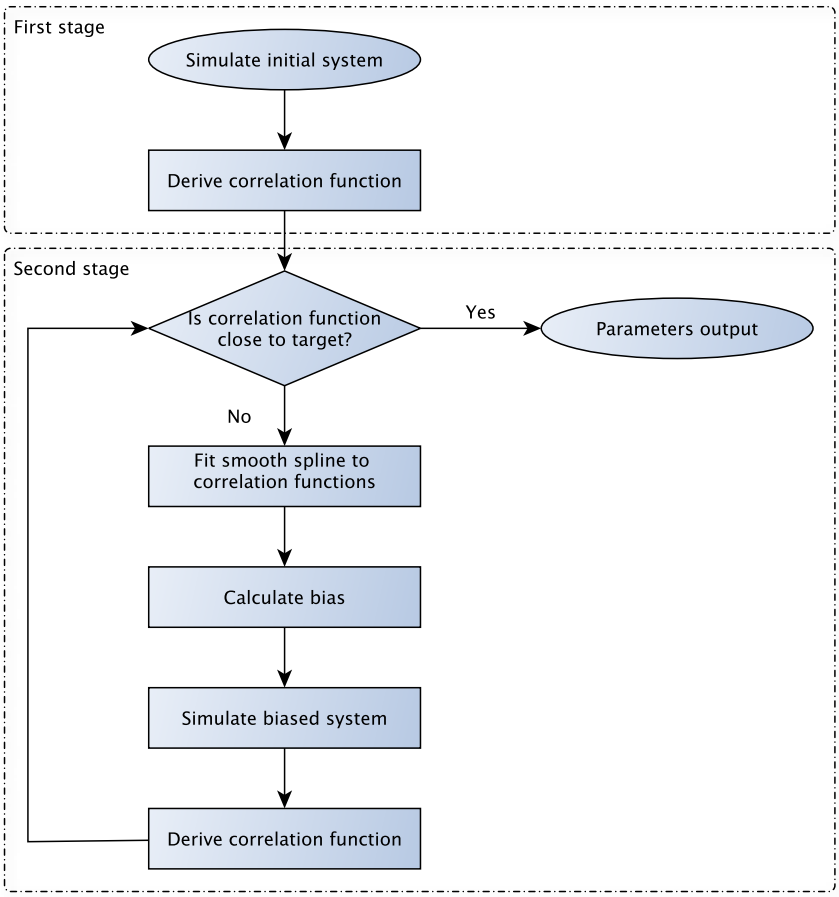


Figure 3.1: Block diagram flowchart of BIDS application.

the location of the second peak that corresponds to the tetrahedral geometry. The ability to cut off potentials also allows the bias to complement long-range corrections or the investigation of the bias derived from a larger simulation box to a smaller one.

3.2 Simulation Overview

The application of BIDS to atomistic simulations of liquid water at ambient temperature were performed in two stages, as shown in Figure 3.1. The first stage was an initial simulation of a standard classical or *ab initio* water model whereas the second stage was the application of BIDS using some target correlation function. Before applying to *ab initio* water, investigations of BIDS in classical water simulations were conducted using force fields: SPC,[44] SPC/Fw,[266] TIP3P,[45] TIP3P/Fs,[263] and TIP4P/2005.[265]

The bias potential obtained from the BIDS scheme should preserve the structural properties of the all-atom system. Wang *et al.*[309] showed that mapping

each water molecule to a CG bead at the position of the oxygen atom gives similar structure and compressibility as the all-atom system. Although we are not coarse-graining in this work, it still stands that the site-site correlation function $g(r)$ most suitable to obtain the corrective bias V for water is naturally the O-O RDF.

Unfortunately, the noise present in raw RDFs leads to a large amount of noise in the bias that is further accentuated in its force from the first derivative. This is further exacerbated in quantum mechanical simulations where the short simulation times lead to poorer statistics. The RDFs were thus smoothed (Section 3.3), as part of the second stage, using cubic basis spline[328] curve-fitting before calculating the resultant bias. The bin size of RDFs can also be optimised (Section 3.4) for the smooth splining procedure to improve their quality.

The calculated bias from the smooth splines will require post-processing (Section 3.5), *i.e.* physically sensible extrapolation to the left, due to the zeros at short range of the RDF. In contrast to a function, the bias is implemented into simulations as a tabulated potential in a series of discrete values that may incur out-of-range errors. The bias hence requires processing, *i.e.* padding with constants, during its calculation to ensure an acceptable input.

The processed bias was then used as a potential that augmented the initial O-O pairwise interaction to correct the initial water model towards the target water model. The procedure was performed iteratively until the O-O RDF of the target system was sufficiently reproduced, as measured by the mean absolute error (MAE) between the system and target RDFs. The error across the iterative simulation is discussed as part of the analysis in Section 4.2 and Section 4.5, where the value that is considered convergent is determined in the latter. The unbiased parameters, including scalar dynamic properties such as the self-diffusion coefficient D_s and the static dielectric constant ϵ_r, as well as the oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H) RDFs, were then outputted to investigate the influence of the bias.

3.3 Smooth Splining of the Radial Distribution Function

An in-house coordinate processing tool was used to generate the O-O RDFs by binning over a range from 0.025 Å to 9.025 Å at 0.050 Å intervals. The chosen number of bins was a compromise between the resolution of the first peak and the degree of noise introduced. A cubic basis spline was fitted using the statistical

computing software R[329] to the RDFs to reduce undesirable noise, notably beyond the first peak. The smoothing algorithm is provided in Appendix A supported by a module of functions in Appendix B.

The objective was to ensure smoothness in each RDF and its first derivative. This fitting was performed on the RDFs rather than the bias because of increased difficulty in distinguishing features from noise in the latter. Since fitting a cubic basis spline is essentially an estimation of the underlying "true" function, it is desirable to maintain magnitudes that must be everywhere zero or positive in the splined RDF to avoid introducing artefacts and/or adding residuals. This echoes the sentiments of Reith *et al.*[107] that the smoothing algorithm has to conserve the real physical features.

Depending on the boundary conditions imposed, a single type of spline may encounter difficulties in satisfying all these conditions, especially in the proximity of the first non-zero data point. Thus, two types of spline methods were employed. The cubic smoothing spline[330–332] was used for the right-hand side of the first peak, where the signal-to-noise ratio[333] was low for a short MD simulation. On the other hand, the monotonic cubic interpolation[334] was performed to the left of the first peak to ensure the function only increases from left to right. A knot[335] was chosen at the first peak to facilitate their coupling, as illustrated in Figure 3.2 (a). This knot is a common junction point with the same boundary conditions for the splines on both sides to ensure continuity.

The dataset for each RDF was simplified by disregarding the leading zeros, except the last two, from the spline fitting process. The cubic smoothing spline was first fitted. The degree of smoothness was controlled by allocating appropriate weights to manage the distinct disparity of the signal-to-noise ratio between the first peak and the remaining RDF. These weights,[336,337] λ, were varied as a position-dependent piecewise function, as illustrated in Figure 3.2 (d). The function transitioned from $\lambda = 1$, $\lambda = 0.001$ to $\lambda = 0.0001$ through a sigmoidal decay from the complementary error function. The knot at the first maximum was defined on the fitted smoothing spline with the corresponding value and gradient as boundary conditions. Then, the monotonic cubic interpolation using Hyman filtering of the Forsythe, Malcolm and Moler (FMM) method[338] was performed through the data points until the knot.

The RDF spline, as illustrated in Figure 3.2 (a), is constructed from the monotone-preserving spline and the latter part of the fitted smoothing spline at the knot. A non-parametric locally weighted regression was performed to facilitate the transition of the first derivative across the knot and to minimise residuals. The previously disregarded leading zeros are appended to the head. The completed

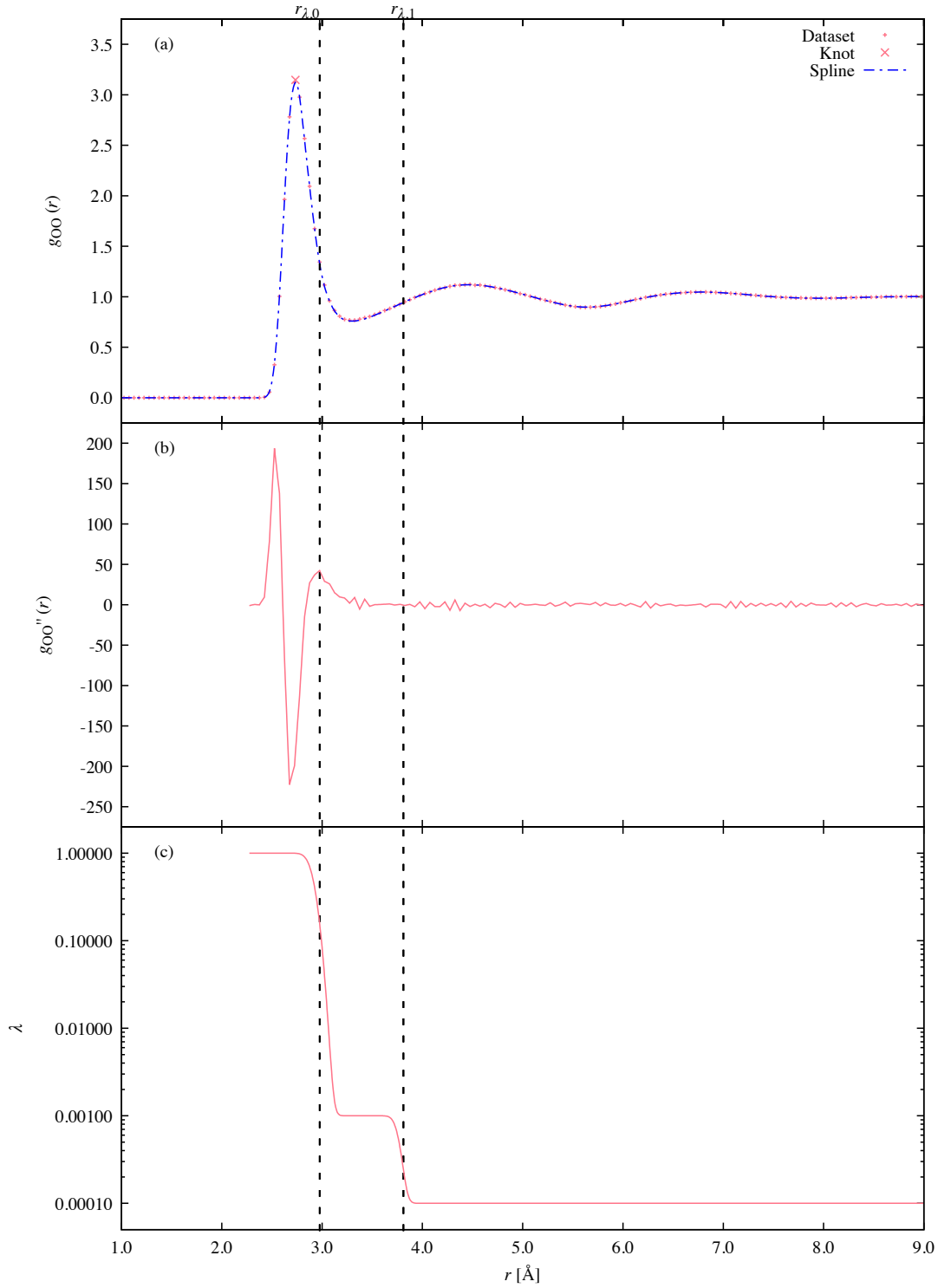


Figure 3.2: (a) Raw dataset of a typical RDF (SPC/Fw) with the position of the knot indicated, (b) second derivative of the raw dataset, and (c) weight allocations. The sigmoidal transition regions for the weight are positioned about the second maximum of the second derivative at $r_{\lambda,0}$ and in between the first valley and the second peak of the RDF at $r_{\lambda,1}$.

Table 3.2: The MAE and RMSD between the fitted smooth splines and their respective raw data. The raw data were the O-O RDFs derived from 10 ns simulations, where the first 0.5 ns was discarded for equilibration.

Model	MAE	RMSD
SPC	0.001565	0.002486
SPC/Fw	0.001568	0.002708
TIP3P	0.001113	0.001846
TIP3P/Fs	0.001476	0.002395
TIP4P/2005	0.001579	0.002734

RDF spline was written to within the same limits from 0.025 Å to 9.025 Å at smaller intervals of 0.010 Å. The root-mean-square deviation (RMSD) and MAE values for the splines were evaluated to ensure goodness of fit after the smoothing process, which are recorded in Table 3.2. The errors for SPC/Fw were used as a reference, since the splining parameters were determined based on SPC/Fw from a 10 ns simulation and as seen in Figure 3.2 (a), has achieved both a satisfactory goodness of fit and degree of smoothness. The measures of errors for the other water force fields have comparable values with SPC/Fw and thus achieved similar goodness of fit.

3.4 Smooth Spline Bin Size Optimisation

The smooth splining algorithm developed is used to generate the smooth RDFs required to produce smooth potential biases. It was then imperative to ensure the RDFs, and by extension the biases, were as smooth as possible when applied in BIDS to prevent any erroneous spikes or dips in the resulting forces. That this could be a problem was more apparent for the shorter production times, *i.e.* 50 ps and 20 ps, which produce noisy statistics. These times are of interest as they are typical of quantum mechanics (QM) simulations that is the intended application for our BIDS method. This is in contrast with the O-O RDF binned to 0.050 Å radial intervals with a production time of 10 ns, *i.e.* a very long MD simulation, used in benchmarking the algorithm. This length of time also allows us to benchmark the BIDS method. Barring a long production time for better statistics, the bin size of the RDF is the parameter that can be optimised at each production time of interest to improve the efficacy of the algorithm. The effect of smoothing on the O-O RDFs from 10 ns, 50 ps and 20 ps simulations using bin sizes from 0.050 Å to 0.185 Å at intervals of 0.005 Å were investigated.

The water models SPC/Fw and TIP3P were selected to gauge the performance of the algorithm with respect to the bin size dr. The former water model was a

Table 3.3: The MAE and RMSD between the raw O-O RDFs of various water models (predicted dataset) and the raw O-O RDF of SPC/Fw (true dataset). Naturally, SPC/Fw has zero errors when compared with itself. Both error measurements showed that the O-O RDF of TIP3P deviated the most from that of SPC/Fw.

Model	MAE	RMSD
SPC	0.060308	0.129698
SPC/Fw	0	0
TIP3P	0.084806	0.165201
TIP3P/Fs	0.041115	0.062123
TIP4P/2005	0.043597	0.130960

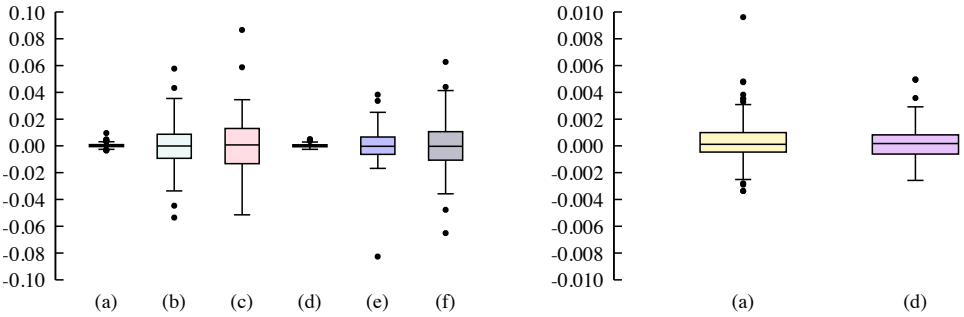


Figure 3.3: Box-and-whisker plot of the deviations of the raw O-O RDFs of: (a) SPC/Fw at 10 ns, (b) SPC/Fw at 50 ps, (c) SPC/Fw at 20 ps, (d) TIP3P at 10 ns, (e) TIP3P at 50 ps, and (f) TIP3P at 20 ps from their reference O-O RDF. The bin sizes used are 0.065 Å, 0.095 Å, 0.100 Å for 10 ns, 50 ps and 20 ps respectively. All spurious leading zeros at short-range were omitted. Although the interquartile range (IQR) box for (a) SPC/Fw at 10 ns was asymmetric, the magnitude of its lower and upper quartiles were below its MAE of 0.001208. About 10% of its deviations were considered outliers. The dissymmetry of the inner fences indicated by the whiskers in (c) SPC/Fw at 20 ps and (e) TIP3P at 50 ps showed greater variation in one direction over the other. The distributions were generally almost symmetric near a zero median and have few outliers, which were features that supported normality.

natural choice as the algorithm was benchmarked against it during development. As shown in Table 3.3, the latter water model was chosen as it produces the O-O RDF most different to that from SPC/Fw. In particular, its O-O RDF has a flatter profile beyond the first peak. The details of the water simulation box are described in Section 4.1.

For each water model, a O-O RDF has to be designated as the reference dataset to allow comparison with different bin sizes and across the production times of interest. The smooth spline of the O-O RDFs with 0.050 Å bin size from 10 ns simulations was chosen over the raw data because the noise was removed. Moreover, the differences between the smooth splines and the raw data were extremely small as shown in Table 3.2. Since noise consists of random errors, it stands that the deviations of the O-O RDFs with other bin sizes and production times from the reference O-O RDFs should be a normal distribution if the chosen smooth splines truly represent the underlying O-O RDF. This appears to be the

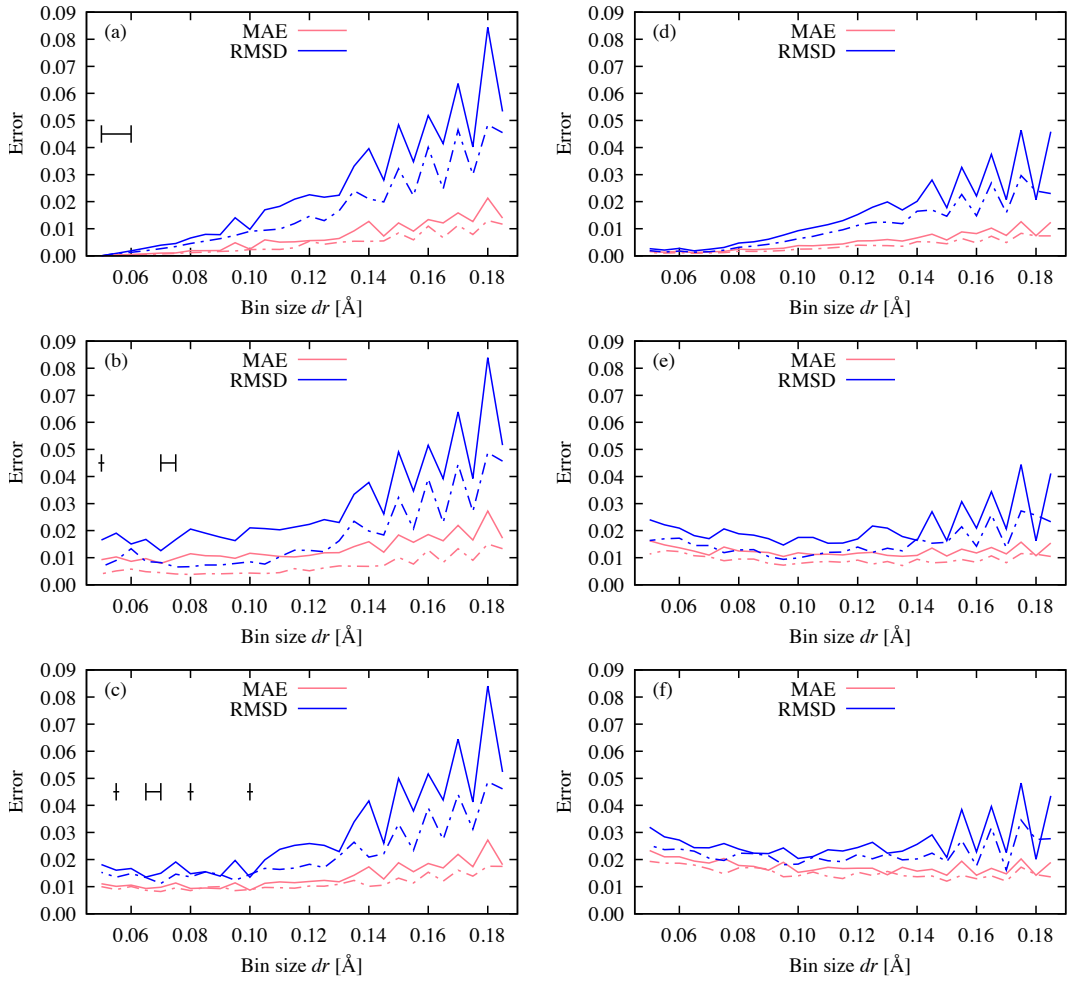


Figure 3.4: The SPC/Fw (solid) and TIP3P (dot-dash) MAE and RMSD values with various bin sizes for: (a) smooth splines at 10 ns, (b) smooth splines at 50 ps, (c) smooth splines at 20 ps, (d) raw data at 10 ns, (e) raw data at 50 ps, and (f) raw data at 20 ps. The domains indicated by the horizontal bars or marks satisfied the criteria that both the MAE and RMSD values were below their respective means across the bin sizes 0.050 Å to 0.130 Å and common to both SPC/Fw and TIP3P. An additional criterion that both their MAE and RMSD were below that of their raw data counterparts, was imposed on those domains of the smooth splines. Any domains greater than 0.130 Å bin size were excluded due to the large errors.

case as illustrated using the box-and-whisker plot[339] in Figure 3.3.

For any simulation of finite time, noise is more pronounced with smaller bin sizes. The prominence of noise when smaller bin sizes were used may contribute to undesired artefacts in the resulting smooth spline. The bin size can typically be increased above the wavelength of the noise to provide a smoother profile. However, the use of larger bin sizes meant that some details, particularly in the first peak, failed to be captured satisfactorily. For each production time of interest, an optimal bin size can be found where the algorithm achieves errors equal to or less than the corresponding raw data when fitting a smooth spline.

When calculating the MAE and RMSD for the resultant smooth splines, the

number of bins used was consistent with the same number of bins as that of the bias potentials determined with 0.050 Å bin size. The errors of the raw data compared to the reference O-O RDF were also calculated. The number of bins used for the reference O-O RDF was aligned with that of the raw data. The left column of Figure 3.4 illustrates that the smooth splining of TIP3P achieved smaller errors than that of SPC/Fw at all bin sizes and production times, which is consistent with the errors in Table 3.2. As the production time was reduced resulting in poorer statistics, the overall errors increased as expected and is more apparent in the domain near the 0.115 Å bin size and below. The difference between the smooth splines at 50 ps and at 20 ps did not appear significant. We note that the zero errors for both water models at 0.050 Å in Figure 3.4 (a) does not mean a perfect smooth spline. They simply represent their respective reference O-O RDFs.

One glaring observation was the significantly greater errors in the larger bin sizes for the smooth splines compared with their raw data counterparts. This was due to the lack of data points to describe the first peak and/or to fit the curve of the first valley, which typically manifested as a shorter first peak and/or a deeper first valley in the smooth spline. The algorithm therefore failed to satisfactorily fit a smooth spline with bin sizes around 0.130 Å and greater.

The errors of the smooth splines generally increased as the bin size increased while the RMSD becomes increasingly erratic with 0.130 Å bin size and greater. It was apparent from the smaller bin sizes that the algorithm tended to perform well given more data points and was capable of averaging out the noise in the shorter production times of 50 ps and 20 ps. A similar increasing trend was also observed in Figure 3.4 (d) due to comparable data point distribution for the raw data from long 10 ns simulations. For shorter production times of 50 ps and 20 ps for the raw data, the errors are higher when using the larger and smaller bin sizes whereas the errors are lower in between, *i.e.* a catenary curve.[340] The same erratic patterns, but inverted, were noticeable in the RMSD at the same domain near the 0.130 Å bin size and greater.

The optimum bin sizes for the algorithm to fit a smooth spline were stipulated as being within the domains where both the MAE and RMSD were below their respective means across the bin sizes 0.050 Å to 0.130 Å for SPC/Fw and TIP3P simultaneously. Additionally, the desired smooth splines should have smaller MAE and RMSD than their raw counterparts. The domains that meet these acceptance criteria are indicated in Figure 3.4 as horizontal bars.

Within the domains indicated by the horizontal bars, some reasonable bin sizes were selected as the desired optimum and tabulated in Table 3.4. It is logical to

presume a shorter simulation provides less statistics resulting in greater noise for the O-O RDF that requires the use of larger bin sizes. Hence, the selection of the optimum bin sizes was such that they increased with shorter production times.

Table 3.4: The optimum bin sizes selected for fitting a smooth spline using the algorithm.

Production time [ps]	Bin size [Å]
10000	0.050
50	0.070
20	0.080

3.5 Processing the Boltzmann Inverted Bias

The bias was calculated from the ratio between the RDF splines, as shown in the second term of Eq. (3.2). Once the bias was generated, the corresponding force can be obtained through its derivative according to Eq. (2.50). A typical corrective bias and its force after processing is shown in Figure 3.5. The smallest non-zero value in the O-O RDFs was found at ∼2.375 Å, which delineates the effective zero-probability region. Although usually only the bias in the region greater than ∼2.375 Å is felt by the system, the separation of oxygen atoms may fall into the zero-probability region, particularly in the equilibration phase given some initial random velocities. Hence, a small offset to at least ∼2.200 Å was used for the bias and a reasonable extrapolation to the left was necessary.

A mathematical analysis of the IBI method by Hanke[341] showed that the PMF is a Lennard-Jones (LJ) type potential. A simplified LJ type function was hence fitted to the short-range repulsion of the constituent PMFs to facilitate an extrapolation:

$$f_{LJ}(r) = \frac{A}{r^n} - \frac{B}{r^{n/2}} + C \tag{3.4}$$

The parameters A, B, C and n were determined through a Gauss-Newton iteration[330,342] to find the nonlinear least-squares estimate. The fitting was first performed for the target PMF with C set to zero to determine the other parameters. The initial values of these parameters were estimated using the usual LJ formulation:

$$A = 4\epsilon\sigma^{12} \tag{3.5a}$$

$$B = 4\epsilon\sigma^{6} \tag{3.5b}$$

$$n = 12 \tag{3.5c}$$

where ϵ is the depth of the first PMF well and σ is the radius prior to the first well at which the PMF is zero. The converged parameters were then fixed in the fitting to the system PMF with C allowed to vary. This ensured that the resulting bias would be a constant at short range. Consequently, this leads to a zero force where the Boltzmann inversion of RDFs gives undefined data. However, coercing the system PMF to conform to the parameters A, B and n of the target PMF causes some artefacts between the right of the constant region r_0 and the left of the defined region r_1, as shown in Figure 3.5 (a). This region was discarded and a cubic-spline interpolation between r_0 and r_1 produced a smooth transition.

The MDF function can be used to taper the effects beyond the two-body interaction. This is accomplished by gradually reducing the bias to zero over a span constrained by the inner cutoff r_m and the outer cutoff r_cut, as described by Eq. (3.3) when $\phi = 0$. The values remained unaffected to the left of r_m whereas zeros were padded to the right of r_cut. The position of r_cut was usually half the cubic box length but can be shorter to truncate the effects of the multi-body contributions. When ϕ is not zero, the preferred r_cut was at a root of the bias. As the MDF function produced an inflection point at r_cut, locating it at a root facilitated smoother tapering to some proportion of the amplitudes. The position of r_m was then located at the extrema prior to r_cut to allow a smoother transition into the tapered region.

The corresponding force was obtained from the differentials of the functions from a cubic basis interpolation passing strictly through all the data points of the bias. The BIDS simulations were performed using the processed bias and its force. Since the head of the bias is padded with some constant while the tail is padded with zeros, the bias can be trivially extended as required to prevent out-of-range errors in simulations.

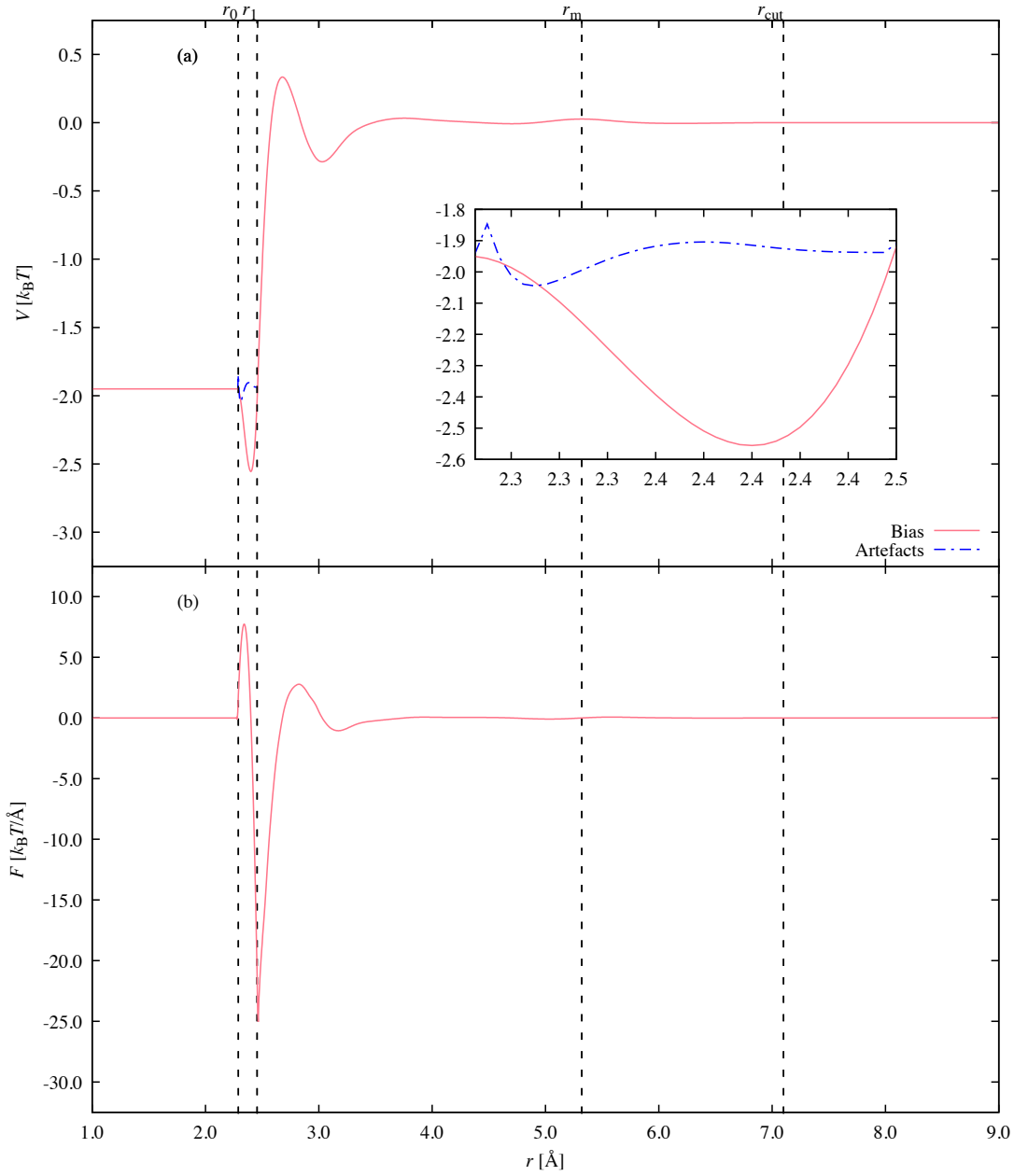


Figure 3.5: The typical: (a) potential and corresponding (b) force for a processed bias (SPC to SPC/Fw). The inset shows the region containing artefacts (dot-dash) from fitting similar LJ-type function to the system RDF and the target RDF. For the values to the left of r_0, the bias was padded with the value at r_0. The bias was interpolated, where the bias was unstable from the LJ-type fitting, between radii r_0 and r_1. Using the MDF function, the bias was decayed to zero between radii r_m and r_{cut}.

References

[44] H. J. C. Berendsen et al. "Interaction Models for Water in Relation to Protein Hydration". In: *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*. Ed. by B. Pullman. Dordrecht: Springer Netherlands, 1981, pp. 331–342. DOI: 10.1007/978-94-015-7658-1_21.

[45] W. L. Jorgensen et al. "Comparison of Simple Potential Functions for Simulating Liquid Water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: 10.1063/1.445869.

[97] A. D. White and G. A. Voth. "Efficient and Minimal Method to Bias Molecular Simulations with Experimental Data". In: *Journal of Chemical Theory and Computation* 10.8 (2014). PMID: 26588273, pp. 3023–3030. DOI: 10.1021/ct500320c.

[107] D. Reith, M. Pütz, and F. Müller-Plathe. "Deriving Effective Mesoscale Potentials from Atomistic Simulations". In: *Journal of Computational Chemistry* 24.13 (2003), pp. 1624–1636. DOI: 10.1002/jcc.10307.

[263] U. W. Schmitt and G. A. Voth. "The Computer Simulation of Proton Transport in Water". In: *The Journal of Chemical Physics* 111.20 (1999), pp. 9361–9381. DOI: 10.1063/1.480032.

[265] J. L. F. Abascal and C. Vega. "A General Purpose Model for the Condensed Phases of Water: TIP4P/2005". In: *The Journal of Chemical Physics* 123.23 (2005), p. 234505. DOI: 10.1063/1.2121687.

[266] Y. J. Wu, H. L. Tepper, and G. A. Voth. "Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties". In: *The Journal of Chemical Physics* 124.2, 024503 (2006). DOI: 10.1063/1.2136877.

[309] H. Wang, C. Junghans, and K. Kremer. "Comparative Atomistic and Coarse-Grained Study of Water: What do we Lose by Coarse-Graining?" In: *The European Physical Journal E* 28.2 (2009), pp. 221–229. DOI: 10.1140/epje/i2008-10413-5.

[318] J. W. Pitera and J. D. Chodera. "On the Use of Experimental Observations to Bias Simulated Ensembles". In: *Journal of Chemical Theory and Computation* 8.10 (2012). PMID: 26592995, pp. 3445–3451. DOI: 10.1021/ct300112v.

[320] A. T. Brünger and M. Nilges. "Computational Challenges for Macromolecular Structure Determination by X-Ray Crystallography and Solution NMR Spectroscopy". In: *Quarterly Reviews of Biophysics* 26.1 (1993), 49?125. DOI: 10.1017/S0033583500003966.

[321] J. Dolenc et al. "Methods of NMR Structure Refinement: Molecular Dynamics Simulations Improve the Agreement with Measured NMR Data of a C-Terminal Peptide of GCN4-p1". In: *Journal of Biomolecular NMR* 47.3 (2010), pp. 221–235. DOI: 10.1007/s10858-010-9425-9.

[322] J. Iwahara, C. D. Schwieters, and G. M. Clore. "Ensemble Approach for NMR Structure Refinement against 1H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule". In: *Journal of the American Chemical Society* 126.18 (2004). PMID: 15125681, pp. 5879–5896. DOI: 10.1021/ja031580d.

[324] A. D. White et al. "Communication: Improved *ab initio* Molecular Dynamics by Minimally Biasing with Experimental Data". In: *The Journal of Chemical Physics* 146.4 (2017), p. 041102. DOI: 10.1063/1.4974837.

[325] P. J. Steinbach and B. R. Brooks. "New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation". In: *Journal of Computational Chemistry* 15.7 (1994), pp. 667–683. DOI: 10.1002/jcc.540150702.

[326] J. Mei, J. W. Davenport, and G. W. Fernando. "Analytic Embedded-Atom Potentials for fcc Metals: Application to Liquid and Solid Copper". In: *Phys. Rev. B* 43 (6 1991), pp. 4653–4658. DOI: 10.1103/PhysRevB.43.4653.

[327] T. Riedrich. "R. G. Bartle, The Elements of Real Analysis. XIV + 447 S. m. Fig. New York/London/Sydney 1964. John Wiley & Sons, Inc. Preis geb. 83 s. net". In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift fr Angewandte Mathematik und Mechanik* 45.5 (), pp. 365–365. DOI: 10.1002/zamm.19650450519.

[328] C. de Boor. *A Practical Guide to Splines*. Applied Mathematical Sciences. Springer New York, 1978.

[329] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2017.

[330] J. Chambers and T. Hastie. *Statistical Models in S*. Wadsworth & Brooks/Cole computer science series. Wadsworth & Brooks/Cole Advanced Books & Software, 1992.

[331] P. Green and B. Silverman. *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1993.

[332] T. Hastie and R. Tibshirani. *Generalized Additive Models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1990.

[333] G. Trejo-Caballero et al. "B-spline Surface Approximation Using Hierarchical Genetic Algorithm". In: *Advances in Soft Computing and Its Applications*. Ed. by F. Castro, A. Gelbukh, and M. González. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 52–63.

[334] F. Fritsch and R. Carlson. "Monotone Piecewise Cubic Interpolation". In: *SIAM Journal on Numerical Analysis* 17.2 (1980), pp. 238–246. DOI: 10.1137/0717021.

[335] H. Prautzsch, W. Boehm, and M. Paluszny. "Smooth curves". In: *Bézier and B-Spline Techniques*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 91–107. DOI: 10.1007/978-3-662-04919-8_7.

[336] K. Höllig. *Finite Element Methods with B-Splines*. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics, 2012.

[337] K. Höllig and J. Hörner. *Approximation and Modeling with B-Splines*. Other Titles in Applied Mathematics. Society for Industrial and Applied Mathematics, 2015.

[338] F. Grund. “Forsythe, G. E. / Malcolm, M. A. / Moler, C. B., Computer Methods for Mathematical Computations. Englewood Cliffs, New Jersey 07632. Prentice Hall, Inc., 1977. XI, 259 S”. In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* 59.2 (1979), pp. 141–142. DOI: 10.1002/zamm.19790590235.

[339] J. Tukey. *Exploratory Data Analysis*. Addison-Wesley series in behavioral science. Addison-Wesley Publishing Company, 1977.

[340] E. Lockwood. *Book of Curves*. A Book of Curves. Cambridge University Press, 2007.

[341] M. Hanke. “Well-Posedness of the Iterative Boltzmann Inversion”. In: *Journal of Statistical Physics* 170.3 (2018), pp. 536–553. DOI: 10.1007/s10955-017-1944-2.

[342] D. Bates and D. Watts. *Nonlinear Regression Analysis and Its Applications*. Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series. Wiley, 1988.

Chapter 4

Development of Boltzmann Inversion Directed Simulation using Force Fields

Characterising the scaling of quantum mechanics (QM) calculations poses the inherent problem of a mixture of scaling regimes due to the complicated many-step nature of computational algorithms.[343] That is to say, the multitude of scaling aspects each exert a degree of influence on the overall scaling. Nonetheless, the effective scaling can be appropriately defined by mathematical observation of the dominant part of the algorithm. The formal scaling of density functional theory (DFT) as conventionally accepted is given as $\mathcal{O}(N^3)$–$\mathcal{O}(N^4)$.[344,345] Consequently, the scaling of computational cost with respect to system size is a more apparent limitation when compared to the $\mathcal{O}(N^2)$–$\mathcal{O}(N \log N)$ scaling of empirical force fields. Moreover, the prefactor to the cost is much higher for DFT. Before application to more resource intensive simulations using DFT, the efficacy of the Boltzmann inversion directed simulation (BIDS) method was first investigated by performing molecular dynamics (MD) simulations using water force fields.

As the BIDS method is a fixed-point iterative scheme, we need to ensure that the target pair correlation function and the corresponding properties of interest reliably converge and remain stable across iterations. Although 10 ns simulations were used for the water force field simulations, such time lengths are difficult to achieve in DFT water simulations due to the much greater computational expense. Hence, the statistics from 50 ps simulations of the former were used as an analog to inform on the convergence and stability of the bulk-phase properties, *i.e.* radial distribution function (RDF), self-diffusion coefficient and static dielectric constant, for the latter.

The short time length leads to poorly converged RDFs that are used to derive the biases. Such RDFs provide poor descriptions of the features, *i.e.* peaks and valleys. Unfortunately, the role of smoothing was to remove noise rather than reproduce features. As the fidelity of the biases were apparently dependent on time length, it was thus also necessary to make certain that these biases iteratively converged to the that of the long simulations. This was accomplished by investigating the error between the biases across iterations from 50 ps simulations to the corresponding bias that provided the greatest improvement to the RDF of each model from 10 ns simulations.

Since long-range dispersion corrections that are commonly used in DFT water simulations are well defined, the bias potential was applied up to the medium-range region. This was achieved by truncating the bias at some cutoff, where the Mei-Davenport-Fernando (MDF) function was employed to taper the potential smoothly to zero and thus provide a smooth force continuity. Its effects on the RDF were examined to find an minimum cutoff value that did not adversely affect the improvements from the bias. The medium-range region is noted to be particularly important for the equilibrium structures of van der Waals (vdW) complexes and the thermodynamic properties of larger molecules.[164,346] However, the overlap of the short- and long-range asymptotic interactions in that region are not well understood.

Another parameter acting to fine tune the bias was the scaling coefficient. It functions to prevent overcorrection in the RDF, but recommended values found in studies using iterative Boltzmann inversion (IBI) were typically given over a range and specific to the system investigated. Thus, a range of scaling coefficient were explored to find the optimal parameter that achieves the lowest error compared to the target RDF. This exploration was carried out across two iterations for both the aforementioned 10 ns and 50 ps time lengths. Additionally, shorter 20 ps scans were considered as a prospective time length to adequately interrogate for the optimal scaling coefficient.

4.1 Simulation Details

The MD program utilised for classical simulations was the large-scale atomic/molecular massively parallel simulator (LAMMPS).[347] This program was used as it readily allows one or more pair interactions to be assigned to each pair of atom types with the *hybrid/overlay* style. When used in tandem with the *table* style, which creates interpolation tables from a file, the bias potential energy and force values as a function of separation can be added on top of the usual potential.

The parameters of the force fields, which use the Lennard-Jones (LJ) potential, for the water models used are reproduced in Table 2.2.

Simulations of 256 water molecules in a periodic cubic box were performed in the canonical ensemble (NVT) at 298 K. A cubic cell length of 19.7 Å was chosen, which leads to a density of 1001.95 kg m^{-3} that was close to the density of water at room temperature. The small system size was adopted for the force field simulations to be indicative of that found in *ab initio* molecular dynamics (AIMD) simulations. In order to maintain the target temperature, the particle velocities were coupled to a chain of 5 Nosé-Hoover thermostats with a relaxation time of 0.1 ps.[348] A timestep of 1 fs was used in the simulations. The same simulation parameters for the standard classical simulations were used in the bias-corrected simulations. The water box was first equilibrated using SPC/Fw water under these conditions for 10 ns before the initial simulations. The simulations were run for 10 ns, including the initial 0.5 ns allowed for equilibration. In order to investigate the typical time lengths in QM, 50 ps simulations were also sampled. Since White *et al.*[324] found 2–5 ps sufficient to converge the bias within a few percent, the initial 5 ps was treated as equilibration.

4.2 Convergence and Stability of Bulk-Phase Properties

The application of BIDS in an iterative sequence requires the potential of mean force (PMF)-generated perturbation to the bias to act as a self-corrective mechanism so that the resulting RDF converges to the target RDF. The bias that reproduces the target RDF is thus a fixed point in the iteration.[349] Previous works based on a similar premise of an iterated potential perturbed by the PMF between two atoms in a fluid have substantiated that the iteration is convergent.[5,107,306,309,350,351]

With the assumption that a set of site-site pairwise additive potentials exists, which can generate the observed site-site pair correlation function, Soper[305] has argued that convergence is established as a consequence of the uniqueness theorem for the site-site pair correlation function in pairwise additive systems. A qualitative analysis of the site-site correlation function as a sum of the term from pair interactions and the term arising from many-body effects was made. The analysis showed that changes in the many-body effects due to the change in the pair potential will not completely negate the changes in the pair interaction when averaged over all r values. Thus, iteration will ensure that any overcorrection or

undercorrection of the pair potential caused by the many-body interactions will be diminished.

In the case of the Weeks-Chandler-Andersen (WCA) liquid[352] and the LJ liquid, Reith *et al.*[107] observed that IBI introduced artefacts below $\pm 0.02\, k_{\mathrm{B}}T$ to a known correct potential without smoothing. By smoothing using a five-point running average,[353] the fluctuations were reduced below $\pm 0.005\, k_{\mathrm{B}}T$. The accumulation of such numerical errors after smoothing was highly unlikely to cause any divergence. When deriving the coarse-grained (CG) potential for a polyisoprene melt, it was reported that the RDF converged after about five iterations based on a weighted merit function that more strongly penalised deviations at small distances.[354] The deviations after that iteration remained stable as they fluctuate within their statistical accuracy and were regarded as noise.

BIDS, using experimental water as the reference, was applied to a variety of LJ water force fields for up to five iterations, as per the finding of Reith *et al.*[107] for the number of iterations required in the case of melts. The force fields in this study were SPC, SPC/Fw, TIP3P, TIP3P/Fs, and TIP4P/2005. As the force fields and their RDFs are varied, this ensures the generality of changes in the bulk-phase properties of the simulated liquid water at room temperature with the application of a bias. Henceforth, the biased water simulations to experiment are identified with the notation BIDS as a prefix to the initial water model. Figure 4.1 shows the trajectory of the properties of interest given a corrective bias, which was scaled by half to improve stability as they have a tendency to be overcorrected according to the observation made by Reith *et al.*[107] The 50 ps simulations are taken to be indicative of how the procedure would work for QM simulations.

The mean absolute error (MAE) between the system RDF from simulation and the target RDF from experiment was used as a measure of error. Instead of a merit function that gives more weight to short distances, the MAE has the same weight at all distances. Improving the RDF beyond the first peak, especially the second peak, was hence as important. Despite different initial errors, the MAE of the biased water force fields rapidly converge to a similar fixed point in the iteration for each time length. The rapid convergence satisfies a criterion of Reith *et al.*[107] for an efficacious iterative solver for the effective potential. As expected, diminishing return was observed across the iterations. The MAEs for the 50 ps simulations reach an error about 0.0120 ± 0.0008 while those for the 10 ns simulations reach a lower error about 0.0048 ± 0.0004, as shown in Figure 4.1 (a) and (d) respectively. Clearly, the MAE approaches a lower threshold limit as the statistical noise reduces as the length of the simulations increases. Overall, BIDS

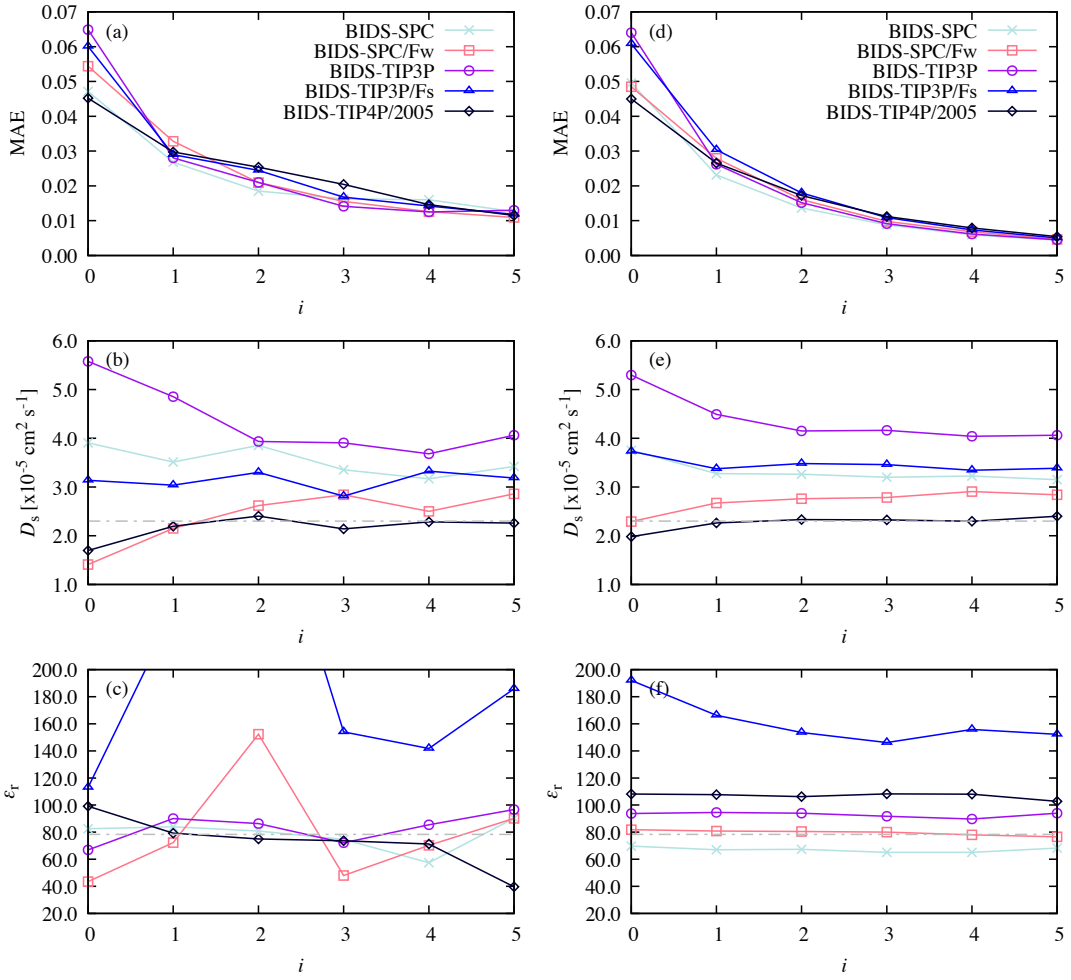


Figure 4.1: The change in properties of water during BIDS from 50 ps simulations: (a) MAE, (b) self-diffusion coefficient, D_s, (c) static dielectric constant, ϵ_r, and from 10 ns simulations: (d) MAE, (e) self-diffusion coefficient, D_s, (f) static dielectric constant, ϵ_r, across the iteration sequence from $i = 0$ to $i = 5$ for various water force fields in comparison to experimental water. The dot-dash lines are the experimental values for self-diffusion coefficient and static dielectric constant. The dielectric constant values 242.22 at $i = 1$ and 356.15 at $i = 2$ for BIDS-TIP3P/Fs are intentionally left out to improve clarity.

can sufficiently reproduce the target RDF within statistical accuracy.

The self-diffusion coefficient D_s is a fundamental quantity in the study of dynamics in fluid systems. Since the diffusion coefficient is a function of the displacement of the spatial position of water molecules with time, a pairwise bias directly affects it through a change in the forces between any two oxygen atoms. Thus, the diffusion coefficient is suitable as a measure of the effects of the bias on the water dynamics. We note that the finite-size effects resulted in a dependence of the self-diffusion on the simulation box size, which provides a slight increment with respect to the inverse of the cell length. Hence, the values reported in this study were slightly less than the true bulk self-diffusion. In Figure 4.1 (e), most of the water force fields from the 10 ns simulations converge to a value closer to

the experimental value in a single step. The change is within a modest 9.57 % to 16.66 % from the initial value. Although the 50 ps simulations in Figure 4.1 (b) have greater statistical errors, they suggest a similar change as the 10 ns simulations. TIP4P/2005 is noteworthy as it achieves a diffusion coefficient close to the experimental value, though the finite-size effects were not accounted for. Unfortunately, the diffusion coefficient may not necessarily converge closer to the target value as shown by SPC/Fw. It is likely that the direct effect of a bias is to shift in a direction, depending on an enhanced or a diminished structure, rather than outright correcting the diffusion coefficient.

The influence of a bias on the static dielectric constant, ϵ_r, is minimal based on the 10 ns simulations, as it maintains a stable value across the iterations. The lack of improvement in the dielectric constant is not unexpected, especially for the water models that are rigid. Being a function of the dipole moments, the dielectric constant is largely dependent on the average O-H bond length and the average H-O-H bond angle, which are not appreciably perturbed by the oxygen-oxygen (O-O) pairwise bias, if at all in the case of rigid molecules. The notable exception is the BIDS-TIP3P/Fs simulation, which changes by an order of magnitude greater than the rest. This is likely due to its flexible and softer bond angle parameter allowing a greater change to the average H-O-H angle. In comparison with the stability of the longer simulations, the dielectric constant from the 50 ps simulations show fluctuations that are likely due to a lack of statistics from the short time length. According to Wu *et al.*[266] relaxation times up to 2 ns were found to be required for the dielectric constant depending on the water force field and the starting structure. The thermostat used also affects the length of the trajectory required for proper convergence. Gereben & Pusztai[355] found that the Nosé-Hoover thermostat took up to 7 ns for SPC/E water with a system size of 216 molecules, while it only took 5 ns for the Berendsen thermostat. They noted that the former thermostat increased the converged value by roughly 5 % from the latter. This is due to thermostats modifying either the equations of motion or the velocities directly, which affects the dynamics and thermodynamics of the system.[356]

4.3 Effects of Time Length on the Bias Potential

The bias potential is derived from the system site-site correlation function, $g_i(r)$, which is dependent on the statistical accuracy. The statistical uncertainties can be reduced with longer simulations. This is routinely afforded in simulations of water force fields, but the computational cost of *ab initio* water is apprecia-

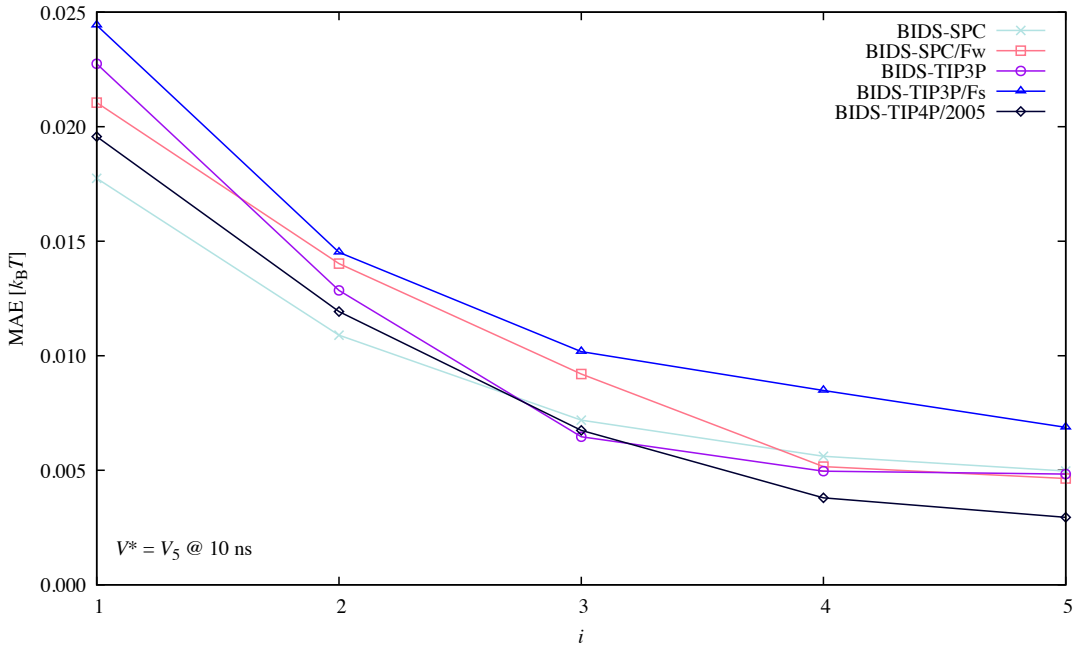


Figure 4.2: The MAE between the biases from 50 ps simulations across each iterative step and the bias from 10 ns simulation at $i = 5$. This demonstrates convergence to the bias that provided the greatest improvement to the RDF of each model.

bly higher. Thus, the reproducibility of the bias from the 10 ns simulation by comparison to the corresponding 50 ps simulation was investigated. When comparing these biases, the short-range extrapolation and any long-range padding were disregarded.

Using the bias from the 10 ns simulation that best reproduces the target RDF as the reference V^*, the trajectory of the bias from the 50 ps simulation to the "true" bias V^* can be obtained as shown in Figure 4.2. The MAEs at the first iteration $i = 1$ are about $0.021\,k_BT$, which is around the magnitude of nonzero artefacts introduced by iterations as observed by Reith *et al.*[107] The trajectory of the biases are convergent to their "true" biases and fall below an error of $0.007\,k_BT$ at $i = 5$. This error is of the same order of magnitude as the WCA potential for a dense liquid after 35 iterations.[107] Hence, the corrective biases can be sufficiently derived from the 50 ps simulations.

The sampling of a short time length simulation does lead to insufficiently converged RDFs. Although this was demonstrably non-detrimental to the convergence of the overall bias, the poor statistics does introduce errors. An example from the BIDS-SPC/Fw iterations is shown in Figure 4.3. The fitting of a smooth spline acts to alleviate the noise. It was, however, incapable of addressing the discrepancies in the intermolecular features, usually in the first valley, of the unconverged RDF. The discrepancies consequently manifested themselves as sys-

tematic increments to the error of the bias potential, which were compounded across the iterations.

4.4 Minimum Bias Potential Cutoff

According to Berweger *et al.*[357] a cutoff for non-bonded interaction energies and pair forces is a strong perturbation and should be regarded as a parameter. In this regard, they reoptimised the LJ radius, σ, and atomic charges, q, for varying cutoff distances of the SPC water force field. They found the cutoff to have negligible effect beyond 7.0 Å and 5.0 Å for those parameters respectively. Here, we investigated the effect of different cutoffs to determine a shorter-range potential that reproduces the target RDF. The directed simulation of SPC/Fw to experimental water was used for this investigation. The original bias has a 9.0 Å cutoff applied. We selected a scaling coefficient for the bias with the lowest MAE when compared to the BIDS-SPC/Fw bias derived from the 10 ns simulation at iteration $i = 5$ in Section 4.2. Based on the errors found in Table 4.1, a scaling coefficient $\alpha = 0.9$ was used.

Table 4.1: The MAE between the biases at various scaling and the original bias.

Scaling coefficient α	MAE
0.5	0.000506
0.6	0.000434
0.7	0.000366
0.8	0.000307
0.9	0.000265
1.0	0.000269

Favourable cutoff distances are located near the RDF minima to ensure fewer particles are subjected to the resultant force change within that region. We elected to take cutoffs at 7.0 Å, 5.0 Å and 3.1 Å. The former two also coincided with the critical points, where the cutoff had negligible effect after, for the σ and q parameters respectively. The latter was used because specifying a cutoff at 3.0 Å would trigger a tolerance check in the smoothing algorithm. This was implemented to prevent the applied tapering function from acting on the first peak of the RDF. Since the generalised MDF tapering function (see Equation 3.3) was applied to smoothly decay the bias to zero at the outer cutoff, *i.e.* $\phi = 0$, the force discontinuity from truncating was avoided. A force discontinuity would result in the energy not being conserved, which can be slightly tempered by decreasing the time step used to integrate the equations of motion.[325] Hence, a truncated bias would be undesirable when applied to the simulations of water.

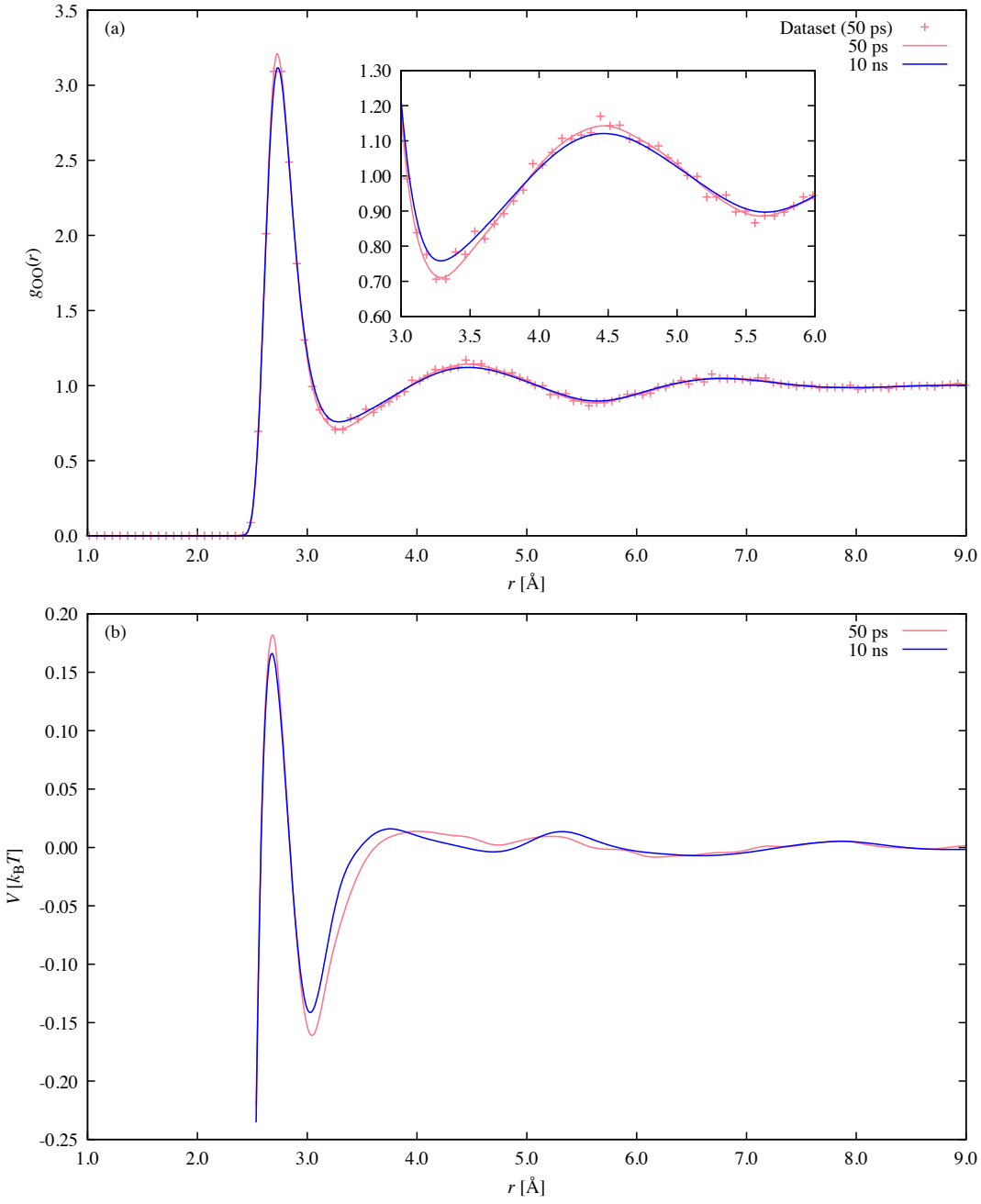


Figure 4.3: The differences between the 50 ps and the 10 ns BIDS-SPC/Fw simulations to experimental water for (a) the RDF splines at $i = 0$ and (b) the bias potentials at $i = 1$. The RDF splines were used in calculating the biases. The poor statistics of the 50 ps simulation lead to a more structured first peak and first valley. Consequently, these discrepancies between the splines are translated to the resulting biases.

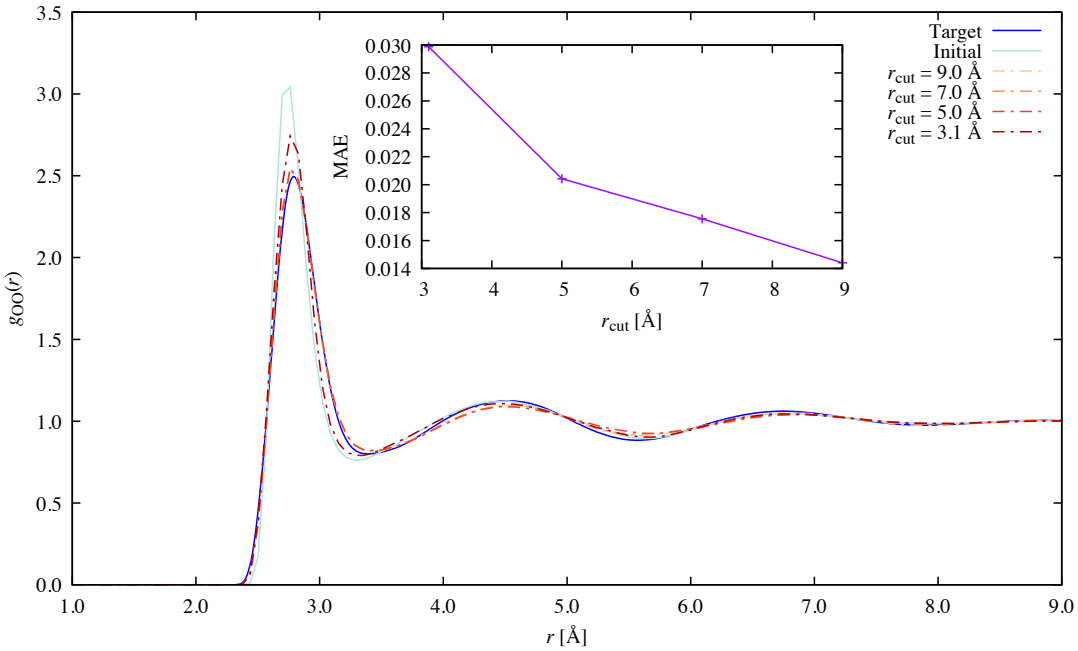


Figure 4.4: The RDFs of the directed simulation of SPC/Fw (initial) to experiment (target) using a bias potential at varying radial cutoffs. The RDFs which were applied a bias with a cutoff above 5 Å are similar. The inset shows the error from the target to more clearly discern the difference in improvements made by the biases with varying cutoffs.

From Figure 4.4, a cutoff equal to or greater than 5.0 Å was demonstrated to yield a similar shape to the RDF from the original bias. This suggests that the water box size when investigating BIDS in QM calculations should at least be greater than 10.0 Å. Otherwise, images of the same water will appear in two places in the RDF and so be intrinsically correlated. Based on the same figure, a cutoff at 7.0 Å for the bias potential was recommended as it was positioned after the problematic medium-range correlation region and the greater of the two cutoffs reported by Berweger *et al.*[357] It was also used by Wang *et al.*[309] when coarse-graining water models using IBI, as most of the structure was contained. The MAE increases as a shorter cutoff was applied, where a significantly greater increase is shown in the error below a 5.0 Å cutoff. This greater error at a 3.1 Å cutoff produced a more structured first peak in the RDF from undercorrecting. Since the tapering effects of the smooth switching cutoff function required some interval, the forces in the short-range region were affected. Beyond the first peak, the RDF remained largely unchanged from the initial one as the bias acted only in the short-range region.

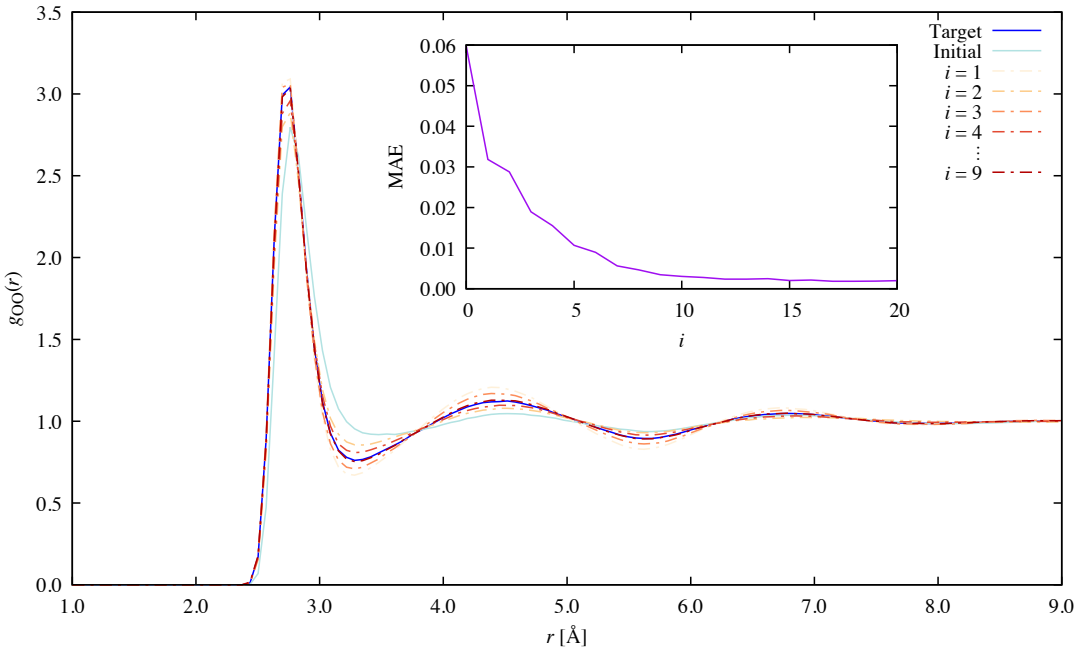


Figure 4.5: The RDFs of the directed simulation of SPC (initial) to SPC/Fw (target) across iterations. The biases were applied with $\alpha = 1$ and $r_{\text{cut}} = 9.0\,\text{Å}$. At each iteration, the bias overcorrects for the target from its previous iteration.

4.5 Exploring the Scaling Coefficient

Reith *et al.*[107] observed that the Boltzmann inverted potential has a tendency to overcorrect, which affects the stability of the convergence. Since the bias is a product of the RDF and hence describes the many-body interactions, applying it as a pairwise interaction introduces excess energy. We likewise observed that the resultant RDFs were overcorrected. The effect was usually more apparent in the medium-range interatomic region between 3.0 Å and 5.0 Å. When applied in the BIDS procedure, this can be observed as damped oscillations of the resultant RDFs about the target RDF as shown in Figure 4.5.

Performing numerous iterations of MD in QM is computationally expensive. The use of a scaling coefficient α in Eq. (3.2) enables the amplitude of the bias to attain an appropriate degree of correction. The number of iterations required to converge to the target RDF can then be lowered to be more favourable for QM calculations. Although proposed values of the coefficient based on the system can be found in the literature,[306,358–362] there is a lack of an *a priori* method to determine the optimum coefficient. The coefficient can be explored through MD scans to sample its effects on the resultant RDF. A number of scanning protocols were tried on correcting BIDS-SPC/Fw and BIDS-TIP3P to experiment, as shown in Figure 4.6. The initial RDFs were taken from 50 ps and 10 ns simulations. Then,

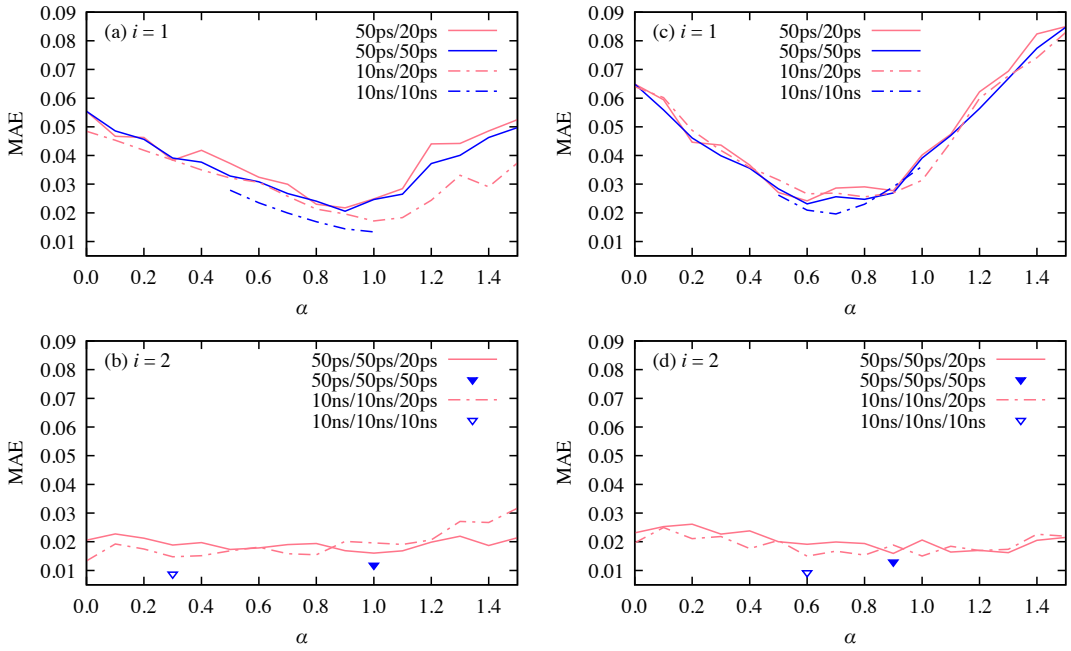
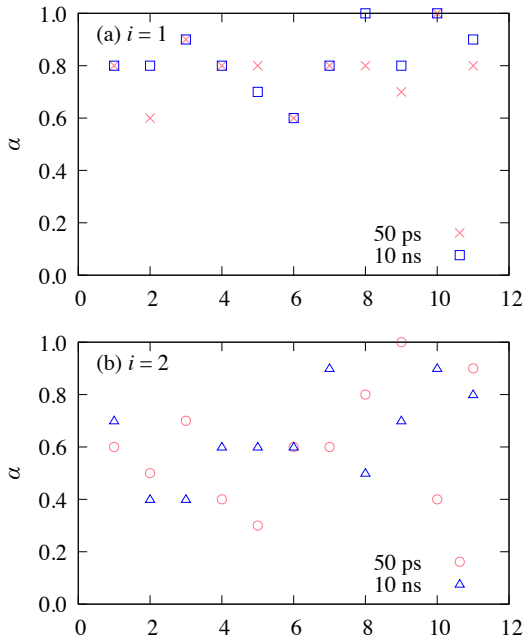


Figure 4.6: The MAE values for: (a) BIDS-SPC/Fw at iteration $i = 1$, (b) BIDS-SPC/Fw at $i = 2$, (c) BIDS-TIP3P at $i = 1$ and (d) BIDS-TIP3P at $i = 2$ across scaling coefficients α from 0.1 to 1.5 at intervals of 0.1 for the 50 ps (solid) and 10 ns (dot-dash/hollow) scans (blue) and their 20 ps scans (red).

the coefficient was scanned using 20 ps and 50 ps for the former, whereas 20 ps and 10 ns were used for the latter. At the optimal value for the longer 50 ps and 10 ns scans, the scanning protocol was repeated.

As Hadley & McCabe[306] pointed out, the coefficient correlates with the stability of the iteration scheme and was always less than 1.0 and never negative. A step interval of 0.1 was set to limit the number of scans when performed in QM calculations. We do note that smaller intervals can be afforded in molecular mechanics (MM) calculations. As a point of reference for changes to the RDF, the MAE at $\alpha = 0.0$ when no bias has been applied is included in Figure 4.6. The MAEs of the scans beyond a coefficient of 1.0 were also shown to test that the coefficient was never greater than 1.0 as per the observation of Hadley & McCabe.

In the first iteration, $i = 1$, of Figure 4.6, the MAEs have a convex behaviour with respect to the scaling coefficient. At low coefficients, the large MAE is due to an undercorrection whereas at high coefficients, the large MAE is due to an overcorrection. The optimum coefficient corresponds to the lowest MAE, which is similar between the 20 ps scan and its longer 50 ps or 10 ns counterparts. The optimum coefficients for BIDS-SPC/Fw are $\alpha = 1.0$ for the 10 ns/10 ns scan and $\alpha = 0.9$ for the 50 ps/50 ps scan, which were in agreement with their 10 ns/20 ps and 50 ps/20 ps scans respectively. Generally, the optimum coefficient from the



Index	Initial	Target
1	SPC	SPC/Fw
2	SPC/Fw	SPC
3	TIP3P	TIP3P/Fs
4	TIP3P	TIP4P/2005
5	TIP3P/Fs	TIP3P
6	TIP4P/2005	TIP3P
7	SPC	Experiment
8	SPC/Fw	Experiment
9	TIP3P	Experiment
10	TIP3P/Fs	Experiment
11	TIP4P/2005	Experiment

Figure 4.7: The optimum scaling coefficients α across two iterations for the water force fields in this study at the two time lengths investigated, 50 ps and 10 ns, where the target RDF was either from another force field or the experimental data. The vertical indices correspond with the various biased water force fields. An outer cutoff of 7.0 Å was applied to the biases here, which was observed to sometimes affect the optimal value by 0.1 when scanning.

20 ps scans was found to be within ± 0.1 of the value from its longer counterparts. This was also the case where the 20 ps scan does not aptly reflect the shape of its longer counterpart, as shown in Figure 4.6 (c) by the 10 ns simulation of BIDS-TIP3P, *i.e* the optimum coefficient was $\alpha = 0.8$ for the 20 ps scan whereas it was $\alpha = 0.7$ for its longer counterpart.

The optimum scaling coefficient was also found to typically be within the range of 0.5 to 1.0 at iteration $i = 1$, as seen in Figure 4.7 (a). The optimisation was hence bounded within this range for this step in the iteration, which reduced the number of scans required. Although the optimum coefficient can conveniently be at 1.0 as observed for BIDS-SPC/Fw, the bulk of the scatter gravitate towards 0.8 supporting the observation of Reith *et al.*[107] that the bias potential tends to overcorrect. The average optimum coefficient is found to be 0.80 ± 0.12 for the first iteration, where the standard deviation was approximately consistent with the ± 0.1 discrepancy between the 20 ps scan and its longer counterpart. However, the distribution of optimum coefficients is at 0.70 ± 0.28 for the second iteration in Figure 4.7 (b). We therefore recommend 0.8 as the scaling coefficient for the BIDS method, should a static value be used across the iterations, in order to maximise the more significant improvement provided in the first iteration.

In the second iteration, $i = 2$, of Figure 4.6, the behaviour of the simulations

are not explicitly explored as it was established that the 20 ps scans adequately reflect their longer counterparts. Furthermore, it is less clear where the optimum coefficient lies, as the previously convex shape of the MAE has flattened out due to minimal improvements to the RDFs compared to the previous iteration. In fact, the shape contains too much noise to discern a meaningful optimal value. The lack of a clear optimum coefficient is likely caused by the bias at $i = 2$ slightly displacing the regions of the RDF that were satisfactorily corrected by the previous bias. Using the coefficient corresponding with the lowest MAE value from the 20 ps scan, the simulations (50 ps/50 ps/50 ps and 10 ns/10 ns/10 ns points) demonstrated smaller improvements than at $i = 1$ but which were nonetheless appreciable.

Considering the trajectory of the MAE value across five iterations in Section 4.2, proceeding with the optimisation past this iteration will unlikely yield any dramatic improvements. In fact, the greatest error after two iterations was 0.0130 for the 50 ps scan of BIDS-TIP3P. This is approximately one standard deviation above the mean, *i.e.* 0.0128, of the previously reported converged MAE for the 50 ps simulations. We consider 0.0130 or below as the converged MAE for the purposes of the iterative scheme for the following reasons. It was a close approximate to the expected error for a 50 ps simulation, whose statistics reflect what is achievable in QM simulations. Thus, results from 10 ns simulations around that error would serve as a better analog to QM simulations. Moreover, the optimisation of the scaling coefficient has shown it is possible to attain that error in two iterations. This limited number of iterations will be beneficial for our QM simulations and avoids the diminishing returns observed in subsequent iterations.

4.6 Summary

The MAEs of the RDFs converged rapidly to a significantly lower error than their initial unbiased systems for both the 50 ps and the 10 ns BIDS simulations. This demonstrated the efficacy of the BIDS method in deriving a bias potential to improve a variety of water force fields to better simulate the liquid structure of water. The dynamics of the water were also found to improve when the distribution of the water molecules better matched experiment. This was indicated by an improvement in the self-diffusion coefficient at the first step of the iteration, which remained stable thereafter. In contrast, the dielectric constant was not directly susceptible to changes in the molecular distribution.

The bias potentials from the 50 ps simulations were found to be similar to the

those of the respective 10 ns simulations. Thus, the biases from short time length simulations with relatively poor statistics were valid and thereby suggest that the applicability of the BIDS method is applicable to QM simulations. The caveat is that smoothing of the RDFs becomes more important in short simulations to ensure the fidelity of the derived bias potential. A cutoff of the bias potential at 7.0 Å was also recommended as this will prevent the bias from interfering with the already well-defined dispersion correction in the long-range region, while preserving the quality of the bias in correcting the RDF. The generalised MDF function was applied to provide a smooth decay of forces up to the specified cutoff.

Since the bias potential obtained was prone to overcorrection, a scaling coefficient was applied to control the amplitude and stabilise the iterations. By performing 20 ps scans with a varying scaling coefficient, we found that the errors behave in a convex manner, at least for the first iteration. Consequently, there exists an optimum coefficient that returns the lowest MAE for the RDF. However, the convex behaviour became less apparent in the second iteration compared to the first iteration. In addition, subsequent iterations provided little improvement to the RDF according to the convergence trajectory of the MAE. We therefore optimised the coefficient across two iterations, where the expected error was found to be around 0.0130, for the application of the BIDS method to the following force field and DFT water simulations in this study. We note that the average optimum coefficient was 0.80 ± 0.12 at the first iteration for the empirical models, based on both 50 ps and 10 ns simulations, that were scanned. The standard deviation was approximately the observed ± 0.1 discrepancy between the 20 ps scan and its longer counterpart. Subsequently, the average optimum coefficient was 0.70 ± 0.28 at the second iteration. When taking a static value across all iterations, we recommend 0.8 for the scaling coefficient in order to maximise the more significant improvement at the first iteration.

References

[5] K. R. Hadley and C. McCabe. "On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties". In: *The Journal of Physical Chemistry B* 114.13 (2010). PMID: 20230012, pp. 4590–4599. DOI: 10.1021/jp911894a.

[107] D. Reith, M. Pütz, and F. Müller-Plathe. "Deriving Effective Mesoscale Potentials from Atomistic Simulations". In: *Journal of Computational Chemistry* 24.13 (2003), pp. 1624–1636. DOI: 10.1002/jcc.10307.

[164] S. Grimme. "Density Functional Theory with London Dispersion Corrections". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.2 (2011), pp. 211–228. DOI: 10.1002/wcms.30.

[266] Y. J. Wu, H. L. Tepper, and G. A. Voth. "Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties". In: *The Journal of Chemical Physics* 124.2, 024503 (2006). DOI: 10.1063/1.2136877.

[305] A. K. Soper. "Empirical Potential Monte Carlo Simulation of Fluid Structure". In: *Chemical Physics* 202.2 (1996), pp. 295 –306. DOI: 10.1016/0301-0104(95)00357-6.

[306] K. R. Hadley and C. McCabe. "A Coarse-Grained Model for Amorphous and Crystalline Fatty Acids". In: *The Journal of Chemical Physics* 132.13, 134505 (2010). DOI: 10.1063/1.3360146.

[309] H. Wang, C. Junghans, and K. Kremer. "Comparative Atomistic and Coarse-Grained Study of Water: What do we Lose by Coarse-Graining?" In: *The European Physical Journal E* 28.2 (2009), pp. 221–229. DOI: 10.1140/epje/i2008-10413-5.

[324] A. D. White et al. "Communication: Improved *ab initio* Molecular Dynamics by Minimally Biasing with Experimental Data". In: *The Journal of Chemical Physics* 146.4 (2017), p. 041102. DOI: 10.1063/1.4974837.

[325] P. J. Steinbach and B. R. Brooks. "New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation". In: *Journal of Computational Chemistry* 15.7 (1994), pp. 667–683. DOI: 10.1002/jcc.540150702.

[343] T. Nagata, D. G. Fedorov, and K. Kitaura. "Mathematical Formulation of the Fragment Molecular Orbital Method". In: *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*. Ed. by R. Zalesny et al. Dordrecht: Springer Netherlands, 2011, pp. 17–64. DOI: 10.1007/978-90-481-2853-2_2.

[344] D. A. Dixon et al. "Computational Chemistry in the Environmental Molecular Sciences Laboratory". In: *High-Performance Computing*. Ed. by R. J. Allan et al. Boston, MA: Springer US, 1999, pp. 215–228. DOI: 10.1007/978-1-4615-4873-7_24.

[345] D. Y. Zubarev, B. M. Austin, and W. A. Lester. "Practical Aspects of Quantum Monte Carlo for the Electronic Structure of Molecules". In: *Practical Aspects of Computational Chemistry I: An Overview of the Last Two Decades and Current Trends*. Ed. by J. Leszczynski and M. K. Shukla. Dordrecht: Springer Netherlands, 2012, pp. 255–292. DOI: 10.1007/978-94-007-0919-5_9.

[346] S. Grimme. "Seemingly Simple Stereoelectronic Effects in Alkane Isomers and the Implications for Kohn?Sham Density Functional Theory". In: *Angewandte Chemie International Edition* 45.27 (), pp. 4460–4464. DOI: 10.1002/anie.200600448.

[347] S. Plimpton. "Fast Parallel Algorithms for Short-Range Molecular Dynamics". In: *Journal of Computational Physics* 117.1 (1995), pp. 1 –19. DOI: 10.1006/jcph.1995.1039.

[348] D. J. Evans and B. L. Holian. "The Nose-Hoover Thermostat". In: *The Journal of Chemical Physics* 83.8 (1985), pp. 4069–4074. DOI: 10.1063/1.449071.

[349] R. Burden and J. Faires. *Numerical Analysis*. Prindle, Weber & Schmidt, 1985.

[350] G. Milano, S. Goudeau, and F. Müller-Plathe. "Multicentered Gaussian-Based Potentials for Coarse-Grained Polymer Simulations: Linking Atomistic and Mesoscopic Scales". In: *Journal of Polymer Science Part B: Polymer Physics* 43.8 (2005), pp. 871–885. DOI: 10.1002/polb.20380.

[351] C. Peter, L. Delle Site, and K. Kremer. "Classical Simulations from the Atomistic to the Mesoscale and Back: Coarse Graining an Azobenzene Liquid Crystal". In: *Soft Matter* 4 (4 2008), pp. 859–869. DOI: 10.1039/B717324E.

[352] M. Allen and D. Tildesley. *Computer Simulation of Liquids*. Oxford science publications. Clarendon Press, 1987.

[353] J. Kenney and E. Keeping. *Mathematics of Statistics*. 3rd. Mathematics of Statistics. Van Nostrand company, 1962.

[354] D. Reith, H. Meyer, and F. Müller-Plathe. "Mapping Atomistic to Coarse-Grained Polymer Models using Automatic Simplex Optimization to Fit Structural Properties". In: *Macromolecules* 34.7 (2001), pp. 2335–2345. DOI: 10.1021/ma001499k.

[355] O. Gereben and L. Pusztai. "On the Accurate Calculation of the Dielectric Constant from Molecular Dynamics Simulations: The Case of SPC/E and SWM4-DP Water". In: *Chemical Physics Letters* 507.1 (2011), pp. 80 –83. DOI: 10.1016/j.cplett.2011.02.064.

[356] J. E. Basconi and M. R. Shirts. "Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations". In: *Journal of Chemical Theory and Computation* 9.7 (2013). PMID: 26583973, pp. 2887–2899. DOI: 10.1021/ct400109a.

[357] C. D. Berweger, W. F. van Gunsteren, and F. Müller-Plathe. "Force Field Parametrization by Weak Coupling. Re-engineering SPC Water". In: *Chemical Physics Letters* 232.5 (1995), pp. 429 –436. DOI: 10.1016/0009-2614(94)01391-8.

[358] H. S. Ashbaugh et al. "Mesoscale Model of Polymer Melt Structure: Self-Consistent Mapping of Molecular Correlations to Coarse-Grained Potentials". In: *The Journal of Chemical Physics* 122.10 (2005), p. 104908. DOI: 10.1063/1.1861455.

[359] D. Bedrov, C. Ayyagari, and G. D. Smith. "Multiscale Modeling of Poly(ethylene oxide)-Poly(propylene oxide)-Poly(ethylene oxide) Triblock Copolymer Micelles in Aqueous Solution". In: *Journal of Chemical Theory and Computation* 2.3 (2006). PMID: 26626667, pp. 598–606. DOI: 10.1021/ct050334k.

[360] J. Elezgaray and M. Laguerre. "A Systematic Method to Derive Force Fields for Coarse-Grained Simulations of Phospholipids". In: *Computer Physics Communications* 175.4 (2006), pp. 264 –268. DOI: 10.1016/j.cpc.2006.01.009.

[361] A. Rakshit and R. C. Picu. "Coarse Grained Model of Entangled Polymer Melts". In: *The Journal of Chemical Physics* 125.16 (2006), p. 164907. DOI: 10.1063/1.2362820.

[362] Q. Sun and R. Faller. "Systematic Coarse-Graining of Atomistic Models for Simulation of Polymeric Systems". In: *Computers & Chemical Engineering* 29.11 (2005). Selected Papers Presented at the Symposium on Modeling of Complex Processes, pp. 2380 – 2385. DOI: 10.1016/j.compchemeng.2005.05.026.

Chapter 5

Transforming One Existing Water Force Field to Another

It would be instructive to investigate the application of Boltzmann inversion directed simulation (BIDS) to one known force field using the radial distribution function (RDF) of another known force field as the target, *e.g.* SPC and TIPnP. This should provide more evidence that transport and thermodynamic properties, such as self-diffusion coefficient and dielectric constant respectively, as well as other correlation functions, such as oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H) RDFs, can be modified to trend towards another potential model through the BIDS methodology. At the same time, some valuable insights into the response of each water model to the Boltzmann inverted bias that is comparable with reparameterisation can be obtained.

The ability of the BIDS method in correcting to the experimental scattering data of water has been established during this thesis. Based on the insights obtained from its development detailed in Chapter 4, we tapered the bias to zero at an outer cutoff of 7.0 Å and optimised the scaling coefficient across two iterations. The former prevents the occurrence of a force discontinuity without being detrimental to the quality of the correction. The latter allows the oxygen-oxygen (O-O) RDF to be sufficiently corrected with the smallest number of iterations.

The simulation details are kept the same as the previous chapter for the water force fields, SPC, SPC/Fw, TIP3P, TIP3P/Fs, and TIP4P/2005, using the canonical ensemble (NVT) with the exception of being limited to 10 ns simulations. The water models in this study can be classified based on: (i) whether they are rigid or flexible, and (ii) the number of sites, disregarding those associated with polarisation effects. The investigation was consequently conducted between the rigid and flexible variants (SPC and SPC/Fw, TIP3P and TIP3P/Fs) and

between the three-site and four-site variants (TIP3P and TIP4P/2005). Since the target is not experimental water, the notation of the biased water simulations, *e.g.* BIDS-SPC, here is modified with the target water model in parentheses as a suffix, *e.g.* BIDS-SPC(SPC/Fw).

5.1 Transforming between Rigid and Flexible Variants

The rigid SPC model is the progenitor of the simple point-charge model, which the flexible SPC/Fw is a variant of. Hence, they have many common parameters including the O-O pair interaction described by the Lennard-Jones (LJ) parameters. The flexibility of SPC/Fw leads to an increase in the molecular dipole moment and hence, stronger attraction to surrounding molecules.[283] Aside from classifying the two as either rigid or flexible, they differ in the parameters, bond length and bond angle, as seen in Table 2.2. The work of Wu *et al.*[266] which went into constructing SPC/Fw, demonstrated that the effect of the bond angle on the RDF is negligible in comparison to that of the bond length. The elongation of the bond length strengthens the hydrogen-bond lifetime, which along with an increased dipole moment contributes to a more structured O-O RDF compared to SPC. The resulting biases from applying BIDS between the two force fields and the consequent RDFs are shown in Figure 5.1 and Figure 5.2.

In a similar vein, TIP3P/Fs is the flexible variant of the rigid TIP3P. According to Table 2.2, they have the same partial charges but vary slightly in the other parameters. TIP3P reproduces approximately the height and location of the first peak in the O-O RDF but is essentially flat beyond this peak. Unfortunately, introducing flexibility in the form of TIP3P/Fs only marginally improves the distinction in the neighbouring peak. This is in contrast to the inclusion of the intramolecular part to the potential surface model by Reimers and Watts,[363] where the second peak arose from the intramolecular relaxation processes. Consider the transferable intermolecular potentials (TIPS) model, which TIP3P was based on, with a similarly flat liquid structure beyond the first peak. Dang & Pettitt[364] reasoned that TIPS is structurally insensitive to the intramolecular motions. However, its second and third solvent shells can be obtained from applying a narrower and deeper hydrogen-bonding well, but this is offset by the first peak becoming too high and the energy becoming too low according to Jorgensen.[257] The resulting biases from applying BIDS between the two force fields and the consequent RDFs are shown in Figure 5.3 and Figure 5.4.

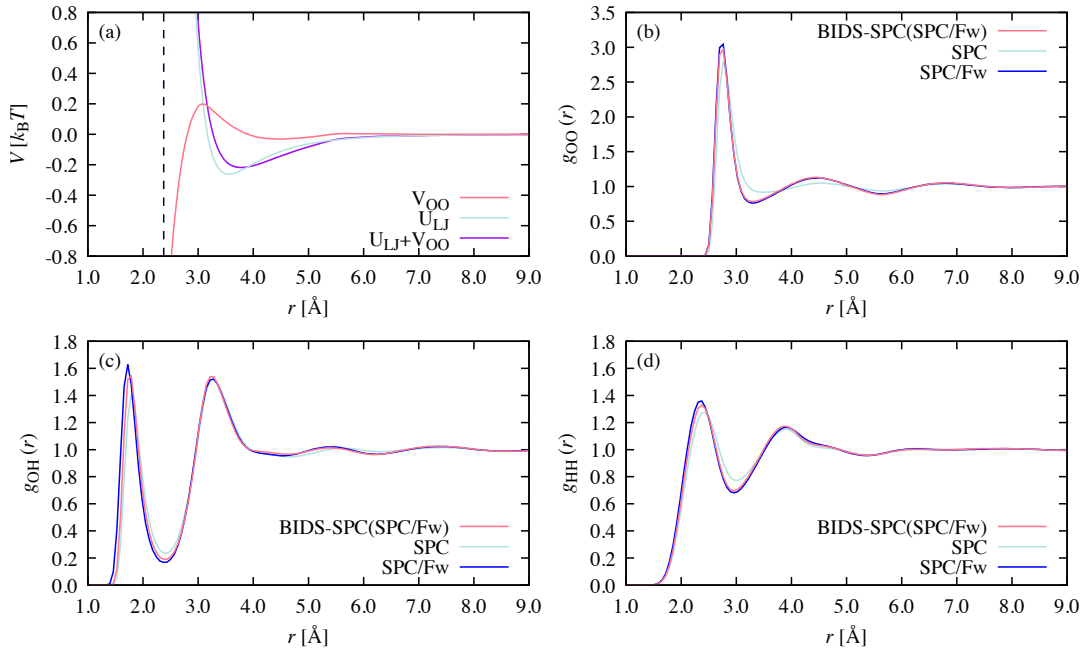


Figure 5.1: The resulting bias and fit to the SPC/Fw target data from the BIDS-SPC(SPC/Fw) simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The fit to the target (b) O-O, (c) O-H and (d) H-H RDFs from a constant NVT simulation of BIDS-SPC(SPC/Fw) using the bias.

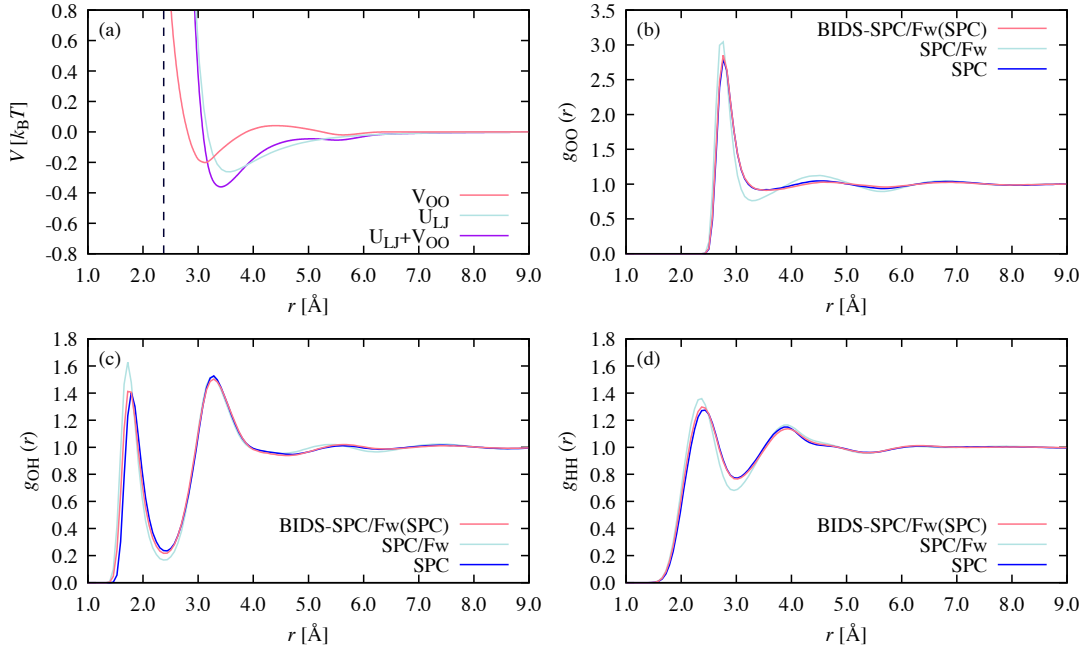


Figure 5.2: The resulting bias and fit to the SPC target data from the BIDS-SPC/Fw(SPC) simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The fit to the target (b) O-O, (c) O-H and (d) H-H RDFs from a constant NVT simulation of BIDS-SPC/Fw(SPC) using the bias.

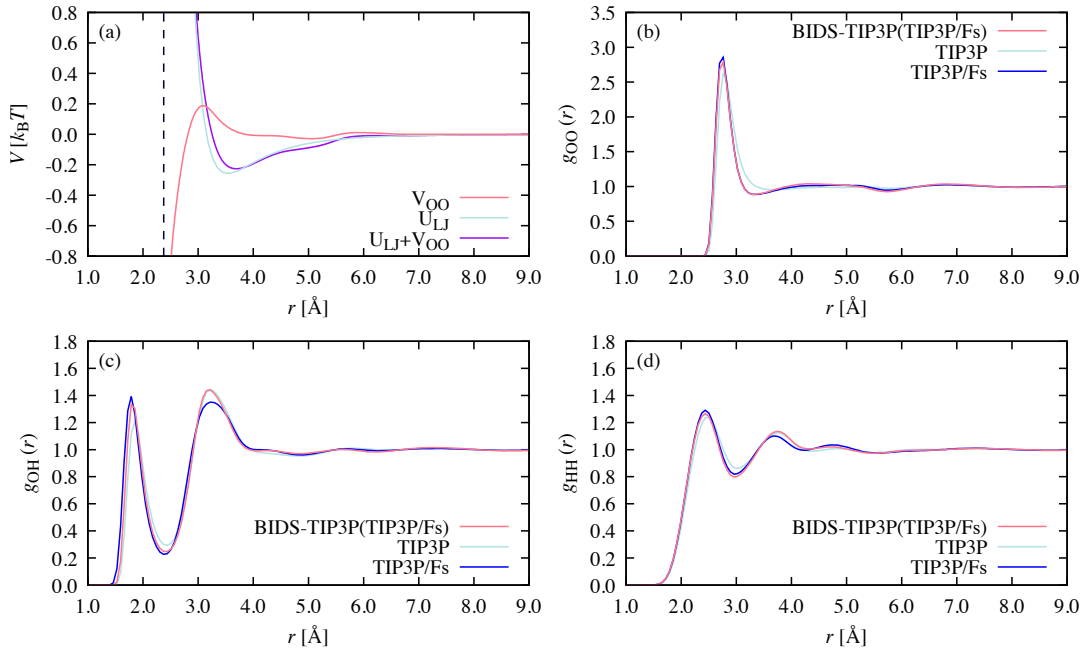


Figure 5.3: The resulting bias and fit to the TIP3P/Fs target data from the BIDS-TIP3P(TIP3P/Fs) simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The fit to the target (b) O-O, (c) O-H and (d) H-H RDFs from a constant NVT simulation of BIDS-TIP3P(TIP3P/Fs) using the bias.

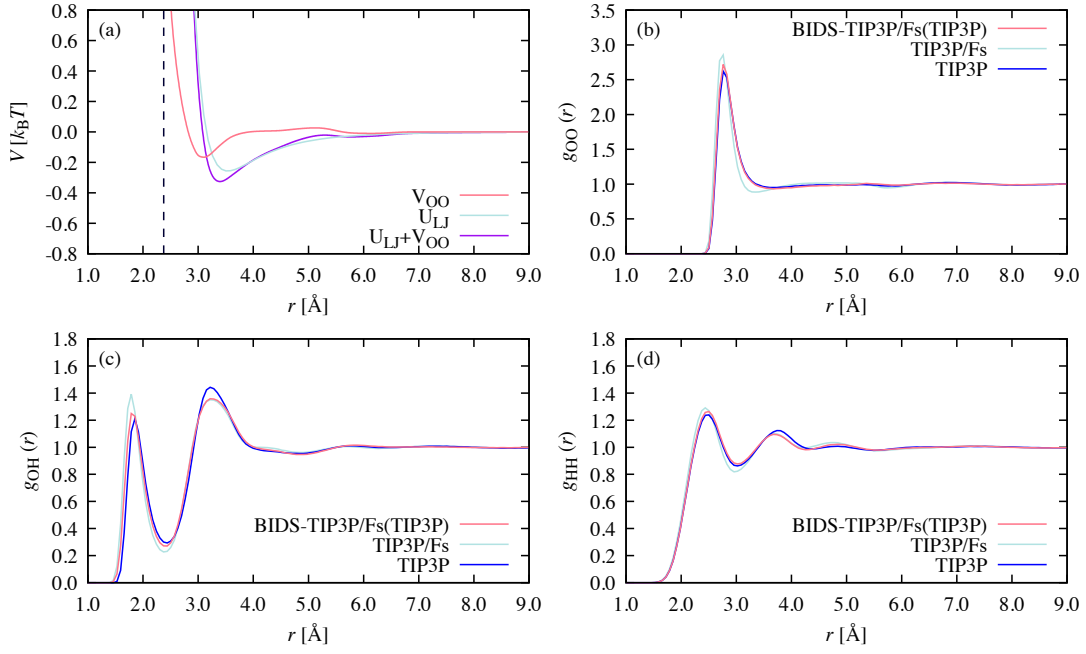


Figure 5.4: The resulting bias and fit to the TIP3P target data from the BIDS-TIP3P/Fs(TIP3P) simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The fit to the target (b) O-O, (c) O-H and (d) H-H RDFs from a constant NVT simulation of BIDS-TIP3P/Fs(TIP3P) using the bias.

Given the differences in the description of the force fields, *i.e.* bond length and flexibility, the biases derived from the BIDS method were able to satisfactorily reproduce the target O-O RDFs in Figure 5.1 to Figure 5.4. The mean absolute error (MAE) values for the O-O RDFs in Table 5.1 were reduced to around 0.0120, which is within the margin of the expected error in two iterations. As observed, these errors correspond to O-O RDFs that satisfactorily reproduced the target data. According to a sensitivity analysis study using SPC and TIP3P flexible models,[365] the LJ interactions are significant at short distances and the parameters have an appreciable influence on the first and second peaks of the site-site RDFs. This holds, in part, for the rigid and the flexible variants for both SPC and TIP3P where the net potentials overall remained LJ-like, as seen in Figure 5.1 (a) to Figure 5.4 (a). These net potentials were perturbed slightly, shifting in the location and depth of the well. Consequently, the repulsive region becomes less repulsive for an enhanced first peak and vice versa. The resultant O-H and H-H RDFs closely matched their targets in the first peaks up to the first valleys. The good fit in their second peaks for the BIDS-SPC variants were, however, due to an already close match, whereas there is a lack of change seen in the BIDS-TIP3P variants. In the same sensitivity analysis study,[365] the increase of either the depth of the well, ϵ, or the LJ radius, σ, had a tendency to diminish the liquid water structure due to an increase in the repulsive forces. Here, the net potential was deeper although slightly broader, corresponding to an enhanced repulsive region, for a less structured O-O RDF and vice versa.

The biases themselves, from Figure 5.1 (a) to Figure 5.4 (a), can arguably be described as typical pairwise potentials although with slight oscillations, where the left of the "well" was attractive to obtain an enhanced first peak or repulsive to obtain a diminished first peak in the O-O RDF. The biases observed are comparable in magnitude and shape such that the parameters to describe them are similar, except the sense is dependent on the direction of the correction. It is convenient to quantify the parameters using the average and the standard deviation. The magnitude of the "well" for the biases were around $0.188\,k_{\mathrm{B}}T$, which is in the same order of magnitude as the well of the LJ potentials at around $0.260\,k_{\mathrm{B}}T$. However, the "well" of the biases were located about 3.095 Å, which is to the left of the LJ wells and thus did not drastically alter the resulting net potential. The shortest distances at which the bias potential between oxygen atoms are zero were located about 2.813 Å. Considering the similarity of the parameters, we can conclude that the biases here are virtually the same and describe the inclusion or the exclusion of the effects of intramolecular flexibility, in terms of the pairwise interaction between oxygen atoms. In fact, they can be

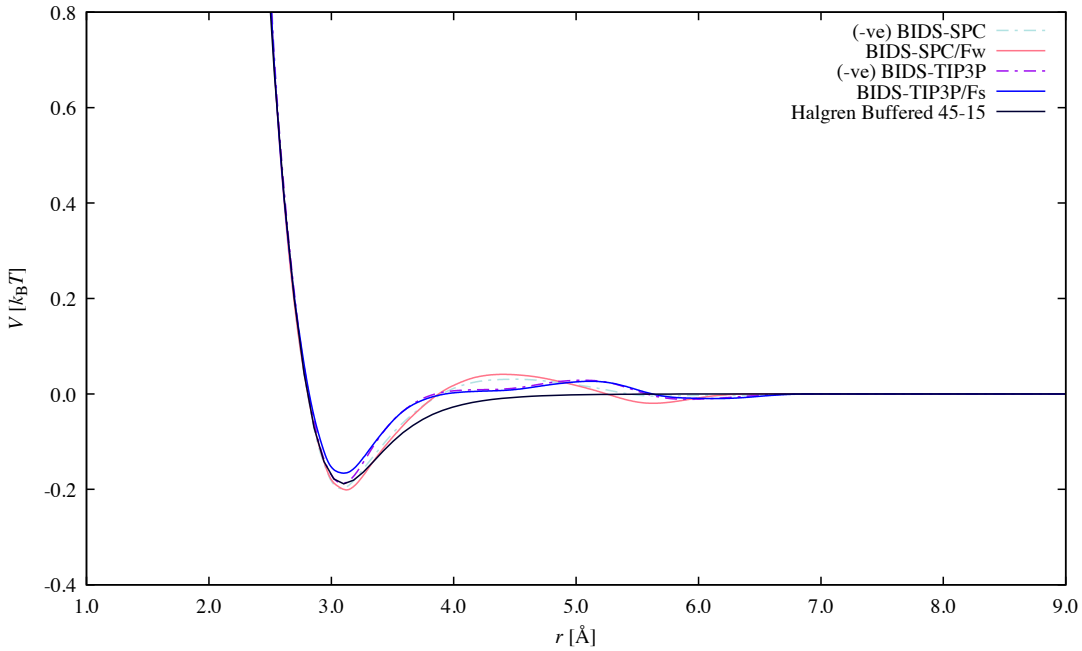


Figure 5.5: Approximating the biases from the BIDS application between rigid and flexible variants with a fitted Halgren buffered 45-15 potential. The sense of BIDS-rigid models were flipped for comparison.

Table 5.1: The O-O RDF MAE, self-diffusion coefficient, D_s [$\times 10^{-5}\mathrm{cm}^2\mathrm{s}^{-1}$], and static dielectric constant, ϵ_r, values for the biased rigid and flexible variants. The zero superscript and the parentheses denote the initial and target values, respectively.

Model	$\mathrm{MAE}^0_{\mathrm{OO}}$	$\mathrm{MAE}_{\mathrm{OO}}$	D^0_s	D_s	ϵ^0_r	ϵ_r
BIDS-SPC(SPC/Fw)	0.0598	0.0105	3.789	2.725 (2.352)	68.680	65.878 (77.915)
BIDS-SPC/Fw(SPC)	0.0603	0.0123	2.352	3.396 (3.789)	77.915	80.412 (68.680)
BIDS-TIP3P(TIP3P/Fs)	0.0476	0.0126	5.410	4.460 (3.711)	99.583	97.651 (173.588)
BIDS-TIP3P/Fs(TIP3P)	0.0476	0.0124	3.711	4.545 (5.410)	173.588	186.362 (99.583)

approximated at the short range up to the "well" using a Halgren buffered 45-15 potential,[366] as seen in Figure 5.5, which can be written as:

$$V_{\mathrm{buff}}(r_{ij}) = \epsilon_{ij}\left(\frac{1+\delta}{\rho_{ij}+\delta}\right)^{n-m}\left(\frac{1+\gamma}{\rho_{ij}^m+\gamma}-2\right) \tag{5.1}$$

where fitted parameters were $n = 45$, $m = 15$, $\delta = 2.145$, and $\gamma = 0.530$. The pairwise separation, r_{ij}, was described via the ratio $\rho_{ij} = r_{ij}/r^0_{ij}$. The values of the well depth, ϵ_{ij}, and the minimum energy distance, r^0_{ij}, were from the respective average values of the biases.

The diffusion coefficient and the dielectric constant from the biased calculations are recorded in Table 5.1. Between the respective variants, the diffusion coefficients D_s changed by similar magnitudes. The values from the BIDS-SPC(SPC/Fw) and BIDS-SPC/Fw(SPC) simulations changed by

$-1.063 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ and $1.044 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$, respectively. The values from the BIDS-TIP3P(TIP3P/Fs) and BIDS-TIP3P/Fs(TIP3P) simulations changed by $-0.950 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ and $0.833 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$, respectively. As demonstrated through the hydrogen-bond lifetime by Wu *et al.*[266] the diffusion coefficient is strongly and inversely correlated with the hydrogen-bond strength. Reducing the repulsive forces in the net potential strengthened the hydrogen bonds and consequently decreased the diffusivity leading to an enhanced liquid structure, whereas increasing these forces had the opposite effect. The dielectric constants, ϵ_r, experienced slight detrimental changes in their values. Except for BIDS-TIP3P/Fs(TIP3P), the changes, *i.e* -2.802, 2.497 and -1.933 in the same sequence as Table 5.1, were of the same order of magnitude as the standard deviation at 298.15 K and 1 atm reported by Wu *et al.*[266] *i.e.* 1.35, 1.62 and 2.20, respectively. The greater change, *i.e.* 12.773, in BIDS-TIP3P/Fs(TIP3P) were due to the greater relaxation of the bond angles, which was softer than those of SPC/Fw.

5.2 Transforming between Three-Site and Four-Site Variants

The TIP4P/2005 model is an alternative parameterisation of the four-site variant of the TIP3P three-site model. In the four-site model, the negative charge is moved off the oxygen and toward the hydrogens on the bisector of the H-O-H angle and was first proposed by Bernal & Fowler.[41] The van der Waals (vdW) potential still acts between the oxygen atoms. Both TIP3P and TIP4P/2005 are rigid and have the same geometry, whereas they differ greatly in the depth of their potential wells and partial charges. This can be seen in Table 2.2. The TIP3P prediction gives too little structure beyond the first peak in the O-O RDFs, which appears to be a general problem for the three-site models, particularly the rigid ones. Four-site models experience a similar effect, but it is less severe, as their O-O RDFs are more structured to begin with. Thus, the TIP4P/2005 reproduces the second and third peaks reasonably well at the correct density.

The resulting biases from applying BIDS between the two force fields and the consequent RDFs are shown in Figure 5.6 and Figure 5.7. Despite the differences in the construction of the force field models, *i.e.* number of sites, the biases derived from the BIDS method were able to satisfactorily reproduce the target O-O RDFs. According to the MAE in Table 5.2, the O-O RDFs closely matched to the target data; well within the expected 0.0130 error. The height and the depth

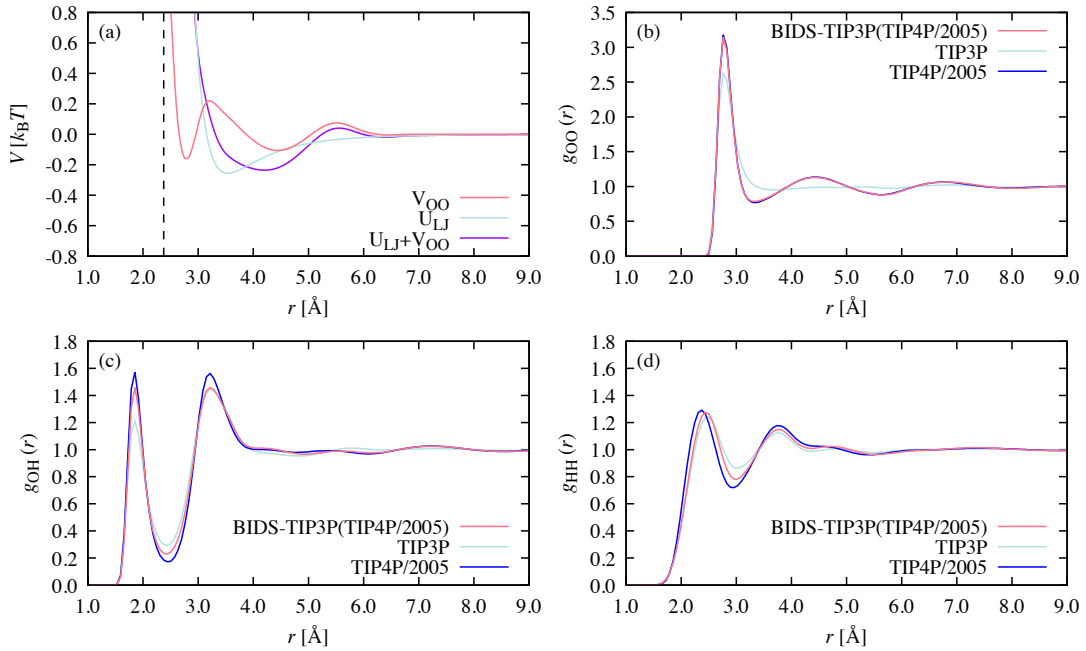


Figure 5.6: The resulting bias and fit to the TIP4P/2005 target data from the BIDS-TIP3P(TIP4P/2005) simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The fit to the target (b) O-O, (c) O-H and (d) H-H RDFs from a constant NVT simulation of BIDS-TIP3P(TIP4P/2005) using the bias.

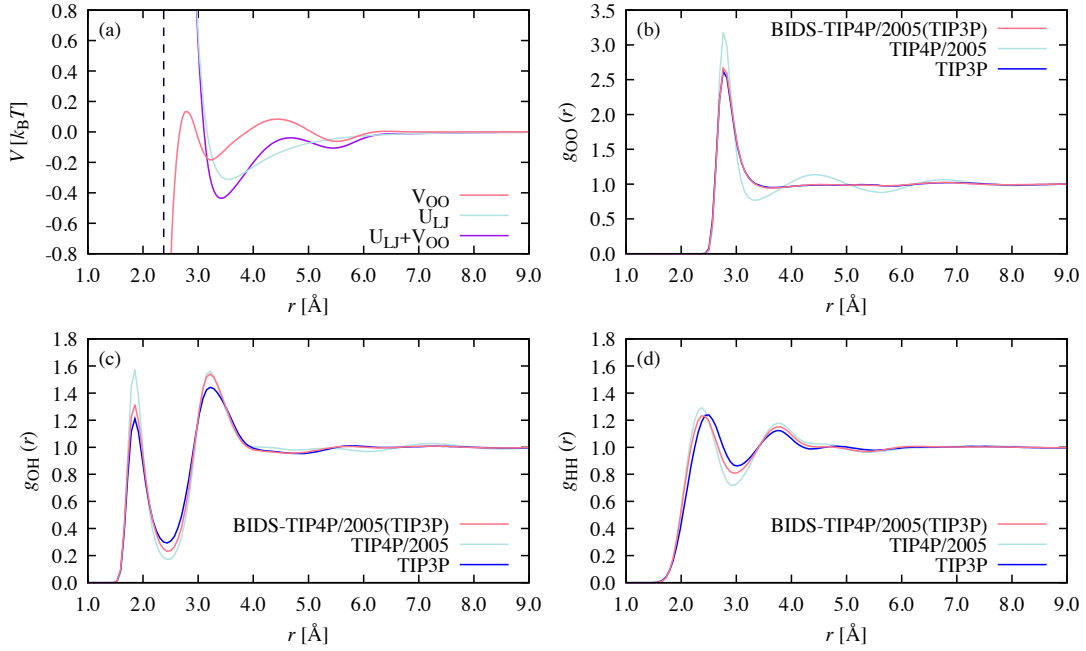


Figure 5.7: The resulting bias and fit to the TIP3P target data from the BIDS-TIP4P/2005(TIP3P) simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The fit to the target (b) O-O, (c) O-H and (d) H-H RDFs from a constant NVT simulation of BIDS-TIP4P/2005(TIP3P) using the bias.

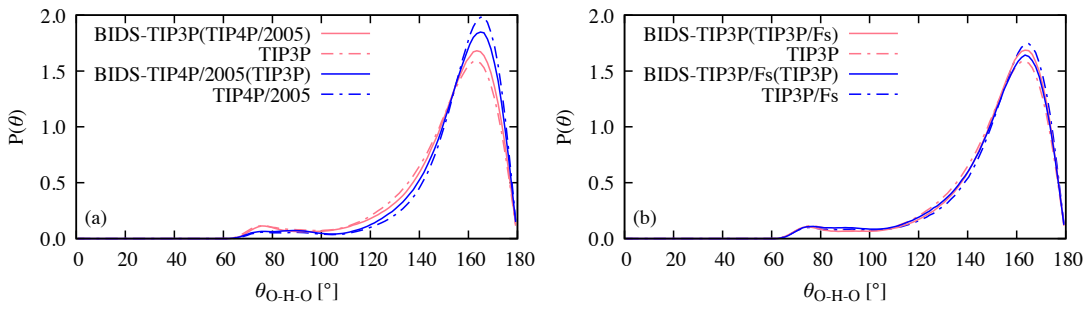


Figure 5.8: The O-H-O angular distributions up to the first solvation shell of liquid water from the BIDS application: (a) between the three-site and four-site TIPnP variants, and (b) between the three-site rigid and flexible TIPnP variants.

of the first peak and first valley, respectively, of the O-H and H-H RDFs were also better emulated. However, these features are clearly not shifted to be in the correct position for the H-H RDFs. This slightly poorer fit in the RDFs, which were not part of the BIDS scheme, relative to that in the investigation between rigid and flexible variants was likely due to the three-site and four-site variants being more distinct water models. The full structure of the liquid is characterised by the orientational pair correlation function between neighbouring molecules, such that assuming that the relative molecular orientations are completely isotropic is inadequate to describe the O-H and H-H partial correlation functions.[367] Given that the geometry, *i.e* O-H bond and H-O-H angle, of the water molecules in the two rigid models were the same, it is likely that the distinct charge distributions resulted in different molecular orientations in the first solvation shell preventing a closer match. As seen in Figure 5.8 (a), the peaks in the O-H-O angular distributions of the three-site TIPnP models are appreciably lower than, and shifted to the left of, the four-site TIPnP models. Here, the biased models are improved towards but do not sufficiently approximate the target models. This is in contrast to the transformation between rigid and flexible TIPnP variants, as seen in Figure 5.8 (b), where the biased models better approximate the O-H-O angular distribution of the target models.

The resulting net potential for BIDS-TIP4P/2005(TIP3P) was consistent with our previous observation for a diminished liquid structure. More interestingly, the BIDS-TIP3P(TIP4P/2005) net potential offered insight into obtaining the second and third solvation shells without the first peak becoming too high. Compared to Figure 5.3 (a), the net potential for BIDS-TIP3P(TIP4P/2005) here has a much broader and flatter well with the LJ radius shifted more significantly to the right. Overall, the repulsive region in the short range was diminished, followed by no appreciable repulsive force in the drawn-out region immediately after. The former

Table 5.2: The O-O RDF MAE, self-diffusion coefficient, D_s [$\times 10^{-5}\mathrm{cm}^2\mathrm{s}^{-1}$], and static dielectric constant, ϵ_r, values for the biased three-site and four-site variants. The zero superscript and the parentheses denote the initial and target values, respectively.

Model	$\mathrm{MAE}^0_{\mathrm{OO}}$	$\mathrm{MAE}_{\mathrm{OO}}$	D^0_s	D_s	ϵ^0_r	ϵ_r
BIDS-TIP3P(TIP4P/2005)	0.0734	0.0066	5.410	3.569 (2.050)	99.583	92.095 (107.990)
BIDS-TIP4P/2005(TIP3P)	0.0730	0.0067	2.050	3.174 (5.410)	107.990	110.686 (99.583)

leads to a more structured first peak in the O-O RDF. The latter occurs in the region between the first valley and second peak contributing to the deepening of the first valley. Remarkably, the attractive region was shifted to the location of the second peak likely enhancing this peak. Such a shape was possible by applying a numerical bias rather than adjusting the parameters for the LJ potential. Perhaps a more comprehensive analytical potential is required in general to sufficiently parameterise the pairwise potential.

The biases, shown in Figure 5.6 (a) and Figure 5.7 (a), were similar but with an opposing sense depending on the direction of the correction. They also have oscillations beyond the "well". As opposed to the biases from the transformation between the rigid and flexible variants, these were less like typical pairwise potentials. Here, we will only look at parameters describing the first "well" and not try to ascribe parameters to the oscillations beyond that. The "wells" of the biases were $-0.159\,k_\mathrm{B}T$ at 2.795 Å and $0.134\,k_\mathrm{B}T$ at 2.785 Å which were in the same order of magnitude as the well of the LJ potentials, $0.257\,k_\mathrm{B}T$ and $0.313\,k_\mathrm{B}T$, of TIP3P and TIP4P/2005 respectively. Undoubtedly, the greater difference of the peaks and valleys between the initial and target RDFs resulted in deeper "wells" and oscillations in the bias potential. Furthermore, the low-probability region to the left of the "wells" does not seem to play a strong role in improving the first peak in the O-O RDFs. Conversely, the preceding small region corresponding with the location of the first peak was consistent, indicating an attraction for an enhanced first peak and a repulsion for a diminished first peak.

The diffusion coefficient and the dielectric constant from the biased calculations are recorded in Table 5.2. The values of the diffusion coefficient, D_s, for BIDS-TIP3P(TIP4P/2005) and BIDS-TIP4P2005(TIP3P) changed by $-1.841 \times 10^{-5}\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ and $1.124 \times 10^{-5}\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, respectively, which were of the same order of magnitude. The change in the diffusion coefficient for BIDS-TIP3P(TIP4P/2005) when correcting to TIP4P/2005 was almost twice that when correcting to TIP3P/Fs. Recall that an enhanced liquid structure was seen in tandem with lowered diffusivity through the strengthening of the hydrogen-bond network. Considering that the O-O RDF of TIP4P/2005 was more structured, with more prominent peaks and valleys compared to TIP3P/Fs, the change was

not unexpected. The dielectric constants, ϵ_r, were slightly worse, changing by -7.489 for BIDS-TIP3P(TIP4P/2005) and 2.696 for BIDS-TIP4P/2005(TIP3P). Both were of the same order of magnitude as the standard deviation at 298.15 K and 1 atm reported by Wu *et al.*[266] Based on the hysteresis of the values between the variants, the thermodynamic properties of TIP4P/2005 were less responsive to changes in the O-O interaction potential. It may be worthwhile to reparameterise the TIP4P/2005 model to reproduce most of the thermodynamic properties of water aside from the liquid structure and then apply the bias.

5.3 Summary

The BIDS method can be extended beyond improving a force field to better match the experimental properties of water. This was demonstrated by biasing one water force field to another. The ability to change the O-O RDF from one water model to another was also found to change the other RDFs that were not directly biased. In particular, transforming between similar water force fields, *i.e.* rigid and flexible variants, also transformed the O-H and H-H RDF up to the first valley. This is in contrast with the transformation between three-site and four-site variants, which were significantly different due to the position of the greater negative charge on an additional fictitious site in the four-site model. This elucidated the role of molecular orientation in enhancing the quality of fit in other RDFs.

The resulting biases were found to have a stationary point that could be fitted by an appropriate pairwise potential. More importantly, the bias provided insight into the shortcomings of the form of the LJ potentials through the net potential. This was highlighted by the flat and broad well of the net potential when TIP3P, which had difficulty predicting the liquid structure of liquid water beyond the first peak, was used to reproduce the second and third O-O RDF peaks of TIP4P/2005. As expected, the low-probability region of the bias had little effect on the first peak of the O-O RDF. This was most evident in the biases transforming between the three-site and four-site variants.

Thus far, the changes to the bulk properties were consistent with the established observations. The corresponding changes in the self-diffusion coefficients from an enhanced or diminished liquid structure of water were achieved. In contrast, the dielectric constants became slightly worse, where the changes were of the same order of magnitude as the standard deviations reported by Wu *et al.*[266] Moreover, slight detrimental changes were previously observed at some of the iterations in the convergence and stability investigation of the dielectric constant

as shown in Figure 4.1 (f). Considering the minor oscillations in the otherwise stable iterations, we can attribute these changes to statistical errors. The exception to this is TIP3P/Fs, which has soft flexible bond angles that allowed the dielectric constant to be more susceptible to the change in the distribution of the water molecules.

References

[41] J. D. Bernal and R. H. Fowler. "A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions". In: *The Journal of Chemical Physics* 1.8 (1933), pp. 515–548. DOI: 10.1063/1.1749327.

[257] W. L. Jorgensen. "Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water". In: *Journal of the American Chemical Society* 103.2 (1981), pp. 335–340. DOI: 10.1021/ja00392a016.

[266] Y. J. Wu, H. L. Tepper, and G. A. Voth. "Flexible Simple Point-Charge Water Model with Improved Liquid-State Properties". In: *The Journal of Chemical Physics* 124.2, 024503 (2006). DOI: 10.1063/1.2136877.

[283] A. Wallqvist and O. Teleman. "Properties of Flexible Water Models". In: *Molecular Physics* 74.3 (1991), pp. 515–533. DOI: 10.1080/00268979100102391.

[363] J. Reimers and R. Watts. "The Structure, Thermodynamic Properties and Infrared Spectra of Liquid Water and Ice". In: *Chemical Physics* 91.2 (1984), pp. 201 –223. DOI: 10.1016/0301-0104(84)80055-5.

[364] L. X. Dang and B. M. Pettitt. "Simple Intramolecular Model Potentials for Water". In: *The Journal of Physical Chemistry* 91.12 (1987), pp. 3349–3354. DOI: 10.1021/j100296a048.

[365] S. Zhu and C. F. Wong. "Sensitivity Analysis of Distribution Functions of Liquid Water". In: *The Journal of Chemical Physics* 99.11 (1993), pp. 9047–9053. DOI: 10.1063/1.465572.

[366] T. A. Halgren. "The Representation of van der Waals (vdW) Interactions in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters". In: *Journal of the American Chemical Society* 114.20 (1992), pp. 7827–7843. DOI: 10.1021/ja00046a032.

[367] A. K. Soper. "Orientational Correlations in Hydrogen Bonded Networks". In: *Hydrogen Bond Networks*. Ed. by M. Bellissent-Funel and J. C. Dore. Dordrecht: Springer Netherlands, 1994, pp. 97–112. DOI: 10.1007/978-94-015-8332-9_11.

Chapter 6

Application to Water Models using Bias from Experimental Data

The Boltzmann inversion directed simulation (BIDS) method has thus far been shown to be able to routinely reproduce a target oxygen-oxygen (O-O) radial distribution function (RDF) from experiment and the various force field models investigated. This was achievable in two iterations, provided that the scaling coefficient was optimised. In this chapter, we will examine whether the experimental liquid water structure can be reproduced by applying the optimally scaled bias to both the water force fields and a density functional theory (DFT) water simulation. The resulting liquid structure and thermodynamic properties will be evaluated against the effects of comparable methods found in literature. This will include the comparison of the atomistic simulations of our biased empirical water models with the effective coarse-grained (CG) water models from single-state iterative Boltzmann inversion (IBI) and multistate iterative Boltzmann inversion (MS IBI).

Wang *et al.*[309] applied the IBI method to coarse-grain three common three-site water models, *i.e* TIP3P, SPC and SPC/E, where one water molecule was mapped to one CG bead at the centre of the oxygen atom. In effect, the non-bonded potential including the Coulomb interaction of the partial charges of the molecule were reduced to an isotropic two-body interaction. The CG simulations contained 10^4 beads in a box adjusted to densities from atomistic simulations with the temperature kept constant at 300 K. The coarse-graining demonstrated a speed-up in the computational time by a factor of 50. The shapes of the CG potentials were found to be dominated by the ratio, γ, between the second well

position and the first well position. This was equivalent to the ratio of two distances $2\sqrt{2} : \sqrt{3}$ or 1.633 due to the geometric constraints of the water clusters. The characteristic lengths were the distance between the reference molecule and the nearest neighbours, as well as the distance between pairs of nearest neighbours. They also studied the optimisation of the tetrahedral packing and the effects of pressure correction. The tetrahedral packing was improved by introducing a higher barrier to enter the first solvation shell. This was accomplished through the effective potential by increasing the height of the peak near the former characteristic length and the depth of the well near the latter characteristic length. At the same time, their positions were maintained to conserve the ratio of the two lengths. However, this results in a more structured RDF. Similarly, there was a compromise when applying a pressure correction as it caused large deviations in the isothermal compressibility preventing simultaneous agreement with the atomistic model.

The centre-of-mass-based mapping in coarse-graining methods are limited to one molecule per bead in liquid water. This is due to the water molecules moving somewhat independently of each other due to their loose association, which requires dynamic reassignment of more than one molecule to the CG bead. Hadley & McCabe[5] used a k-means clustering algorithm[368,369] for a dynamic scheme to allow mapping of multiple water molecules to a single CG bead in order to determine the optimal clustering with respect to solvation. The coarse-graining method used to determine the CG force field for TIP3P was the single-state IBI method. A box of 901 water molecules was equilibrated to a comparable density to the experimental value, first in the canonical ensemble (NVT) and then in the isothermal-isobaric ensemble (NPT). In the range of one- to nine-water beads investigated, the four-water bead model demonstrated the optimal balance between computational efficiency and accurate structural properties. The four-water bead model provided an increase in computational speed by a factor of 254, while achieving the most accurate water density at a difference of 0.1 % from the atomistic value. This difference was an order of magnitude less than the other multi-water bead models. The four-water bead model was also able to properly solvate the amphiphilic solute 1-pentanol, unlike the nine-water bead model, while providing better agreement with the water-water RDF from the water-pentanol mixture compared to the one-water bead model.

Following the success of the k-means algorithm, Moore *et al.*[113] used MS IBI to derive a CG water model with the optimal four water molecules mapped to one spherically symmetric bead. TIP3P was chosen as per Hadley & McCabe,[5] although a greater number of water molecules, 5832, was simulated. By using

MS IBI, the state dependence and the structural artefacts, which were otherwise present in the single-state IBI, were reduced through the simultaneous optimisation to three distinct states. These states were the bulk NVT at 1.0 g/mL and 305 K, bulk NPT at 305 K and 1.0 atm, and NVT droplet at 305 K. The use of both ensembles allowed the density-pressure relationship to be satisfied and provided a means to account for pressure without a correction. To further constrain the optimisation, they considered the Chiu *et al.*[370] Morse potential from a CG water force field that was optimised for the surface tension as an initial guess. Consequently, the potential had to be modified with a softer repulsion to allow the iterative scheme to update the smaller separations. The derived potential accurately reproduced the bulk properties, density and structural correlations, of the atomistic model at 305 K and 1.0 atm. The effect of softening of the initial potential was also observed as a reduced surface tension from that of the original potential that matched the experiment. However, the reduced value was a reasonable approximation to the atomistic model at 300 K. At the same time, the derived potential was able to reproduce the stable interface in simulations of droplets. In contrast with the single-state IBI potential, it maintained a stable interface in the simulation of droplets in agreement with the atomistic model. Moreover, the potential was not so strong as to readily crystallise at physiological conditions even in the presence of a nucleation site.

In the case of our biased DFT water simulation, the minimal bias approach of the experiment directed simulation (EDS) method applied to the *ab initio* molecular dynamics (AIMD) simulation of DFT water by White *et al.*[324] was most relevant as a benchmark. The bias is minimal in that it changes the statistical ensemble the least, based on the maximum entropy argument. The bias thus only weakly perturbs other unbiased properties. Using this bias, that incorporated the experimental correlation data, they attempted to address the inaccuracies in the underlying density functional without explicitly accounting for the nuclear quantum effects (NQE) or by increasing the simulation temperature. In addition to simulations of pure water, the diffusion of an excess proton in water was also investigated. A 15.5118 Å cubic box containing 128 water molecules, where the starting configuration was equilibrated with SPC, was simulated in an NVT ensemble at 300 K to derive the bias. The chosen density functional was Becke exchange with Lee-Yang-Parr correlation (BLYP) with a TZV2P basis set. Three 20 ps replicates of the EDS method found the converged bias, which was then applied at no noticeable cost to the computational resources in a 40 ps production run in the microcanonical ensemble (NVE) at a temperature of 300 K. With reference to the unbiased BLYP, the structural properties of BLYP-EDS was im-

proved as the O-O RDF was a near match to the experimental data. This also led to improvements in the unbiased observables including the oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H) RDFs where the over-structuring was diminished. While the self-diffusion coefficient of their BLYP simulation was two orders of magnitude lower than the experimental value, BLYP-EDS increased it by an order of magnitude. The form of the bias potential required for these improvements provided insight into the shortcomings of the density functional. It added repulsion in order to overcome the unphysical degree of over-polarisation and the anomalous charge transfer at short range, which contributed to the over-structured and slowly diffusing water. Furthermore, the transferability of the bias was tested in the AIMD simulation of a hydrated excess proton in water. The ratio of proton to water diffusion for DFT water, *e.g.* 70:1 for BLYP, greatly disagrees with the 4:1 for experiment due to the slow diffusivity of water. The BLYP-EDS water was shown to improve the excess proton diffusivity, which along with the increased water diffusion led to a better ratio 10:1 of proton to water diffusion. The same EDS method was also applied to BLYP with the D3 dispersion correction, which similarly resulted in improved structure and density. This demonstrated that both the short-range bias potential and the long-range dispersion correction were complementary corrections.

6.1 Improving Molecular Dynamics Simulations of Water Force Fields

The details of the water force field simulations were the same as for all the other simulations in the previous chapters. NVT simulations for the water force fields, SPC, SPC/Fw, TIP3P, TIP3P/Fs and TIP4P/2005, were performed within the open source large-scale atomic/molecular massively parallel simulator (LAMMPS). A density of $1001.95\,\mathrm{kg\,m^{-3}}$ was reproduced with a cubic box of length 19.7 Å containing 256 water molecules, where the initial configuration was obtained from a 10 ns equilibration with SPC/Fw. The temperature was maintained at 298 K by a chain of 5 Nosé-Hoover thermostats with a relaxation time of 0.1 ps. The systems were integrated at a timestep of 1 fs for a time length of 10 ns. The BIDS method was applied in two iterations with the scaling coefficient optimised and a bias cutoff of 7.0 Å. The target data was the experimental O-O RDF by Soper.[21]

The resulting biases applied to the respective water force fields produced the corresponding pair correlations, as seen in Figure 6.1 to Figure 6.5. Their MAE for

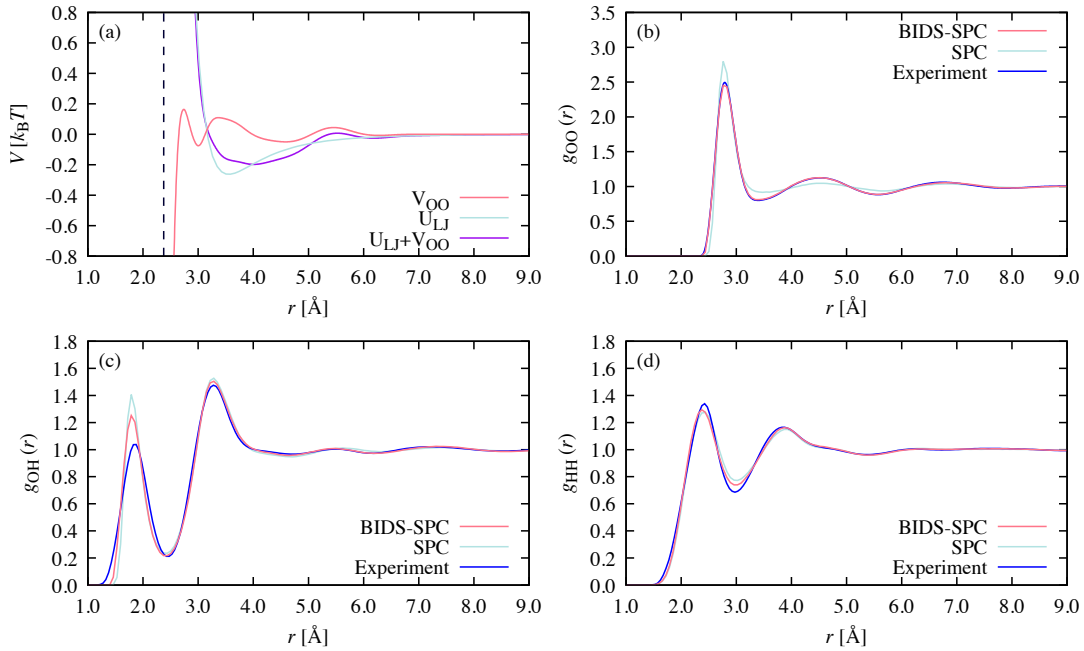


Figure 6.1: The resulting bias and fit to experimental target data from the BIDS-SPC simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The (b) O-O, (c) O-H and (d) H-H RDFs from an NVT simulation for the initial (green) and biased (red) SPC models and for the experiment (blue).

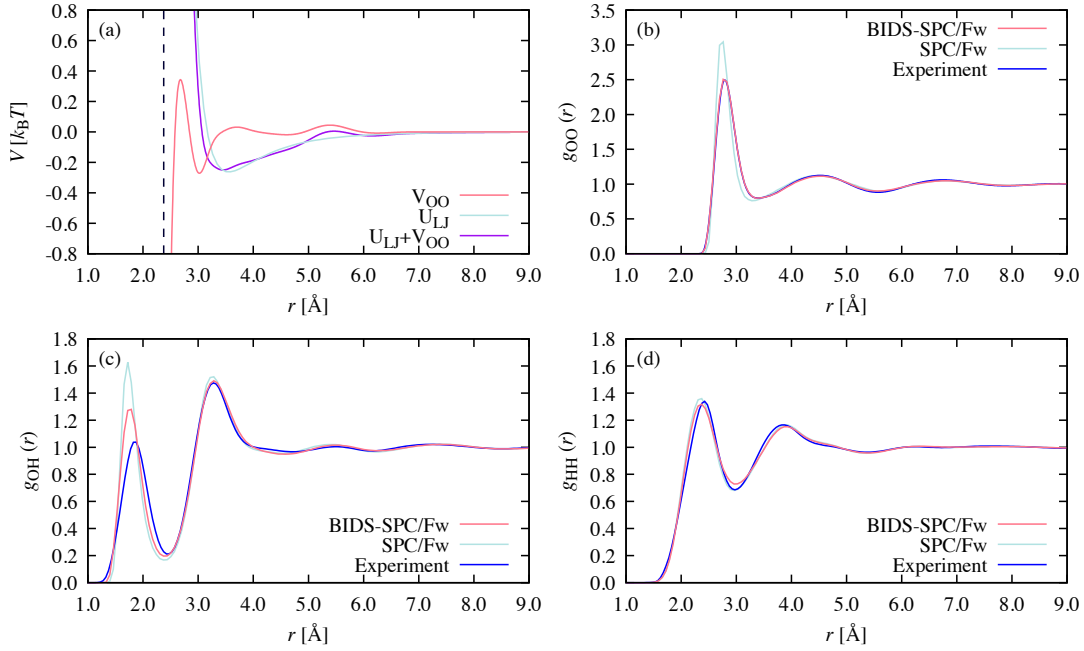


Figure 6.2: The resulting bias and fit to experimental target data from the BIDS-SPC/Fw simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The (b) O-O, (c) O-H and (d) H-H RDFs from an NVT simulation for the initial (green) and biased (red) SPC/Fw models and for the experiment (blue).

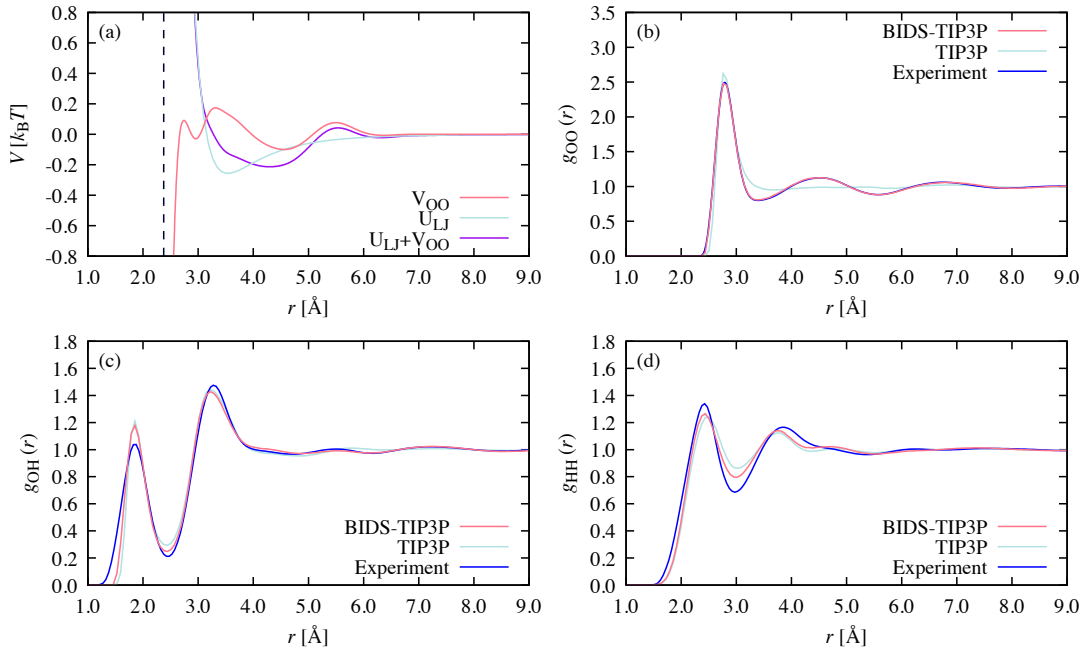


Figure 6.3: The resulting bias and fit to experimental target data from the BIDS-TIP3P simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The (b) O-O, (c) O-H and (d) H-H RDFs from an NVT simulation for the initial (green) and biased (red) TIP3P models and for the experiment (blue).

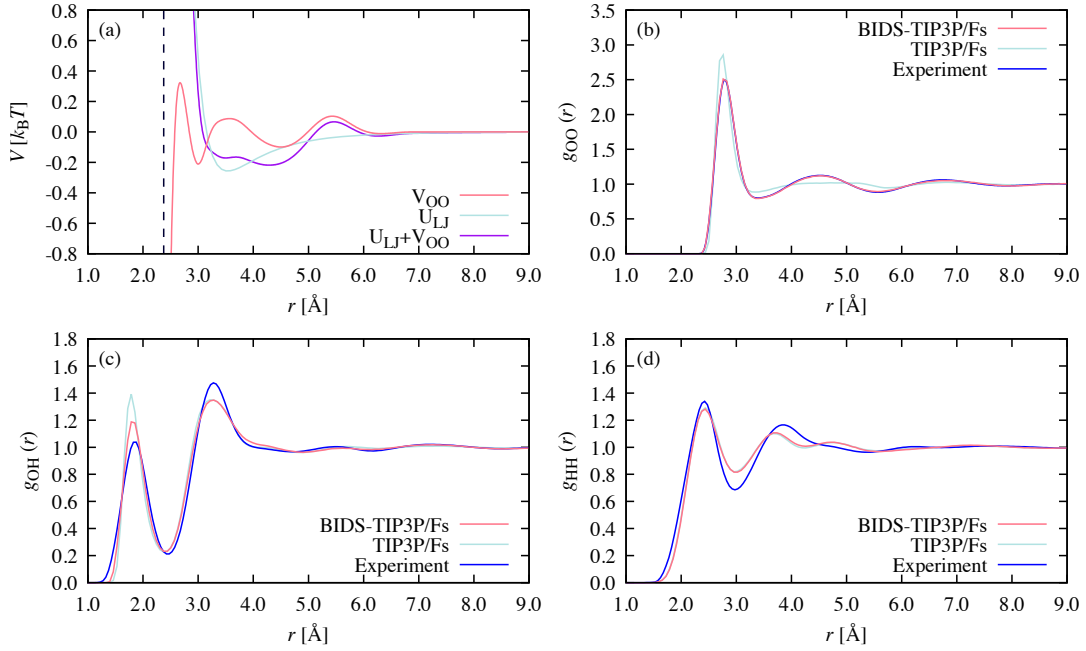


Figure 6.4: The resulting bias and fit to experimental target data from the BIDS-TIP3P/Fs simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The (b) O-O, (c) O-H and (d) H-H RDFs from an NVT simulation for the initial (green) and biased (red) TIP3P/Fs models and for the experiment (blue).

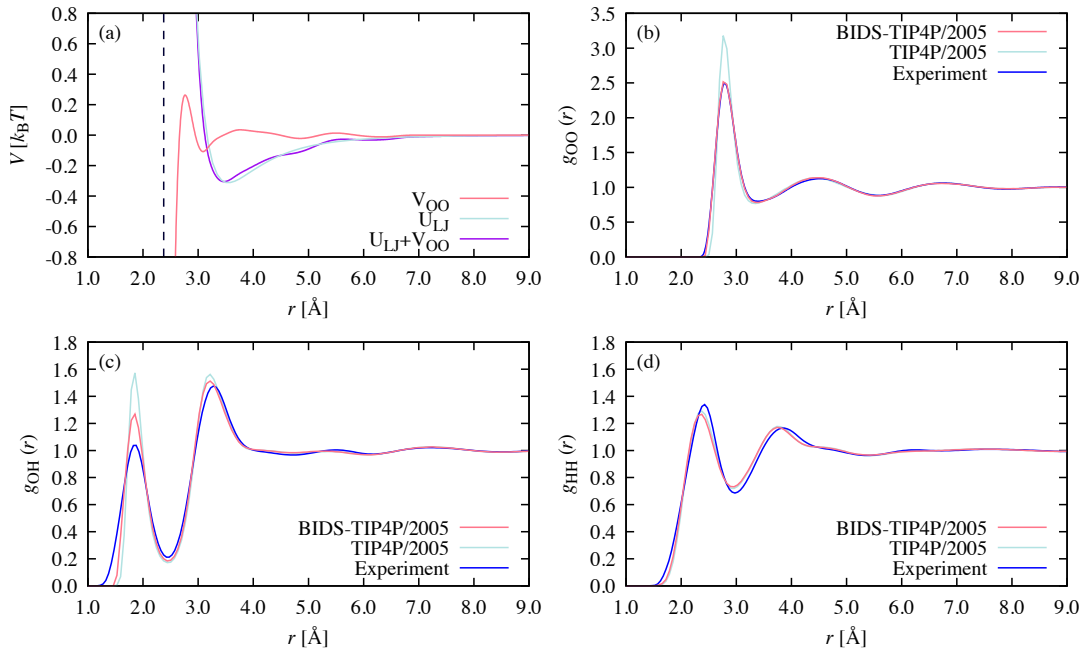


Figure 6.5: The resulting bias and fit to experimental target data from the BIDS-TIP4P/2005 simulation. (a) The bias applied and the resulting net potential, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The (b) O-O, (c) O-H and (d) H-H RDFs from an NVT simulation for the initial (green) and biased (red) TIP4P/2005 models and for the experiment (blue).

Table 6.1: The O-O RDF mean absolute error (MAE), self-diffusion coefficient, D_s [$\times 10^{-5}$cm^{2}s^{-1}], and static dielectric constant, ϵ_r, values for the experimentally biased water force fields. The zero superscript denotes the initial values. The diffusion coefficient and dielectric constant from experiment are $2.299 \times 10^{-5}\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ and 78.405, respectively. The mapping of the number of water molecules to a single bead and the diffusion coefficients are also given for the CG water models, where the parentheses denote the target values.

Model	$\mathrm{MAE}^0_{\mathrm{OO}}$	$\mathrm{MAE}_{\mathrm{OO}}$	D^0_s	D_s	ϵ^0_r	ϵ_r
BIDS-SPC	0.0489	0.0067	3.789	3.244	68.680	65.697
BIDS-SPC/Fw	0.0480	0.0084	2.352	2.820	77.915	76.807
BIDS-TIP3P	0.0639	0.0082	5.410	4.027	99.583	89.343
BIDS-TIP3P/Fs	0.0598	0.0064	3.711	3.406	173.588	154.459
BIDS-TIP4P/2005	0.0446	0.0096	2.050	2.288	107.990	111.110
1:1 CG IBI-SPC[309]				17.975 (4.437)		
1:1 CG IBI-TIP3P[309]				19.390 (5.932)		
4:1 CG MS IBI-TIP3P[113]				16.070 (3.050)		

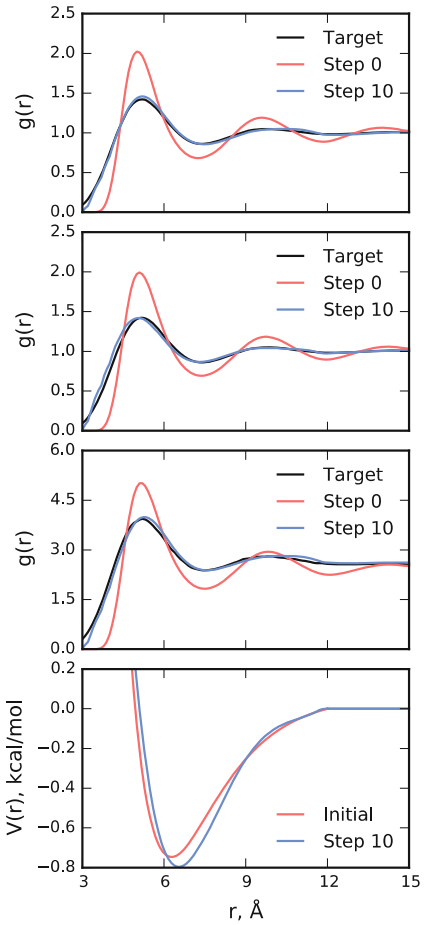


Figure 6.6: The RDFs for bulk NVT (top), bulk NPT (middle-top) and droplet NVT (middle-bottom) and potential (bottom) derived from the MS IBI method, adapted from Moore *et al.*[113] Reprinted by permission from Springer Nature Customer Service Centre GmbH: T. C. Moore, C. R. Iacovella, and C. McCabe. "Development of a Coarse-Grained Water Forcefield via Multistate Iterative Boltzmann Inversion". In: *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*. Ed. by R. Q Snurr, C. S. Adjiman, and D. A. Kofke. Singapore: Springer Singapore, 2016, pp. 37–52.

the O-O RDF and the investigated bulk thermodynamic properties are recorded in Table 6.1. The observable that was biased for, the O-O RDFs, were in good agreement with the target experimental data as expected. This was reflected in the MAEs which reduced by one order of magnitude to about 0.0079 ± 0.0038. As reported in the literature, the respective target correlation functions were also well matched in the single-state IBI[5,309] and MS IBI[113] CG water models. Moreover, the MS IBI method was able to reproduce their target RDFs from bulk NVT, bulk NPT and droplet NVT water, albeit with discernible but minor discrepancies, as shown in Figure 6.6 from Moore *et al.*[113] It is clear that the methods based on the iterative Boltzmann inversion scheme provide a straightforward route to deriving a potential that reproduces the target correlation function in both atomistic and CG systems.

Given the above success in reproducing biased quantities, it would be more telling to look to the unbiased observables, such as the O-H and H-H RDFs. Although the locations were not improved, the height of the first peak of the O-H RDFs all reduced becoming closer to the experimental data. The exception was for BIDS-TIP3P in Figure 6.3, where the height remained unchanged. As the first peak of the TIP3P O-O RDF was already close to the experiment, the resulting bias was weak in that region. The distribution of the water molecules in the first solvation shell was consequently not significantly perturbed by the bias, if at all. Based on our previous study in Chapter 5 on transformations between like-water force fields, we also expected to see some improvements in the first peak of the H-H RDF. However, they were overall weakly perturbed by the bias which was most apparent for BIDS-TIP3P/Fs in Figure 6.4. The weak response was not detrimental to most of the cases here. Excluding the three-site TIPnP variants, the initial water force fields reproduced H-H RDFs that were already fairly close to the experiment. As mentioned in Chapter 4, the BIDS method does not outright correct unbiased observables, but indirectly induces a response depending on the enhancing or diminishing of the liquid structure. This can cause deviations for properties that were already well approximated. For example, the self-diffusion coefficient of BIDS-SPC/Fw in the convergence and stability investigation in Chapter 4 deviated from the experimental value, when the initial SPC/Fw model already closely approximated the experimental value. We do note that the first valley of the H-H RDF for BIDS-TIP3P was noticeably improved from enhancing the liquid structure beyond the first peak of the O-O RDF.

The potentials from coarse-graining TIP3P into clusters of different numbers of molecules were shown in Figure 2.4 to have a form similar to a Lennard-

Table 6.2: The ratio, γ, between the second well position and the first well position for the biases of the various water force fields. The perfect ratio 1.633 describes the tetrahedral packing in liquid water. The first well positions, r_1^γ, and the second well positions, r_2^γ, are provided. The ratio found in the potential of mean force (PMF) of the experimental O-O RDF from Soper[21] is presented. The mapping of the number of water molecules to a single bead and the ratio of distances are also given for the CG water models.

Model	r_1^γ [Å]	r_2^γ [Å]	γ
Experimental PMF	2.790	4.530	1.624
BIDS-SPC	3.005	4.585	1.526
BIDS-SPC/Fw	3.020	4.605	1.525
BIDS-TIP3P	2.955	4.560	1.543
BIDS-TIP3P/Fs	3.000	4.500	1.500
BIDS-TIP3P/2005	3.090	4.850	1.570
1:1 CG IBI-SPC[309]			1.636
1:1 CG IBI-TIP3P[309]			1.617

Jones (LJ) potential, except for the one-water bead.[5] The potential of the one-water bead has multiple wells and an erratic shape, which was due to the high degree of structuring and hydrogen bonding features in the target RDF. Based on the location of the first well and second well, the erratic shape was most likely a product of the dominant ratio of the characteristic distances, 1.633, at the short range. The number from the experimental PMF shows that this is a good approximation. The one-water bead CG IBI-SPC and CG IBI-TIP3P with a 7.0 Å cutoff presented in Figure 2.3 were observed to have similar ratios 1.636 and 1.617, respectively.[309] In order to preserve the tetrahedral packing in liquid water, the ratio of the potentials should not deviate significantly from the dominant ratio. The bias potentials derived in this study were shown in Figure 6.1 (a) to Figure 6.5 (a) and their respective ratios were recorded in Table 6.2. Our biases were more stable in shape, usually with well-defined peaks and valleys in the short range and diminishing oscillations at medium to long range. Since the biases were dependent on the difference between the RDFs, they were not as susceptible to the degree of structuring in the target RDF. In effect, the biases had a shorter peak of $0.109\,k_\mathrm{B}T$ at 3.365 Å for BIDS-SPC and $0.173\,k_\mathrm{B}T$ at 3.315 Å for BIDS-TIP3P. This is in comparison to the peak of $\sim 0.882\,k_\mathrm{B}T$ ($\sim 2.2\,\mathrm{kJ\,mol^{-1}}$) and $\sim 0.682\,k_\mathrm{B}T$ ($\sim 1.7\,\mathrm{kJ\,mol^{-1}}$) at ~3.1 Å for the one-water bead CG IBI-SPC and CG IBI-TIP3P, respectively. Likewise, the difference between the second peak and second well that acts as a barrier of entry into the first solvation shell was relatively small for the biases. This barrier was slightly greater where the peaks in the O-O RDF beyond the first shell were enhanced. Consequently, the tetrahedral packing becomes more pronounced. The ratios of our biases at about 1.533 were a reasonable approximation to the dominant ratio, further contributing to the

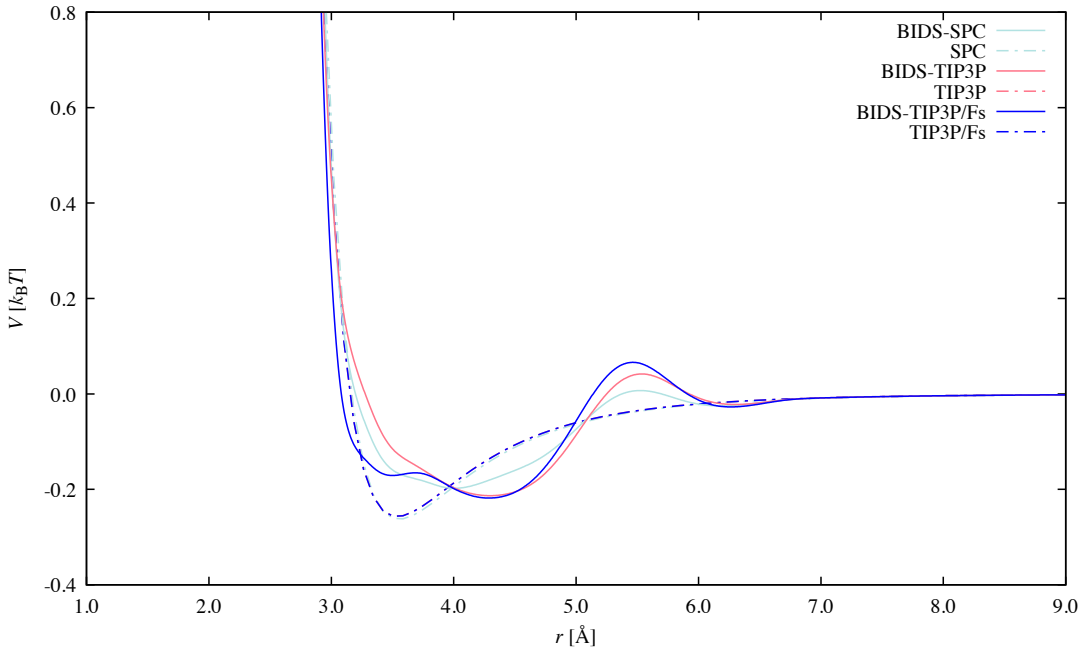


Figure 6.7: The resulting net potential from the BIDS application to water force fields that produced a flattened structure beyond the first peak of their RDFs. The three unbiased potentials are within the same line thickness.

tetrahedral packing of water. Additionally, the resulting net potentials possessed forms that can be described as typical of a pairwise interaction due to the greater influence of the LJ potential.

Generally, the three-site rigid models were observed to give too little structure beyond the first peak of the O-O RDF. [45] One way of enhancing the structure beyond the first peak was to introduce flexibility. This was effective for SPC/Fw but at the cost of the first peak becoming too high (Figure 6.2 (b)). Furthermore, TIP3P/Fs only marginally improved for the neighbouring peaks but still experienced the undesired accompanying increase in the first peak (Figure 6.4 (b)). The net potentials from BIDS-SPC, BIDS-TIP3P and BIDS-TIP3P/Fs provide insight into the shortcomings of the LJ potential. As seen in Figure 6.7, the net potentials overall have wells that are appreciably broader and shallower than the unbiased potentials, with the minima at a greater radial distance. In fact, the net potentials are arguably within a good tolerance of each other, such that there may be a common potential function that can be parameterised to fit them. This, along with the Halgren buffered 45-15 potential function fitted to the BIDS biases from different systems in Figure 5.5, suggest that a multi-system optimisation scheme may be viable for selected sets of water models.

Calculations of the properties of the biased water force fields produced self-diffusion coefficients and dielectric constants that were comparable to the con-

verged values in Figure 4.1. Since the dielectric constant was established to be resilient to the bias without softening of the bond angle, the diffusion coefficient was of more interest. Consistent with the converged values, the diffusion improved when the water distribution was corrected for, except for BIDS-SPC/Fw. The first peak in the O-O RDF of SPC/Fw was diminished when the bias was applied. This led to the weakening of the hydrogen bond network and consequently, an increase in the lateral diffusive motion of the water molecules. The same response in the diffusion was also observed for BIDS-TIP4P/2005. BIDS-SPC/Fw and BIDS-TIP4P/2005 increased by $0.468 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ and $0.239 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$, respectively. Rather than an outright correction, the diffusion thus increases with diminished liquid structure and vice versa. The other biased water force fields show decreased diffusivity, where BIDS-TIP3P changed by $-1.383 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ while BIDS-SPC and BIDS-TIP3P/Fs changed by $-0.545 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ and $-0.306 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$, respectively. There was an order of magnitude difference due to the interplay between the first peak in the O-O RDF and the peaks beyond that. For BIDS-TIP3P, the liquid structure was enhanced beyond the first peak whereas the first peak itself was not perturbed significantly. Similarly for BIDS-SPC and BIDS-TIP3P/Fs, the liquid structure beyond the first peak was enhanced. However, the first peak itself was diminished and thereby provided an opposing effect. While the diffusion has predominantly been described in terms of the intensity of the first peak, this also demonstrated the importance of the subsequent peaks to the diffusional dynamics of water. When compared to the CG models derived from both IBI and MS IBI, the changes in our diffusion coefficients were not as drastic. There was a speed-up factor of 3–5 times in diffusion for the CG models. The effect of coarse-graining clearly resulted in diffusional dynamics that did not reflect the atomistic behaviour. This was due to the effective time scale from the use of coarse-graining being larger, which also affects all other dynamics present in the system.[371] Furthermore, the 4:1 CG bead model did not necessarily represented the same water molecules through time due to the dynamic mapping of the k-means clustering.[113]

6.2 Improving *Ab Initio* Molecular Dynamics Simulations

The application of the BIDS method to the AIMD simulation of liquid water at 298 K was carried out using the BLYP-D3 functional. The choice of functional facilitate comparison with the result of White *et al.*[324] At room temperature, the

light hydrogen atoms of liquid water deviate significantly from the classical behaviour assumed to allow the approximation of the nuclear equations of motion in most AIMD simulations.[83–85] This approximation neglects the NQE, which is required for an accurate description of the microscopic structure of liquid water. Consequently, the predicted water RDFs becomes over-structured. A crude treatment for the NQE is to elevate the temperature by 30 K to mimic the effect of quantum nuclei on the O-O RDF of liquid water. By maintaining a temperature of 298 K in this study, the softening of the liquid structure using a two-body bias potential can thus be investigated. Furthermore, the dispersion-corrected density functional, BLYP-D3, was chosen to ensure that the resulting bias provided a complementary correction. The dielectric constant was not converged given the statistics of the 50 ps simulation and has shown to respond weakly to the bias according to the water force field simulations. Therefore, we only investigate the self-diffusion coefficient. The simulation was conducted with the Quickstep engine in CP2K, which unfortunately does not readily allow the bias potential from a file to be assigned to a pair of atoms. Hence, modifications to the program were required to create a branch in CP2K to facilitate the application of the BIDS method, as seen in Appendix C. An example input file specifying the required keyword nested within the generic potential section in the input structure in order to read the bias potential energy as a function of separation from a file is provided in Appendix D.

6.2.1 Simulation Details

The AIMD simulation of 96 water molecules contained in a 14.2 Å cubic box that was first equilibrated using SPC/Fw, was performed in the NVT ensemble. The number of water molecules was lower than in the atomistic simulations due to the higher cost of the DFT calculation. Concurrently, the cubic box was reduced in size to reproduce the 1001.95 $\mathrm{kg\,m^{-3}}$ density. The temperature was controlled at 298 K using the canonical sampling through velocity rescaling (CSVR) thermostat. The BLYP-D3 density functional along with the Gaussian and Plane Waves (GPW) method was used. The simple functional was chosen to enable comparison with the results from the application of EDS using the same level of DFT.[324] The DZVP basis sets were used with a 400 Ry cutoff for the auxiliary density. The core electron states were treated with Goedecker-Teter-Hutter (GTH) pseudopotentials. The molecular dynamics (MD) integration time step was 0.5 fs for a simulation time of 50 ps. The initial 5 ps was taken to be the equilibration period. The bias was added as a generic potential between the oxygen atoms through a

Table 6.3: The O-O RDF MAE and self-diffusion coefficient, D_s [$\times 10^{-5} cm^2 s^{-1}$], values for the BIDS-BLYP-D3 water compared to experiment. The zero superscript denotes the initial values. The diffusion is $2.299 \times 10^{-5}\,cm^2\,s^{-1}$ from experiment. The diffusion coefficients are also given for BLYP-EDS water, where the subscript denotes the replicate.

Model	MAE^0_{OO}	MAE_{OO}	D^0_s	D_s
BIDS-BLYP-D3	0.0497	0.0136	0.936	0.993
BLYP-EDS$_1$[324]			0.070	0.510
BLYP-EDS$_2$[324]			0.070	0.260
BLYP-EDS$_3$[324]			0.070	0.980

Hamiltonian defined by a mixing of forces. The parameters of the BIDS method remained the same. The target data was the experimental O-O RDF by Soper.[21] The scaling coefficients for two steps in the iteration scheme were scanned using 20 ps simulations, as was shown to be effective in Section 6.1.

6.2.2 Results and Discussion

The experimental O-O RDF was fitted using the bias derived from the BIDS method for the BLYP-D3 water model, as shown in Figure 6.8. Similarly, the results of the EDS method in the supplementary material of White *et al.*[324] provided a comparable correction to the O-O RDF as seen in Figure 6.9, but a closer inspection showed slight discrepancies in the first and second peaks of the liquid structure. This was similar to the MS IBI where the CG potential was constrained to multiple targets. Likewise, correcting for the pressure in single-state IBI also resulted in some loss of the quality of fit in the RDF. In the EDS method, the bias was constrained to maximise the entropy on top of the achieving the target RDF. Much like any optimisation scheme, it was thus likely that a small drop in the quality of the target occurs when other variables were incorporated.

The corresponding MAE for the O-O RDF was recorded in Table 6.3. The error of the fitted O-O RDF was 0.0136, which was slightly greater than the expected 0.0130 error for a 50 ps simulation using water force fields as mentioned in Section 4.5. This was due to the reduction in the number of points describing the curve of the O-O RDF to half of the smaller cubic box size. Moreover, the O-O RDF lacked the long-range region beyond 7.0 Å that was inherently small in error, which contributed to lower MAE values. The expected error becomes 0.0153 by recalculating the value using the same set of O-O RDFs but truncated at the cutoff of 7.0 Å. This was well above the error achieved for the BIDS-BLYP-D3 water model. Thus, the parameters used in the BIDS method achieved an acceptable accuracy to the target experimental data.

The structures of the unbiased observables were diminished, more so for the

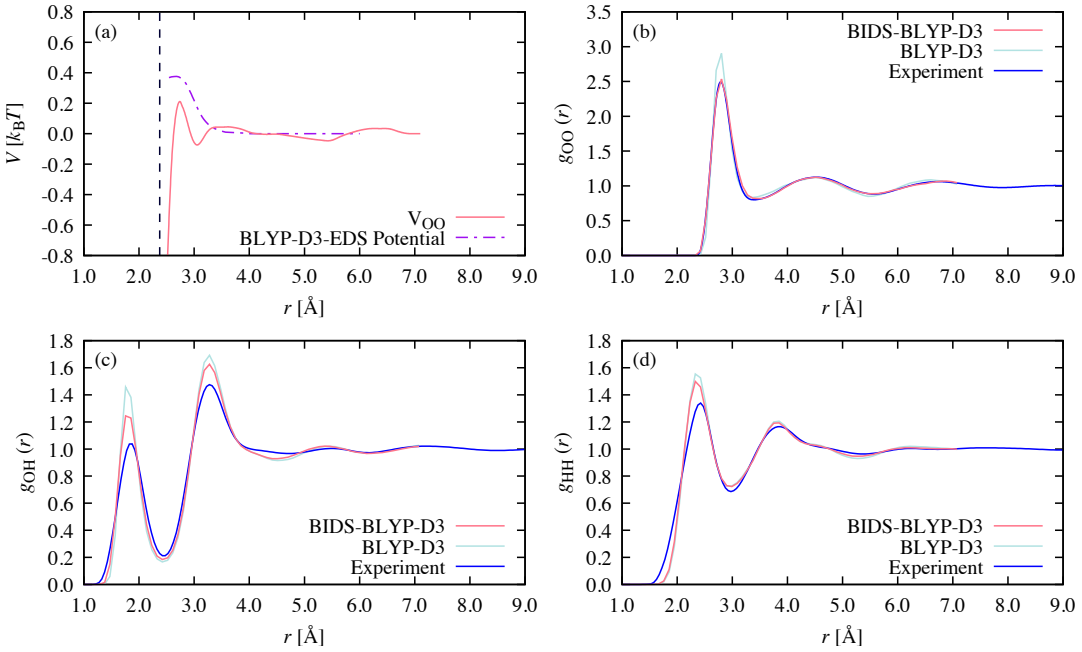


Figure 6.8: The resulting bias and fit to experimental target data from the BIDS-BLYP-D3 simulation. (a) The bias applied, where the vertical dashed line delineates the effectively zero-probability region in the O-O RDF. The BLYP-D3-EDS potential[324] is superimposed to facilitate comparison on the same scale. The (b) O-O, (c) O-H and (d) H-H RDFs from an NVT simulation for the initial (green) and biased (red) BLYP-D3 models and for the experiment (blue).

O-H RDF than the H-H RDF. This was consistent with the weakening of the hydrogen bond network from the diminished O-O RDF. The first and second peaks of the O-H RDF were improved, while the first peak of the H-H RDF was improved. Additionally, the first peak of the O-H RDF was improved to a greater degree than the first peak of the H-H RDF. The same observation was present in the structure of the unbiased observables using the EDS method. In contrast, the H-H RDF first valley of the BIDS-BLYP-D3 water model was not perturbed, whereas that of the BLYP-D3-EDS water model experienced a noticeable improvement. This was due to the greater correction to the more structured O-O RDF at that region produced by the initial water model used for the EDS method, compared to that used for the BIDS method which closely approximated the experimental water beyond the first peak.

Similar to the EDS method, the bias derived from the BIDS method has been demonstrated to act as a complementary supplement for empirical dispersion corrections. Conversely, the EDS biases were in the form of a short-range added repulsion from applying a unit step function that provided curvature predominantly between 2.5 Å and ∼4.0 Å, whereas the BIDS-BLYP-D3 bias had a well-defined first peak and first well followed by slight oscillations. Our bias was

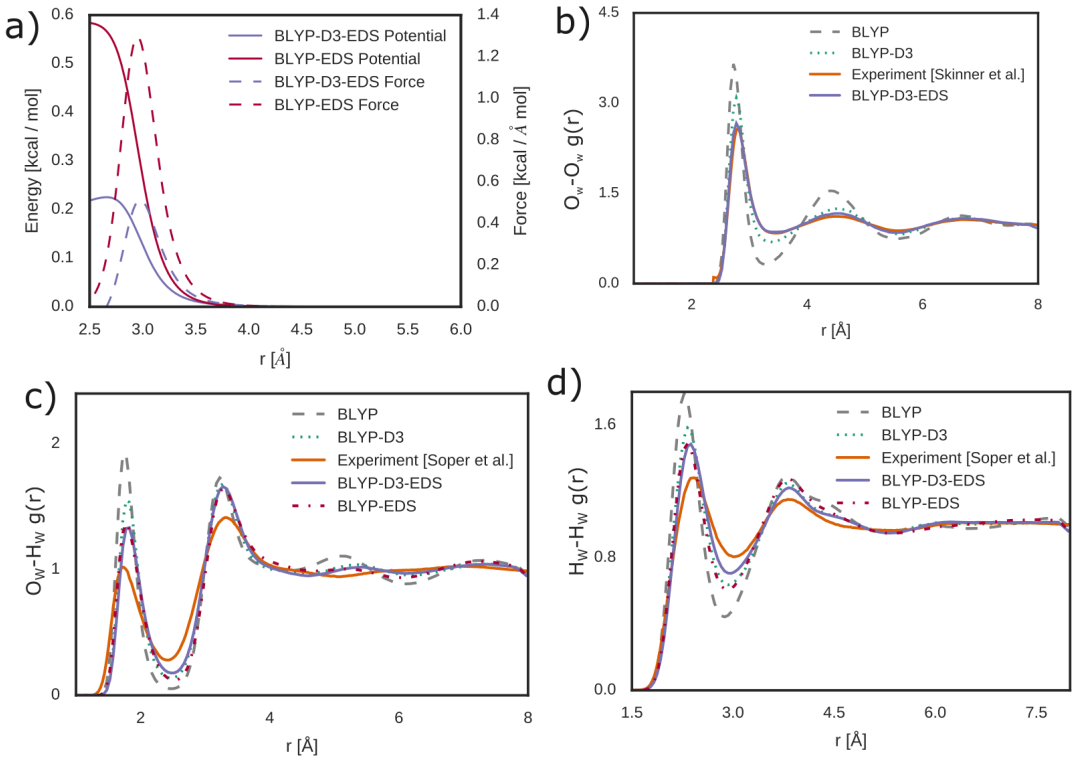


Figure 6.9: The resulting bias and fit to experimental target data from the EDS AIMD simulations.[324] (a) The equilibrated bias potential and force for BLYP-EDS (red) and BLYP-D3-EDS (blue) as a projection onto the distance between oxygen atoms. A smaller bias is required to converge when using the D3 dispersion correction. The (b) O-O, (c) O-H and (d) H-H RDFs from an NVE simulation using the equilibrated bias and experimental data from Soper *et al.*[21] The simultaneous use of EDS and D3 show better performance than either independently. Reprinted from A. D. White et al. "Communication: Improved *ab initio* Molecular Dynamics by Minimally Biasing with Experimental Data". In: *The Journal of Chemical Physics* 146.4 (2017), p. 041102, with the permission of AIP Publishing.

also repulsive in a smaller overlapping region with the EDS biases from 2.740 Å to 3.055 Å. While the BLYP-EDS bias was fairly high, approximately $0.98\,k_{\mathrm{B}}T$ ($0.58\,\mathrm{kcal\,mol^{-1}}$) at about 2.5 Å, the BLYP-D3-EDS bias was significantly smaller, approximately $0.38\,k_{\mathrm{B}}T$ ($0.22\,\mathrm{kcal\,mol^{-1}}$) at about 2.7 Å. This height acted as a barrier of entry into the first solvation shell in the liquid structure. The difference between the first peak and first well of the BIDS-BLYP-D3, which contained the repulsive region, also provided a barrier comparable to the smaller of the two EDS biases with a magnitude of $0.285\,k_{\mathrm{B}}T$. Conversely, the bias was attractive to the left of the aforementioned repulsive region. This would usually contribute to an enhanced structure in the short range. However, this part of the bias occupied the low-probability region. The ratio of the characteristic distances, which described the tetrahedral packing in liquid water, of the BIDS-BLYP-D3 bias was 1.412 that was a significant deviation from the perfect 1.633 ratio. Additionally, the second well was not as well defined as the biases for the water force fields. The bias was hence not crucial in enforcing the tetrahedral packing in the BLYP-D3 water model.

In addition to the improvements in the unbiased static properties, the self-diffusion coefficient also responded favourably to the bias from the BIDS method, as shown in Table 6.3. The diffusivity of the water molecules increased slightly by $0.057 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ from an initial value of $0.936 \times 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ in accordance with the weakening of the hydrogen bond. This improvement was, however, an order of magnitude lower than those observed in the water force fields investigations. Unfortunately, White *et al.*[324] did not calculate the diffusivities when the simulation time was below 40 ps, as in the case for the BLYP-D3-EDS simulations. However, those for the BLYP-EDS was provided for the three replicates that were simulated. As observed, the third replicate reproduced the diffusivity that was closest to the BIDS-BLYP-D3 water model. It was likely that this diffusion value was approximately the upper limit reproducible with the BLYP water model where the distribution of the water molecules was corrected from a two-body potential, whether it was simply a bias or a combination with the dispersion interaction, without accounting for the NQE. We again note the presence of finite-size effects decreasing the diffusivity of the system, particularly for a smaller cubic box of 96 water molecules.

6.3 Summary

The two-body potential between the water molecules can be optimised to closely reproduce the target O-O correlation function from the experimental data of

Soper,[21] for both water force fields and DFT water. The potential can also be optimised for other properties, *i.e.* pressure correction, MS IBI and maximal entropy, albeit slightly reducing the quality of the O-O RDF. Compared to the CG one-water bead potential from the IBI method, the application of the BIDS method in water force fields led to a bias that was less susceptible to the degree of structuring in the target O-O RDF. The bias also maintained a reasonably good ratio of characteristic distances that are important to the tetrahedral packing in liquid water. Due to the underlying LJ potential of the initial system, the resulting net potential maintained a form that was typical of pairwise interactions. Furthermore, the net potential provides some valuable insight into the shortcomings of the form of the LJ function. Most notably, the form of the potential required to reproduce the structure beyond the first peak of the O-O RDF in three-site models without the first peak becoming too high. The atomistic simulations used in the BIDS method were also necessary to retain the order of magnitude of the self-diffusion coefficient, as coarse-graining resulted in a speed up by a factor of 3–5 times.

The response of the BLYP-D3 water to the bias derived from the BIDS method was similar to that of the water force fields. The experimental O-O RDF was matched which subsequently improved the fit of the unbiased O-H and H-H RDFs in the short-range region to the experiment. A similar observation can be seen in the resultant pair correlation functions for the EDS method of White *et al.*[324] In fact, the biases obtained from both methods for the BLYP model with D3 dispersion correction provided a comparable barrier of entry into the first solvation shell, though the BIDS-BLYP-D3 bias does have a smaller barrier. We also note that our bias deviated significantly from the perfect ratio between the second well position and the first well position, on top of having a second well that was less defined. Thus, the bias was not responsible for the tetrahedral packing in the BLYP-D3 water model. Additionally, the self-diffusion coefficients calculated in both the BIDS and the EDS method provided insight into the upper limit achievable using a two-body potential in the BLYP-based water models. Considering the success of the MS IBI method, we note that the transferability of the BIDS biases can subsequently be improved through multiple states to accommodate the range of conditions explorable by DFT water models. As the EDS biases were not derived from the Boltzmann inversion of pair correlation functions and their repulsive form was predominantly constrained to the short range region, it is not known whether extending this method to multiple states will be convergent or achieve the desired outcome.

References

[5] K. R. Hadley and C. McCabe. "On the Investigation of Coarse-Grained Models for Water: Balancing Computational Efficiency and the Retention of Structural Properties". In: *The Journal of Physical Chemistry B* 114.13 (2010). PMID: 20230012, pp. 4590–4599. DOI: 10.1021/jp911894a.

[21] A. K. Soper. "The Radial Distribution Functions of Water as Derived from Radiation Total Scattering Experiments: Is There Anything We Can Say for Sure?" In: *ISRN Physical Chemistry* 2013.279463 (2013), p. 67. DOI: 10.1155/2013/279463.

[45] W. L. Jorgensen et al. "Comparison of Simple Potential Functions for Simulating Liquid Water". In: *The Journal of Chemical Physics* 79.2 (1983), pp. 926–935. DOI: 10.1063/1.445869.

[83] B. Chen et al. "Hydrogen Bonding in Water". In: *Phys. Rev. Lett.* 91 (21 2003), p. 215503. DOI: 10.1103/PhysRevLett.91.215503.

[84] J. A. Morrone and R. Car. "Nuclear Quantum Effects in Water". In: *Phys. Rev. Lett.* 101 (1 2008), p. 017801. DOI: 10.1103/PhysRevLett.101.017801.

[85] M. Ceriotti et al. "Nuclear Quantum Effects and Hydrogen Bond Fluctuations in Water". In: *Proceedings of the National Academy of Sciences* 110.39 (2013), pp. 15591–15596. DOI: 10.1073/pnas.1308560110.

[113] T. C. Moore, C. R. Iacovella, and C. McCabe. "Development of a Coarse-Grained Water Forcefield via Multistate Iterative Boltzmann Inversion". In: *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*. Ed. by R. Q Snurr, C. S. Adjiman, and D. A. Kofke. Singapore: Springer Singapore, 2016, pp. 37–52. DOI: 10.1007/978-981-10-1128-3_3.

[309] H. Wang, C. Junghans, and K. Kremer. "Comparative Atomistic and Coarse-Grained Study of Water: What do we Lose by Coarse-Graining?" In: *The European Physical Journal E* 28.2 (2009), pp. 221–229. DOI: 10.1140/epje/i2008-10413-5.

[324] A. D. White et al. "Communication: Improved *ab initio* Molecular Dynamics by Minimally Biasing with Experimental Data". In: *The Journal of Chemical Physics* 146.4 (2017), p. 041102. DOI: 10.1063/1.4974837.

[368] L. M. L. Cam and J. Neyman. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Biology and Problems of Health*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Held at the Statistical Laboratory, University of California, June 21-July 18, 1965 and December 27, 1965-January 7, 1966. University of California Press, 1967.

[369] D. Steinley. "K-Means Clustering: A Half-Century Synthesis". In: *British Journal of Mathematical and Statistical Psychology* 59.1 (2006), pp. 1–34. DOI: 10.1348/000711005X48266.

[370] S. Chiu, H. L. Scott, and E. Jakobsson. "A Coarse-Grained Model Based on Morse Potential for Water and n-Alkanes". In: *Journal of Chemical Theory and Computation* 6.3 (2010). PMID: 26613312, pp. 851–863. DOI: 10.1021/ct900475p.

[371] S. J. Marrink, A. H. de Vries, and A. E. Mark. “Coarse Grained Model for Semiquantitative Lipid Simulations”. In: *The Journal of Physical Chemistry B* 108.2 (2004), pp. 750–760. DOI: 10.1021/jp036508g.

Chapter 7

Conclusions

We have developed a robust and straightforward Boltzmann inversion directed simulation (BIDS) method to derive a bias potential that incorporates the pair correlation function, which follows the Boltzmann distribution. The method can be utilised in atomistic simulations using the experimental pair correlation data. In this study, we have demonstrated its utility in liquid water simulations at ambient temperature by deriving the oxygen-oxygen (O-O) bias. The derived biases enabled all the water models studied here to reproduce the O-O radial distribution function (RDF) from experiment. This was accomplished through the application of the iterative Boltzmann inversion scheme to optimise the bias as an addition to the system potential. The bias was implemented as an empirical correction in the same fashion as the long-range dispersion correction in density functional theory (DFT). In practice, the computational scaling was thus maintained at a similar level to the theory of the simulated system. This was particularly advantageous for *ab initio* molecular dynamics (AIMD) simulations, which not only produced over-structured pair correlation functions from omitting the nuclear quantum effects but also scaled unfavourably with the system size.

In the BIDS method, the amplitude of the bias potential was optimised through a scaling coefficient at each step in the iteration scheme. This prevented overcorrection in the biased pair correlation function and stabilised the convergence. A sufficiently good fit to the target O-O RDF was usually achieved within two iterations. A smooth switching function was applied to the bias in order to prevent a force discontinuity at the cutoff at 7.0 Å. The quality of fit of the resulting O-O RDF was not adversely affected at the selected cutoff. Moreover, the cutoff allowed the bias to complement the long-range dispersion correction usually implemented in DFT. The short-range region of the bias, where the probability of finding an oxygen was low, had little effect on the overall liquid structure. However, ensuring a smooth transition of forces to the zero probability

region was important to accommodate the equilibration phase of the molecular dynamics (MD) simulation. In this phase, the separation between particles may fall into the low probability region. The constituent RDFs required to derive the bias were treated with a smooth splining algorithm augmented by a weighting function parameterised for the sharper and narrower first peak relative to the subsequent peaks. This was required to remove artefacts that would otherwise be translated into the bias due to the lack of statistics, particularly from the short AIMD simulations.

The bias potential derived from the BIDS method was found to be convergent and effective in reproducing the experimental O-O RDF for both force field and DFT simulations of liquid water at ambient temperature. Moreover, the BIDS method was not limited to the experimental data and can be used for any target pair correlation functions, including the O-O RDFs of various water models. Concurrently, the other pair correlation functions, *i.e.* oxygen-hydrogen (O-H) and hydrogen-hydrogen (H-H) RDFs, required for a full description of the liquid structure were usually improved despite not being directly biased. The improvement occurred in the short-range region up to the second peak for both the O-H and H-H RDFs. In addition, the dynamics of the system were also affected by the changes in the liquid structure. The self-diffusion coefficient was found to increase when the structure of the O-O RDF was diminished in accordance with the weakening of the hydrogen-bond network and vice versa. Although this typically led to an improved diffusivity, this was not always the case as demonstrated by SPC/Fw that already well approximated the experimental value. In contrast, the dielectric constant or relative permittivity were not significantly affected. The changes in the dielectric constant depended on the softness of the bond angle. In our investigation, TIP3P/Fs was the only water force field with a sufficiently soft bond angle for BIDS to elicit a noticeable change. Compounding on the lack of response, the dielectric constant was also not sufficiently converged for the time scale afforded by AIMD simulations.

The use of force fields in the water simulations allowed insight into the resultant interaction between water molecules. The greatest change can be observed around the wells of the net potential from the initial Lennard-Jones (LJ) potential and the bias potential. In addition to the location and the depth of the well, the breadth of the well was also affected. This led to changes in the forces experienced in the short-range region, which became less repulsive for an enhanced first peak in the O-O RDF and vice versa. For the net potential, the widening of the breadth of the well with a flat curvature was observed to contribute to the deepening of the first valley in the O-O RDF. A right shift in position of the

attractive region accompanied this widening, which contributed to an enhanced structure beyond the first peak in the O-O RDF. The bias itself has a typically defined "well", whose sense is dependent on the required change to the liquid structure. This was usually followed with subsequent oscillations from the difference between the O-O RDFs of the target and the system. Since the shape of the net potential well was usually difficult to reproduce with the LJ form, a more comprehensive analytical potential should be considered.

The effects of adding a bias derived from the BIDS method was evaluated against coarse-graining a potential using the iterative Boltzmann inversion (IBI) method. Both methods solve for the pair potential in the inverse Henderson problem that reproduces the correlation function, which obeys the Boltzmann distribution. However, the former adds a bias to perturb the system whereas the latter derives a novel potential for the coarse-grained (CG) system. Due to the underlying LJ potential of the system in the BIDS method, the resulting net potentials with the biases maintained a form that was typical of a pairwise interaction. In contrast, the one-water bead CG potentials derived in the literature from the IBI method was susceptible to the degree of structuring in the target pair correlation function. This caused large oscillations in the one-water bead CG potential that was unusual for a pairwise interaction. Furthermore, coarse-graining led to the unphysical increase in the self-diffusion coefficient by a factor of 3–5 times due to the change in the effective time scale. Approximate values for the ratio of the well positions describing the tetrahedral packing of water clusters that were observed in the CG potentials were also found in the BIDS biases.

In the case of the Becke exchange with Lee-Yang-Parr correlation (BLYP) water model with D3 dispersion correction, the ratio of the well positions was underestimated and the one of the wells was ill defined for the bias potential. As such, the bias had little effect in enforcing the tetrahedral packing in the DFT water model investigated. The short-range repulsive region of the bias had a smaller barrier of entry into the first solvation shell compared to that derived in literature from the experiment directed simulation (EDS) method. The respective biases led to a satisfactory fit to the experimental O-O RDF of liquid water for BLYP-D3. The corresponding improved fit to the unbiased O-H and H-H RDFs were also similar between both methods. The self-diffusion coefficient achieved from emulating the local packing of water molecules from experiment was similar between BLYP-D3 corrected by the BIDS method and BLYP corrected by the EDS method. This suggested an upper limit to improving the diffusivity as a result of correcting the O-O RDF via a two-body bias potential. Moreover, the improvement to the self-diffusion was not as significant as those observed in the

application of the BIDS method to water force fields.

7.1 Recommendations

The bias potentials thus far have been optimised to reproduce the liquid structure for water at room temperature. As observed in this work, the bias potential achieves this single target rather well according to an average error measure or the mean absolute error (MAE). Here, the use of an average error is sufficient with an approximately equal reproduction of the experimental liquid structure as an end goal. In the case where reporting the intermediate convergence is important, a weighted error function will however be more suitable.

The BIDS method is recommended to investigate any force field water simulations that do not reproduce the experimental liquid structure. The resulting correction to the O-O RDF causes a measurable and usually desirable change to other properties. In contrast, the BIDS method is better used to drive structural refinement of an approximately correct DFT water to that consistent with experimental water without detrimentally perturbing other properties. This is because unlike for force fields, the BIDS corrective bias does not translate into proportional improvements to other properties in DFT water simulations. This is somewhat expected, as although the corrective bias is implemented similarly to a two-body dispersion, it is not fundamentally built to emulate a physical interaction like dispersion but rather to bias a system towards structural conformity.

As a result of optimising to a single thermodynamic state, the bias also lacks transferability to other thermodynamics states that are of interest. It would be interesting to explore the optimisation of the bias to multiple states, similar to that employed in multistate iterative Boltzmann inversion (MS IBI). On top of improving transferability to other thermodynamic states, the MS IBI method can be extended to ensemble configurations as well as interfacial systems. We do note the unfortunate and significant increase to the initial computational cost required to derive a converged bias, particularly for the DFT water. This is due to MS IBI requiring three states to be simulated per iteration until convergence is achieved with each RDF. In fact, approximately 50 iterations, as mentioned in the theory chapter, has been reported for the potential to be well converged.

The bias can also be constrained to satisfy the Ornstein-Zernike equation in order to reconcile with statistical fluid mechanics. Clearly, a more physically meaningful bias should simultaneously satisfy the convergence criteria to the target RDF within some set tolerance and the Ornstein-Zernike (OZ) equation. The challenge in this endeavour is in the approximation of the bridge function in the

formally exact closure relation required to complete the Ornstein-Zernike equation. Similar to any optimisation procedure, the caveat of optimising for more than one target variable was a reduction in the simultaneous quality of fit. However, this was observed to be quite small in the application of the MS IBI method in coarse-graining a water force field.

Appendix A

Smooth Splining Algorithm

```
#!/usr/bin/Rscript
# ****************************************************************
#> \brief Smooth splining function for correlation functions
#> \param g ... correlation function
#> \param wd ... working directory
#> \param gt ... target correlation function
#> \param fine ... fine binning
#> \retval gofr ... smooth correlation function
#> \author
#>    T. Ling 2015 <tiongtze.ling@postgrad.curtin.edu.au>
# ****************************************************************
gofr<-function(g,wd,gt=g,fine=FALSE) {

  # Call functions in working directory
  source(file.path(wd,"function.R"))

  # Initialise parameters
  phi<-(1+sqrt(5))/2
  lower.itv<-0.05
  lower.span<-2/300
  upper.itv<-0.185
  upper.span<-4/300

  # Determine binning for smooth spline based on target
  #  correlation function, or use fine binning
  if (fine==FALSE) {
    dp<-as.integer(10^min(max(num.dp(gt$V1),2),log10(200)))
  } else {
    dp<-as.integer(2000)
  }

  # Remove leading zeros from the correlation functions
```

```
lower<-left(gt$V1)
g<-tail(g,dim(g)[1]-last.zero(g$V2)+2)
gt<-tail(gt,dim(gt)[1]-last.zero(gt$V2)+2)

# Determine curves for the correlation function and its second
#  derivative based on the binning
g.fun<-splinefun(data.frame(g[c("V1","V2")]))
g.f<-curveless(g.fun(x,deriv=0),left(gt$V1),right(gt$V1),
               vspan(gt$V1,dp))
g.h<-curveless(g.fun(x,deriv=2),left(gt$V1),right(gt$V1),
               vspan(gt$V1,dp))

# Determine the weighing function for smooth splining based on
#  the correlation function curve and its second derivative
i1<-x.axis(g.h$y,which.min(g.h$y[1:tail(which(g$V2!=0),1)]))
i2<-max2(g.h$y[1:(2*i1)])
mu<-mean(c(g.f$x[i1],g.f$x[i2]))
sigma<-(g.f$x[i2]-g.f$x[i1])/2
erfc<-function(x) (1-0.001)*pnorm(x,mu,sigma,lower=FALSE)+0.001
g.w1<-curveless(erfc(x),left(gt$V1),right(gt$V1),
                vspan(gt$V1,dp))

i3<-head(which(g.w1$y==0.001),1)
i4<-i3+round((i3-i2)/2)
mu<-g.f$x[i4]
sigma<-1/(phi*pi)*(g.f$x[i4]-g.f$x[i3])
erfc<-function(x) (0.001-0.0001)*pnorm(x,mu,sigma,lower=FALSE)
                  +0.0001
g.w2<-curveless(erfc(x),g.f$x[i3],right(gt$V1),xspan(g.f$x[i3],
                right(gt$V1),dp))

g.w<-data.frame(x=c(head(g.w1$x,i3-1),g.w2$x),
                y=c(head(g.w1$y,i3-1),g.w2$y))

# Fit initial smooth spline to correlation function curve based
#  on the weighing function
gspline<-smooth.spline(g.f$x,g.f$y,g.w$y,
                       control.spar=list(low=0.25,high=0.5))
gspline.fun<-splinefun(data.frame(gspline[c("x","y")]))
gspline.f<-curveless(gspline.fun(x,deriv=0),left(gt$V1),
                     right(gt$V1),vspan(gt$V1,dp))

# Determine location of the knot on the initial smooth spline
imax<-which.max(gspline.f$y)
ihyman<-tail(which(g$V1<gspline.f$x[imax]),1)
if (hat.all.equal(g$V1[ihyman],gspline.f$x[imax]))
```

```
    {ihyman<-ihyman-1}

  # Fit monotonic spline to initial smooth spline before knot
  hyman.fun<-splinefun(data.frame(c(g$V1[1:ihyman],
                       gspline.f$x[imax]),c(g$V2[1:ihyman],
                       gspline.f$y[imax]),method="hyman"))
  hyman.f<-curveless(hyman.fun(x,deriv=0),left(gt$V1),
                     gspline.f$x[imax],xspan(left(gt$V1),
                     gspline.f$x[imax],dp))

  # Complete smooth correlation function using monotonic spline
  #  and initial smooth spline after knot
  gofr<-data.frame(x=gspline.f$x,
                   y=c(hyman.f$y,
                       gspline.f$y[(imax+1):length(gspline.f$y)]))

  # Localised regression of smooth correlation function and zero
  #  negatives that occur
  mu<-lower.itv+(upper.itv-lower.itv)/2
  sigma<-1/(phi*pi)*(upper.itv-lower.itv)/2
  erfc<-function(x) (upper.span-lower.span)
                     *pnorm(x,mu,sigma,lower=TRUE)+lower.span
  dspan<-erfc(1/dp)
  izero<-tail(which(hyman.f$x<g$V1[which.max(g$V2)]),1)
  for (i in 1:2) {
    gofr<-suppressWarnings(loess.smooth(gofr$x,gofr$y,span=dspan,
                                        degree=1,
                                        family="gaussian",
                                        evaluation=vspan(gofr$x,
                                                         dp)))
    flip<-TRUE
    for (j in izero:1) {
      if (gofr$y[j]<0) {flip<-FALSE}
      if (flip==FALSE) {gofr$y[j]<-0}
    }
  }

  # Pad leading zeros to the smooth correlation function
  xpad<-head(seq(lower,left(gofr$x),1/dp),(left(gofr$x)-lower)*dp)
  ypad<-rep(0,(left(gofr$x)-lower)*dp)
  gofr<-data.frame(V1=c(xpad,gofr$x),V2=c(ypad,gofr$y))

  return(gofr)
}
```

Appendix B

Module for Functions

```
#!/usr/bin/Rscript
# ****************************************************************
#> \brief Module for functions
#> \author
#>    T. Ling 2015 <tiongtze.ling@postgrad.curtin.edu.au>
# ****************************************************************

# ****************************************************************
#> \brief Determine equally spaced points over a range for a
#         function. This was modified from "curve" to remove
#         drawing of a plot.
#         Evaluate the points for a function from the spline
#         interpolation of a pair correlation function and its
#         derivatives.
#> \param expr ... function name
#> \param from ... range lower bound
#> \param to ... range upper bound
#> \param n ... number of bins in range
#> \param xname ... character name with respect to which
#                   derivative is computed
# ****************************************************************
curveless<-function (expr, from = NULL, to = NULL, n = 101,
                     xname = "x", ...)
{
  sexpr <- substitute(expr)
  if (is.name(sexpr)) {
    expr <- call(as.character(sexpr), as.name(xname))
  }
  else {
    if (!((is.call(sexpr) || is.expression(sexpr)) && xname %in%
          all.vars(sexpr)))
      stop(gettextf("'expr' must be a function, or a call or an
```

```
                    expression containing '%s'", xname),
            domain = NA)
    expr <- sexpr
  }
  if (is.null(from) || is.null(to)) {
    xl <- c(0, 1)
    if (is.null(from))
      from <- xl[1L]
    if (is.null(to))
      to <- xl[2L]
  }
  x <- seq.int(from, to, length.out = n)
  ll <- list(x = x)
  names(ll) <- xname
  y <- eval(expr, envir = ll, enclos = parent.frame())
  if (length(y) != length(x))
    stop("'expr' did not evaluate to an object of length 'n'")
  invisible(list(x = x, y = y))
}

# ************************************************************
#> \brief Returns a boolean where two arrays are element-wise
#         equal. Compare two indexed positions to ensure knot was
#         not duplicated, when forming the final smooth spline
#         from the monotonic-preserving spline and the cubic basis
#         spline.
#> \param x ... position i
#> \param y ... position j
# ************************************************************
hat.all.equal<-Vectorize(function(x,y) {isTRUE(all.equal(x,y))})

# ************************************************************
#> \brief Count number of decimal places for the radial distances
#         from the correlation function.
#> \param x ... radial distance array
# ************************************************************
num.dp<-function(x) {
  stopifnot(class(x)=="numeric")
  x<-sub("0+$","",x)
  x<-sub("^.+[.]","",x)
  nchar(x)
}

# ************************************************************
#> \brief Find left position in the correlation function.
#> \param x ... radial distance array
```

```
# ****************************************************************
left<-function(x) {x[1]}

# ****************************************************************
#> \brief Find right position in the correlation function.
#> \param x ... radial distance array
# ****************************************************************
right<-function(x) {tail(x,1)}

# ****************************************************************
#> \brief Count length of the correlation function.
#> \param x ... radial distance array
#> \param dp ... exponent number of decimal places
# ****************************************************************
vspan<-function(x,dp) {round((tail(x,1)-x[1])*dp+1)}

# ****************************************************************
#> \brief Find position in the region of the correlation
#         function.
#> \param x1 ... position 1
#> \param x2 ... position i
#> \param dp ... exponent number of decimal places
# ****************************************************************
xspan<-function(x1,x2,dp) {round((x2-x1)*dp+1)}

# ****************************************************************
#> \brief Find position of zero before the first nonzero in the
#         correlation function.
#> \param xt ... array i
#> \param xi ... array j
# ****************************************************************
last.zero<-function(xt,xi=xt) {
  xt<-head(which(xt!=0),1)-1
  xi<-head(which(xi!=0),1)-1
  max(xt,xi)
}

# ****************************************************************
#> \brief Find first maximum after first minimum in the second
#         derivative of the correlation function.
#> \param x ... array
# ****************************************************************
max2<-function(x) {
  min.glb<-which.min(x)
  min.glb-1+which(x[min.glb:length(x)]==max(x[min.glb:length(x)]))
}
```

```
# ****************************************************************
#> \brief Find x-axis intercept of the second derivative of the
#         correlation function.
#> \param x ... array
#> \param i ... index
# ****************************************************************
x.axis<-function(x,i=1) {
  sign.flip<-i-1+head(which(x[i:length(x)]/abs(x[i:length(x)])!=
                            x[i]/abs(x[i])),1)

  if(length(sign.flip)!=0 && abs(x[sign.flip])>abs(x[sign.flip-1])
    ){sign.flip<-sign.flip-1}

  sign.flip<-max(0,sign.flip)

  return(sign.flip)
}
```

Appendix C

Branch in CP2K

```
"cp2k/src/force_fields.F"
28,29d27
<    ! Call Boltzmann inversion directed simulation utility module
<    USE bids_util,                        ONLY: read_splinit
124d121
<     CALL read_splinit ! Call initialise read spline file

"cp2k/src/pair_potential.F"
19,22d18
<
! Call function to
<
! get unique index
<                                                  ! number
<                                                  get_unit_number,&
67,72c63
<              get_nonbond_storage, &
<              bids_store_read ! Declare global variable to store
<                              ! read spline
<
<    ! Declare global variable type
<    INTEGER, DIMENSION(:), ALLOCATABLE :: bids_store_read
---
>              get_nonbond_storage
91,98d81
<       !> \subbrief Initialise spline file
<       ! ..................................................
<       INTEGER                                            :: &
<          bids_i
<       CHARACTER(len=4)                                   :: &
<          bids_func
<       INTEGER, DIMENSION(:), ALLOCATABLE                 :: &
```

```
<          bids_temp_read
112,113d94
<       ALLOCATE (bids_temp_read(ngp))
<       bids_i = 0
122,130c103
<                   bids_func = TRIM(pot%set(k)%gp%potential)
<                   IF (bids_func == "READ") THEN
<                      bids_i = bids_i+1
<                      bids_temp_read(bids_i) = ngp
<                   ELSE
<                      CALL parsef(ngp, TRIM( &
<                                  pot%set(k)%gp%potential), &
<                                  pot%set(k)%gp%parameters)
<                   END IF
---
>                   CALL parsef(ngp, TRIM( &
>                      pot%set(k)%gp%potential), &
>                      pot%set(k)%gp%parameters)
135,140d107
<       IF (ALLOCATED(bids_store_read)) THEN
<          DEALLOCATE (bids_store_read)
<       END IF
<       ALLOCATE (bids_store_read(bids_i))
<       bids_store_read(:) = bids_temp_read(1:bids_i)
<       ! ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
276,282c243,244
<                                  ! Setup read spline cutoff
<               pot%spl_f%cutoff = pot%spl_f%cutoff* &
<                                     pot%spl_f%fscale(1)- &
<                                  ener_pot(pot, hicut0, &
<                                     0.0_dp, &
<                                     SIZE(bids_store_read), &
<                                     bids_store_read)
---
>               pot%spl_f%cutoff = pot%spl_f%cutoff* &
>                                     pot%spl_f%fscale(1)- &
>                                  ener_pot(pot, hicut0, 0.0_dp)
340,342c302
<             ! Include the cutoff for the read spline
<             e = ener_pot(pot, x, energy_cutoff, &
<                          SIZE(bids_store_read), bids_store_read)
---
>             e = ener_pot(pot, x, energy_cutoff)
411,416d370
<       !> \subbrief Real generation of read spline
<       ! ...................................................
```

```
<       INTEGER                                                 :: &
<          bids_unit_x, bids_unit_y, err
<       REAL(KIND=dp), DIMENSION(:), ALLOCATABLE                :: &
<          bids_xstore
431,441d384
<
<       bids_unit_x = get_unit_number()
<       bids_unit_y = get_unit_number()
<
<       CALL open_file(file_name="spl_x", file_status="UNKNOWN", &
<          file_form="FORMATTED", file_action="WRITE", &
<          file_position="APPEND", unit_number=bids_unit_x)
<       CALL open_file(file_name="spl_y", file_status="UNKNOWN", &
<          file_form="FORMATTED", file_action="WRITE", &
<          file_position="APPEND", unit_number=bids_unit_y)
<
443,448d385
<          IF (ALLOCATED(bids_xstore)) THEN
<             DEALLOCATE (bids_xstore)
<          END IF
<          ALLOCATE (bids_xstore(npoints+1), STAT=err)
<          IF (err/=0) STOP "*** Not enough memory for BIDS ***"
<
456,457c393
<             e = ener_pot(pot, x, energy_cutoff, &
<                          SIZE(bids_store_read), bids_store_read)
---
>             e = ener_pot(pot, x, energy_cutoff)
462d397
<             bids_xstore(jx) = x
466,469d400
<          WRITE (bids_unit_x, *) spline_data%id_nr, &
<                 pot%set(1)%gp%myid, bids_xstore
<          WRITE (bids_unit_y, *) spline_data%id_nr, &
<                 pot%set(1)%gp%myid, spline_data%y
479,480c410
<             e = ener_pot(pot, x, energy_cutoff, &
<                          SIZE(bids_store_read), bids_store_read)
---
>             e = ener_pot(pot, x, energy_cutoff)
505,507d434
<       CALL close_file(unit_number=bids_unit_x)
<       CALL close_file(unit_number=bids_unit_y)
<       DEALLOCATE (bids_xstore)
566,568c493
<                e = ener_pot(pot, x, energy_cutoff, &
```

```
<                   SIZE(bids_store_read), bids_store_read)
<       ! ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
---
>                e = ener_pot(pot, x, energy_cutoff)
1115,1117c1040
<             ! Include the cutoff for read spline
<             pot%spl_f%cutoff = ener_pot(pot, hicut0, 0.0_dp, &
<                SIZE(bids_store_read), bids_store_read)
---
>             pot%spl_f%cutoff = ener_pot(pot, hicut0, 0.0_dp)

"cp2k/src/pair_potential.F"
14,17d13
<    ! Call Boltzmann inversion directed simulation utility module
<    USE bids_util,                       ONLY: bids_read_init, &
<                                               read_splfile, &
<                                               bids_bias
26,27d21
<    ! Call spline methods module for interpolation
<    USE splines_methods,                 ONLY: spline_value
49,50c43
<    FUNCTION ener_pot(pot, r, energy_cutoff, bids_size_read, &
<                      bids_store_read) RESULT(value)
---
>    FUNCTION ener_pot(pot, r, energy_cutoff) RESULT(value)
54,60d46
<       !> \subbrief Evaluates the nonbond potential energy for
<       !            the bias
<       ! ..................................................
<       INTEGER, OPTIONAL, INTENT(IN)                      :: &
<          bids_size_read
<       INTEGER, DIMENSION(:), OPTIONAL, INTENT(IN)        :: &
<          bids_store_read
156,168c142
<             IF (ANY(pot%set(j)%gp%myid == bids_store_read)) THEN
<                IF (bids_read_init /= pot%set(j)%gp%myid) THEN
<                   CALL read_splfile( &
<                      pot%set(j)%gp%potential(5:), &
<                      pot%set(j)%gp%myid)
<                   bids_read_init = pot%set(j)%gp%myid
<                END IF
<                lvalue = spline_value(bids_bias, r)
<             ELSE
<                lvalue = evalf(pot%set(j)%gp%myid, &
<                               pot%set(j)%gp%values)
<             END IF
```

```
<       ! ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
---
>             lvalue = evalf(pot%set(j)%gp%myid, &
>                            pot%set(j)%gp%values)

"cp2k/src/bids_util.F"
!-----------------------------------------------------------------------!
!> \brief Boltzmann inversion directed simulation utility
!-----------------------------------------------------------------------!

MODULE bids_util

   USE cp_files,                        ONLY: close_file, &
                                              get_unit_number, &
                                              open_file
   USE kinds,                           ONLY: dp
   USE splines_methods,                 ONLY: init_spline
   USE splines_types,                   ONLY: spline_data_type
#include "./base/base_uses.f90"

   IMPLICIT NONE
   PRIVATE
   CHARACTER(len=*), PARAMETER, PRIVATE :: moduleN = 'bids_util'
   PUBLIC :: read_splinit, bids_read_init, read_splfile, bids_bias

   INTEGER :: bids_read_init
   TYPE(spline_data_type), POINTER :: bids_bias

CONTAINS

! *****************************************************************
!> \brief Initialize read spline file
! *****************************************************************

   SUBROUTINE read_splinit

      bids_read_init = -1

   END SUBROUTINE read_splinit

! *****************************************************************
!> \brief Read spline file
!> \param spl_filename
! *****************************************************************

   SUBROUTINE read_splfile(spl_filename, myid)
```

```
      CHARACTER(len=*), INTENT(IN)              :: spl_filename
      INTEGER, INTENT(IN)                       :: myid

      CHARACTER(len=*), PARAMETER :: routineN = 'read_splfile', &
         routineP = moduleN//':'//routineN

      REAL(KIND=dp), DIMENSION(:), ALLOCATABLE :: vx
      INTEGER
:: unit_nr, ierr, &
                                                  io, ix, nn

      IF (ALLOCATED(vx)) THEN
         DEALLOCATE (vx)
      END IF

      IF (ASSOCIATED(bids_bias)) THEN
         DEALLOCATE (bids_bias)
      END IF

      ALLOCATE (bids_bias)

      ! null information
      bids_bias%ref_count = 0
      bids_bias%id_nr = 0

      unit_nr = get_unit_number()
      CALL open_file (file_name=TRIM(spl_filename), file_status= &
         "OLD", file_action="READ", unit_number=unit_nr)

      !Determine the number of lines in file
      nn = 0
      DO WHILE (.TRUE.)
         READ (unit_nr, *, iostat=io)
         IF (io /= 0) EXIT
         nn = nn+1
      END DO
      REWIND (unit_nr)
      bids_bias%n = nn

      ALLOCATE (vx(1:nn))
      ALLOCATE (bids_bias%y(1:nn))
      ALLOCATE (bids_bias%y2(1:nn))

      DO ix = 1, nn
         READ (unit_nr, *) vx(ix), bids_bias%y(ix)
```

```
      END DO

      bids_bias%h = vx(2)-vx(1)
      bids_bias%x1 = vx(1)

      ! Calculates y2 and other spline parameters
      CALL init_spline (bids_bias, bids_bias%h)

      CALL close_file (unit_nr)

   END SUBROUTINE read_splfile

END MODULE bids_util
```

Appendix D

CP2K Input File

```
&GLOBAL
  PROJECT H2O-DS_BLYP_D3
  PRINT_LEVEL MEDIUM
  RUN_TYPE MD
&END GLOBAL
&MULTIPLE_FORCE_EVALS
   FORCE_EVAL_ORDER 2 3
&END
&FORCE_EVAL
  METHOD MIXED
  &MIXED
    &PRINT
      &PROGRAM_RUN_INFO MEDIUM
      &END
    &END
    MIXING_TYPE GENMIX
    &GENERIC
      ERROR_LIMIT 1.0E-7
      MIXING_FUNCTION X+Y
      VARIABLES X Y
    &END GENERIC
    GROUP_PARTITION 1 1
  &END
  &SUBSYS
    &CELL
      ABC 14.2 14.2 14.2
    &END CELL
     &COORD
@INCLUDE 'COORD'
     &END COORD
     &VELOCITY
@INCLUDE 'VELOCITY'
```

```
     &END VELOCITY
    &KIND H
      BASIS_SET  DZVP-GTH
      POTENTIAL  GTH-BLYP-q1
    &END KIND
    &KIND O
      BASIS_SET  DZVP-GTH
      POTENTIAL  GTH-BLYP-q6
    &END KIND
  &END SUBSYS
&END FORCE_EVAL
&FORCE_EVAL
  METHOD Quickstep
  &DFT
    BASIS_SET_FILE_NAME  ./GTH_BASIS_SETS
    POTENTIAL_FILE_NAME  ./GTH_POTENTIALS
    CHARGE 0
    &MGRID
      CUTOFF  400
      REL_CUTOFF  50
    &END MGRID
    &POISSON
      PERIODIC XYZ
    &END POISSON
    &PRINT
      &MULLIKEN OFF
      &END
      &HIRSHFELD OFF
      &END
    &END
    &QS
      METHOD GPW
      EPS_DEFAULT 1.0E-10
    &END QS
    &SCF
      SCF_GUESS RESTART
      EPS_SCF 1.0E-7
      MAX_SCF 500
      &OT
        MINIMIZER DIIS
        PRECONDITIONER FULL_KINETIC
      &END
    &END SCF
    &XC
      &XC_FUNCTIONAL BLYP
      &END XC_FUNCTIONAL
```

```
      &VDW_POTENTIAL
        DISPERSION_FUNCTIONAL PAIR_POTENTIAL
        &PAIR_POTENTIAL
          TYPE DFTD3
          PARAMETER_FILE_NAME ./dftd3.dat
          REFERENCE_FUNCTIONAL BLYP
        &END PAIR_POTENTIAL
      &END
    &END XC
  &END DFT
&END FORCE_EVAL
&FORCE_EVAL
  METHOD FIST
  &MM
    &PRINT
      &DIPOLE
      &END
    &END
    &FORCEFIELD
      &NONBONDED
        &GENPOT
          ATOMS O O
          ! Keyword READ specifies the functional form by reading
          ! the potential from a spline file
          FUNCTION READ v
          VARIABLES
          PARAMETERS
          VALUES
          RCUT 7.1
          RMIN 0.
          RMAX 7.1
        &END GENPOT
        &GENPOT
          ATOMS H H
          FUNCTION A*X
          VARIABLES X
          PARAMETERS A
          VALUES  0.
          RCUT 7.1
          RMIN 0.
          RMAX 7.1
        &END GENPOT
        &GENPOT
          ATOMS O H
          FUNCTION A*X
          VARIABLES X
```

```
            PARAMETERS A
            VALUES  0.
            RCUT 7.1
            RMIN 0.
            RMAX 7.1
          &END GENPOT
        &END NONBONDED
        &CHARGE
          ATOM O
          CHARGE 0.0
        &END CHARGE
        &CHARGE
          ATOM H
          CHARGE 0.0
        &END CHARGE
      &END FORCEFIELD
      &POISSON
        &EWALD
          EWALD_TYPE none
        &END EWALD
      &END POISSON
    &END MM
  &END FORCE_EVAL
  &MOTION
    &PRINT
      &TRAJECTORY
        &EACH
          MD 5
        &END
      &END
    &END
    &MD
      ENSEMBLE NVT
      STEPS 400000
      TIMESTEP 0.5
      COMVEL_TOL 1E-12
      TEMPERATURE 298.0
      &THERMOSTAT
        TYPE CSVR
        &CSVR
          TIMECON 10.0
        &END CSVR
      &END
    &END MD
  &END MOTION
```

Appendix E

Copyright Permissions

Creative Commons Legal Code

Attribution-NonCommercial-ShareAlike 4.0 International

Official translations of this license are available in other languages.
Creative Commons Corporation (“Creative Commons”) is not a law firm and does not provide legal services or legal advice. Distribution of Creative Commons public licenses does not create a lawyer-client or other relationship. Creative Commons makes its licenses and related information available on an “as-is” basis. Creative Commons gives no warranties regarding its licenses, any material licensed under their terms and conditions, or any related information. Creative Commons disclaims all liability for damages resulting from their use to the fullest extent possible.

Using Creative Commons Public Licenses

Creative Commons public licenses provide a standard set of terms and conditions that creators and other rights holders may use to share original works of authorship and other material subject to copyright and certain other rights specified in the public license below. The following considerations are for informational purposes only, are not exhaustive, and do not form part of our licenses.

Considerations for licensors: Our public licenses are intended for use by those authorized to give the public permission to use material in ways otherwise restricted by copyright and certain other rights. Our licenses are irrevocable. Licensors should read and understand the terms and conditions of the license they choose before applying it. Licensors should also secure all rights necessary before applying our licenses so that the public can reuse the material as expected. Licensors should clearly mark any material not subject to the license. This includes other CC-licensed material, or material used under an exception or limitation to copyright.

Considerations for the public: By using one of our public licenses, a licensor grants the public permission to use the licensed material under specified terms and conditions. If the licensor’s permission is not necessary for any reason–for example, because of any applicable exception or limitation to copyright–then that use is not regulated by the license. Our licenses grant only permissions under copyright and certain other rights that a licensor has authority to grant. Use of the licensed material may still be restricted for other reasons, including because others have copyright or other rights in the material. A licensor may make special requests, such as asking that all changes be marked or described. Although not required by our licenses, you are encouraged to respect those requests where reasonable.

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License

By exercising the Licensed Rights (defined below), You accept and agree to be bound by the terms and conditions of this Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License ("Public License"). To the extent this Public License may be interpreted as a contract, You are granted the Licensed Rights in consideration of Your acceptance of these terms and conditions, and the Licensor grants You such rights in consideration of benefits the Licensor receives from making the Licensed Material available under these terms and conditions.

Section 1 – Definitions.

1. **Adapted Material** means material subject to Copyright and Similar Rights that is derived from or based upon the Licensed Material and in which the Licensed Material is translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor. For purposes of this Public License, where the Licensed Material is a musical work, performance, or sound recording, Adapted Material is always produced where the Licensed Material is synched in timed relation with a moving image.
2. **Adapter's License** means the license You apply to Your Copyright and Similar Rights in Your contributions to Adapted Material in accordance with the terms and conditions of this Public License.
3. **BY-NC-SA Compatible License** means a license listed at creativecommons.org/compatiblelicenses, approved by Creative Commons as essentially the equivalent of this Public License.
4. **Copyright and Similar Rights** means copyright and/or similar rights closely related to copyright including, without limitation, performance, broadcast, sound recording, and Sui Generis Database Rights, without regard to how the rights are labeled or categorized. For purposes of this Public License, the rights specified in Section 2(b)(1)-(2) are not Copyright and Similar Rights.
5. **Effective Technological Measures** means those measures that, in the absence of proper authority, may not be circumvented under laws fulfilling obligations under Article 11 of the WIPO Copyright Treaty adopted on December 20, 1996, and/or similar international agreements.
6. **Exceptions and Limitations** means fair use, fair dealing, and/or any other exception or limitation to Copyright and Similar Rights that applies to Your use of the Licensed Material.
7. **License Elements** means the license attributes listed in the name of a Creative Commons Public License. The License Elements of this Public License are Attribution, NonCommercial, and ShareAlike.
8. **Licensed Material** means the artistic or literary work, database, or other material to which the Licensor applied this Public License.
9. **Licensed Rights** means the rights granted to You subject to the terms and conditions of this Public License, which are limited to all Copyright and Similar Rights that apply to Your use of the Licensed Material and that the Licensor has authority to license.
10. **Licensor** means the individual(s) or entity(ies) granting rights under this Public License.
11. **NonCommercial** means not primarily intended for or directed towards commercial advantage or monetary compensation. For purposes of this Public License, the exchange of the Licensed Material for other material subject to Copyright and Similar Rights by digital file-sharing or similar means is NonCommercial provided there is no payment of monetary compensation in connection with the exchange.

12. **Share** means to provide material to the public by any means or process that requires permission under the Licensed Rights, such as reproduction, public display, public performance, distribution, dissemination, communication, or importation, and to make material available to the public including in ways that members of the public may access the material from a place and at a time individually chosen by them.
13. **Sui Generis Database Rights** means rights other than copyright resulting from Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, as amended and/or succeeded, as well as other essentially equivalent rights anywhere in the world.
14. **You** means the individual or entity exercising the Licensed Rights under this Public License. **Your** has a corresponding meaning.

Section 2 – Scope.

1. **License grant**.
    1. Subject to the terms and conditions of this Public License, the Licensor hereby grants You a worldwide, royalty-free, non-sublicensable, non-exclusive, irrevocable license to exercise the Licensed Rights in the Licensed Material to:
        1. reproduce and Share the Licensed Material, in whole or in part, for NonCommercial purposes only; and
        2. produce, reproduce, and Share Adapted Material for NonCommercial purposes only.
    2. Exceptions and Limitations. For the avoidance of doubt, where Exceptions and Limitations apply to Your use, this Public License does not apply, and You do not need to comply with its terms and conditions.
    3. Term. The term of this Public License is specified in Section 6(a).
    4. Media and formats; technical modifications allowed. The Licensor authorizes You to exercise the Licensed Rights in all media and formats whether now known or hereafter created, and to make technical modifications necessary to do so. The Licensor waives and/or agrees not to assert any right or authority to forbid You from making technical modifications necessary to exercise the Licensed Rights, including technical modifications necessary to circumvent Effective Technological Measures. For purposes of this Public License, simply making modifications authorized by this Section 2(a)(4) never produces Adapted Material.
    5. Downstream recipients.
        1. Offer from the Licensor – Licensed Material. Every recipient of the Licensed Material automatically receives an offer from the Licensor to exercise the Licensed Rights under the terms and conditions of this Public License.
        2. Additional offer from the Licensor – Adapted Material. Every recipient of Adapted Material from You automatically receives an offer from the Licensor to exercise the Licensed Rights in the Adapted Material under the conditions of the Adapter's License You apply.
        3. No downstream restrictions. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, the Licensed Material if doing so restricts exercise of the Licensed Rights by any recipient of the Licensed Material.
    6. No endorsement. Nothing in this Public License constitutes or may be construed as permission to assert or imply that You are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor or others designated to receive attribution as provided in Section 3(a)(1)(A)(i).
2. **Other rights**.

1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise.
2. Patent and trademark rights are not licensed under this Public License.
3. To the extent possible, the Licensor waives any right to collect royalties from You for the exercise of the Licensed Rights, whether directly or through a collecting society under any voluntary or waivable statutory or compulsory licensing scheme. In all other cases the Licensor expressly reserves any right to collect such royalties, including when the Licensed Material is used other than for NonCommercial purposes.

Section 3 – License Conditions.

Your exercise of the Licensed Rights is expressly made subject to the following conditions.

1. **Attribution**.
    1. If You Share the Licensed Material (including in modified form), You must:
        1. retain the following if it is supplied by the Licensor with the Licensed Material:
            1. identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated);
            2. a copyright notice;
            3. a notice that refers to this Public License;
            4. a notice that refers to the disclaimer of warranties;
            5. a URI or hyperlink to the Licensed Material to the extent reasonably practicable;
        2. indicate if You modified the Licensed Material and retain an indication of any previous modifications; and
        3. indicate the Licensed Material is licensed under this Public License, and include the text of, or the URI or hyperlink to, this Public License.
    2. You may satisfy the conditions in Section 3(a)(1) in any reasonable manner based on the medium, means, and context in which You Share the Licensed Material. For example, it may be reasonable to satisfy the conditions by providing a URI or hyperlink to a resource that includes the required information.
    3. If requested by the Licensor, You must remove any of the information required by Section 3(a)(1)(A) to the extent reasonably practicable.
2. **ShareAlike**.
   In addition to the conditions in Section 3(a), if You Share Adapted Material You produce, the following conditions also apply.
    1. The Adapter's License You apply must be a Creative Commons license with the same License Elements, this version or later, or a BY-NC-SA Compatible License.
    2. You must include the text of, or the URI or hyperlink to, the Adapter's License You apply. You may satisfy this condition in any reasonable manner based on the medium, means, and context in which You Share Adapted Material.
    3. You may not offer or impose any additional or different terms or conditions on, or apply any Effective Technological Measures to, Adapted Material that restrict exercise

of the rights granted under the Adapter's License You apply.

Section 4 – Sui Generis Database Rights.

Where the Licensed Rights include Sui Generis Database Rights that apply to Your use of the Licensed Material:

1. for the avoidance of doubt, Section 2(a)(1) grants You the right to extract, reuse, reproduce, and Share all or a substantial portion of the contents of the database for NonCommercial purposes only;
2. if You include all or a substantial portion of the database contents in a database in which You have Sui Generis Database Rights, then the database in which You have Sui Generis Database Rights (but not its individual contents) is Adapted Material, including for purposes of Section 3(b); and
3. You must comply with the conditions in Section 3(a) if You Share all or a substantial portion of the contents of the database.

For the avoidance of doubt, this Section 4 supplements and does not replace Your obligations under this Public License where the Licensed Rights include other Copyright and Similar Rights.

Section 5 – Disclaimer of Warranties and Limitation of Liability.

1. **Unless otherwise separately undertaken by the Licensor, to the extent possible, the Licensor offers the Licensed Material as-is and as-available, and makes no representations or warranties of any kind concerning the Licensed Material, whether express, implied, statutory, or other. This includes, without limitation, warranties of title, merchantability, fitness for a particular purpose, non-infringement, absence of latent or other defects, accuracy, or the presence or absence of errors, whether or not known or discoverable. Where disclaimers of warranties are not allowed in full or in part, this disclaimer may not apply to You.**
2. **To the extent possible, in no event will the Licensor be liable to You on any legal theory (including, without limitation, negligence) or otherwise for any direct, special, indirect, incidental, consequential, punitive, exemplary, or other losses, costs, expenses, or damages arising out of this Public License or use of the Licensed Material, even if the Licensor has been advised of the possibility of such losses, costs, expenses, or damages. Where a limitation of liability is not allowed in full or in part, this limitation may not apply to You.**
3. The disclaimer of warranties and limitation of liability provided above shall be interpreted in a manner that, to the extent possible, most closely approximates an absolute disclaimer and waiver of all liability.

Section 6 – Term and Termination.

1. This Public License applies for the term of the Copyright and Similar Rights licensed here. However, if You fail to comply with this Public License, then Your rights under this Public License terminate automatically.
2. Where Your right to use the Licensed Material has terminated under Section 6(a), it reinstates:

    1. automatically as of the date the violation is cured, provided it is cured within 30 days of Your discovery of the violation; or

2. upon express reinstatement by the Licensor.

   For the avoidance of doubt, this Section 6(b) does not affect any right the Licensor may have to seek remedies for Your violations of this Public License.

3. For the avoidance of doubt, the Licensor may also offer the Licensed Material under separate terms or conditions or stop distributing the Licensed Material at any time; however, doing so will not terminate this Public License.
4. Sections 1, 5, 6, 7, and 8 survive termination of this Public License.

Section 7 – Other Terms and Conditions.

1. The Licensor shall not be bound by any additional or different terms or conditions communicated by You unless expressly agreed.
2. Any arrangements, understandings, or agreements regarding the Licensed Material not stated herein are separate from and independent of the terms and conditions of this Public License.

Section 8 – Interpretation.

1. For the avoidance of doubt, this Public License does not, and shall not be interpreted to, reduce, limit, restrict, or impose conditions on any use of the Licensed Material that could lawfully be made without permission under this Public License.
2. To the extent possible, if any provision of this Public License is deemed unenforceable, it shall be automatically reformed to the minimum extent necessary to make it enforceable. If the provision cannot be reformed, it shall be severed from this Public License without affecting the enforceability of the remaining terms and conditions.
3. No term or condition of this Public License will be waived and no failure to comply consented to unless expressly agreed to by the Licensor.
4. Nothing in this Public License constitutes or may be interpreted as a limitation upon, or waiver of, any privileges and immunities that apply to the Licensor or You, including from the legal processes of any jurisdiction or authority.

Creative Commons is not a party to its public licenses. Notwithstanding, Creative Commons may elect to apply one of its public licenses to material it publishes and in those instances will be considered the “Licensor.” The text of the Creative Commons public licenses is dedicated to the public domain under the CC0 Public Domain Dedication. Except for the limited purpose of indicating that material is shared under a Creative Commons public license or as otherwise permitted by the Creative Commons policies published at creativecommons.org/policies, Creative Commons does not authorize the use of the trademark “Creative Commons” or any other trademark or logo of Creative Commons without its prior written consent including, without limitation, in connection with any unauthorized modifications to any of its public licenses or any other arrangements, understandings, or agreements concerning use of licensed material. For the avoidance of doubt, this paragraph does not form part of the public licenses.

Creative Commons may be contacted at creativecommons.org.

AIP PUBLISHING LICENSE
TERMS AND CONDITIONS

Jan 11, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and AIP Publishing ("AIP Publishing") consists of your license details and the terms and conditions provided by AIP Publishing and Copyright Clearance Center.

License Number	4505730084283
License date	Jan 11, 2019
Licensed Content Publisher	AIP Publishing
Licensed Content Publication	Journal of Chemical Physics
Licensed Content Title	Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range
Licensed Content Author	Lawrie B. Skinner, Congcong Huang, Daniel Schlesinger, et al
Licensed Content Date	Feb 21, 2013
Licensed Content Volume	138
Licensed Content Issue	7
Type of Use	Thesis/Dissertation
Requestor type	Student
Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	1
Title of your thesis / dissertation	Improving Simulations of Aqueous Systems through Experimental Bias
Expected completion date	Jan 2019
Estimated size (number of pages)	165
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Billing Type	Invoice
Billing Address	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, Australia 6102 Attn: TiongTze Ling (500.1105)
Total	**0.00 AUD**
Terms and Conditions	

AIP Publishing -- Terms and Conditions: Permissions Uses

AIP Publishing hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted from [FULL CITATION], with the permission of AIP Publishing." For an article, the credit line and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP Publishing and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP Publishing, with permission, from another source. Under no circumstances does AIP Publishing purport or intend to grant permission to reuse material to which it does not hold appropriate rights.
   You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes.
3. The foregoing license shall not take effect unless and until AIP Publishing or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
4. AIP Publishing or Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP Publishing, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
5. AIP Publishing makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP Publishing, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
6. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP Publishing. This license may not be amended except in a writing signed by the party to be charged.
7. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

V1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 11, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4505740236075
License date	Jan 11, 2019
Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	Principles of Soil and Plant Water Relations
Licensed Content Author	M.B. Kirkham
Licensed Content Date	Jan 1, 2005
Licensed Content Pages	14
Start Page	27
End Page	40
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier chapter?	No
Will you be translating?	No
Original figure numbers	Figure 3.1
Title of your thesis/dissertation	Improving Simulations of Aqueous Systems through Experimental Bias
Expected completion date	Jan 2019
Estimated size (number of pages)	165
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Publisher Tax ID	GB 494 6272 12
Total	**0.00 AUD**
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions

established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement

between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog

    - by updating a preprint in arXiv or RePEc with the accepted manuscript
    - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
    - directly by providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
    - via non-commercial hosting platforms such as their institutional repository
    - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above:
Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third

AIP PUBLISHING LICENSE
TERMS AND CONDITIONS

Jan 11, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and AIP Publishing ("AIP Publishing") consists of your license details and the terms and conditions provided by AIP Publishing and Copyright Clearance Center.

License Number	4505740438607
License date	Jan 11, 2019
Licensed Content Publisher	AIP Publishing
Licensed Content Publication	Journal of Chemical Physics
Licensed Content Title	A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions
Licensed Content Author	J. D. Bernal, R. H. Fowler
Licensed Content Date	Aug 1, 1933
Licensed Content Volume	1
Licensed Content Issue	8
Type of Use	Thesis/Dissertation
Requestor type	Student
Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	1
Title of your thesis / dissertation	Improving Simulations of Aqueous Systems through Experimental Bias
Expected completion date	Jan 2019
Estimated size (number of pages)	165
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Billing Type	Invoice
Billing Address	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, Australia 6102 Attn: TiongTze Ling (500.1105)
Total	**0.00 AUD**
Terms and Conditions	

AIP Publishing -- Terms and Conditions: Permissions Uses

AIP Publishing hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted from [FULL CITATION], with the permission of AIP Publishing." For an article, the credit line and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP Publishing and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP Publishing, with permission, from another source. Under no circumstances does AIP Publishing purport or intend to grant permission to reuse material to which it does not hold appropriate rights.
   You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes.
3. The foregoing license shall not take effect unless and until AIP Publishing or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
4. AIP Publishing or Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP Publishing, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
5. AIP Publishing makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP Publishing, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
6. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP Publishing. This license may not be amended except in a writing signed by the party to be charged.
7. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

V1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 17, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4511180130523
License date	Jan 17, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	The Concept of the Potential Energy Surface
Licensed Content Author	Errol G. Lewars
Licensed Content Date	Jan 1, 2016
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Improving Simulations of Aqueous Systems through Experimental Bias
Institution name	n/a
Expected presentation date	Jan 2019
Portions	Figure 2.6
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Billing Type	Invoice
Billing Address	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, Australia 6102 Attn: TiongTze Ling (500.1105)
Total	**0.00 AUD**
Terms and Conditions	

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Nature Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and

requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/ book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION**

(Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), **[COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.1

ELSEVIER LICENSE
TERMS AND CONDITIONS

Jan 17, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4511210218097
License date	Jan 17, 2019
Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	Theory of Simple Liquids
Licensed Content Author	Jean-Pierre Hansen,Ian R. McDonald
Licensed Content Date	Jan 1, 2013
Licensed Content Pages	43
Start Page	105
End Page	147
Type of Use	reuse in a thesis/dissertation
Intended publisher of new work	other
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier chapter?	No
Will you be translating?	No
Original figure numbers	Equation 4.6.1
Title of your thesis/dissertation	Improving Simulations of Aqueous Systems through Experimental Bias
Expected completion date	Jan 2019
Estimated size (number of pages)	165
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Publisher Tax ID	GB 494 6272 12
Total	**0.00 AUD**
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions

established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at http://myaccount.copyright.com).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:
"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."
4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.
5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.
6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.
7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.
10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.
11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.
12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).
13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement

between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.
14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:
15. **Translation**: This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.
16. **Posting licensed content on any Website**: The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.
Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at http://www.elsevier.com . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website posting.
17. **For journal authors:** the following clauses are applicable in addition to the above:
Preprints:
A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).
Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).
If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.
Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications.
Authors can share their accepted author manuscript:

- immediately
    - via their non-commercial person homepage or blog

    - by updating a preprint in arXiv or RePEc with the accepted manuscript
    - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
    - directly by providing copies to their students or to research collaborators for their personal use
    - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period
    - via non-commercial hosting platforms such as their institutional repository
    - via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.
Policies for sharing publishing journal articles differ for subscription and gold open access articles:
Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.
Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.
If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.
Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.
Please refer to Elsevier's posting policy for further information.
18. **For book authors** the following clauses are applicable in addition to the above:
Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. **Posting to a repository:** Authors are permitted to post a summary of their chapter only in their institution's repository.
19. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions
You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third

AIP PUBLISHING LICENSE
TERMS AND CONDITIONS

Jan 25, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and AIP Publishing ("AIP Publishing") consists of your license details and the terms and conditions provided by AIP Publishing and Copyright Clearance Center.

License Number	4515750442963
License date	Jan 25, 2019
Licensed Content Publisher	AIP Publishing
Licensed Content Publication	Journal of Chemical Physics
Licensed Content Title	Derivation of coarse-grained potentials via multistate iterative Boltzmann inversion
Licensed Content Author	Timothy C. Moore, Christopher R. Iacovella, Clare McCabe
Licensed Content Date	Jun 14, 2014
Licensed Content Volume	140
Licensed Content Issue	22
Type of Use	Thesis/Dissertation
Requestor type	Student
Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	1
Title of your thesis / dissertation	Improving Simulations of Aqueous Systems through Experimental Bias
Expected completion date	Jan 2019
Estimated size (number of pages)	165
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Billing Type	Invoice
Billing Address	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, Australia 6102 Attn: TiongTze Ling (500.1105)
Total	**0.00 AUD**
Terms and Conditions	

AIP Publishing -- Terms and Conditions: Permissions Uses

AIP Publishing hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted from [FULL CITATION], with the permission of AIP Publishing." For an article, the credit line and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP Publishing and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP Publishing, with permission, from another source. Under no circumstances does AIP Publishing purport or intend to grant permission to reuse material to which it does not hold appropriate rights.
   You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes.
3. The foregoing license shall not take effect unless and until AIP Publishing or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
4. AIP Publishing or Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP Publishing, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
5. AIP Publishing makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP Publishing, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
6. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP Publishing. This license may not be amended except in a writing signed by the party to be charged.
7. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

V1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 25, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4515751511700
License date	Jan 25, 2019
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Development of a Coarse-Grained Water Forcefield via Multistate Iterative Boltzmann Inversion
Licensed Content Author	Timothy C. Moore, Christopher R. Iacovella, Clare McCabe
Licensed Content Date	Jan 1, 2016
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Improving Simulations of Aqueous Systems through Experimental Bias
Institution name	n/a
Expected presentation date	Jan 2019
Portions	Figure 2
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Billing Type	Invoice
Billing Address	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, Australia 6102 Attn: TiongTze Ling (500.1105)
Total	**0.00 AUD**
Terms and Conditions	

Springer Nature Terms and Conditions for RightsLink Permissions

Springer Nature Customer Service Centre GmbH (the Licensor) hereby grants you a non-exclusive, world-wide licence to reproduce the material and for the purpose and requirements specified in the attached copy of your order form, and for no other use, subject to the conditions below:

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), **[COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.1

AIP PUBLISHING LICENSE
TERMS AND CONDITIONS

Jan 25, 2019

This Agreement between Curtin University -- Tiong Tze Ling ("You") and AIP Publishing ("AIP Publishing") consists of your license details and the terms and conditions provided by AIP Publishing and Copyright Clearance Center.

License Number	4515760590862
License date	Jan 25, 2019
Licensed Content Publisher	AIP Publishing
Licensed Content Publication	Journal of Chemical Physics
Licensed Content Title	Communication: Improved ab initio molecular dynamics by minimally biasing with experimental data
Licensed Content Author	Andrew D. White, Chris Knight, Glen M. Hocky, et al
Licensed Content Date	Jan 28, 2017
Licensed Content Volume	146
Licensed Content Issue	4
Type of Use	Thesis/Dissertation
Requestor type	Student
Format	Print and electronic
Portion	Figure/Table
Number of figures/tables	1
Title of your thesis / dissertation	Improving Simulations of Aqueous Systems through Experimental Bias
Expected completion date	Jan 2019
Estimated size (number of pages)	165
Requestor Location	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, WA 6102 Australia Attn: TiongTze Ling (500.1105)
Billing Type	Invoice
Billing Address	Curtin University Level 2 Reception, Building 500 Kent Street Bentley Perth, Australia 6102 Attn: TiongTze Ling (500.1105)
Total	**0.00 AUD**
Terms and Conditions	

AIP Publishing -- Terms and Conditions: Permissions Uses

AIP Publishing hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted from [FULL CITATION], with the permission of AIP Publishing." For an article, the credit line and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP Publishing and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP Publishing, with permission, from another source. Under no circumstances does AIP Publishing purport or intend to grant permission to reuse material to which it does not hold appropriate rights.
   You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes.
3. The foregoing license shall not take effect unless and until AIP Publishing or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
4. AIP Publishing or Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP Publishing, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
5. AIP Publishing makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP Publishing, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
6. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP Publishing. This license may not be amended except in a writing signed by the party to be charged.
7. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

V1.2

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.