universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Tracing the solid-liquid coexistence line of water based on an ab initio potential“

verfasst von / submitted by

Nils Clees

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2018 / Vienna, 2018

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 876
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Physik
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Christoph Dellago

Acknowledgments

First and foremost, I would like to thank Prof. Christoph Dellago for supervising this thesis and guiding me through my academic path. The opportunity to work in the research group of computational physics was an enriching experience and I want to thank everyone for the interesting discussions and insightful meetings.
A special thanks goes to Clemens Moritz, who was always there to share his knowledge and experience. Next, I want to thank Andreas Singraber for his patience and dedication when explaining technical details. Also, I would like to thank Gyorgy Hantal for giving me feedback on my thesis.
Furthermore, I would like to thank my fellow master students, Jakob Michl, Matthias Kiss and Igor Kraut, as they contributed to a great working environment.
Moreover, I want to thank all my friends, who supported me during my Master's thesis and ensured that I enjoyed my time in Vienna.
Overall, many people shaped my academic life and thus helped me to achieve this degree. A special thanks, however, goes to Prof. Tanja Schilling, who guided me during my Bachelor's thesis and inspired me to pursue a degree of Master of Science.
Last but not least, I am very grateful for the love and support my family has given me, especially my parents who are always there for me.

Abstract

The goal of the Master's Thesis is to trace the coexistence line between solid ice Ih and liquid water from molecular dynamics simulation. This is achieved by integrating the Clausius-Clapeyron equation, starting from a coexistence point which was previously obtained from the Interface Pinning method [1]. The procedure that is used, known as Gibbs-Duhem integration [2],[3], requires the evaluation of the enthalpy and the density difference between the solid and the liquid phase. In order to obtain these two observables, molecular dynamics simulations of water are performed using a Neural Network Potential (NNP). This NNP has been trained to the RPBE and BLYP density functionals with and without Van der Waals corrections. The advantage of using a NNP is that one obtains an accurate ab initio potential energy surface of water without facing high computational costs. Previous research projects have shown that Van der Waals interactions determine some unique properties of water, i.e. the density peak of the liquid phase [1]. Thus it is very interesting to see how the inclusion of them will influence the phase boundary of ice and liquid water.

Zusammenfassung

Das Ziel dieser Masterarbeit ist die Koexistenz-Linie zwischen Eis Ih und Wasser durch Molekular Dynamik Simulationen zu bestimmen. Dies gelingt durch die Integration der Clausius-Clapeyron Gleichung an einem Koexistenz-Punkt, welcher bereits durch die "Interface Pinning" Methode lokalisiert wurde. Die angewandte Prozedur, bekannt als Gibbs-Duhem Integration [2],[3], benötigt die Berechnung der Enthalpie und Dichte Differenz des festen und flüssigen Zustandes. Um diese zwei Observablen zu erhalten werden Molekular Dynamik Simulationen mit einem Neuronalen Netz Potential (NNP) ausgeführt. Dieses NNP wurde trainiert anhand des RPBE und BLYP Dichte Funktionals mit und ohne Van der Waals Korrektur. Der Vorteil eines NNP liegt darin, dass man die genaue ab-initio potentielle Energie-Oberfläche von Wasser durch geringen Rechenaufwand bestimmen kann. Vorherige Forschungsarbeiten haben erwiesen, dass die Van der Waals Interaktionen eine entscheidende Rolle in der Bestimmung der einzigartigen Eigenschaften von Wasser haben, i.e. die Wiedergabe der Dichteanomalie des Wassers [1]. Folglich ist es sehr interessant herauszufinden, wie die Einrechnung dieser Interaktionen die Phasengrenze zwischen Eis und Wasser beeinflusst.

Contents

Water is the driving force of all nature.

Leonardo da Vinci

1 Introduction

A small drop of water is a fascinating creation of nature. Accumulated it forms streams, rivers and oceans, covering two-thirds of the earth. Biologists argue that all life on this planet has emerged from water, marking its relevance as a medium where biological processes occur. Furthermore, water is crucial for the climate and geology of the earth and is also indispensable for society's domestic and industrial processes. The presence of water on earth is thus fundamentally essential. Interestingly it is also the only substance which exists in all of its three states on Earth.

How good is the understanding of this abundant substance? Intensive studies showed many anomalies caused by this simple molecule. The most popular being the positive volume of freezing and the density peak of liquid water at 4°C, which explain why ice floats [4]. Another intriguing property of water is its rich phase diagram, containing many crystalline forms and three amorphous ices [5]. The only solid form existing on Earth, however, is low-density hexagonal ice, whose phase boundary to liquid water displays yet another curious behaviour. Specifically, the slope of the liquid-solid coexistence line being negative, such that the melting point decreases with increasing pressure [6].

The understanding of water has been extended beyond experimental knowledge, with the help of computer simulation techniques [7],[8]. The use of simple water models allowed an investigation of many properties of water during the last decades [9],[10],[11],[12],[13]. Despite their success, empirical water potentials lack transferability, because of their parametrization to experimental data. As an alternative, ab-initio molecular dynamics has been developed, where the forces governing the interactions between atoms are directly obtained from quantum theory [14]. The combination of density functional theory and molecular dynamics has thus been able to accurately describe some properties of water based on first principles [15],[1]. This method, however, is computationally expensive, leading to the development of a neural network potential [16],[17]. This analytical function is able to accurately represent the ab-initio potential energy surface of water, making it possible to perform large scale molecular dynamics simulations at a fraction of the previous costs.

Hence, this thesis investigates the impact of van der Waals interactions on the thermodynamic anomalies of water. Previous studies already discovered that interactions in liquid water are governed by a fine balance between strong directional Hydrogen bonds and weak non-directional van der Waals forces [1]. This research will focus on the coexistence line between ice Ih and liquid water. The phase boundary will be traced by using the Gibbs-Duhem integration method [2],[3] by performing molecular dynamics simulations based on the RPBE and BLYP density functional with and without dispersion corrections [18]. Ultimately the thermodynamic description delivered by the phase envelope will be a severe test for the underlying density functionals to describe water at different pressures and temperatures.

The structure of this thesis is built up as follows. In the first chapter, the thesis will review some basic concepts of molecular dynamics simulation. The next chapter will highlight the Gibbs-Duhem integration method, which will be used to trace the melting line of ice Ih. Then a discussion on existing water models will follow, specifically the results of the phase diagram for rigid non-polarizable models. Thereafter the thesis will give an insight in the theory of ab-initio molecular dynamics and will elaborate on how van der Waals interactions are explicitly incorporated. Lastly, the thesis will outline how the neural network is used to evaluate the energies and forces of the simulations. The results of this work will then be presented and discussed. Finally, some concluding remarks will summarise our findings and give an outlook on future research.

2 Molecular Dynamics

The use of computer simulation to solve many-body problems that are too complex to determine an analytic solution, began in 1953. Back then, the Monte Carlo technique was applied to obtain the equation of state for rigid spheres [7]. This success led to the development of the method of Molecular Dynamics (MD). In MD simulations one studies the dynamics of a classical many-body system and determines its static and transport properties.
The procedure of a MD simulation is quite similar to the one followed in real experiments in the laboratory. First, the system is prepared in an initial state, which specifies the initial positions and momenta of all particles in the system. Next, one follows approximately the natural time evolution of this state by solving Newton's equations of motion. It is important to know that this is a classical approximation. Since the hydrogen atoms within the water molecule are very light, quantum effects will be relevant. So the true trajectory should be obtained by solving the Schrödinger equation. The computational costs, however, would drastically increase, so one commonly uses the classical approximation. Nevertheless, the quantum nature of nuclear motion should be taken into account, especially when studying rare events, like transient autoprotolysis. This has been achieved by Ceriotti et al. [19], who have shown that the quantum treatment of the motion of the hydrogen atom has a qualitative impact on the hydrogen bond fluctuations in water.
In MD simulations, the forces between the particles are typically calculated using an empirical interatomic potential. In our research, however, a neural network based on ab-initio calculations is used to determine the energy of each atom in the system. The system will thus evolve in time, and while it is doing so one measures the quantity of interest. The result will be given as a statistical average, where the error gets smaller the longer one simulates. If the system under study satisfies the ergodic hypothesis, macroscopic thermodynamic observables can be determined from these averages.

MD simulations were first successfully applied by Alder and Wainwright in 1956 to study collisions between hard disks [8]. Later, in 1964, Rahman was the first to use the Lennard-Jones potential to simulate liquid Argon and evaluate its coefficient of self-diffusion [20]. The first MD simulation of liquid water was achieved by Rahman and Stillinger in 1971 [9]. They used a rigid water model subject to an effective pair potential, that was able to mimic hydrogen bonding. Using this model they were able to examine static structural properties and kinetic behaviour of liquid water. The technique of MD has since then been applied to study protein folding, and many other chemical and biological processes.

This thesis shall give a review of some of the basic concepts used in MD. First Newton's equations of motion are discussed, followed by a numerical integration method to solve them. Next the technique of molecular dynamics in the isothermal-isobaric ensemble is outlined. Finally two methods to determine structural and dynamical properties of the simulated system are discussed.

2.1 Equations of motion

The goal of a MD simulation is to predict thermodynamic properties of an N-particle system based on the knowledge of the positions and momenta of the particles in the system. In order to measure a macroscopic observable, like temperature, volume, pressure, one considers the microstates that are compatible with the given macrostate. A microstate is defined as a snapshot of the system consisting of all the particles' coordinates and momenta at a given time t:

$$\mathbf{r}^N(t) = \{\mathbf{r}_1(t), \mathbf{r}_2(t), ..., \mathbf{r}_N(t)\}$$

$$\mathbf{p}^N(t) = \{\mathbf{p}_1(t), \mathbf{p}_2(t), ..., \mathbf{p}_N(t)\}$$

It is useful to describe the microstate by its potential energy U and its kinetic energy K:

$$U(\mathbf{r}^N) = U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) \quad (1)$$

$$K(\mathbf{p}^N) = \sum_{i=1}^{N} \frac{\mathbf{p}_i}{2m_i} \quad (2)$$

The idea of molecular dynamics is then to solve the laws of motion of classical mechanics, and measure the quantity of interest while the system evolves in time. In Newtonian form, the equation of motion of a single particle i reads:

$$m_i \ddot{\mathbf{r}}_i(t) = \dot{\mathbf{p}}_i(t) = \mathbf{F}_i(\mathbf{r}^N(t)) \quad (3)$$

where the dot notation is used for the derivative with respect to time $\frac{d}{dt}$. The force acting on particle i is obtained from the negative gradient of the potential energy:

$$\mathbf{F}_i(\mathbf{r}^N(t)) = -\nabla_i U(\mathbf{r}^N(t)) = -\begin{pmatrix} \frac{\partial U(\mathbf{r}^N)}{\partial x_i} \\ \frac{\partial U(\mathbf{r}^N)}{\partial y_i} \\ \frac{\partial U(\mathbf{r}^N)}{\partial z_i} \end{pmatrix} \quad (4)$$

Thus Newton's equation of motion for particle i reads:

$$\dot{\mathbf{r}}_i(t) = \frac{\mathbf{p}_i(t)}{m_i} \quad (5)$$

$$\dot{\mathbf{p}}_i(t) = -\nabla_i U(\mathbf{r}^N(t)) \quad (6)$$

In classical mechanics there are different ways to express the equations of motion. One can use, for instance, the fundamental Hamiltonian form to describe the dynamics of the system:

$$\dot{\mathbf{r}}_i(t) = \frac{\partial H}{\partial \mathbf{p}_i} \quad (7)$$

$$\dot{\mathbf{p}}_i(t) = -\frac{\partial H}{\partial \mathbf{r}_i} \quad (8)$$

where the Hamiltonian is given by the total energy of the system:

$$H(\mathbf{r}^N, \mathbf{p}^N) = K(\mathbf{p}^N) + U(\mathbf{r}^N) \quad (9)$$

For a system of N particles, equations (7) and (8) become a set of $6N$ coupled ordinary differential equations. So for a given initial condition of a microstate $\{\mathbf{r}^N(0), \mathbf{p}^N(0)\}$ one can solve this set of equations numerically using discrete timesteps. Thus, one obtains a time series of microstates, which corresponds to a trajectory in phase space $\{\mathbf{r}^N(t), \mathbf{p}^N(t)\}$. The quantities of interest can be computed along this trajectory and output as statistical average:

$$\bar{A} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau dt A(\mathbf{r}^N(t), \mathbf{p}^N(t)) \tag{10}$$

Here τ is the simulation time, which is of course finite in a computer simulation, but must be chosen long enough for the system to explore phase space. The system is called ergodic if the dynamics is such that any possible configuration can be reached in a finite time. If the ergodicity hyptothesis holds, then the time average of an observable $A(\mathbf{r}^N, \mathbf{p}^N)$ matches the ensemble average:

$$\bar{A} = \langle A \rangle \tag{11}$$

where the ensemble average is defined by the probability density $f(\mathbf{r}^N, \mathbf{p}^N)$ to observe a particular microscopic state:

$$\langle A \rangle = \frac{\int d\mathbf{r}^N d\mathbf{p}^N A(\mathbf{r}^N, \mathbf{p}^N) f(\mathbf{r}^N, \mathbf{p}^N)}{\int d\mathbf{r}^N d\mathbf{p}^N f(\mathbf{r}^N, \mathbf{p}^N)} \tag{12}$$

Newton's equations of motion have a few important properties. First the time evolution of the system is deterministic, i.e. for a given initial microstate the trajectory will always be the same. In practice, however, for technical reasons, one never obtains the same trajectory. Next the total energy and momentum of the system are conserved. Since the MD simulations use periodic boundary conditions, total angular momentum is not conserved. In addition the equations of motion are time reversible, i.e. if one reverses the sign of the momenta of all particles at a given time t, the system will evolve backwards on its trajectory in phase space. Finally Liouville's Theorem holds for these equations. This means that phase space volumes will be conserved along the trajectories following the evolution of the system:

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \sum_{i=1}^{N} \frac{\partial \rho}{\partial \mathbf{r}_i} \dot{\mathbf{r}}_i + \frac{\partial \rho}{\partial \mathbf{p}_i} \dot{\mathbf{p}}_i = 0 \tag{13}$$

where $\rho = \rho(\mathbf{r}^N, \mathbf{p}^N)$ denotes the phase space density.

2.2 Integration Methods

In computer simulations, Newton's equations of motion can only be integrated by using a finite timestep Δt. The numerical method commonly employed is the velocity Verlet algorithm [21]. This algorithm is easily derived by taking the Taylor expansion up to the 3^{rd} order of the coordinates of a single particle i:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\Delta t + \frac{\ddot{\mathbf{r}}(t)}{2!}\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{14}$$

One can easily identify physical quantities in this equation, like the velocity $\mathbf{v}(t) = \dot{\mathbf{r}}(t)$ and the force $\mathbf{F}(t) = m\ddot{\mathbf{r}}(t)$ acting on the particle of mass m:

$$\boxed{\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m}\Delta t^2 + \mathcal{O}(\Delta t^3)} \tag{15}$$

The last equation yields the new position of the particle upon taking a time step Δt, with an error of order $\mathcal{O}(\Delta t^3)$. One can use the time reversibility of equation (15) to write the position of particle i after reversing its momentum:

$$\mathbf{r}(t) = \mathbf{r}(t+\Delta t) - \mathbf{v}(t+\Delta t)\,\Delta t + \frac{\mathbf{F}(t+\Delta t)}{2m}\Delta t^2 \tag{16}$$

Insert equation (16) back into (15) to get:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t+\Delta t) - \mathbf{v}(t+\Delta t)\,\Delta t + \frac{\mathbf{F}(t+\Delta t)}{2m}\Delta t^2 + \mathbf{v}(t)\Delta t + \frac{\mathbf{F}(t)}{2m}\Delta t^2 \tag{17}$$

So the velocities after each time step are obtained:

$$\boxed{\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t) + \mathbf{F}(t+\Delta t)}{2m}\Delta t} \tag{18}$$

The boxed equations above represent the velocity Verlet algorithm [21], which has the advantage that both position and velocity, and thus potential and kinetic energy, of the new time step can be known at the same time.
The velocity Verlet algorithm conserves phase space volumes and is thereby also highly accurately energy conserving. Furthermore it is self-starting, unlike other integration algorithms which need a time step from the past $t - \Delta t$. As can be seen in the velocity Verlet algorithm, the forces acting on the particle need to be evaluated in order to obtain the position and velocity after a time step Δt. The force acting on particle i is given as the negative gradient with respect to the position of particle i of the total potential energy of the system as shown in equation (4). So one needs to know the total potential energy of the system, which is associated to all the atoms positions relative to each other. In many cases empirical potentials have been used, we shall briefly discuss them in section 4. A more fundamental way to evaluate $U(\mathbf{r}^N)$ is by solving the Schrödinger equation for the ground state of the electrons of each atom. This is exactly what is done in density functional theory [22], which will be elaborated on in section 5.1.

2.3 NpT Ensemble

The velocity Verlet algorithm yields the trajectory of the system in phase space while conserving the total energy E with high accuracy. Since the system of N particles is simulated in a box of volume V, we reproduce the microcanonical ensemble NVE. Thus time averages calculated during the simulation will correspond to NVE ensemble averages, given that the dynamics of the system are ergodic. In experiments, however, it is often more practical to consider the system at constant temperature and pressure. Especially in our study of the melting line of ice Ih the isothermal-isobaric ensemble NpT is much more convenient to perform the simulation. The question we will answer in this section is how one simulates this ensemble in computer simulations.

Thermostat In order to control the temperature of the system one needs to introduce a thermostat. There are several mechanisms for thermostatting, like for example the Andersen thermostat [23] that uses stochastic steps along the trajectory of the particles to reassign their velocity and thus sample the desired temperature. The relation between the kinetic energy K and temperature T is given by the equipartition theorem:

$$K(\mathbf{p}^N) = \frac{3N}{2} k_B T \tag{19}$$

Here N is the total number of particles, and k_B is the Boltzmann constant. Although this thermostat is often used, it causes decorrelations of the particle velocities. When one is interested in transport properties like for example the diffusion coefficient one shouldn't use this thermostat as it disturbs the dynamics.
The thermostat we will use in our simulations is the Nosé-Hoover thermostat, which was originally developed by Nosé [24] and later improved by Hoover [25]. Instead of relying on stochastic collisions, it modifies the equations of motion such that they continuously sample the canonical NVT ensemble. This is done by introducing additional variables which extend the physical Hamiltonian of the system to get the Nosé-Hoover Hamiltonian [25]:

$$H_{NH}(r^N, p^N, \xi, s) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r^N}) + \frac{\xi^2 Q}{2} + 3Nk_B T \ln(s) \tag{20}$$

The first two terms represent the kinetic and potential energy of the physical Hamiltonian. The variable $\xi = \frac{sp_s}{Q}$ was introduced to modify the equations of motion and thereby rescale the velocities to sample the desired temperature. It's constituents s and p_s are so called thermostat variables. The parameter Q, often called thermostat mass, determines how quickly the thermostat reacts to get the desired temperature T. The microcanonical distribution sampled by the extended Hamiltonian H_{NH} will correspond to a canonical distribution in physical phase space.

The equations of motion generated by equation (20) are:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} \tag{21}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i - \xi\mathbf{p}_i \tag{22}$$

$$\frac{d\xi}{dt} = \frac{1}{Q}\Big(\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i} - 3Nk_BT\Big) \tag{23}$$

$$\frac{d\ln(s)}{dt} = \xi \tag{24}$$

A closer look at equation (23), shows how the instantaneous kinetic energy is compared to the desired temperature. Depending on the sign of the resulting difference between the two terms on the right, the slope of ξ will be positive or negative. This in turn will increase or decrease the value of ξ in equation (22) and adjust the velocity of the particles. If the instantaneous kinetic energy is higher or lower than the target temperature, the velocities decrease or increase respectively. It is important to note that it is the time derivative of the ξ-variable which reacts to deviations from the desired temperature. One needs to integrate equation (23) to get the value of ξ. This mechanism is called integral feedback.
It has been shown that in some cases the equations of motion outlined in (21) - (24) fail to reproduce ergodic dynamics [26]. In order to help sample the canonical distribution, Martyna et al. proposed the use of so-called Nosé-Hoover chains [27]. In this method the equations of motion are complemented by a set of variables $\xi_1, \xi_2, ...\xi_M$:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} \tag{25}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i - \xi_1\mathbf{p}_i \tag{26}$$

$$\frac{d\xi_1}{dt} = \frac{1}{Q_1}\Big(\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i} - 3Nk_BT\Big) - \xi_2\xi_1 \tag{27}$$

$$\vdots$$

$$\frac{d\xi_k}{dt} = \frac{1}{Q_k}\big(\xi_{k-1}^2Q_{k-1} - k_BT\big) - \xi_{k+1}\xi_k \tag{28}$$

$$\vdots$$

$$\frac{d\xi_M}{dt} = \frac{1}{Q_M}\big(\xi_{M-1}^2Q_{M-1} - k_BT\big) \tag{29}$$

$$\frac{d\ln(s)}{dt} = \xi_k \qquad k = 1...M \tag{30}$$

One can see how each thermostat variable ξ_k is forced by the next one ξ_{k+1} to follow the canonical distribution.

Barostat In order to control the pressure of the system one needs to introduce a barostat. The method follows the procedure of Andersen to sample the NpH ensemble [23]. In this technique the volume is introduced as dynamical variable V, it's conjugated momentum being p_V. One can write the extended Hamiltonian:

$$H_A(r^N, p^N, V, p_V) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r^N}) + \frac{p_V^2}{2W} + pV \tag{31}$$

Here the parameter W, which can be viewed as the mass of a piston driving the volume fluctuations, determines the rate of reaction of the barostat. The corresponding equations of motion are given by:

$$\frac{d\mathbf{r}_i}{dt} = \frac{\mathbf{p}_i}{m_i} + \frac{\dot{V}}{3V}\mathbf{r}_i \tag{32}$$

$$\frac{d\mathbf{p}_i}{dt} = \mathbf{F}_i - \frac{\dot{V}}{3V}\mathbf{p}_i \tag{33}$$

$$\frac{dV}{dt} = \frac{p_V}{W} \tag{34}$$

$$\frac{dp_V}{dt} = \frac{1}{3V}\sum_{i=1}^{N}\left(\frac{\mathbf{p}_i^2}{m_i} + \mathbf{F}_i \cdot \mathbf{r}_i\right) - p \tag{35}$$

In the first two equations (32) and (33), the additional terms on the right hand side, adapt positions and momenta of the particles due to the scaling of the simulation box. The two terms on the right hand side of equation (35) are the instantaneous and target pressure, respectively. If the instantaneous pressure of the system is higher or lower than the target pressure, the slope of p_V will be positive or negative. The value of p_V will thus increase or decrease the volume V (see equation (34)), which will in turn decrease or increase the pressure. Just like the Nosé-Hoover thermostat, these equations of motion correspond to an integral feedback mechanism that reacts with some delay. The time scale of the motion of the volume is determined by the value of W, which acts as a kind of "mass". As in the case of Nosé Hoover chains, one can also use chains coupled to the barostat.
It is important to note that the barostat discussed so far is coupled to the overall box volume, thus controlling the pressure through isotropic scaling. In the case of structural transformations in solids under external stress it is favorable to use a flexible simulation box. Under those circumstances, the barostat is coupled to the individual dimensions x, y, z and the xy, yz, xz tilt dimensions. The method of Parrinello and Rahman [28] does exactly this by dynamically changing the vectors spanning the simulation box. Their technique uses a Lagrangian that is extended with an additional strain energy, in order to control the time evolution of these vectors.

Isothermal-Isobaric In order to simulate the NpT ensemble the previously discussed barostat equations need to be coupled to the Nosé-Hoover thermostat equations. This is exactly what is done in the molecular dynamics program LAMMPS [29]. In LAMMPS simulations, the NpT ensemble is sampled by using the equations of motion proposed by Shinoda et al. [30]. These equations combine the hydrostatic equations introduced by Martyna et al. [31], with the strain energy proposed by Parrinello and Rahman [28]. An important feature of these equations is that they use two different chains for the thermostat and the barostat. Each chain can thus be optimized to the time scale corresponding to the temperature and volume fluctuations respectively.
The time integration schemes used in LAMMPS simulations, closely follow the time-reversible measure-preserving Verlet and rRESPA integrators derived by Tuckerman et al. [32]. These efficient integration algorithms, derived using the Liouville formalism, are able to closely conserve the phase space volume. The pressure and temperature will be fluctuating around the target values p and T.

2.4 Radial Distribution Function

In order to study the structural properties of the system in a MD simulation, one often considers the radial distribution function $g(r)$ (also called pair correlation function). This function describes the average local density at a distance r around a reference particle i. Formally, it is defined by the deviation of the number of particles, inside a radius r, from a homogenous distribution (ideal gas):

$$g(r) = \frac{1}{4\pi\rho r^2 N} \langle \sum_{j(\neq i)} \delta(r - r_{ij}) \rangle \tag{36}$$

where ρ is the average number density of particles, N is the total number of particles in the system, δ is the delta function and r_{ij} is the distance between the reference particle i and another particle j. The radial distribution function is thus able to characterize the local structure around an atom by counting its neighbours up to a given cutoff radius. Experimentally, $g(r)$ can be determined from x-ray or neutron scattering experiments [33]. For systems governed by additive pair interactions, thermodynamic properties are completely determined by $g(r)$.
The typical behaviour of $g(r)$ for a solid, liquid and gaseous structure can be seen in figure 1. One can see that up to $r/\sigma = 1$ the function is zero, because no two particles of radius σ can occupy the same position. The first peak represents the shell of nearest neighbours. While there are typically a few lower order peaks for the liquid phase, the solid phase presents more peaks depending on the particular structure of the crystal. For a dilute gas there is one peak, corresponding to the minimum of the pair potential. In the limit of large r the structural correlations between particles vanish, leading to a homogenous distribution:

$$\lim_{r\to\infty} g(r) = 1 \tag{37}$$

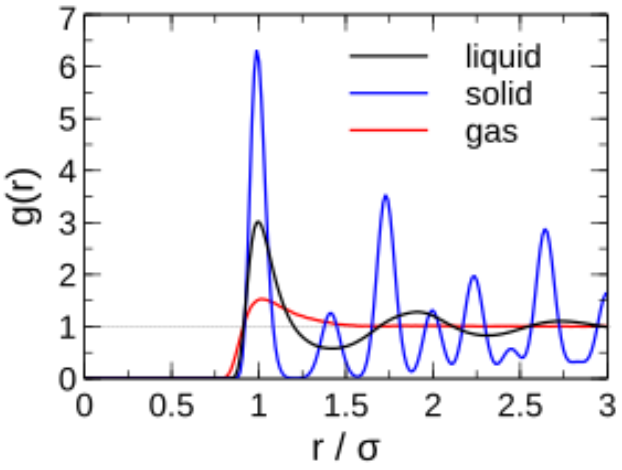


Figure 1: The radial distribution functions of solid (T =50 K), liquid (T =80 K), and gaseous argon (T =300 K). The radii are given in reduced units of the molecular diameter ($\sigma = 3.822$ Å). (from Wikimedia [34])

2.5 Mean Squared Displacement

The dynamical properties of a substance can easily be determined from molecular dynamics simulation. Often one is interested in the diffusion coefficient, which can be obtained from the mean squared displacement (MSD) of the particles in the system. In general, the evolution of the concentration $c(\mathbf{r}, t)$ of particles diffusing in a liquid is governed by the diffusion equation:

$$\frac{\partial c(\mathbf{r}, t)}{\partial t} - D\nabla^2 c(\mathbf{r}, t) = 0 \tag{38}$$

The solution to this equation with initial starting condition $c(\mathbf{r}, 0) = \delta(\mathbf{r})$ is:

$$c(\mathbf{r}, t) = \frac{1}{(4\pi Dt)^{3/2}} \exp\left(-\frac{\mathbf{r}^2}{4Dt}\right) \tag{39}$$

To obtain the relation between diffusion constant and MSD, one first multiplies the diffusion equation by $\mathbf{r}^2$ and integrates over space to get:

$$\begin{aligned}
\int d\mathbf{r} \frac{\partial c}{\partial t} \mathbf{r}^2 &= D \int d\mathbf{r} (\nabla^2 c) \mathbf{r}^2 \\
\frac{\partial}{\partial t} \langle \mathbf{r}^2 \rangle &= D \left(\int d\mathbf{r} \nabla (\mathbf{r}^2 \nabla c) - \int d\mathbf{r}\, \nabla \mathbf{r}^2 \nabla c \right) \\
&= D \left(\int d\mathbf{s} (\mathbf{r}^2 \nabla c) - \int d\mathbf{r}\, 2\mathbf{r}\, \nabla c \right)
\end{aligned}$$

In the last line Gauss's law is used to transform the volume integral into a surface integral. One can choose an arbitrary big surface such that $\nabla c(\mathbf{r}, t) = 0$. Hence one gets:

$$\begin{aligned}
\frac{\partial}{\partial t} \langle \mathbf{r}^2 \rangle &= -D \int d\mathbf{r}\, 2\mathbf{r}\, \nabla c \\
&= -2D \left(\int d\mathbf{r}\, \nabla (\mathbf{r}\, c) - \int d\mathbf{r} (\nabla \mathbf{r}) c \right) \\
&= -2D \left(\int d\mathbf{s}\, (\mathbf{r}\, c) - \int d\mathbf{r}\, c \right)
\end{aligned}$$

Where the first term inside the bracket on the right hand side vanishes and the second term yields the dimension in which the substance diffuses.

$$\frac{\partial}{\partial t} \langle \mathbf{r}^2 \rangle = 6D \tag{40}$$

Finally, the Einstein relation is obtained:

$$\langle \mathbf{r}^2 \rangle = 6D\, t \tag{41}$$

According to this equation the mean squared displacement grows linearly in time. One can thus obtain the diffusion coefficient by plotting the MSD averaged over all

particles against time. The resulting plot will look something like the graph seen in figure 2. At the beginning of an MD simulation collision events in a liquid phase are not well separated in time, causing a small peak in mean squared displacement. After some short time the linear regime is entered and the diffusion coefficient can be determined from the slope:

$$D = \frac{1}{6} \lim_{t\to\infty} \frac{d}{dt} \frac{1}{N} \sum_{i=1}^{N} \langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \rangle \tag{42}$$

where N denotes the total number of particles in the system. Notice that the MSD in equation (42) is a time-correlation function:

$$\langle |\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2 \rangle = \langle \mathbf{r}_i^2(t) \rangle + \langle \mathbf{r}_i^2(0) \rangle - 2 \langle \mathbf{r}_i(t) \cdot \mathbf{r}_i(0) \rangle \tag{43}$$

Where the last term on the right hand side is the position autocorrelation function, which can efficiently be computed by using the Fast Fourier Transform [35].

Alternatively to the MSD approach, one can also use the velocity autocorrelation function to evaluate the diffusion coefficient. This method uses the Green-Kubo relation for the diffusion constant:

$$D = \frac{1}{3} \int_0^\infty dt \, \langle \mathbf{v}(0)\mathbf{v}(t) \rangle \tag{44}$$

which is outlined in more detail in the book of Frenkel and Smit [36].

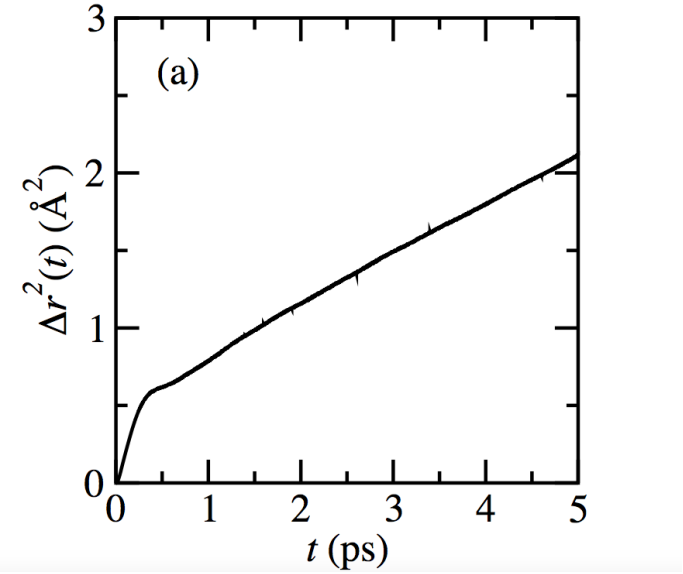
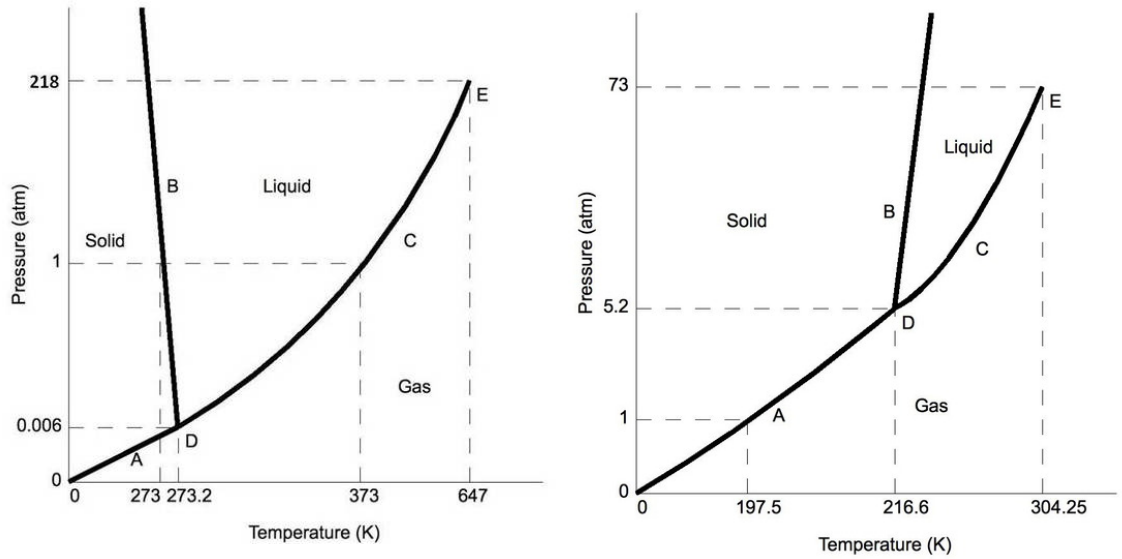


Figure 2: Mean-square displacement for a particular model of heavy water at a temperature of 300 K. (from Tuckerman [37])

3 Gibbs Duhem Integration

This thesis investigates the phase diagram of water, specifically the coexistence line between ice Ih and liquid. The phase diagram of water is of particular interest due to its negative slope in the solid-liquid coexistence line, as can bee seen in figure 3a. For most substances the slope is positive, e.g. for CO_2 as shown in figure 3b, implying that the melting point increases with pressure. Yet for water this is not the case, because with rising pressure liquid water remains denser than it's solid counterpart. This anomalous behaviour is analyzed by tracing the phase boundary between ice Ih and liquid water using the Gibbs-Duhem method [2],[3]. This method uses an initial coexistence point to integrate the Clausius-Clapeyron equation and thus obtain the saturation line. The melting point of ice Ih has already been determined [1] by using the interface pinning method [38]. In this context, this chapter shall first discuss how phase coexistence can be determined. Next the underlying methodology of the Gibbs-Duhem integration is outlined. Moreover, the Runge Kutta scheme that was used to integrate the Clausius-Clapeyron equation is discussed.



(a) Phase diagram of H_2O from experiment (b) Phase diagram of CO_2 from experiment

Figure 3: from Chemistry LibreTexts [39]

3.1 Phase Coexistence

In equilibrium thermodynamics the stable phase of a substance, at a particular temperature T and pressure p, is given by the lowest Gibbs free energy. For a pure substance the Gibbs free energy is defined by the chemical potential μ. So, one can write the thermodynamic condition at the phase transition as follows:

$$p^{\alpha} = p^{\beta} \qquad T^{\alpha} = T^{\beta} \qquad \mu^{\alpha} = \mu^{\beta} \tag{45}$$

Here the superscripts α and β denote two phases, such that p^{α} and p^{β}, T^{α} and T^{β}, μ^{α} and μ^{β} represent the pressure, temperature and chemical potential in each phase, respectively. There are two popular methods for the determination of phase coexistence using computer simulations. One is a direct approach which simulates both phases in a single cell, the other is an indirect approach using two separate simulations.

The indirect approach requires the calculation of the chemical potential μ via simulations using the Widom insertion method [40]. This particle insertion method works well for dilute phases but is not reliable in dense liquids or solids.
Alternatively, the Gibbs free energy of each phase can be evaluated by thermodynamic integration (see the topical review of Vega et al. [41]). In this method, one integrates along a thermodynamic path to a state of known Gibbs free energy, for example an Einstein solid [42], a harmonic solid [43] or an ideal gas [44]. This procedure, however, is not advisable in the study of phase boundaries, because the integration path may contain first order transitions and will hence not be reversible.

The direct approach, where both phases evolve in the same cell, was long deemed impossible due to the long time duration of phase transformation. Furthermore, the presence of a first order phase transition gives rise to an interface separating both phases. The surface tension of the interface creates a free energy barrier which will likely leave the system in a metastable phase. The work of Panagiotopoulos brought a great advance in the simulation technique [45]. In this technique, called Gibbs Ensemble Method, the interface is removed and replaced by Monte Carlo moves. These moves allow the exchange of particles and volume between the boxes, using acceptance criterions which ensure equilibration of the chemical potential and pressure, respectively. Although the Gibbs Ensemble Method has since its discovery been widely adopted for direct simulation of phase equilibria, a different method has been used.
This method, called Interface Pinning (IP) method, introduced by Pedersen et al. [46], has been used to accurately compute the melting temperature of ice Ih [1] from first-principles calculations. The technique simulates both phases in an elongated box and separates them by an explicit interface. As the system evolves in the NpT ensemble one computes the average force needed to *pin* the interface via a harmonic bias potential. In order to distuingish between the phases, one defines a suitable order parameter Q, e.g. the collective density field or the Steinhard order parameter [47], which couples to the external field. The shifting interface can thus be described as an equilibrium process, where the Gibbs free energy difference ΔG can

be computed directly. The Newton-Raphson root finding method, described in Pedersen's paper [38] , is then used to find a coexistence point, that is the temperature at which $\Delta G = 0$. The IP method is very efficient and flexible, and can easily be combined with ab-initio molecular dynamics to yield accurate free energy differences.

3.2 Clausius-Clapeyron Equation

Once an accurate coexistence point has been found, one can use the Gibbs-Duhem integration (GDI) introduced by Kofke [2],[3] to determine the whole coexistence line. This method uses the Clausius-Clapeyron equation to permit direct evaluation of phase coexistence along the saturation line. This equation shall first be derived.

Fundamentally the Gibbs free energy of two phases α and β, along the coexistence line in the p-T diagram, must remain the same:

$$G^{\alpha}(p,T) = G^{\beta}(p,T) \tag{46}$$

Hence,

$$\left(\frac{\partial G^{\alpha}}{\partial T}\right)_p dT + \left(\frac{\partial G^{\alpha}}{\partial p}\right)_T dp = \left(\frac{\partial G^{\beta}}{\partial T}\right)_p dT + \left(\frac{\partial G^{\beta}}{\partial p}\right)_T dp \tag{47}$$

where dT and dp denote infinitesimal changes in temperature T and pressure p along the coexistence line. Here one further considers that the total number of particles N in the system is fixed, such that $dN = 0$. In equation (47) one can recognize the natural variables $\left(\frac{\partial G}{\partial p}\right)_T = V$ as the volume and $\left(\frac{\partial G}{\partial T}\right)_p = -S$ as the entropy:

$$-S^{\alpha}dT + V^{\alpha}dp \;\; = \;\; -S^{\beta}dT + V^{\beta}dp \tag{48}$$

One can consider the molar entropy $s = \frac{S}{N}$ and the molar volume $v = \frac{V}{N}$ as natural variables:

$$s^{\alpha}dT - v^{\alpha}dp = s^{\beta}dT - v^{\beta}dp \tag{49}$$

$$\Longleftrightarrow \;\; \frac{dp}{dT} = \frac{s^{\beta} - s^{\alpha}}{v^{\beta} - v^{\alpha}} = \frac{\Delta s}{\Delta v} \tag{50}$$

The chemical potential in each phase is given by:

$$\mu^{\alpha} = h^{\alpha} - Ts^{\alpha} \qquad\qquad \mu^{\beta} = h^{\beta} - Ts^{\beta} \tag{51}$$

Since at coexistence both phases have the same chemical potential $\mu^{\alpha} = \mu^{\beta}$:

$$(s^{\alpha} - s^{\beta})T = h^{\alpha} - h^{\beta} \tag{52}$$

$$\Delta s = \frac{\Delta h}{T} \tag{53}$$

Finally the Clausius-Clapeyron equation is obtained by inserting (53) into (50):

$$\boxed{\frac{dp}{dT} = \frac{\Delta h}{T\Delta v}} \tag{54}$$

This relation describes how the pressure adjusts to a change in temperature, such that two coexisting phases remain in equilibrum while having the same Gibbs free energy.[1] One can say that by integrating this equation one gets the "coexistence pressure" for the selected temperature. It was in 1993 that Kofke realized that numerically integrating the Clausius Clapeyorn equation, starting from a previously known coexistence point, would yield the whole coexistence line between two phases [2],[3]. If the coexistence line has a steep slope in the in the p-T diagram it is more convenient to integrate the Clausius-Clapeyron equation differently:

$$\boxed{\frac{dT}{dp} = \frac{T\Delta v}{\Delta h}} \tag{55}$$

Especially when the coexistence line is known to present reentrant behaviour (i.e. one temperature corresponds to two possible pressures) one should take special care in choosing the equation to integrate.

3.3 Runge-Kutta Method

The Clausius-Clapeyron equation (55) is an ordinary differential equation (ODE) of first order. Let us denote the pressure as independent variable x, the temperature as $y = y(x)$ and the integrand as function of both $f = f(x, y)$, such that:

$$\frac{dy}{dx} = f(x, y) \tag{56}$$

In general, one can use an explicit Runge-Kutta method to solve ODEs:

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i \tag{57}$$

where the next value y_{n+1} is obtained from the present value y_n plus a stepsize h weighted by several function evaluations k_i:

$$\begin{aligned} k_1 &= f(x_n, y_n) \\ k_2 &= f(x_n + c_2 h, y_n + h(a_{21} k_1)) \\ k_3 &= f(x_n + c_3 h, y_n + h(a_{31} k_1 + a_{32} k_2)) \\ &... \\ k_m &= f(x_n + c_m h, y_n + h(a_{m1} k_1 + a_{m2} k_2 + ... + a_{m,m-1} k_{m-1})) \end{aligned}$$

The number of stages is given by m and the coefficients a_{ij}, b_i and c_j are given in the so-called Butcher tableau. For our purposes we use the Dormand-Prince technique [49] within the dopri5 package of SciPy [50]. This adaptive Runge-Kutta technique uses an additional lower order step to estimate the error:

$$\epsilon = y_{n+1} - y^*_{n+1} = y_n + h \sum_{i=1}^{s} (b_i - b^*_i) k_i \tag{58}$$

[1] The Gibbs-Duhem method can be extended to consider different "field" variables, e.g. fugacity fraction or a parameter defining the intermolecular potential, instead of pressure and temperature. Furthermore one can also consider additional phases to reproduce for example three phase coexistence lines. The interested reader is refered to [48].

This error is then used to control the step size. The extended Butcher tableau, given in table 1, gives the c_j coefficients in the outer left column, the b_i and b_i^* coefficients in the two lower rows and the a_{ij} coefficients span the matrix in between.

c_j	a_{ij}						
0							
1/5	1/5						
3/10	3/40	9/40					
4/5	44/45	-56/15	32/9				
8/9	19372/6561	-25360/2187	64448/6561	-212/729			
1	9017/3168	-355/33	46732/5247	49/176	-5103/18656		
1	35/384	0	500/1113	125/192	-2187/6784	11/84	
b_j	35/384	0	500/1113	125/192	-2187/6784	11/84	0
b_j^*	5179/57600	0	7571/16695	393/640	-92097/339200	187/2100	1/40

Table 1: Butcher tableau for Dormand-Prince method

The first row of b_i coefficients gives the 5^{th} order solution, and the second row of b_i^* coefficients is used to get the 4^{th} order solution. Note that even though the Dormand-Prince method has seven stages it only needs six function evaluations per step, because the last stage of a step corresponds to the first stage of the next step. This is known as the FSAL (First Same As Last) property and can easily be seen in the Butcher tableau 1.

3.4 Implementation

The Gibbs-Duhem method requires the evaluation of the integrand $\frac{dT}{dp}$ by performing molecular dynamics (MD) simulations. Here the simulation program LAMMPS [29] is chosen, which can be invoked from a Python script. The intrinsic difficulty of the problem that is obtaining the density and enthalpy of two phases to compute the slope, suggests the use of parallel computing. So two simulations are set up simultaneously with MPI [51],[52],[53]; an isotropic NpT-simulation for the liquid phase and an anisotropic NpT-simulation for the solid phase. The density and enthalpy in each phase are computed as running average, in order to increasingly damp their fluctuations. After 200'000 MD timesteps (corresponding to 0.1 ns), where 50% are discarded for the purpose of equilibration, the final values of density and enthalpy are used to determine the integrand. Next the Runge Kutta scheme, outlined in section 3.3 is used to find the subsequent coexistence point. From this state point, defined by temperature T and pressure p, a new NpT-simulation can be started for each phase and thus a new slope can be computed. The procedure is repeated until a final predefined pressure/temperature is reached. We emphasize again that the Gibbs-Duhem method is not self-starting. The initial point to start the integration has previously been determined within our research group [1].

In general the GDI follows this procedure:

1. Determine a coexistence point (e.g. by IP method [46])
   i) Start isotropic/anisotropic NpT simulations for the liquid/solid phase
   ii) Measure the enthalpy and density in each phase
   iii) Take the difference in enthalpy and density between the two phases
   iv) Calculate the slope $s = \frac{T\Delta v}{\Delta h}$
   v) Determine a new state point by integrating the slope
   vi) Repeat at i) m times (m = function evaluations of RK method)
2. Integrate with RK method
3. Find new coexistence point
4. Repeat at i)

There are many parameters to consider when performing a Gibbs-Duhem integration. First one should choose the integrand such that it is slowly varying. Second the integration step size needs to be set, on one hand guaranteeing stability and accuracy of the integration method, on the other hand defining the range between coexistence points. Next, the integrator must be chosen; higher-order integrators will be more accurate but need more input data and might be less stable. Kofke proposed the use of predictor-corrector algorithms [2],[3], but Runge Kutta schemes have grown more popular [54]. Moreover, care must be taken on the possible departures and reentrant behaviours from the correct coexistence line. Furthermore, the system size used for the MD simulation is an important consideration. Last but not least, the method relies on an accurate initial coexistence point. The interface pinning method [46] is the method of choice for delivering this data. Ultimately, the level of accuracy of the GDI will depend on the underlying interatomic interaction used to evaluate the forces acting on the particles.

4 Water Models

Despite its simple molecular form, water has been very difficult to model. This is due to its ability to form strong directional H-Bonds, which are generally acknowledged to cause many anomalies that distinguish water from other liquids. In addition, the oxygen atom in the water molecule causes via its strong electronegativity a pronounced dipole, which is particularly difficult to model. This dipole is responsible for strong non-additive inductive effects, that in turn further increase the polarization within the water molecule.
Many water models have been proposed since the work of Bernal and Fowler, who developed the BF model in 1933 [55]. One can categorize the subsequent models based on their description of the water monomer and the treatment of the interactions between the molecules. The water monomer is either treated as rigid or flexible molecule and one distinguishes between polarizable and non-polarizable models. The interactions between N molecules can be resolved into many body terms [9],

$$U(\mathbf{r}^N) = \sum_i u_1(\mathbf{r}_i) + \sum_{i<j} u_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_{i<j<k} u_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + ... \tag{59}$$

where $\mathbf{r}_i$ is the configurational vector of the i^{th} molecule. Further classification of water potentials often follows the level of approximation of this expansion. Finally the parametrization of the model can be achieved by using either experimental data or ab initio calculations. As one can see there are many different approaches to model water. The historical development of water models has been reviewed i.a. by Ouyang and Bettens [56] and by Guillot [57].
We will focus our attention to rigid non-polarizable models that have been used to estimate the phase diagram of water. Those are namely the rigid models SPC/E, TIP4P, TIP4P/2005 and TIP4P/Ice. Their predictions will, along with experimental data, serve as reference to assess the quality of our results using the Neural Network Potential. The SPC/E model will serve as state of the art example to obtain the phase diagram of water from computer simulations. Next, the TIP4P model will be presented with a much better description of the phase diagram. In this part, we will also discuss the existence of reentrant behaviour. Furthermore, we will show the results of the TIP4P/2005 model, which was designed to include several target properties. Here the importance of accurate densities will be emphasized to obtain a qualitatively good phase diagram. In addition, the TIP4P/2005 model delivers a description of the phase diagram at negative pressures. Finally, our discussion will be completed by the TIP4P/Ice model, which delivers the best phase diagram results.

4.1 Rigid Water Models

Early treatments of water models often used rigid molecules, motivated by the fact that bond vibrations are typically of high frequency and thus much stronger than interactions between neighbors. For rigid models the first term of equation (59) vanishes, as it represents the energy required to deform the individual molecule. One further approximates that higher order ($n > 2$) terms have no signficant contribution to the total interaction energy:

$$U(\mathbf{r}^N) \approx \sum_{i<j} u_2(\mathbf{r}_i, \mathbf{r}_j) \tag{60}$$

In the case of rigid molecules the configurational vector $\mathbf{r}_i$ has six components: three components specify the center-of-mass position and three Euler angles are used to specify the orientation of the molecule. The pair potential approximation in equation (60) has been widely used under the form of the Lennard-Jones potential [58]:

$$u_{LJ}(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{61}$$

Note that the potential energy depends on the relative distance $\mathbf{r}_{ij}$ between two particles i and j. The parameters ϵ and σ represent the depth of the potential well and the molecular separation for zero interaction energy, respectively. To account for the electrostatic interactions caused by the positive and negative charges within the water molecule one adds a Coulomb term:

$$u_{ab}(r_{ij}) = 4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] + \frac{q_a q_b}{r_{ij}} \tag{62}$$

Where q_a and q_b model the charge distribution of atom a and b. Having this simple site-site interaction one needs to specify the geometry of the water monomer to know where to place the sites. One uses a fixed bond distance r_{OH} and a fixed bond angle α, to specify the position of the oxygen and the hydrogen atoms. The interaction site can then be placed on the atom's position or like in the case of the TIPnP-family on an additional position relative to the water molecule. The set of parameters $\epsilon, \sigma, q_a, q_b, r_{OH}, \alpha$ can then be adapted to reproduce some important properties of water. A fundamental goal is to reproduce the experimental phase diagram of water shown in figure 4.

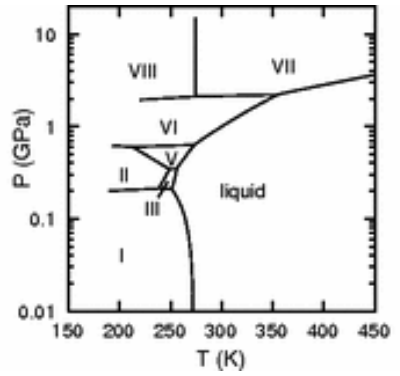
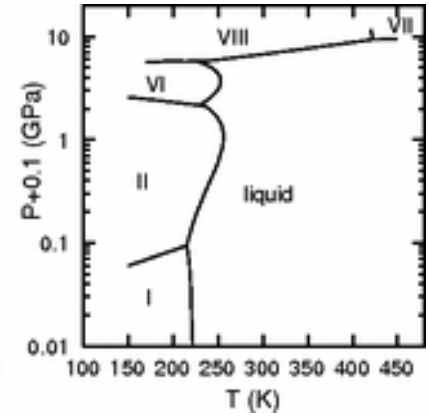


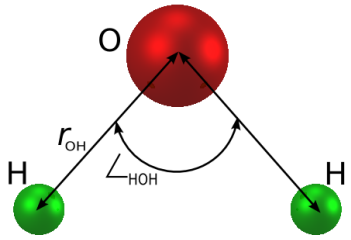
Figure 4: Experimental phase diagram of water
(from Sanz et al. [54])

4.1.1 SPC/E

The SPC/E model first proposed by Berendsen *et al.* [10], is a slight modification of the Single Point Charge (SPC) model. Its geometry, shown in figure 5b , is specified by $r_{OH} = 1\,\text{Å}$ and the tetrahedral angle $\alpha = 109.47°$. The Lennard-Jones interaction site is placed along with the negative charge on the oxygen atom. The positive charges are placed on the hydrogen atoms. The target properties for this model were the density and vaporization enthalpy of water at ambient conditions. The resulting effective potential, where the parameters are given in table 5b, gave good approximations of the radial distribution function, the diffusion coefficient and density of liquid water. The increase of the value of the charges with respect to the SPC model, simulate a higher dipole moment. Thus induction effects are enhanced, leading to improved values of the before mentioned properties.
The melting temperature of ice I for the SPC/E model, was determined to be $T_m = 217(7)$ K at $p = 1$ bar by Sanz, Vega, Abascal and Dowell [54]. In their work, Sanz et al. computed several melting temperatures of different ices. Using the Gibbs-Duhem method [2][3], they were able to get the phase diagram of the SPC/E model shown in figure 5a. As one can see the predictions of the SPC/E model are rather poor.



(a) Phase diagram of water for the SPC/E model. The coexistence pressures have been shifted by 0.1 GPa to include results for ice I (which appears at slightly negative pressures for the SPC/E model). Only the stable phases are shown in the diagram. (from Sanz et al. [54])



(b) SPC/E model (fig. from Sklogwiki [59])

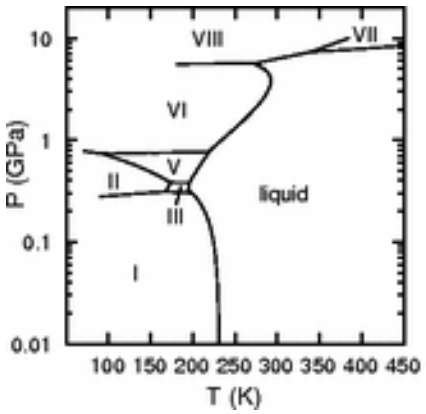
r_{OH}(Å)	$\alpha(°)$	$\epsilon(\frac{kJ}{mol})$
1.0	109.47	0.650

σ (Å^6)	q_H (e)	q_O (e)
3.166	0.4238	−0.8476

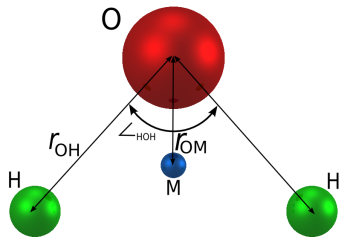
Figure 5: Phase diagram and structure of the SPC/E model

4.1.2 TIP4P

The TIP4P model was first proposed by Jorgensen *et al.* in 1983 [11]. It introduces a massless site M carrying the negative charge, located at 0.15 Å from the oxygen atom on the H-O-H bisector as can be seen in figure 6b. This particular structure was already introduced by Bernal and Fowler [60] in 1933. Moving the negative charge towards the positive charges, reduces the dipole and thereby gives additional control over inductive effects. The angle between the three atoms is $\alpha = 104.52°$, corresponding to the value obtained from experiments [4]. The parameters were chosen to reproduce the density of liquid water at room temperature and the vaporization enthalpy (see table in figure 6b).
The melting temperature of ice I, for the TIP4P model, was located at $T_m = 232(5)$ K for $p = 1$ bar by Sanz et al. [61]. In their studies, other coexistence points were found for stable ices, and they applied the Gibbs-Duhem method to obtain the phase diagram shown in figure 6a. The result reproduces many features of the experimental phase diagram.



(a) Phase diagram of water for the TIP4P model. Only the stable phases are shown in the diagram. (from Sanz et al. [54])



(b) TIP4P model (fig. from Sklogwiki [59])

r_{OH}(Å)	$\alpha(°)$	r_{OM}(Å)	
0.9572	104.52	0.15	

$\epsilon(\frac{kJ}{mol})$	σ(Å^6)	q_H(e)	q_O(e)
0.6480	3.15365	0.5200	−1.0400

Figure 6

An interesting characteristic of the model is that the slope of the liquid - ice VI, - ice V, - ice III and - ice I lines, is changing its sign in the p-T diagram. This so called reentrant behaviour is better seen in the enlarged figure 7 from Sanz et al. [54]. The prominent chemist-physicist Tammann speculated with this possibility at the beginning of the 20th century (quoted by Bridgmann in Ref.[62] on page 553). The origin of this particular behaviour of the slope can easily be understood. The slope of a coexistence line is determined by the ratio of enthalpy difference Δs and volume difference Δv, as shown in equation (50). Since the entropy of a liquid phase s_l is greater than the entropy of a solid phase s_s, the difference $\Delta s = s_l - s_s$ will always be negative. Consequently, the change in slope is governed by a change in the sign of Δv. The physical explanation of the reentrant behaviour is that liquid

water is far more compressible at high pressures than the low density ices, namely I, III, V, VI. Thus for positive pressures the change in slope occurs when the density of water exceeds the density of the coexisting low density ice. Similarly, in the negative pressure regime, the reentrant behaviour of the liquid-ice I line, occurs when the density of ice I surpasses the density of liquid water. Consequently, one should obtain a liquid phase either by applying pressure to ice I, III and V at very low temperatures or by applying negative pressure to ice I close to the triple point. It is important to note that this reentrant behaviour has never been observed in nature, because a new ice polymorph emerges upon approaching the reentrant point, as can be seen in figure 4. This makes the study of the coexistence line between ice I and liquid water even more interesting, because it is the only *upper* part of a reentrant line which lies within a thermodynamically stable region.

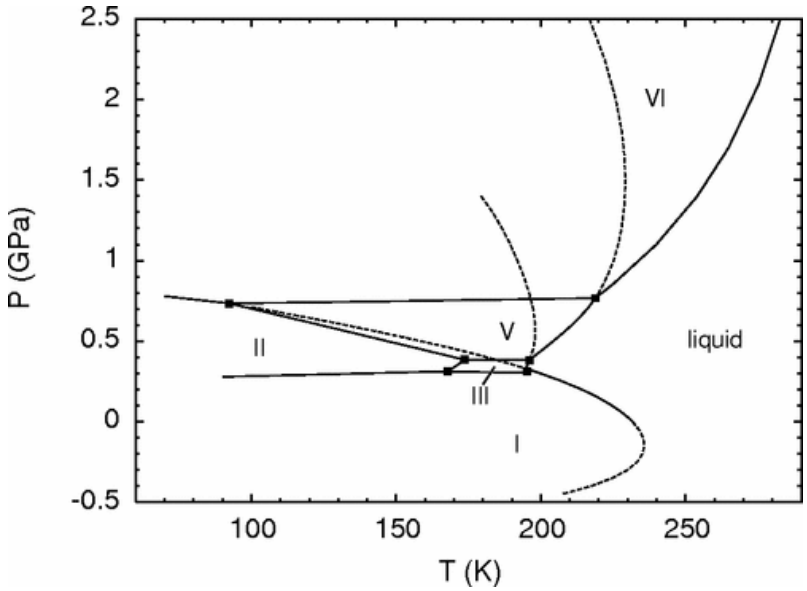


Figure 7: Detailed view of the central part of the phase diagram of TIP4P water. The solid lines correspond to thermodynamically stable coexistence lines. The dashed lines correspond to metastable coexistence lines.
Reentrant behaviour is observed for the liquid-ice III and liquid-ice V coexistence lines, although the upper part of the reentrant lines lie in metastable regions.
The liquid-ice I coexistence line also exhibits reentrant behaviour. In this case, however, the upper part of the reentrant corresponds to the thermodynamically stable region. The change in slope occurs at negative pressures (about $p = -0.15$ GPa and $T = 235$ K). (from Sanz et al. [54])

4.1.3 TIP4P/2005

In 2005, Abascal and Vega proposed a general purpose model for the condensed phases of water [12]. The model, namely TIP4P/2005, uses more target properties, including the temperature of maximum density and melting temperature of hexagonal ice Ih. Paschek already emphasized that a water model should reproduce the correct T_{md}, in order to account for hydrophobic effects [63]. The temperature dependence of density obtained from the TIP4P/2005 model can be seen in figure 8. The model closely reproduces the anomalous behaviour of water's density.

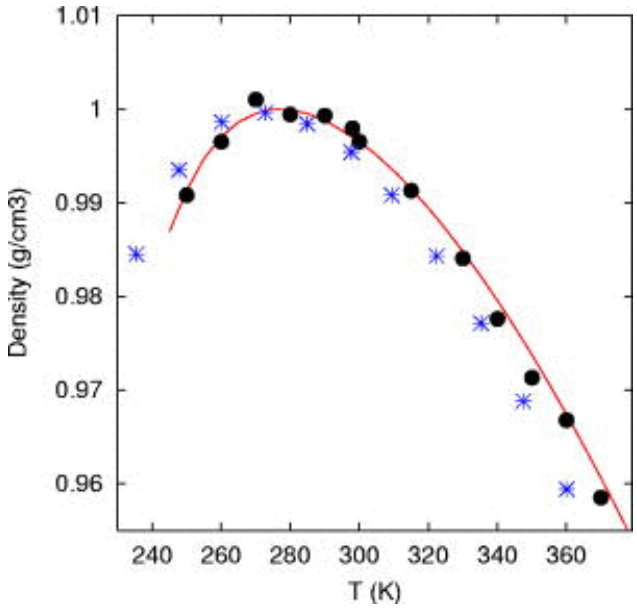
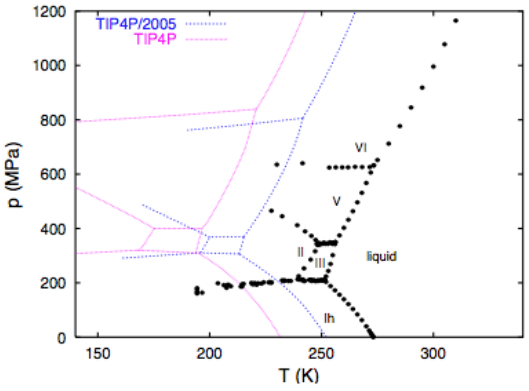


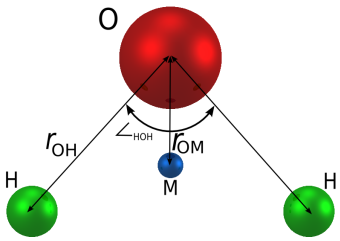
Figure 8: Densities of the TIP4P/2005 model (circles) at $p = 1$ bar compared to the values for TIP4P/Ew (stars) [64] and experimental data (full line). (from Abascal and Vega [12])

The predictions of the phase diagram using this model are shown in figure 9a obtained by Abascal and Vega [12]. The results are better than the ones obtained from the previous TIP4P model, although the melting temperature of ice I is still underestimated. Nevertheless the TIP4P/2005 model emphasizes the importance of accurate densities to reproduce the experimental phase diagram.

Furthermore the TIP4P/2005 model has been used by Conde, Vega, Tribello and Slater to compute the phase diagram at negative pressures [66]. Besides Ice Ih, Conde et al. introduced so called virtual ices, sI, sII and sH to occupy some negative pressure regions. The primary interest is focused on their results for the negative pressure coexistence line between Ice Ih and liquid water. Notice that close to the reentrant point the virtual ice polymorph sII emerges.



(a) Phase diagram of TIP4P (purple dots) and TIP4P/2005 (blue dots) models, compared to experimental values (black circles). (from Vega et al. [65])



(b) TIP4P model (fig. from Sklogwiki [59])

r_{OH} (Å)	$\alpha(^\circ)$	r_{OM} (Å)	
0.9572	104.52	0.1546	
$\epsilon(\frac{kJ}{mol})$	σ (Å^6)	q_H (e)	q_O (e)
0.7749	3.1589	0.5564	−1.1128

Figure 9

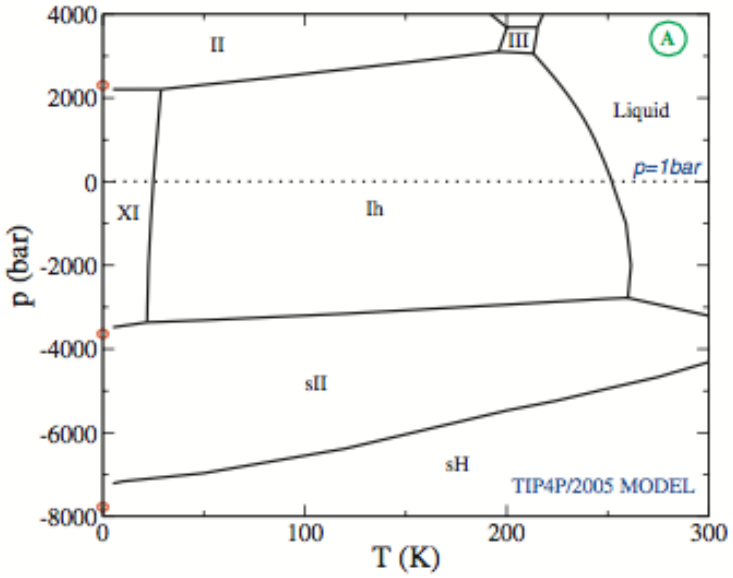
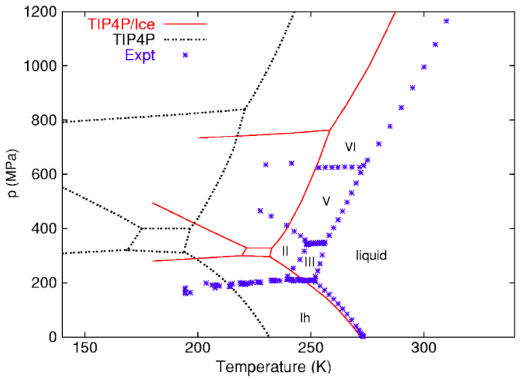


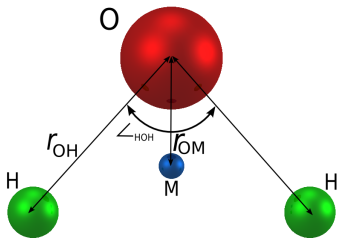
Figure 10: Phase diagram in the region of negative pressures obtained for the TIP4P/2005 model. The dashed line indicates room pressure, $p = 1$ bar. The open circles correspond to the values of coexistence pressure at $T = 0$ K, estimated using the zero-order approximation method [67], which uses the differences of energies and volumes between ices at 0 K to estimate the transition pressure.
(from Conde et al. [66])

4.1.4 TIP4P/Ice

The TIP4P/Ice model was designed by Abascal et al. [13] to reproduce the solid phases of water. While this potential model fails for the very dense ices (VII and VIII) it yields quite accurate results for low density ices, particulary ice Ih. The phase diagram of the TIP4P/Ice is the most accurate, because the model has been parametrized to fit selected points of coexistence lines. Most important is the correct prediction of the melting temperature of ice Ih to be 272.2 K at 1 bar.



(a) The phase diagram of TIP4P/Ice (red lines) and TIP4P (black dashed lines), compared to experimental values (blue stars). (from Abascal et al. [13])



(b) TIP4P model (fig. from Sklogwiki [59])

r_{OH} (Å)	$\alpha(^\circ)$	r_{OM} (Å)	
0.9572	104.52	0.1577	
$\epsilon(\frac{kJ}{mol})$	σ (Å^6)	q_H (e)	q_O (e)
0.8822	3.1668	0.5897	−1.1794

Figure 11

The TIP4P/Ice model points out the current limits of how accurately the phase diagram can be obtained from a rigid, non-polarizable water model. It is remarkable how accurate the coexistence lines can be retraced from the TIP4P/Ice model. The ultimate goal, however, remains to obtain a water potential which is not only valid within the phase diagram region, to which it has been parametrized to.

5 Ab Initio Molecular Dynamics

In traditional molecular dynamics (MD) simulations the forces acting on the atoms are derived from a pre-specified empirical potential, which is parametrized to reproduce experimental or ab initio data. The drawback of this approach, is that the resulting empirical potential is often restricted to regions of the phase diagram corresponding to the initial fitting information. Within this approach charges are often considered as static, thereby neglecting polarization effects. There have been attempts to include polarization effects in so-called polarizable models (see the review of Ouyang and Bettens [56]), but these models still lack transferability. Another shortcoming of classical MD is the use of models with predefined connectivity between atoms. Chemical bond breaking and forming processes are hence unable to be simulated. Semi-empirical techniques have been proposed, e.g. empirical valence bond method [68], to deal with this problem. The issue of transferability, however, persists because such methods require the reparametrization of each type of chemical reaction.
The idea of ab initio molecular dynamics (AIMD) is to compute the interatomic forces 'on-the-fly' from electronic structure calculations. Since the electronic variables are considered as active degrees of freedom, electronic polarization and the breaking/forming of bonds can be accurately described. Any electronic structure method can be used for AIMD, with density functional theory (DFT) being the most popular approach. Since the pioneering work of Car and Parrinello in 1985 [14], the concept of AIMD has been widely developed (see [69]), and different electronic structure representations have been proposed [70] [71].
In the next subsections the Kohn-Sham formulation of DFT will be described, followed by a short description of methods to include dispersion interactions.

5.1 Density Functional Theory

In Density Functional Theory (DFT) one is interested in solving the ground state energy of the atoms in a system. One uses the Born-Oppenheimer approximation to treat the motion of the nuclei separately from the motion of the electrons around it. The Schrödinger Equation needs to be solved to find the ground state for the electrons:

$$\Big\{ -\frac{\hbar^2}{2m}\sum_i^N \nabla_i + \sum_i^N V_i(\mathbf{r}_i) + \sum_i^N \sum_{j<i} U(\mathbf{r}_i, \mathbf{r}_j) \Big\} \Psi = E\Psi \tag{63}$$

The first term accounts for the kinetic energy of each electron, the second one contributes to the interaction energy between electrons and nuclei, the third represents the interaction energy between all electrons, and E on the right hand side represents the minimum energy for the electrons. The electronic wave function Ψ in the above equation can be decomposed into single electron wave functions by using the Hartree approximation:

$$\Psi(\mathbf{r}) = \Psi_1(\mathbf{r}_1)\Psi_2(\mathbf{r}_2)...\Psi_N(\mathbf{r}_N) \tag{64}$$

In 1964, Hohenberg and Kohn introduced two theorems which form the theoretical foundation of DFT [22]. The first theorem states that there exists a *unique* energy

functional that yields the ground state energy from the Schrödinger equation 63. This ground state energy is a unique functional of the electron density $E = E\big[n(r)\big]$. The second Hohenberg and Kohn theorem states that the electron density that minimizes the energy of the functional is the true electron density of the solution of the Schrödinger equation. So the problem reduces from a $3N$- to a 3-dimensional problem, that consists in finding the right electron density $n(\mathbf{r})$ that minimizes the energy functional. This is done by solving the Kohn-Sham equations [72] for each single electron:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{XC} \right] \Psi_i = \epsilon_i \Psi_i \tag{65}$$

Here $V(\mathbf{r})$ represents the electron's energy contribution due to the interaction with other nuclei. $V_H(\mathbf{r})$ is the energy contribution due to the electron's interaction with all the electrons, namely the electron density $n(\mathbf{r})$. Note here that $V_H(r)$ already requires a given electron density $n(r)$. So one uses a recursive algorithm which defines an initial trial $n(r)$ to solve the Kohn Sham equations, in order to obtain a new estimate of the electron density. The additional contribution $V_{XC}(\mathbf{r})$ is called exchange correlation potential. It was introduced to correct for the electron's self-interaction in the V_H term. The exact functional form of V_{XC} remains, to this day, unknown. Yet there are a number of approximations that have been proposed. For example, the generalized gradient approximation (GGA) which uses information about the local electron density and the local gradient in electron density. There are many distinct GGA functionals. In our research we use the RPBE[73] and BLYP[74][75] density functionals. The predictions of these two functionals for the energy of water clusters are reasonably good. When used for simulations of liquid water, however, the results for the structure, the diffusion coefficient and the density of the liquid are rather poor [76],[77],[78],[79],[80]. The inclusion of van der Waals interactions significantly improve the properties of simulated liquid water [1]. The origin of these forces and their incorporation into DFT-GGA will be discussed in the next section.

5.2 Van der Waals Interactions

The interaction between water molecules in condensed matter phases is governed by a fine balance between strongly directional H-bonds and weak isotropic van der Waals (vdW) forces. Describing vdW interactions, which fundamentally occur due to correlations of temporary fluctuations in electron density between molecules, is extremly challenging with DFT. There are various approaches, like the empirical correction method (DFT-D [81],[82],[83]), the nonlocal vdW functional (vdW-DF [84],[85],[86]), dispersion-corrected atom-centered potentials (DCACP) and pure density functionals (DF). The inclusion of these weak interations in MD simulations of water is of great importance. It was shown, for instance, that the use of DFT-D and DCACP in the BLYP density functional significantly improved structure and dynamical properties of liquid water [87]. Other studies used the vdW-DF method to undermine the importance of vdW interactions [88],[89],[90]. Most notable is the DFT-D method, which has been widely applied in research [91],[92],[93],[94]. The

latter approach, is based on an atom pairwise sum over C_6R^{-6} potentials, which yield the correct asymptotic behaviour for large intermolecular distances. For simulations of bulk water, which incorporates many long-range interactions, this method is thus strongly advisable. Recently the DFT-D3 method proposed by Grimme [18], was applied to the RPBE and BLYP functionals to describe the effect of VdW forces on the density of water [1] (see figure 12). Within the D3 approach one considers pair interactions of the form:

$$E_{vdW}^{(2)} = -\sum_{A<B}^{N_{pairs}} \left(\frac{C_6^{AB}(CN)}{r_{AB}^6} f_{d,6}(r_{AB}) + s_8 \frac{C_8^{AB}(CN)}{r_{AB}^8} f_{d,8}(r_{AB}) \right) \tag{66}$$

Here r_{AB} denotes the distance between two atom pairs A and B. The 6^{th} and 8^{th}-order dispersion coefficients $C_{6/8}^{AB}$ depend on the chemical environment via fractional coordination numbers CN. The damping functions $f_{d,n}$ control the range of the vdW interactions by screening them to zero at short distances and thus avoid near singularities:

$$f_{d,n}(r_{AB}) = \left(1 + 6 \left(\frac{r_{AB}}{s_{r,n} R_0^{AB}} \right)^{-\alpha_n} \right)^{-1} \tag{67}$$

The vdW corrections, depend through the s_8 and $s_{r,6}$ parameters, on the density functional. Consequently the inclusion of the correction term may have a greater or lesser influence for different DFTs.

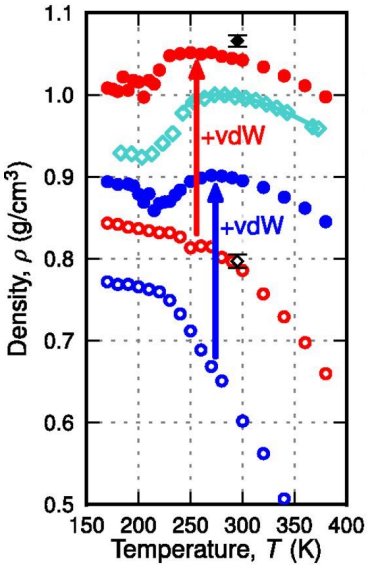


Figure 12: Density isobars at $p = 1$ bar obtained from NNP simulations based on the RPBE (blue empty circles), RPBE-D3 (blue filled circles), BLYP (red empty circles) and BLYP-D3 (red filled circles) density functionals. Results are compared to experimental data for bulk (cyan line) and confined water [95] (cyan diamonds), and data from AIMD simulations with the BLYP (black empty diamond) and BLYP-D3 (black filled diamond) density functional [96]. (from Morawietz et al. [1])

6 Neural Network Potential

The major disadvantage of ab initio molecular dynamics is the computational cost of the simulations for large system sizes and long time scales. In this research a neural network potential (NNP) is used, which is able to accurately reproduce the ab initio potential energy surface (PES) of water, while keeping computational costs low. The NNP, which was trained for the BLYP and RPBE with and without van der Waals interactions, was already successfully used to reproduce water's density maximum and melting temperature [1]. The PES was constructed based on the high-dimensional NNP approach introduced by Behler and Parrinello [16], where the total energy E is expressed as sum of individual atomic contributions E_i,

$$E = \sum_i E_i \tag{68}$$

Atom-centered symmetry functions [97] are used to describe the local environment of the atoms in the system. This section will give a comprehensive review of how the neural network (NN) evaluates the total potential energy of a system of water molecules.

6.1 Feed Forward Neural Network

In contrast to traditional physically motivated potentials, the neural network approach is a purely mathematical approach. The method, inspired from biology, makes use of reference data, in order to approximate an unknown function, here the shape of the PES. An analytically constructed NN is constituted of many interconnected neurons, like the nerve cells in the brain. Such an artificial neuron is denoted by a node i which is connected to other nodes j via weighted links a_{ji}. Each node represents a real number y_i which is obtained from a linear combination of the weighted values of the incoming nodes:

$$y_i = f_a(\sum_j a_{ji} y_j) \tag{69}$$

Here f_a is an activation function, which allows the representation of arbitrary PES and is typically a non-linear function. There are many different neural networks types which have been suggested in literature [98]. For our purposes, a feed forward neural network has been used.
A feed forward NN consists of nodes arranged in layers as shown in figure 13. The evaluation proceeds from the input layer, through two hidden layers to the output layer, each node being connected to all the other nodes in the subsequent column. The input layer describes the atomic environment with so called symmetry functions, which will be explained in more detail in section 6.2. The two hidden layers, in our case containing $N_{HL} = 25$ nodes each, have no physical meaning, yet are crucial to allow the representation of complicated functions and thus determine the flexibility of the NN. The output layer yields the energy of the reference atom E_{atom}. The information is propagated from node to node via weighted links, which are the fitting

parameters of the NN. The value of each node in the hidden layer is calculated as follows:

$$y_i^n = f_a^n(b_i^n + \sum_j a_{ji}^{n-1,n} y_j^n) \tag{70}$$

Here the superscript n denotes the layer of the node and b_i^n is a bias weight used as adjustable offset for the activation function. The architecture of the NN shown in figure 13 can be specified by the short notation $M - 25 - 25 - 1$ ttl, where M is the number of input symmetry functions. The letters t and l represent the use of a hyperbolic $t$angent and $l$inear functions in each layer:

$$f^1(x) = \tanh(x) \tag{71}$$
$$f^2(x) = \tanh(x) \tag{72}$$
$$f^3(x) = x \tag{73}$$

The choice of an output *linear* function is to avoid any restriction on the output energy value,which is finally given by:

$$E_{atom} = f_1^3 \left\{ b_1^3 + \sum_{k=1}^{N_{HL}^2} a_{k1}^{23} f_k^2 \left[b_k^2 + \sum_{j=1}^{N_{HL}^1} a_{jk}^{12} f_j^1 \left(b_j^1 + \sum_{i=1}^{M} a_{ij}^{01} G_i \right) \right] \right\} \tag{74}$$

Here G_i denotes the input symmetry functions, which will be explained in more detail in the next section 6.2. The "training" of the NN occurs by iteratively optimizing the fitting parameters $a_{ij}^{n-1,n}$ and b_i^n with respect to reference total energies and forces obtained from DFT. The total number of weights used is given by:

$$N_w = \sum_{k=1}^{M_{HL}+1} (N_{k-1} N_k + N_k) \tag{75}$$

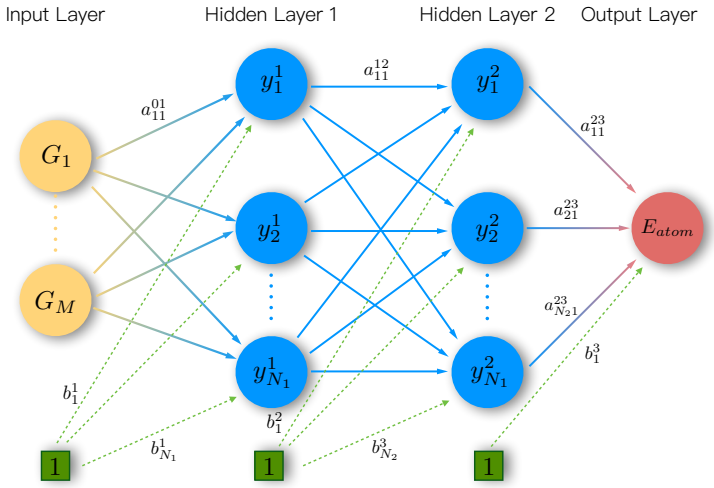


Figure 13: Representation of a Feed Forward Neural Network

6.2 Symmetry Functions

When constructing a neural network it is of utmost importance that the output value, the total energy of the system, satisfies physical symmetries. This means that any transformation that does not change the structure of the system, should leave the total energy invariant. Those transformations are for example translation or rotation of the system and interchange of like atoms. The question arises which input functions one should use to represent the local atomic environment. Cartesian coordinates are a natural choice, however these absolute values have no physical meaning, because the potential energy depends only on the relative distances between atoms. The neural network should thus be constructed by input functions which are:

(i) Continuous in value and slope

(ii) Invariant with respect to translation and rotation of the system

(iii) Invariant with respect to permutation of atoms of the same element

(iv) Able to represent arbitrary system sizes

In order to satisfy these properties one introduces so called "Symmetry functions". These functions encode the local environment of each atom and ensure that the output energy is invariant with respect to all operations that do not change the relative atomic configuration of the system.
This work will use atom-centered symmetry functions [97], where the radial extension of the local environment of atom i is restricted by a cutoff function:

$$f_c(r_{ij}) = \begin{cases} \tanh^3(1 - \frac{r_{ij}}{r_c}) & \text{with } r_{ij} < r_c \\ 0 & \text{with } r_{ij} > r_c \end{cases} \tag{76}$$

Here, r_{ij} denotes the distance between the reference atom i and its neighbors j. At the cutoff radius r_c this function makes sure that the symmetry function is zero in value and in slope. A cutoff radius of $r_c = 6.36\,\text{Å}$ is chosen, which is sufficiently large to include several nearest neighbors. To describe the atomic environment, radial and angular symmetry functions of type 2 and 4 respectively are used [97]. Both functions describe the neighbor distribution of a reference atom and thus depend on all atomic coordinates within the given cutoff sphere.
Radial symmetry functions are obtained from the summation of distance dependent Gaussians multiplied by the cutoff function:

$$G_i^2 = \sum_{j \neq i} e^{-\eta(r_{ij} - r_s)^2} f_c(r_{ij}) \tag{77}$$

Summing over all neighbors j gives the radial symmetry function the physical interpretation of a coordination number. The parameters η and r_s allow to respectively change the width and position of the Gaussian, which replicates the decaying atomic interactions. An example of radial symmetry functions of type 2 is depicted in figure 15.

Angular symmetry functions are constructed from a term containing the angle of all triplets of atoms $\alpha_{ijk} = \frac{\mathbf{R}_{ij} \cdot \mathbf{R}_{ik}}{R_{ij} R_{ik}}$, where i is the central atom and $\mathbf{R}_{ij} = \mathbf{R}_i - \mathbf{R}_j$ is the Cartesian distance between atom i and j:

$$G_i^4 = 2^{1-\zeta} \sum_{j \neq i} \sum_{k \neq i,j} \left[\left(1 + \lambda cos(\alpha_{ijk})\right)^{\zeta} e^{-\eta(r_{ij}^2 + r_{ik}^2 + r_{jk}^2)} f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk}) \right] \tag{78}$$

Here the parameters ζ, λ can be adjusted to fit different angular distributions. One uses the cosine function to account for the periodicity of the potential with respect to the angle α_{ijk}. Furthermore the angular part must be symmetric with respect to α_{ijk}. So the parameter $\lambda \in \{+1, -1\}$ is used to shift the maximum of the cosine to $\alpha_{ijk} = 0°$ or $\alpha_{ijk} = 180°$, as seen in figure 14. The angular resolution is controlled by the parameter ζ and can be investigated at different distances from the reference atom by varying η.
The symmetry functions are suitable for describing the local environment of an atom in order to determine its energy contribution. However, they can also be employed to identify various ice polymorphs, as shown by Geiger and Dellago [99].

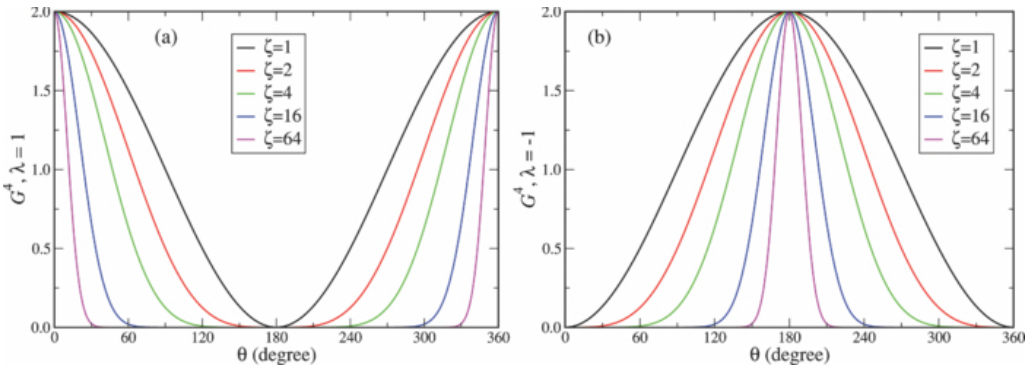


Figure 14: Angular contributions of angular symmetry functions of type G^4. The plotted angular part corresponds to a system of three atoms. In many-atom systems, the angular terms for all triples of atoms are summed. (from Behler [97])

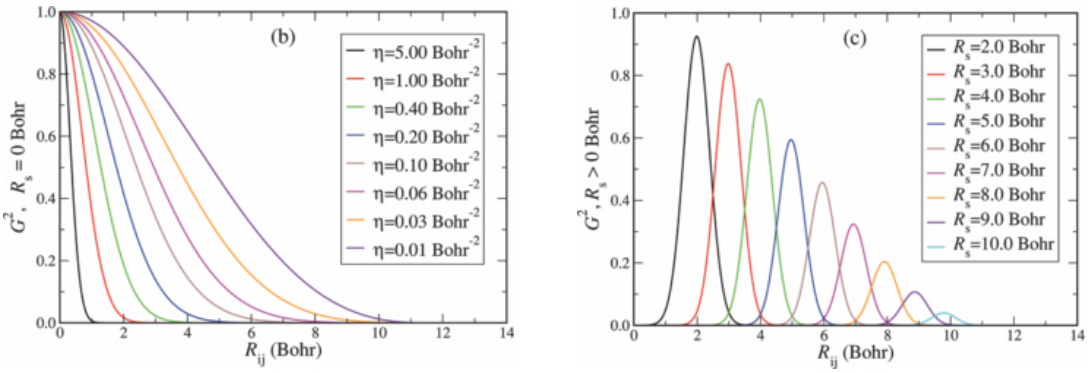


Figure 15: Radial symmetry function of type G^2 for an atom with one neighbor only. In many-atom systems, the radial symmetry function value of an atom is obtained by adding the contributions of each neighbour. (from Behler [97])

6.3 High Dimensional NNP

So far this chapter has introduced the structure of a feed forward neural network and the atom-centered symmetry functions. By using the symmetry functions suggested in section 6.2 one of the major challenges involving the evaluation of a NNP has been achieved; that is the potential energy of the system being invariant with respect to translation and rotation of the system. Hence properties (i) and (ii) outlined in section 6.2 are fulfilled. However points (iii) and (iv), namely the permutation symmetry and the scalability of the NN to high-dimensional systems of arbitrary size, are not satisfied yet.
The high dimensional NNP approach introduced by Behler and Parrinello [16] solves these remaining issues. This approach calculates the total potential energy of the system by using a set of Feed Forward Neural Networks instead of a single one. The structure of such a network can be seen in figure 16.

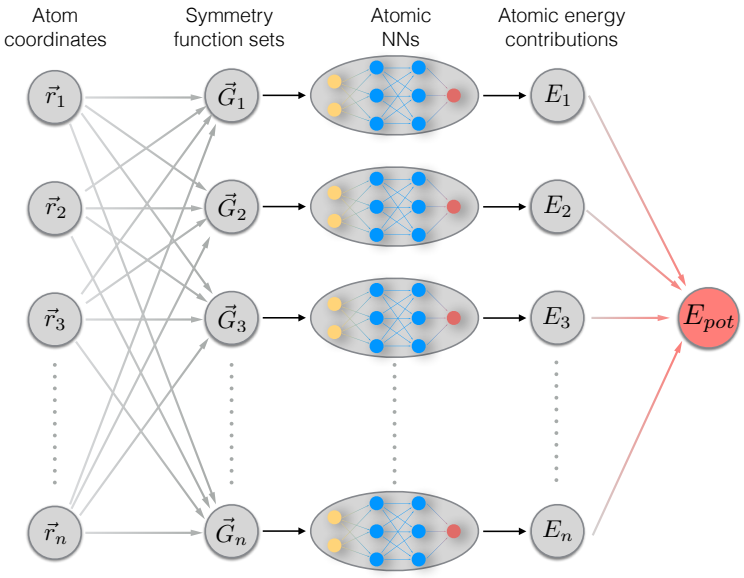


Figure 16: Representation of a High Dimensional Neural Network

First the Cartesian coordinates of all the atoms are used to create a set of symmetry functions $\mathbf{G}_i$ for each atom in the system. A set of $M = 30$ and $M = 27$ symmetry functions for each oxygen and hydrogen atom respectively has been used. Next each row uses a standard feed forward NN with two hidden layers of $N_{HL} = 25$ nodes each, to evaluate the atomic energy contribution of atom i. The total energy is obtained by summing the output of the individual atomic NNs, for each element:

$$E = \sum_{i=1}^{N^{hydrogen}} E_i^{hydrogen} + \sum_{j=1}^{N^{oxygen}} E_i^{oxygen} \tag{79}$$

The output of each atomic NN is then given by equation 74, which is for the oxygen and hydrogen atoms:

$$E^{hydrogen} = f_1^3 \left\{ b_1^3 + \sum_{k=1}^{25} a_{k1}^{23} f_k^2 \left[b_k^2 + \sum_{j=1}^{25} a_{jk}^{12} f_j^1 \left(b_j^1 + \sum_{i=1}^{27} a_{ij}^{01} G_i \right) \right] \right\} \tag{80}$$

$$E^{oxygen} = f_1^3 \left\{ b_1^3 + \sum_{k=1}^{25} a_{k1}^{23} f_k^2 \left[b_k^2 + \sum_{j=1}^{25} a_{jk}^{12} f_j^1 \left(b_j^1 + \sum_{i=1}^{30} a_{ij}^{01} G_i \right) \right] \right\} \tag{81}$$

Notice that no explicit long-range electrostatics are considered here. An electrostatic term could be included based on environment dependent charges [100]. This, however, is not necessary for a cutoff radius of $r_c = 6.32\,\text{Å}$ as stated in the supporting information of Morawietz et al.[1]. Consequently Ewald sums can be avoided, leading to linear scaling of computational cost with system size. From the functional forms of the potential energies, one can take the negative gradient to obtain the forces acting on the individual atoms. The force component F_{α_k}, resulting from N atomic interactions, acts on atom k with respect to the Cartesian coordinate $\alpha = x, y$ or z and is given by:

$$F_{\alpha_k} = -\frac{\partial E}{\partial \alpha_k} = -\sum_{i=1}^{N} \frac{\partial E_i}{\partial \alpha_k} = -\sum_{i=1}^{N} \sum_{j=1}^{M_i} \frac{\partial E_i}{\partial G_{i,j}} \frac{\partial G_{i,j}}{\partial \alpha_k} \tag{82}$$

Note that $\frac{\partial E_i}{\partial G_{i,j}}$ is given by the architecture of the feed forward NN and $\frac{\partial G_{i,j}}{\partial \alpha_k}$ is defined by the functional form of the symmetry functions. M_i denotes the number of symmetry functions of atom i.

Hitherto, this chapter explained how the energies and forces are evaluated by the Neural Network. Furthermore the high-dimensional NNP approach, which can be reviewed in more detail here [17], satisfies the last two conditions (iii) and (iv). Since for each element the *same* type of symmetry functions are used, and since the resulting energy given by equation 79 is independent of the order of summation, permutation invariance and scalability of the NN to arbitrary system sizes is guaranteed. Ultimately the HDNN approach outlined in this section, is able to reproduce ab-initio quality PESs at low costs.

The actual "training" of the NN has already been achieved [1], and the resulting NNP representing the RPBE and BLYP density functionals with and without van der Waals interactions are used in this thesis. The VdW corrections were implemented using the D3 method proposed by Grimme [18]. Evaluating the NNP is computationally less demanding than computing the correction term. Consequently, the forces and energies obtained from DFT calculation were rectified with the D3 method before training the NN. It is noteworthy that the fitting parameters a_{ij}^{kl} and b_j^i are determined by using the total energy and the forces, not the individual atomic energy, as target quantity. After comparison with energies and forces obtained from DFT, the total number $N_w = 2827$ (see equation 75) of weights are optimized using an extended Kalman filtering method. This method uses statistical noise from the time series of observables in order to estimate unknown variables. For a detailed description of Kalman filtering we refer to [101]. Last but not least, one should mention that the NNPs were trained using the code RuNNer, developed by Behler [102].

7 Results

7.1 Simulation Details

All molecular dynamics simulations have been performed with the program LAMMPS [29] extended with four neural network potentials (NNP), developed in our research group [1] and based on the high-dimensional approach of Behler [16]. The NNPs were trained to represent the RPBE and BLYP density functionals with and without van der Waals corrections using Grimme's D3 method [18]. The local environment of the hydrogen and oxygen atoms was described using a total of 27 and 30 atom centered symmetry functions [97], respectively. The cutoff radius was set to $r_c = 6.36$ Å, which is long enough not to require explicit consideration of long-range electrostatics [1]. Consequently, Ewald sums can be omitted, leading to linear scaling of the computational costs. The melting temperatures of ice Ih at $p = 1$ bar for the different NNPs were previously determined within our research group [1], by using the interface pinning method [46] and can be seen in table 2. For each coexistence point the Gibbs-Duhem method [2],[3] was applied to obtain the whole melting curve. The Clausius-Clapeyron equation $\frac{dT}{dp} = \frac{T\Delta v}{\Delta h}$ was integrated with a step size in pressure of $dp = 250$ atm using a 5^{th} order Runge Kutta scheme [49]. The coexistence line is traced for positive as well as negative pressures. A system size of 360 molecules (1080 atoms) with periodic boundary conditions was used to simulate bulk water. Simulations were carried out in the NpT ensemble using the equations of motion of Shinoda et al.[30] with a time step of $dt = 0.5$ fs. Anisotropic and isotropic-NpT simulations were used for the solid and the liquid phase, respectively. The enthalpy and density were calculated as running average during 0.1 ns, where 0.05 ns were discarded for equilibration. The use of parallel computing allowed to simulate each phase simultaneously and thus efficiently determine the integrand of equation (55).

	RPBE	RPBE-D3	BLYP	BLYP-D3
T_m(K)	265 ± 2	269 ± 3	321 ± 3	272 ± 2

Table 2: Melting temperatures of ice Ih at $p = 1$bar, obtained from interface pinning simulations from the work of Morawietz et al. [1]

7.2 Coexistence lines between ice Ih and liquid water

The Gibbs-Duhem integration was applied to the four NNPs, with the melting temperature of ice Ih at $p = 1$ bar serving as initial state. The enthalpy and density of the solid and liquid phase have been computed along the coexistence line. The results of the melting point for the different functionals are compared to rigid models and experimental values in table 3.
One can see that the NNP trained on the BLYP-D3 functional delivers the best melting temperature. It should be noted that the melting temperatures have been obtained from NNP interface pinning calculations. The melting temperature of the reference DFT method is slightly higher, due to small differences between DFT and NNP energies (see Appendix of Ref.[1]). The enthalpy of fusion is overestimated by both D3 corrected functionals. Nevertheless these ab-initio quality results are better than rigid model results, which underestimate the enthalpy of fusion. The densities from the functionals with Van der Waals corrections, which reproduce the density peaks (see figure 12), are a bit off from experimental values. Anyhow, the DFTs with dispersion correction drastically improve thermodynamic properties of water compared to the non-corrected functionals. Most importantly the density of liquid water is estimated to be greater than the density of ice Ih. Note here that the Clausius-Clapeyron equation evaluates the difference in molar volume, not the difference in density. Since the molar volume is proportional to the inverse of the density, Δv will have the opposite sign of $\Delta\rho$. This results in the negative slope, seen in the outer right column of table 3, which provokes the melting line to curve to lower temperatures with increasing pressure. Since for the non-corrected functionals the density of liquid water is underestimated, the slope is positive. The RPBE-D3 functional overestimates the initial slope, while the BLYP-D3 underestimates it.
In general the initial state point is best approximated by the TIP4P/Ice model. However, this model has been parametrized to reproduce the experimental properties, and is thus, unlike the density functionals, not transferable to different regions of the phase diagram. Furthermore the rigid models do not allow breaking and formation of bonds, this encourages the use of an ab-initio quality high dimensional neural network potential.

Table 3: Melting properties of ice Ih at $p = 1$ bar for different models. T_m is the melting temperature; H_l and H_s are the enthalpies of liquid water and ice (the $3RT$ term arising from the translational and rotational kinetic terms is not included for the rigid models); $\Delta H = H_l - H_s$ is the melting enthalpy; ρ_l and ρ_s are the densities of the liquid and the solid phase; the difference in density $\Delta\rho = \rho_l - \rho_s$: dp/dT is the slope of the coexistence curve. - Data for the rigid models taken from [12],[13] and [103]. Melting temperatures for the density functionals taken from [1]

Model	T_m (K)	H_l (kcal/mol)	H_s (kcal/mol)	ΔH (kcal/mol)	ρ_l (g/cm^3)	ρ_s (g/cm^3)	$\Delta\rho$ (g/cm^3)	dp/dT (bar/K)
SPC/E	215.0	-12.49	-13.23	0.74	1.011	0.950	0.061	-126
TIP4P	232.0	-10.98	-12.03	1.05	1.002	0.940	0.062	-160
TIP4P/2005	252.1	-12.17	-13.33	1.16	0.993	0.921	0.072	-135
TIP4P/Ice	272.2	-11.66	-12.95	1.29	0.985	0.906	0.079	-120
Experiment	273.15			1.44	0.999	0.917	0.082	-137
RPBE	265	-436.404	-438.523	2.12	0.672	0.786	-0.114	97
RPBE-D3	269	-440.346	-441.962	1.617	0.901	0.863	0.038	-282.72
BLYP	321	-430.785	-432.545	1.85	0.759	0.842	-0.083	99
BLYP-D3	272	-435.139	-436.621	1.482	1.051	0.914	0.138	-88.49

First the coexistence points obtained from the GDI shall be discussed. Tables 4 to 9 show the state points at coexistence for the different functionals.

The data collected for the RPBE-D3 functional for positive pressures can be seen in table 4. One can see that each increment in pressure $dp = 250$ atm results in a lower temperature. As the pressure goes up the enthalpy of fusion decreases. Through out the GDI, the density of the liquid phase remains higher than the solid phase. As the density difference and enthalpy difference between the two phases increases and decreases respectively, the slope of the coexistence line decreases.

The results of the GDI for the RPBE-D3 functional at negative pressures are shown in table 5. The first thing to notice is the change in sign of the density difference between $p = -1012.25$ bar and $p = -1265.56$ bar. The density of the liquid phase is decreasing faster than it's solid counterpart, and thus causes a change in slope of the coexistence line. This so called reentrant behaviour has already been observed for the TIP4P model around $p = -1500$ bar [54]. It suggests that a liquid phase can be obtained from a solid phase by applying negative pressure. It is the first time that such a behaviour for the melting line of ice Ih has been reproduced from ab-initio quality calculations. Our findings encourage the experimental search for this effect.

The values obtained for the BLYP-D3 functional are reported in table 6 and 7 for positive and negative pressures respectively. Compared to the RPBE-D3 functional, the density difference between the two phases is much higher. As a result the slope of the coexistence line is lower than the one obtained for the RPBE-D3 based NNP. Again the reentrant behaviour is observed but at much lower pressures $p = -2278.81$ bar and $p = -2532.13$ bar.

For completeness this report includes the data obtained for the RPBE and BLYP functionals without the D3-correction term in table 8 and 9. It has been shown that these two functionals do not reproduce the density maximum for liquid water [1] (see figure 12 in section 5.2). Hence it is of no surprise that they do not reproduce the correct densities, resulting in a negative density difference between liquid and solid phase. Consequently the slope of the melting line is always positive.

Table 4: Thermodynamic properties along the coexistence line between liquid water and ice Ih: Results for the **RPBE-D3** functional at *positive pressures*

T (K)	p (bar)	H_l (kcal/mol)	H_s (kcal/mol)	ΔH (kcal/mol)	ρ_l (g/cm^3)	ρ_s (g/cm^3)	$\Delta\rho$ (g/cm^3)	dp/dT (bar/K)
269.00	1.00	-440.346	-441.962	1.617	0.901	0.863	0.038	-282.72
268.11	254.31	-440.250	-441.858	1.608	0.909	0.867	0.042	-263.09
267.08	507.62	-440.146	-441.762	1.616	0.924	0.870	0.053	-211.81
265.84	760.94	-440.008	-441.673	1.665	0.934	0.873	0.061	-196.15
264.62	1014.25	-439.977	-441.583	1.606	0.940	0.876	0.064	-181.34
263.20	1267.56	-439.934	-441.490	1.556	0.950	0.879	0.071	-162.39
261.62	1520.87	-439.847	-441.416	1.569	0.960	0.882	0.077	-152.46
259.95	1774.19	-439.726	-441.336	1.610	0.966	0.886	0.081	-153.05
258.21	2027.50	-439.707	-441.248	1.541	0.975	0.889	0.086	-139.34
256.36	2280.81	-439.637	-441.176	1.539	0.980	0.891	0.089	-136.66
254.31	2534.12	-439.627	-441.100	1.474	0.994	0.894	0.101	-118.82
252.24	2787.44	-439.532	-441.033	1.501	1.002	0.898	0.105	-118.97
250.09	3040.75	-439.551	-440.960	1.409	1.004	0.900	0.103	-114.44
247.75	3294.06	-439.501	-440.896	1.395	1.014	0.904	0.110	-109.42
245.31	3547.37	-439.439	-440.829	1.391	1.023	0.907	0.115	-105.94
242.89	3800.69	-439.434	-440.765	1.332	1.033	0.909	0.124	-96.55
240.25	4054.00	-439.411	-440.699	1.288	1.037	0.912	0.125	-94.44
237.56	4307.31	-439.418	-440.645	1.227	1.047	0.915	0.132	-87.21
234.69	4560.62	-439.438	-440.586	1.148	1.046	0.918	0.128	-85.25
231.41	4813.94	-439.353	-440.539	1.186	1.063	0.920	0.142	-81.91
228.17	5067.25	-439.437	-440.483	1.045	1.060	0.924	0.137	-76.36

Table 5: Thermodynamic properties along the coexistence line between liquid water and ice Ih: Results for the **RPBE-D3** functional at *negative pressures*

T (K)	p (bar)	H_l (kcal/mol)	H_s (kcal/mol)	ΔH (kcal/mol)	ρ_l (g/cm^3)	ρ_s (g/cm^3)	$\Delta\rho$ (g/cm^3)	dp/dT (bar/K)
269.00	1.00	-440.346	-441.962	1.617	0.901	0.863	0.038	-282.72
269.71	-252.31	-440.420	-442.065	1.644	0.887	0.861	0.026	-411.74
270.28	-505.63	-440.541	-442.159	1.618	0.880	0.857	0.023	-455.86
270.68	-758.94	-440.727	-442.270	1.543	0.864	0.854	0.011	-929.43
271.02	-1012.25	-440.707	-442.382	1.675	0.857	0.851	0.006	-1702.29
271.08	-1265.56	-440.838	-442.488	1.650	0.846	0.847	-0.001	14851.78
270.89	-1518.88	-441.023	-442.624	1.602	0.832	0.845	-0.013	745.73
270.37	-1772.19	-441.165	-442.744	1.578	0.824	0.841	-0.018	537.02
269.56	-2025.50	-441.318	-442.870	1.552	0.797	0.837	-0.040	224.97

Table 6: Thermodynamic properties along the coexistence line between liquid water and ice Ih: Results for the **BLYP-D3** functional at *positive pressures*

T (K)	p (bar)	H_l (kcal/mol)	H_s (kcal/mol)	ΔH (kcal/mol)	ρ_l (g/cm^3)	ρ_s (g/cm^3)	$\Delta\rho$ (g/cm^3)	dp/dT (bar/K)
272	1	-435.139	-436.621	1.482	1.051	0.914	0.138	-88.49
269.08	254.31	-435.184	-436.570	1.386	1.059	0.917	0.142	-81.85
266.02	507.62	-435.154	-436.525	1.371	1.071	0.921	0.150	-78.86
262.52	760.94	-435.164	-436.484	1.320	1.081	0.924	0.157	-74.49
259.16	1014.25	-435.164	-436.433	1.269	1.096	0.926	0.169	-68.22
255.33	1267.56	-435.171	-436.401	1.230	1.127	0.930	0.197	-59.62
250.95	1520.87	-435.201	-436.379	1.178	1.127	0.933	0.194	-59.23
246.29	1774.19	-435.241	-436.358	1.117	1.130	0.936	0.194	-57.54
241.21	2027.50	-435.355	-436.339	0.984	1.152	0.939	0.213	-48.29
235.47	2280.81	-435.395	-436.343	0.947	1.179	0.942	0.237	-43.89
229.25	2534.12	-435.548	-436.359	0.810	1.161	0.945	0.216	-41.75
223.29	2787.44	-435.562	-436.366	0.804	1.180	0.949	0.232	-40.48

Table 7: Thermodynamic properties along the coexistence line between liquid water and ice Ih: Results for the **BLYP-D3** functional at *negative pressures*

T (K)	p (bar)	H_l (kcal/mol)	H_s (kcal/mol)	ΔH (kcal/mol)	ρ_l (g/cm^3)	ρ_s (g/cm^3)	$\Delta\rho$ (g/cm^3)	dp/dT (bar/K)
272	1	-435.139	-436.621	1.482	1.051	0.914	0.138	-88.49
274.81	-252.31	-435.167	-436.682	1.515	1.038	0.912	0.126	-96.66
277.32	-505.63	-435.203	-436.740	1.537	1.020	0.908	0.112	-106.37
279.69	-758.94	-435.262	-436.812	1.550	0.995	0.905	0.090	-128.52
281.81	-1012.25	-435.316	-436.873	1.558	0.992	0.902	0.090	-128.27
283.74	-1265.56	-435.380	-436.949	1.568	0.985	0.899	0.086	-131.67
285.15	-1518.88	-435.432	-437.037	1.604	0.968	0.896	0.072	-158.64
286.55	-1772.19	-435.483	-437.121	1.638	0.944	0.893	0.051	-220.33
287.31	-1999.01	-435.530	-437.190	1.660	0.923	0.891	0.032	-346.08
287.63	-2278.81	-435.753	-437.328	1.575	0.911	0.887	0.024	-423.32
288.19	-2532.13	-435.842	-437.429	1.586	0.880	0.884	-0.003	2950.45
287.86	-2785.44	-436.093	-437.549	1.455	0.837	0.881	-0.044	196.89

Table 8: Thermodynamic properties along the coexistence line between liquid water and ice Ih: Results for the **RPBE** functional

T (K)	p (bar)	H_l (kcal/mol)	H_s (kcal/mol)	ΔH (kcal/mol)	ρ_l (g/cm^3)	ρ_s (g/cm^3)	$\Delta\rho$ (g/cm^3)	dp/dT (bar/K)
261.98	-252.31	-436.690	-438.706	2.016	0.661	0.782	-0.121	76.35
265.00	1.00	-436.404	-438.523	2.119	0.672	0.786	-0.114	86.13
267.72	254.31	-436.335	-438.338	2.003	0.690	0.788	-0.098	96.67
270.17	507.62	-436.180	-438.179	1.999	0.704	0.793	-0.089	108.11
272.40	760.94	-435.974	-438.007	2.033	0.712	0.795	-0.083	118.18
274.38	1014.25	-435.771	-437.848	2.077	0.721	0.798	-0.078	130.14
276.25	1267.56	-435.644	-437.687	2.043	0.732	0.802	-0.069	145.83
277.96	1520.87	-435.451	-437.536	2.085	0.737	0.806	-0.068	151.23
279.49	1774.19	-435.297	-437.374	2.077	0.748	0.808	-0.060	173.28
280.92	2027.50	-435.213	-437.233	2.019	0.756	0.811	-0.056	183.74
282.25	2280.81	-434.124	-436.175	2.051	0.760	0.812	-0.053	198.36
283.43	2534.12	-434.872	-436.932	2.060	0.770	0.815	-0.046	232.72
284.50	2787.44	-434.683	-436.796	2.114	0.776	0.819	-0.043	255.85
285.46	3040.75	-434.634	-436.660	2.027	0.784	0.824	-0.039	270.71
286.32	3294.06	-434.394	-436.518	2.124	0.787	0.826	-0.038	292.14
287.12	3547.37	-434.309	-436.380	2.071	0.796	0.829	-0.033	336.42
287.83	3800.69	-434.155	-436.257	2.102	0.801	0.831	-0.030	374.69
288.43	4054.00	-434.103	-436.117	2.014	0.809	0.835	-0.026	421.65
288.98	4307.31	-433.956	-435.994	2.038	0.814	0.837	-0.022	497.37
289.47	4560.62	-433.814	-435.863	2.049	0.821	0.840	-0.019	599.75
289.85	4813.94	-433.601	-435.743	2.142	0.824	0.843	-0.019	617.99
290.20	5067.25	-433.521	-435.626	2.105	0.832	0.847	-0.016	755.16

Table 9: Thermodynamic properties along the coexistence line between liquid water and ice Ih: Results for the **BLYP** functional

T (K)	p (bar)	H_l (kcal/mol)	H_s (kcal/mol)	ΔH (kcal/mol)	ρ_l (g/cm^3)	ρ_s (g/cm^3)	$\Delta\rho$ (g/cm^3)	dp/dT (bar/K)
311.57	-758.94	-431.228	-433.078	1.850	0.713	0.834	-0.121	68.15
315.17	-505.63	-431.114	-432.884	1.770	0.736	0.835	-0.100	80.61
318.27	-252.31	-430.838	-432.715	1.877	0.742	0.839	-0.097	88.07
321.00	1.00	-430.785	-432.545	1.759	0.759	0.842	-0.082	98.72
323.55	254.31	-430.698	-432.357	1.658	0.770	0.842	-0.072	107.14
326.00	507.62	-430.385	-432.204	1.820	0.773	0.846	-0.073	116.91
328.19	760.94	-430.255	-432.045	1.790	0.786	0.849	-0.063	134.37
330.12	1014.25	-429.985	-431.880	1.896	0.784	0.850	-0.066	135.45
331.90	1267.56	-429.892	-431.739	1.847	0.793	0.853	-0.061	144.40
333.57	1520.87	-429.722	-431.589	1.868	0.797	0.856	-0.059	150.18
335.20	1774.19	-429.752	-431.438	1.686	0.809	0.858	-0.050	163.55
336.64	2027.50	-429.275	-431.291	2.016	0.809	0.860	-0.052	187.58
337.92	2280.81	-429.212	-431.142	1.930	0.816	0.862	-0.046	202.72
339.08	2534.12	-429.033	-431.006	1.973	0.821	0.865	-0.044	219.35
340.20	2787.44	-428.928	-430.873	1.945	0.828	0.868	-0.040	239.31

The major result of this thesis is the visualization of the coexistence lines within the phase diagram in figure 17. One can see the non-corrected functionals RPBE and BLYP in blue and red empty circles, respectively. The corrected functionals RPBE-D3 and BLYP-D3 are drawn in blue and red filled circles, respectively.
The two functionals without D3 correction yield melting lines with a positive slope, leading to higher coexisting temperatures for increasing pressures. These coexistence lines are in poor agreement with experimental values. The BLYP functional produces by far the worst description. Since for this functional the melting point has been largely overestimated ($T = 321$ K), the resulting melting curve is found at much higher temperatures. The RPBE functional closely reproduces the melting temperature at ambient conditions ($T = 265$ K), yet the curvature of the line is still tilted in the wrong direction. Only a few coexistence points have been obtained in the negative pressure regime, because these conditions lie outside of the training region of the neural network.
The incorporation of Van der Waals interactions considerably improves the coexistence lines. The NNP based on the RPBE-D3 functional slightly overestimates the slope of the melting line. Coexistence points are obtained up into the region of ice V. One could have integrated further, yet this was enough to visualize the curvature of the melting line. In the negative pressure regime one can see the reentrant behaviour. The turning point in the sign of the slope happens earlier than for the TIP4P and TIP4P/2005 models. The results for the BLYP-D3 functional slightly underestimate the slope of the coexistence line. Here data points have been collected until the region of ice II. For negative pressures the reentrant point is obtained later than for the RPBE-D3 functional.
Error bars are included for the dispersion corrected functionals. First, the error on the slope for each coexistence point is evaluated, by calculating the standard deviation of block averages of the density and enthalpy.

$$\delta s_i = |s_i| \sqrt{\left(\frac{\delta \Delta \rho}{\Delta \rho}\right)^2 + \left(\frac{\delta \Delta H}{\Delta H}\right)^2} \tag{83}$$

Next the absolute error in temperature for each coexistence point was obtained by comparing relative errors:

$$\frac{\delta T}{dT} = \frac{\delta s_i}{s_i} \tag{84}$$

$$\iff \delta T = |\frac{\delta s_i}{s_i}(T_1 - T_0)| \tag{85}$$

Finally, by assuming that each point on the melting line is independent from the others, all Gaussian errors are summed up to obtain an error bar on the last point.

$$\delta s = \sum_i \delta s_i \tag{86}$$

This error thus serves as measure to assess the precision of the Gibbs-Duhem integration. Error bars range from 1.9 K to 5.1 K. These values can be reduced by choosing longer equilibration times during the GDI.

Ultimately the Gibbs-Duhem method applied to the four density functionals, results in an ab-initio quality melting lines of ice Ih at positive and negative pressures. It is astonishing to see how powerful the impact of the van der Waals interactions is on the phase boundaries. Most notable is the change in curvature due to the addition of dispersion corrections. Furthermore the correction of the initial coexistence point, namely the melting temperature at $p = 1$ bar, improves the results. The amplitude of the D3 correction depends on the functional (see equation (66)), therefore the results for the BLYP functional suffer a greater change than the RPBE functional when D3 corrections are applied.
The lines of density maxima (LDM), computed by Singraber et al. [104] for the RPBE-D3 and BLYP-D3 functionals are shown in figure 17, in dark blue and dark red, respectively. In both cases, the LDM suffers a change of slope, causing the line to bend back to colder temperatures for sufficiently negative pressures. This change of slope occurs around pressures of $p \approx -20$ MPa and $p \approx -200$ MPa for the RPBE-D3 and BLYP-D3 functional, respectively. It is emphasized again that the coexistence line is determined by the density difference of the liquid and the solid phase. The positive slope of the dispersion corrected melting lines results from the fact that the D3 corrections reproduce higher densities for the liquid phase. As one can see in figure 17, the upper part of the RPBE-D3 line of density maxima closely follows the curvature of the RPBE-D3 coexistence line. This indicates that the density of the liquid phase along the coexistence line is close to the maximum density of water based on the RPBE-D3 functional.

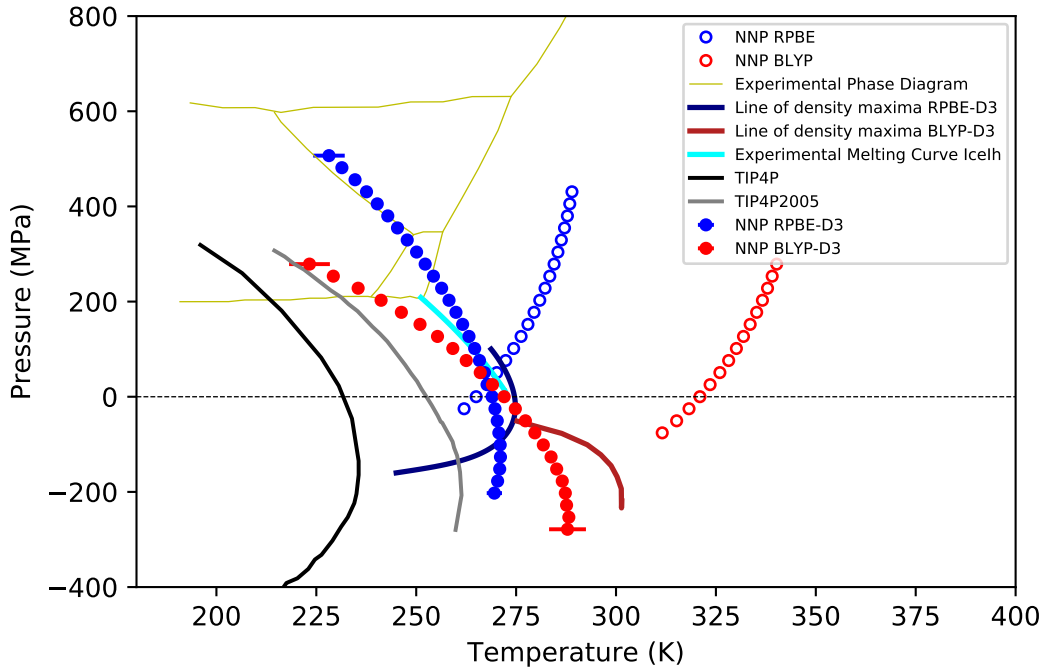


Figure 17: Phase diagram of water obtained from NNP simulations based on the RPBE (blue empty circles), RPBE-D3 (blue filled circles), BLYP (red empty circles) and BLYP-D3 (red filled circles) density functionals. Errors bars have been included for the RPBE-D3 and BLYP-D3 functionals. The experimental phase diagram (golden lines) is from Sanz et al.[54]. The melting curve of ice Ih (cyan) is from the IAPWS [105]. The melting line of ice Ih for the TIP4P (black line) and TIP4P/2005 (grey line) models are from Sanz [61] and Abascal [12], respectively. The lines of density maxima based on the RPBE-D3 (dark blue line) and BLYP-D3 (dark red line) were determined by Singraber et al. [104]

7.3 Structural Properties

In order to assess the liquid nature at each coexistence point this thesis analyzes the structure of the configuration by evaluating the radial distribution function (RDF) between oxygen atoms. The cutoff distance has been chosen at 7 Å and the number of bins was set to 360. The results for $g_{OO}(r)$ for the state points along the melting line are seen in figure 18 and figure 19 for the RPBE-D3 and BLYP-D3 functionals, respectively. One can see that the first and second peak are shifted to the left as one goes along the coexistence line. This is expected, because for lower temperatures and higher pressures the structure becomes more packed. The distance of the nearest neighbours thus becomes smaller.
The experimental oxygen-oxygen RDF of water at ambient conditions has been extracted from X-ray scattering data by Hura et al. [33]. The results for the BLYP-D3 functional do show some unexpected behaviour. First the RDF for $T = 287$ K, $p = -1999$ atm shows some disturbances. These are due to simulation times being not long enough for this particular state point. Next, in the high pressure regime and at low temperatures, (for $T = 223$ K, $p = 2750$ atm and $T = 235$ K, $p = 2250$ atm), the second order peak shrinks considerably. It is shifted partially to the left, leading to the formation of a peak much closer to the first order one, and partially to the right, increasing the height of the third order peak. The origin of this behaviour might be related to short equilibration times.
Beside these shortcomings the RDFs along the coexistence lines of the RPBE-D3 and BLYP-D3 functionals are qualitatively good.

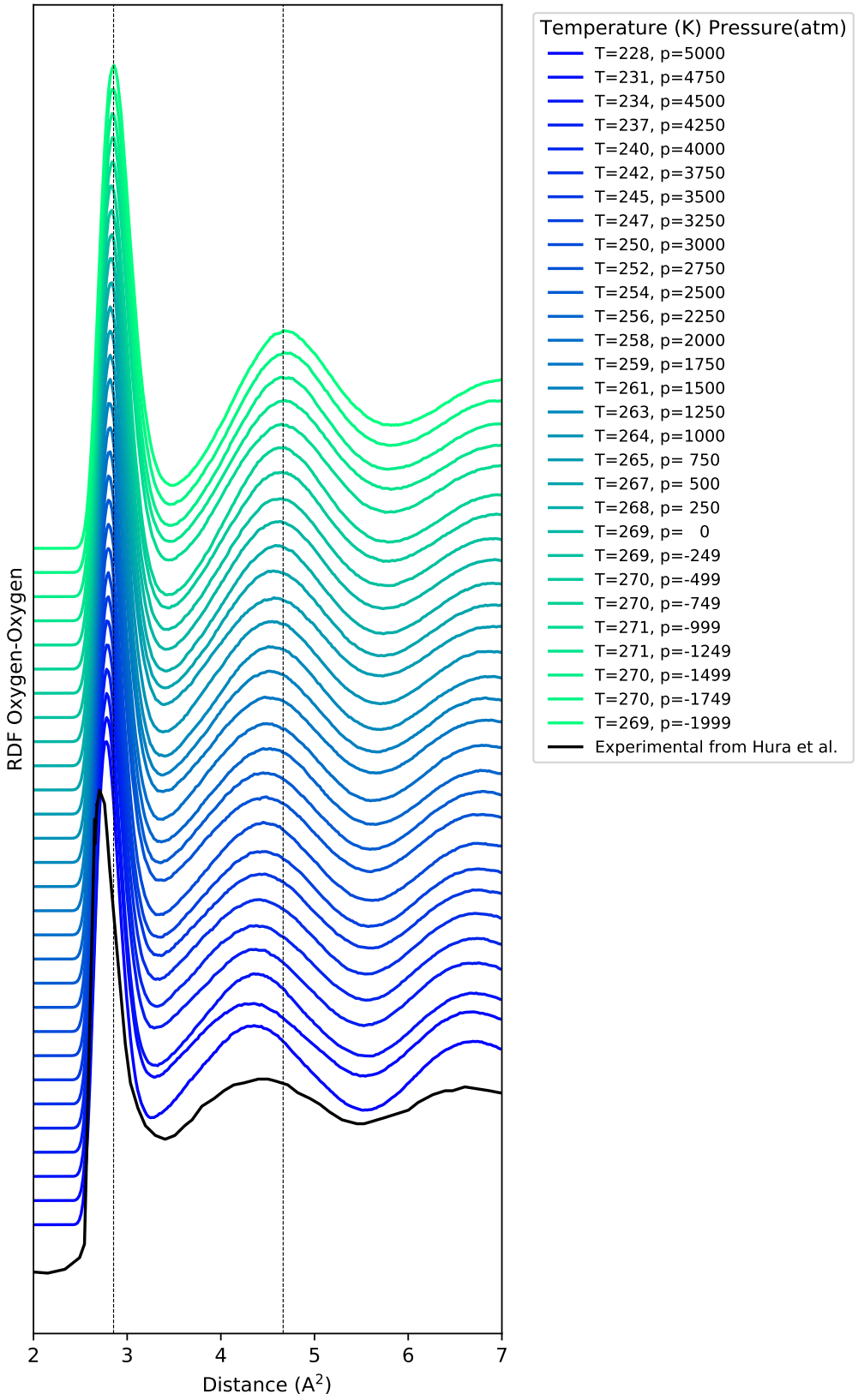


Figure 18: Radial distribution function of the Oxygen-Oxygen atoms along the coexistence line of RPBE-D3

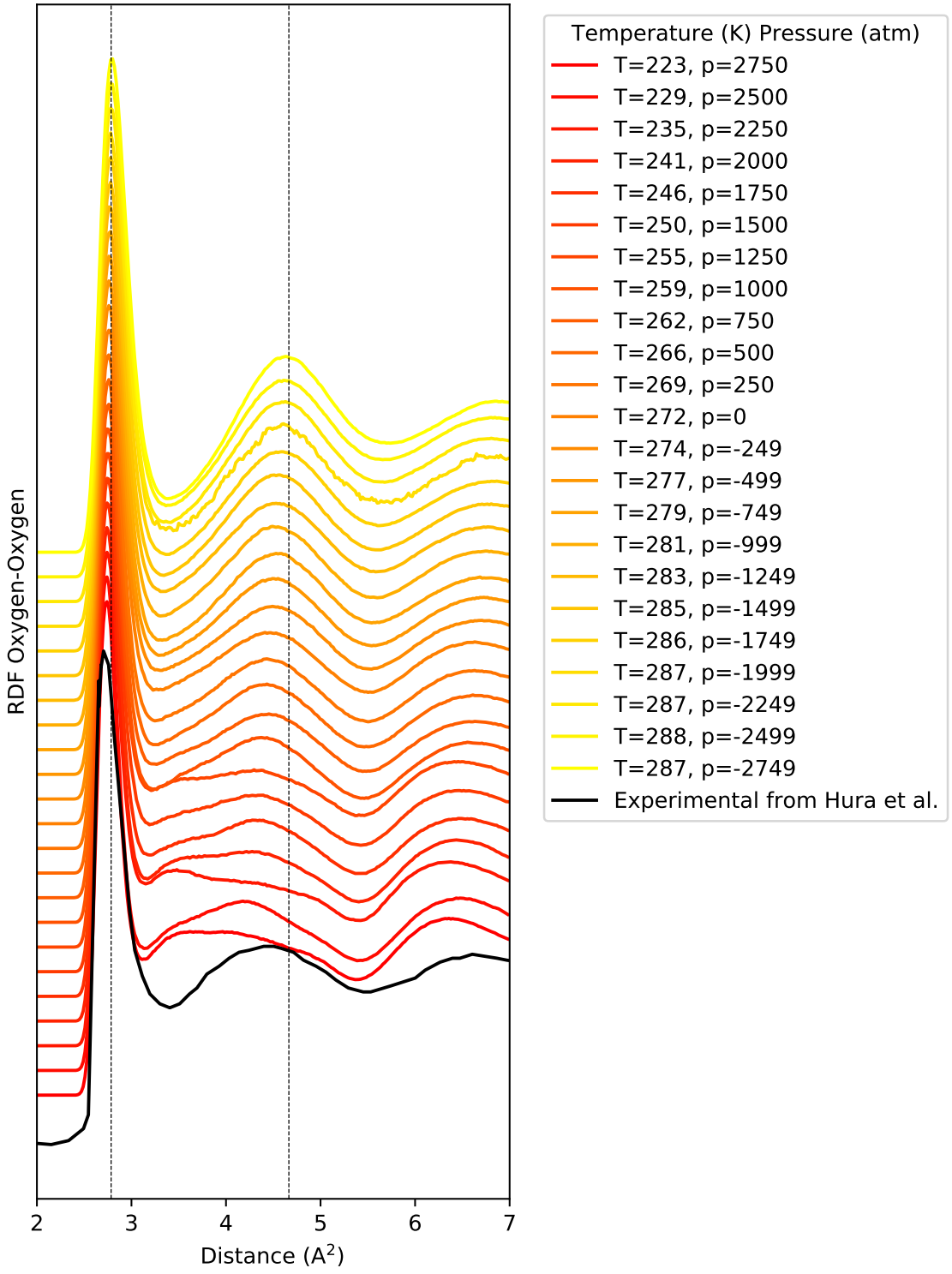


Figure 19: Radial distribution function of the Oxygen-Oxygen atoms along the coexistence line of BLYP-D3

7.4 Dynamical Properties

The previous section showed that the structure of the coexistence points is indeed liquid-like. To verify that dynamical properties of liquid water are fulfilled, we calculate the mean squared displacement (MSD) of the oxygen atoms. An efficient computation of the MSD is achieved by using the Fast Fourier Transform [35]. The result for the MSD at positive pressures, using the RPBE-D3 functional is given in figure 20. The slope decreases along the coexistence line, indicating that the movement of the particles reduces along the coexistence line. The small oscillations seen in the lines of figure 20 correspond to the damping time of the thermostat.

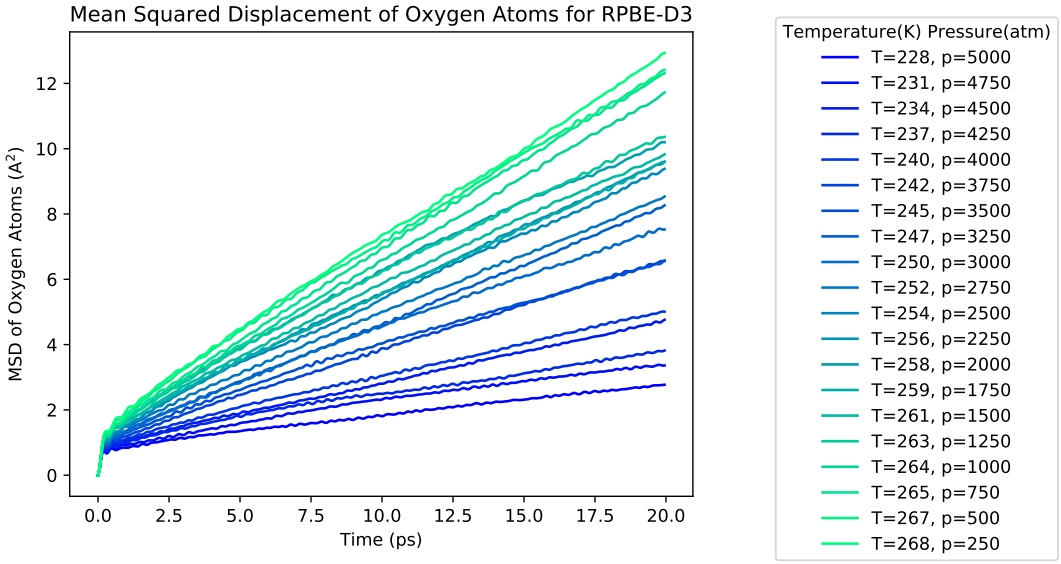


Figure 20: MSD of the Oxygen atoms at positive pressures for the RPBE-D3 functional

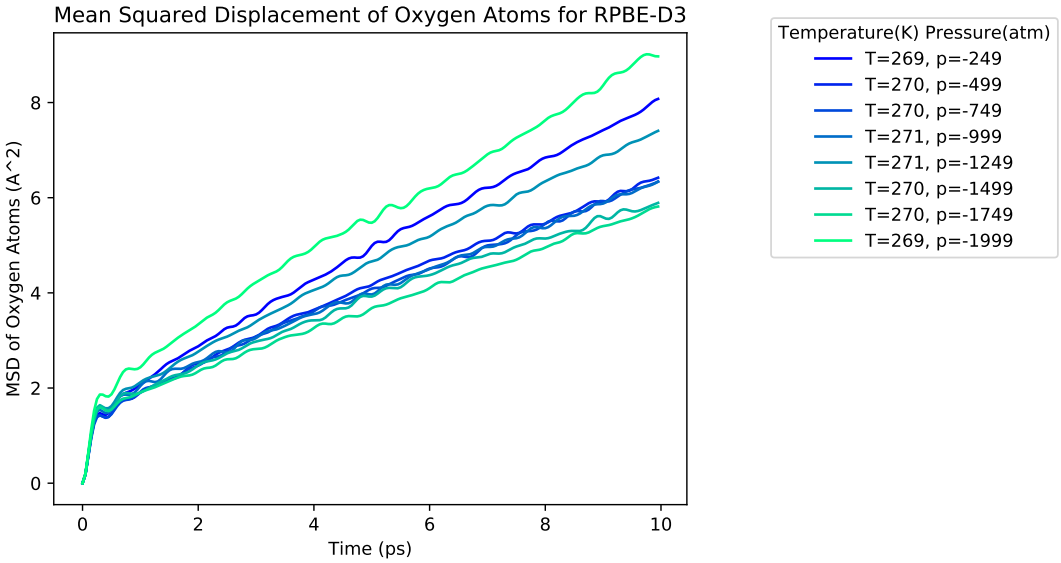


Figure 21: MSD of the Oxygen atoms at negative pressures for the RPBE-D3 functional

The diffusion coefficient can be computed from the slope of the MSD by using the Einstein diffusion equation 41. Next one can plot the pressure against the diffusion coefficient as seen in figure 22. It can be seen that the diffusion coefficient decreases with rising pressure along the coexistence line. In the negative pressure regime the line is strongly discontinuous, as already expected from figure 21. The origin of this behaviour remains unknown. It is important to emphasize that the results for the diffusion coefficient should be corrected for finite size effects:

$$D = D_{PBC} + 2.837297 \frac{k_B T}{6\pi\eta L_{box}} \tag{87}$$

where D_{PBC} is the diffusion coefficient before finite size correction, L_{box} is the length of the simulation box with periodic boundary conditions, k_B is the Boltzmann constant, T is the temperature and η is the viscosity. The viscosity can be computed form the stress autocorrelation function using the Green-Kubo formula. The aim of this thesis, however, is not to compute exact values of the diffusion coefficient, but rather to show that the liquid phase simulated along the coexistence line still behaves like a liquid. Hence by comparing figure 20 and 18, one sees that the first peak in the RDF is around 2.7 Å and that after 20 ps the atoms moved $\sqrt{12} \approx 3.46$ Å. Consequently oxygen atoms do move further than their nearest neighbours. As one goes along the melting line this movement decreases.

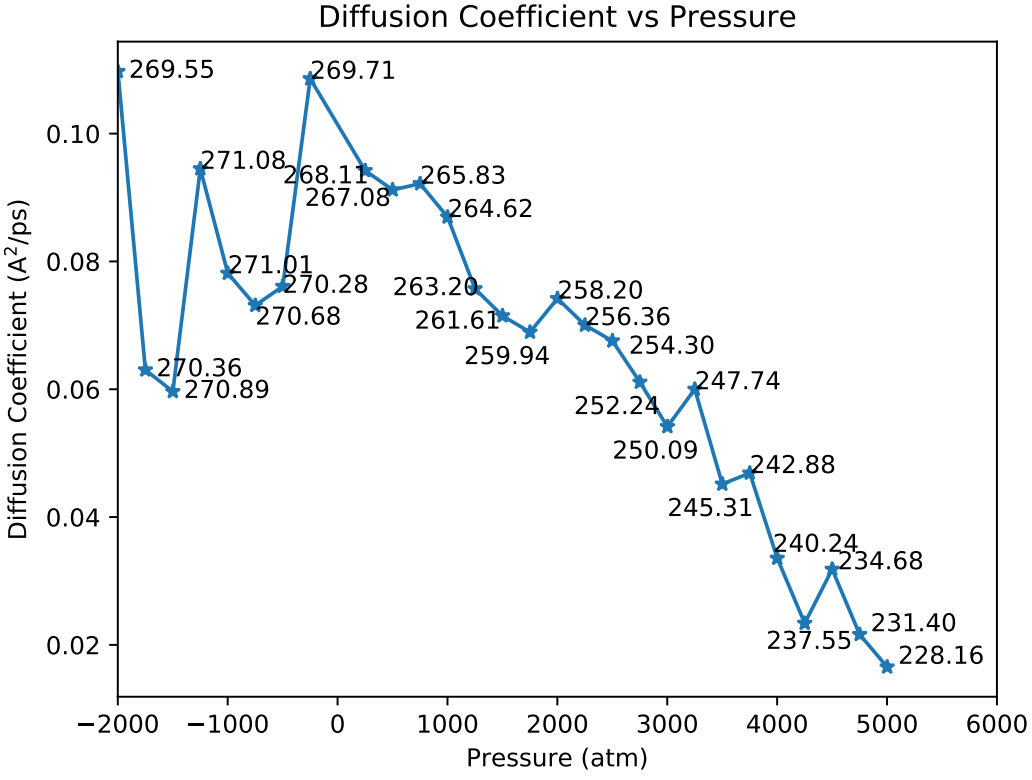


Figure 22: Pressure dependence of the diffusion coefficient for the RPBE-D3 functional. The melting temperature (in Kelvin) is annotated at each point.

The results for the MSD based on the BLYP-D3 functional can be seen in figure 23. The diffusion coefficient has been plotted against the pressure in figure 24, exhibiting less discontinuity as the RPBE-D3 functional. The values for the diffusion coefficient are higher for the BLYP-D3 than for the RPBE-D3 funtional. Exact values, however, should again be obtained by correcting for finite size effects.

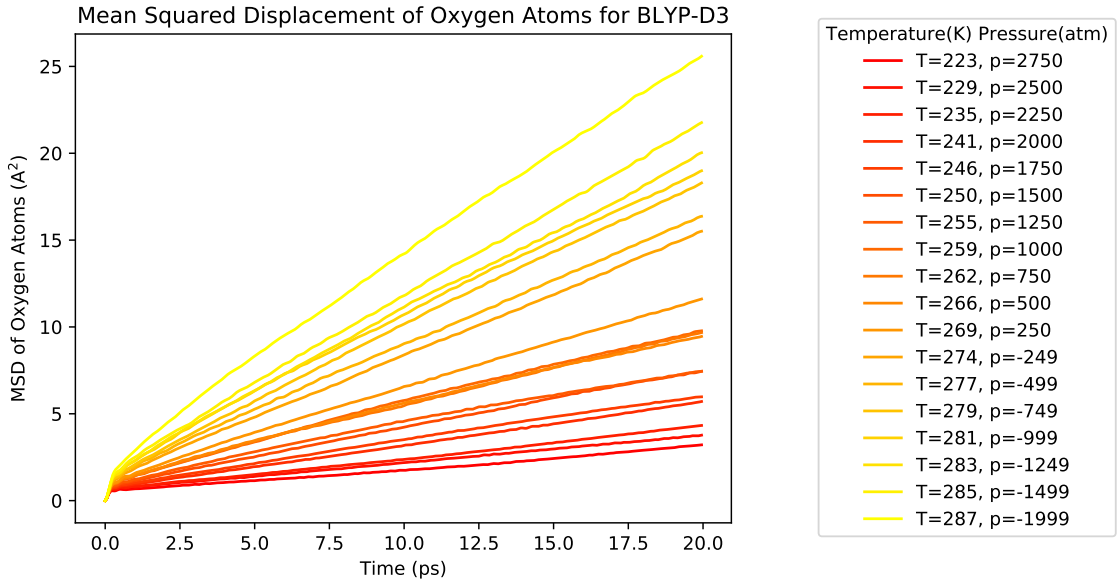


Figure 23: MSD of Oxygen atoms for the BLYP-D3 functional

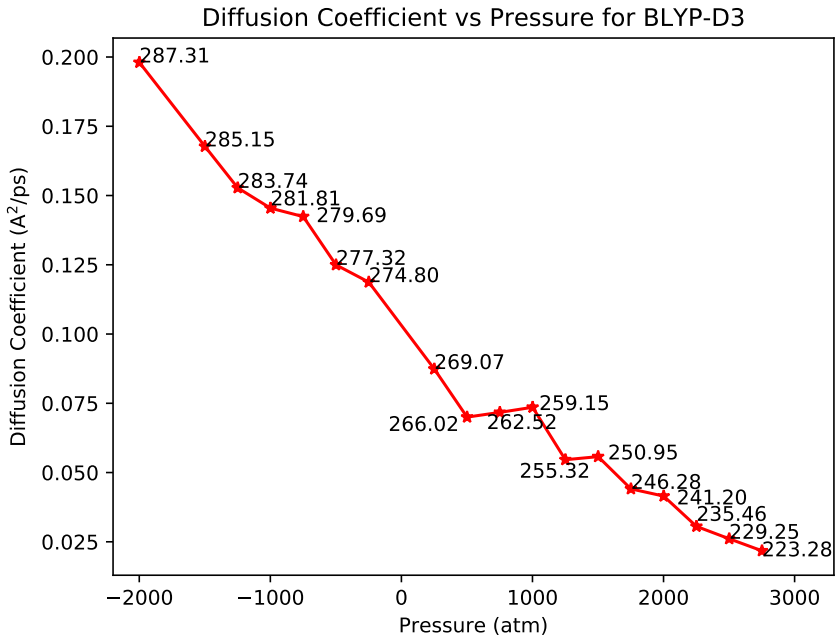


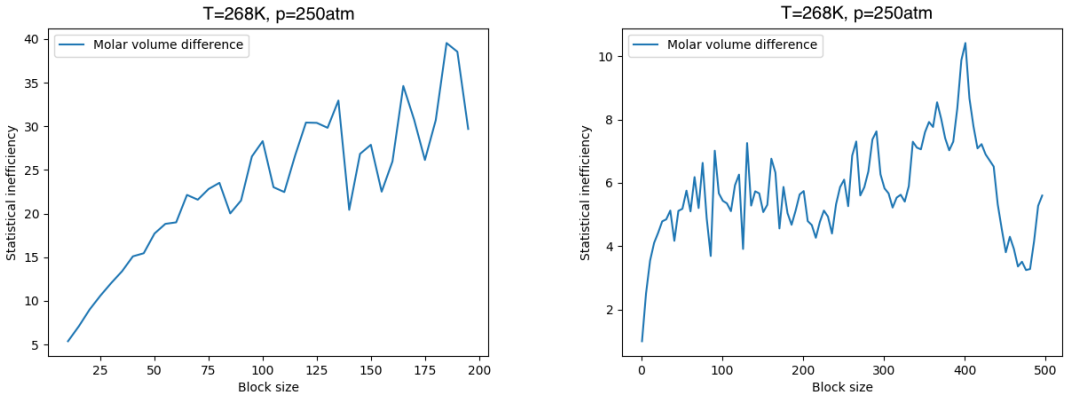
Figure 24: Pressure dependence of the diffusion coefficient for the RPBE-D3 functional. The melting temperature (in Kelvin) is annotated at each point.

7.5 Statistical Inefficiency

When considering statistical errors in molecular dynamics simulations, one often evaluates block averages. The idea of this method is to divide the entire sample of an observable of size M into blocks of size M_B. Next one computes the average of each block and evaluates the variance of these averages $\mathrm{Var}(\bar{A})$. The correlation length of the sample is then given by the product

$$P(M_B) = \frac{M_B \mathrm{Var}(\bar{A})}{\mathrm{Var}(A)} \tag{88}$$

where $\mathrm{Var}(A)$ is the variance of the entire sample. In the limit of large block sizes, the quantity $S = \lim_{M_B \to \infty} P(M_B)$, called statistical inefficiency, should reach a plateau if the simulation is long enough.
In figure 25, the statistical inefficiency of the molar volume difference for two simulation runs is compared. The first run, seen in figure 25a, corresponds to a total simulation time of 0.1 ns. One can see that a plateau is starting to form for bigger block sizes. The second run, seen in figure 25b, corresponds to a longer simulation of 1 ns. For the latter, a plateau is reached and one can say that the simulation was long enough. Since the molecular dynamics simulations carried out during the GDI have been only 0.1 ns, we suggest that longer times should be chosen.



(a) Simulation run of 0.1 ns (b) Simulation run of 1 ns

Figure 25: Comparison of the statistical inefficiency of a short and a long simulation run

8 Conclusion

The use of a neural network potential allowed us to perform molecular dynamics simulations based on first principles. The Gibbs-Duhem integration method was successfully used to trace the saturation line between ice Ih and liquid water. Four coexistence lines have been obtained, representing the RPBE and BLYP density functionals with an without Grimme's D3 correction. The inclusion of van der Waals interactions, has been crucial to reproduce the correct curvature of the melting line within the phase diagram. The RPBE-D3 and BLYP-D3 slightly over- and underestimate the experimental line, respectively. The accurate performance of these two functionals is due to their ability to reproduce the correct densities of liquid water. In the negative pressure regime reentrant behaviour is observed, suggesting the experimental possibility to melt ice Ih by stretching it. Structural and dynamical properties along the coexistence lines have been analyzed and found to satisfy liquid properties. The quality of the results could further be improved by choosing longer simulation times.

The whole phase diagram can potentially be drawn by determining further coexistence points between stable phases. The method of choice to obtain an initial starting point for the integration of the Clausius-Clapeyron equation, is the interface pinning method. The GDI procedure can then be used to get the whole saturation line. By combining the intersections of the resulting lines one could achieve an ab-initio quality phase diagram of water.

The benefits of a neural network potential are, besides immense reduction of computational costs, the allowance of breaking and forming of bonds. As an outlook one could then achieve the determination of the pH-values of water to assess the quality of the underlying density functional. Eventually one can also develop NNPs for CO_2, Cu_2S or other substances and obtain their phase diagram, following the same procedure as in this thesis.

References

[1] T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How van der waals interactions determine the unique properties of water," *Proceedings of the National Academy of Sciences*, vol. 113, no. 30, pp. 8368–8373, 2016.

[2] D. A. Kofke, "Gibbs-duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation," *Molecular Physics*, vol. 78, pp. 1331–1336, April 1993.

[3] D. A. Kofke, "Direct evaluation of phase coexistence by molecular simulation via integration along the saturation line," *The Journal of Chemical Physics*, vol. 98, no. 5, pp. 4149–4162, 1993.

[4] L. Ralf, "Water: From clusters to the bulk," *Angewandte Chemie International Edition*, vol. 40, no. 10, pp. 1808–1827.

[5] V. Petrenko and R. Whitworth, *Physics of Ice*. OUP Oxford, 2002.

[6] D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water*. Oxford Classic Texts in the Physical Sciences, OUP Oxford, 2005.

[7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[8] B. J. Alder and T. E. Wainwright, "Studies in molecular dynamics. i. general method," *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, 1959.

[9] A. Rahman and F. H. Stillinger, "Molecular dynamics study of liquid water," *The Journal of Chemical Physics*, vol. 55, no. 7, pp. 3336–3359, 1971.

[10] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, "The missing term in effective pair potentials," *The Journal of Physical Chemistry*, vol. 91, no. 24, pp. 6269–6271, 1987.

[11] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of Chemical Physics*, vol. 79, no. 2, pp. 926–935, 1983.

[12] J. L. F. Abascal and C. Vega, "A general purpose model for the condensed phases of water: Tip4p/2005," *The Journal of Chemical Physics*, vol. 123, no. 23, p. 234505, 2005.

[13] J. L. F. Abascal, E. Sanz, R. G. Fernández, and C. Vega, "A potential model for the study of ices and amorphous water: Tip4p/ice," *The Journal of Chemical Physics*, vol. 122, no. 23, p. 234511, 2005.

[14] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density-functional theory," *Phys. Rev. Lett.*, vol. 55, pp. 2471–2474, Nov 1985.

[15] 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 perdewburkeernzerhof and beckeleeyangparr functionals," *The Journal of Chemical Physics*, vol. 130, no. 22, p. 221102, 2009.

[16] J. Behler and M. Parrinello, "Generalized neural-network representation of high-dimensional potential-energy surfaces," *Phys. Rev. Lett.*, vol. 98, p. 146401, Apr 2007.

[17] J. Behler, "Representing potential energy surfaces by high-dimensional neural network potentials," *Journal of Physics: Condensed Matter*, vol. 26, no. 18, p. 183001, 2014.

[18] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (dft-d) for the 94 elements h-pu," *The Journal of Chemical Physics*, vol. 132, no. 15, p. 154104, 2010.

[19] M. Ceriotti, J. Cuny, M. Parrinello, and D. E. Manolopoulos, "Nuclear quantum effects and hydrogen bond fluctuations in water," *Proceedings of the National Academy of Sciences*, vol. 110, no. 39, pp. 15591–15596, 2013.

[20] A. Rahman, "Correlations in the motion of atoms in liquid argon," *Phys. Rev.*, vol. 136, pp. A405–A411, Oct 1964.

[21] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *The Journal of Chemical Physics*, vol. 76, no. 1, pp. 637–649, 1982.

[22] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Phys. Rev.*, vol. 136, pp. B864–B871, Nov 1964.

[23] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *The Journal of Chemical Physics*, vol. 72, no. 4, pp. 2384–2393, 1980.

[24] S. Nosé, "A unified formulation of the constant temperature molecular dynamics methods," *The Journal of Chemical Physics*, vol. 81, no. 1, pp. 511–519, 1984.

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

[26] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna, "Non-hamiltonian molecular dynamics: Generalizing hamiltonian phase space principles to non-hamiltonian systems," *The Journal of Chemical Physics*, vol. 115, no. 4, pp. 1678–1702, 2001.

[27] G. J. Martyna, M. L. Klein, and M. Tuckerman, "Nosé-hoover chains: The canonical ensemble via continuous dynamics," *The Journal of Chemical Physics*, vol. 97, no. 4, pp. 2635–2643, 1992.

[28] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, no. 12, pp. 7182–7190, 1981.

[29] S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *Journal of Computational Physics*, vol. 117, no. 1, pp. 1 – 19, 1995.

[30] W. Shinoda, M. Shiga, and M. Mikami, "Rapid estimation of elastic constants by molecular dynamics simulation under constant stress," *Phys. Rev. B*, vol. 69, p. 134103, Apr 2004.

[31] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," *The Journal of Chemical Physics*, vol. 101, no. 5, pp. 4177–4189, 1994.

[32] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna, "A liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermalisobaric ensemble," *Journal of Physics A: Mathematical and General*, vol. 39, no. 19, p. 5629, 2006.

[33] G. Hura, J. M. Sorenson, R. M. Glaeser, and T. Head-Gordon, "A high-quality x-ray scattering experiment on liquid water at ambient conditions," *The Journal of Chemical Physics*, vol. 113, no. 20, pp. 9140–9148, 2000.

[34] C. Rowley, "Simulated radial distribution functions for solid, liquid, and gaseous argon," 2015. CC BY-SA 4.0.

[35] Calandrini, V., Pellegrini, E., Calligari, P., Hinsen, K., and Kneller, G.R., "nmoldyn - interfacing spectroscopic experiments, molecular dynamics simulations and models for time correlation functions," *JDN*, vol. 12, pp. 201–232, 2011.

[36] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*. Computational science series, Elsevier Science, 2001.

[37] M. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*. Oxford Graduate Texts, OUP Oxford, 2010.

[38] U. R. Pedersen, "Direct calculation of the solid-liquid Gibbs free energy difference in a single equilibrium simulation," *The Journal of Chemical Physics*, vol. 139, no. 10, p. 104102, 2013.

[39] M. McKinnell, J. Verhein, P. Yu, L. K. Chan, J. Dhaliwal, S. Bhela, and C. Wong-Sing. https://chem.libretexts.org/Textbook_Maps/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)

/Physical_Properties_of_Matter/States_of_Matter/
Phase_Transitions/Phase_Diagrams.

[40] B. Widom, "Some topics in the theory of fluids," *The Journal of Chemical Physics*, vol. 39, no. 11, pp. 2808–2812, 1963.

[41] C. Vega, E. Sanz, J. L. F. Abascal, and E. G. Noya, "Determination of phase diagrams via computer simulation: methodology and applications to water, electrolytes and proteins," *Journal of Physics: Condensed Matter*, vol. 20, no. 15, p. 153101, 2008.

[42] J. Q. Broughton and G. H. Gilmer, "Molecular dynamics investigation of the crystalfluid interface. i. bulk properties," *The Journal of Chemical Physics*, vol. 79, no. 10, pp. 5095–5104, 1983.

[43] W. G. Hoover, S. G. Gray, and K. W. Johnson, "Thermodynamic properties of the fluid and solid phases for inverse power potentials," *The Journal of Chemical Physics*, vol. 55, no. 3, pp. 1128–1136, 1971.

[44] W. G. Hoover and F. H. Ree, "Melting transition and communal entropy for hard spheres," *The Journal of Chemical Physics*, vol. 49, no. 8, pp. 3609–3617, 1968.

[45] A. Z. Panagiotopoulos, "Direct determination of phase coexistence properties of fluids by monte carlo simulation in a new ensemble," *Molecular Physics*, vol. 61, no. 4, pp. 813–826, 1987.

[46] U. R. Pedersen, F. Hummel, G. Kresse, G. Kahl, and C. Dellago, "Computing gibbs free energy differences by interface pinning," *Phys. Rev. B*, vol. 88, p. 094101, Sep 2013.

[47] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," *Phys. Rev. B*, vol. 28, pp. 784–805, Jul 1983.

[48] S. J. Ilja, *Monte Carlo Methods for Simulating Phase Equilibria of Complex Fluids*, pp. 405–443. Wiley-Blackwell, 2007.

[49] J. Dormand and P. Prince, "A family of embedded runge-kutta formulae," *Journal of Computational and Applied Mathematics*, vol. 6, no. 1, pp. 19 – 26, 1980.

[50] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," 2001–. [Online; accessed 21-June-2018].

[51] L. D. Dalcin, R. R. Paz, P. A. Kler, and A. Cosimo, "Parallel distributed computing using python," *Advances in Water Resources*, vol. 34, no. 9, pp. 1124 – 1139, 2011. New Computational Methods and Software Tools.

[52] L. Dalcin, R. Paz, M. Storti, and J. Delia, "Mpi for python: Performance improvements and mpi-2 extensions," *Journal of Parallel and Distributed Computing*, vol. 68, no. 5, pp. 655 – 662, 2008.

[53] L. Dalcin, R. Paz, and M. Storti, "Mpi for python," *Journal of Parallel and Distributed Computing*, vol. 65, no. 9, pp. 1108 – 1115, 2005.

[54] E. Sanz, C. Vega, J. L. F. Abascal, and L. G. MacDowell, "Phase diagram of water from computer simulation," *Phys. Rev. Lett.*, vol. 92, p. 255701, Jun 2004.

[55] J. D. Bernal and R. H. Fowler, "A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions," *The Journal of Chemical Physics*, vol. 1, no. 8, pp. 515–548, 1933.

[56] "Modelling water: A lifetime enigma," *CHIMIA International Journal for Chemistry*, vol. 69, no. 3, 2015.

[57] B. Guillot, "A reappraisal of what we have learnt during three decades of computer simulations on water," *Journal of Molecular Liquids*, vol. 101, no. 1, pp. 219 – 260, 2002. Molecular Liquids. Water at the New Millenium.

[58] "On the determination of molecular fields. —ii. from the equation of state of a gas," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 106, no. 738, pp. 463–477, 1924.

[59] C. McBride, "SklogWiki, open edit encyclopedia," 2009. [Online; accessed 21-June-2018].

[60] J. D. Bernal and R. H. Fowler, "A theory of water and ionic solution, with particular reference to hydrogen and hydroxyl ions," *The Journal of Chemical Physics*, vol. 1, no. 8, pp. 515–548, 1933.

[61] E. Sanz, C. Vega, J. L. F. Abascal, and L. G. MacDowell, "Tracing the phase diagram of the four-site water potential (tip4p)," *The Journal of Chemical Physics*, vol. 121, no. 2, pp. 1165–1166, 2004.

[62] P. W. Bridgman, "Water, in the liquid and five solid forms, under pressure," *Proceedings of the American Academy of Arts and Sciences*, vol. 47, no. 13, pp. 441–558, 1912.

[63] D. Paschek, "Temperature dependence of the hydrophobic hydration and interaction of simple solutes: An examination of five popular water models," *The Journal of Chemical Physics*, vol. 120, no. 14, pp. 6674–6690, 2004.

[64] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, "Development of an improved four-site water model for biomolecular simulations: Tip4p-ew," *The Journal of Chemical Physics*, vol. 120, no. 20, pp. 9665–9678, 2004.

[65] C. Vega, J. L. F. Abascal, M. Conde, and J. Aragones, "What ice can teach us about water interactions: A critical comparison of the performance of different water models," vol. 141, pp. 251–76; discussion 309, 02 2009.

[66] M. M. Conde, C. Vega, G. A. Tribello, and B. Slater, "The phase diagram of water at negative pressures: Virtual ices," *The Journal of Chemical Physics*, vol. 131, no. 3, p. 034510, 2009.

[67] J. L. Aragones, E. G. Noya, J. L. F. Abascal, and C. Vega, "Properties of ices at 0 k: A test of water models," *The Journal of Chemical Physics*, vol. 127, no. 15, p. 154518, 2007.

[68] A. Warshel and R. M. Weiss, "An empirical valence bond approach for comparing reactions in solutions and in enzymes," *Journal of the American Chemical Society*, vol. 102, no. 20, pp. 6218–6226, 1980.

[69] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, 2009.

[70] Z. Liu, L. E. Carter, and E. A. Carter, "Full configuration interaction molecular dynamics of na2 and na3," *The Journal of Physical Chemistry*, vol. 99, no. 13, pp. 4355–4359, 1995.

[71] B. D. Martino, M. Celino, and V. Rosato, "An high performance fortran implementation of a tight-binding molecular dynamics simulation," *Computer Physics Communications*, vol. 120, no. 2, pp. 255 – 268, 1999.

[72] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, pp. A1133–A1138, Nov 1965.

[73] B. Hammer, L. B. Hansen, and J. K. Nørskov, "Improved adsorption energetics within density-functional theory using revised perdew-burke-ernzerhof functionals," *Phys. Rev. B*, vol. 59, pp. 7413–7421, Mar 1999.

[74] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Phys. Rev. A*, vol. 38, pp. 3098–3100, Sep 1988.

[75] C. Lee, W. Yang, and R. G. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, pp. 785–789, Jan 1988.

[76] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, "The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water," *The Journal of Chemical Physics*, vol. 122, no. 1, p. 014515, 2005.

[77] I.-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello, "Liquid water from first principles: investigation of different sampling approaches," *The Journal of Physical Chemistry B*, vol. 108, no. 34, pp. 12990–12998, 2004.

[78] M. J. McGrath, J. I. Siepmann, I. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack, "Isobaricisothermal monte carlo simulations from first principles: Application to liquid water at ambient conditions," *ChemPhysChem*, vol. 6, no. 9, pp. 1894–1901.

[79] Y. A. Mantz, B. Chen, and G. J. Martyna, "Structural correlations and motifs in liquid water at selected temperatures: ab initio and empirical model predictions," *The Journal of Physical Chemistry B*, vol. 110, no. 8, pp. 3540–3554, 2006. PMID: 16494410.

[80] T. D. Kühne, M. Krack, and M. Parrinello, "Static and dynamical properties of liquid water from first principles by a novel carparrinello-like approach," *Journal of Chemical Theory and Computation*, vol. 5, no. 2, pp. 235–241, 2009. PMID: 26610101.

[81] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras, "Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment," *The Journal of Chemical Physics*, vol. 114, no. 12, pp. 5149–5155, 2001.

[82] G. Stefan, "Accurate description of van der waals complexes by density functional theory including empirical corrections," *Journal of Computational Chemistry*, vol. 25, no. 12, pp. 1463–1473.

[83] J. Petr, Černý Jiří, H. Pavel, and S. D. R., "Density functional theory augmented with an empirical dispersion term. interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations," *Journal of Computational Chemistry*, vol. 28, no. 2, pp. 555–569.

[84] Y. Andersson, D. C. Langreth, and B. I. Lundqvist, "van der waals interactions in density-functional theory," *Phys. Rev. Lett.*, vol. 76, pp. 102–105, Jan 1996.

[85] D. C. Langreth, M. Dion, H. Rydberg, E. Schrder, P. Hyldgaard, and B. I. Lundqvist, "Van der waals density functional theory with applications," *International Journal of Quantum Chemistry*, vol. 101, no. 5, pp. 599–610.

[86] T. Sato, T. Tsuneda, and K. Hirao, "Van der waals interactions studied by density functional theory," *Molecular Physics*, vol. 103, no. 6-8, pp. 1151–1164, 2005.

[87] I.-C. Lin, A. P. Seitsonen, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger, "Importance of van der waals interactions in liquid water," *The Journal of Physical Chemistry B*, vol. 113, no. 4, pp. 1127–1131, 2009. PMID: 19123911.

[88] J. Wang, G. Román-Pérez, J. M. Soler, E. Artacho, and M.-V. Fernández-Serra, "Density, structure, and dynamics of water: The effect of van der waals interactions," *The Journal of Chemical Physics*, vol. 134, no. 2, p. 024516, 2011.

[89] C. Zhang, J. Wu, G. Galli, and F. Gygi, "Structural and vibrational properties of liquid water from van der waals density functionals," *Journal of Chemical Theory and Computation*, vol. 7, no. 10, pp. 3054–3061, 2011. PMID: 26598149.

[90] A. Møgelhøj, A. K. Kelkkanen, K. T. Wikfeldt, J. Schiøtz, J. J. Mortensen, L. G. M. Pettersson, B. I. Lundqvist, K. W. Jacobsen, A. Nilsson, and J. K. Nørskov, "Ab initio van der waals interactions in simulations of water alter structure from mainly tetrahedral to high-density-like," *The Journal of Physical Chemistry B*, vol. 115, no. 48, pp. 14149–14160, 2011. PMID: 21806000.

[91] J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy, "Isobaricisothermal molecular dynamics simulations utilizing density functional theory: An assessment of the structure and density of water at near-ambient conditions," *The Journal of Physical Chemistry B*, vol. 113, no. 35, pp. 11959–11964, 2009. PMID: 19663399.

[92] R. Jonchiere, A. P. Seitsonen, G. Ferlat, A. M. Saitta, and R. Vuilleumier, "Van der waals effects in ab initio water at ambient and supercritical conditions," *The Journal of Chemical Physics*, vol. 135, no. 15, p. 154503, 2011.

[93] 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," *The Journal of Chemical Physics*, vol. 137, no. 4, p. 044506, 2012.

[94] K. Forster-Tonigold and A. Gro, "Dispersion corrected rpbe studies of liquid water," *The Journal of Chemical Physics*, vol. 141, no. 6, p. 064501, 2014.

[95] F. Mallamace, C. Branca, M. Broccio, C. Corsaro, C.-Y. Mou, and S.-H. Chen, "The anomalous behavior of the density of water in the range 30 $K < T <$ 373 K," *Proceedings of the National Academy of Sciences*, vol. 104, no. 47, pp. 18387–18391, 2007.

[96] M. Del Ben, J. Hutter, and J. VandeVondele, "Probing the structural and dynamical properties of liquid water with models including non-local electron correlation," *The Journal of Chemical Physics*, vol. 143, no. 5, p. 054506, 2015.

[97] J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *The Journal of Chemical Physics*, vol. 134, no. 7, p. 074106, 2011.

[98] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford University Press, Inc., 1995.

[99] P. Geiger and C. Dellago, "Neural networks for local structure detection in polymorphic systems," *The Journal of Chemical Physics*, vol. 139, no. 16, p. 164105, 2013.

[100] T. Morawietz, V. Sharma, and J. Behler, "A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges," *The Journal of Chemical Physics*, vol. 136, no. 6, p. 064103, 2012.

[101] S. Haykin, *Kalman Filtering and Neural Networks*. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control, Wiley, 2004.

[102] "Runner - a neural network code for high-dimensional potential-energy surfaces (lehrstuhl für theoretische chemie, ruhr-universität bochum),"

[103] C. Vega, E. Sanz, and J. L. F. Abascal, "The melting temperature of the most common models of water," *The Journal of Chemical Physics*, vol. 122, no. 11, p. 114507, 2005.

[104] A. Singraber, T. Morawietz, J. Behler, and C. Dellago, "Density anomaly of water at negative pressures from first principles," *Journal of Physics: Condensed Matter*, vol. 30, no. 25, p. 254005, 2018.

[105] "International association for the properties of water and steam, an6-16:," tech. rep.