Improving *ab initio* molecular dynamics of liquid water

ZHAOFENG LI

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
PHYSICS

ADVISOR: ROBERTO CAR

SEPTEMBER 2012

Abstract

Water is arguably the most important liquid on the earth, yet the physics behind many water properties is still poorly understood. Being able to describe on the fly breaking and formation of chemical bonds, cooperative polarization effects, proton transfer, etc., ab initio molecular dynamics (AIMD) within Kohn-Sham density functional theory (DFT) is an ideal theoretical tool for investigating liquid water structure. To treat water from first-principle quantum mechanical theory, one needs both a good description of the potential energy surface for the nuclei and a quantum mechanical treatment of nuclear dynamics. The nuclear potential energy surface is derived from the quantum mechanical ground-state of the electrons following the Born-Oppenheimer adiabatic approximation of nuclei and electron dynamics.

In this thesis we focus on the improved description of the nuclear potential energy surface that can be achieved by more accurate functional approximations. So far water simulations have used the generalized gradient approximation (GGA). This approximation is affected by the so-called self-interaction error that causes an excessive delocalization of the protons in hydrogen bonds (H-bonds). Moreover, it neglects the non-local correlations responsible for van der Waals (vdW) interactions. These interactions are weaker than H-bonds, but a number of recent papers and the work discussed in this thesis show that vdW interactions play a sizeable role in the water structure.

We adopt the hybrid functional approximation which mixes a fraction of exact exchange into the GGA functional to mitigate the spurious self-interaction error, leading to a better description of H-bond and vibrational properties. In addition, we adopt a non-local functional approximation that accounts for vdW interactions, which are crucial to describe correctly the structure

and the equilibrium density of water. Both hybrid functional and non-local correlation are found to be indispensable, resulting in an improved liquid water structure compared to the GGA approximation. The improvement is manifest in several properties, including the radial distribution functions, the bond angle distribution, the broken H-bond statistics, etc.

The statistical effects of quantized nuclear dynamics can be treated in an exact numerical way through path-integral (PI) AIMD simulation. But this calculation is computationally expensive. Due to the large computational cost of the improved DFT functional used in this study, we adopt a cheap approximation for quantum nuclear motion. According to this approximation, the nuclei move classically at a temperature $\sim$ 30K higher than desired temperature. This choice was motivated by a recent work comparing PI AIMD simulation with classical AIMD simulation at ambient conditions with the same DFT functional. This work found that the oxygen-oxygen radial distribution function resulting from classical trajectories mimicked closely the corresponding quantity from the path-integral study when the temperature of the classical simulation was $\sim$ 30K higher than that of quantum simulation. This approximate treatment of quantum nuclei restricts our investigation to the analysis of the oxygen distribution functions, as observables that depend more directly on the proton distribution are more strongly quantum mechanical. We find that the oxygen distribution resulting from our improved DFT approximations agrees well with the available experimental data.

An analysis of the O site distribution using an order parameter that differentiates between sites representative of a high density liquid (HDL) environment and sites representative of a low density liquid (LDL) environment is particularly illuminating, as it allows us to better understand the competing role of H-bonds and vdW interactions in liquid water. This analysis shows that

HDL and LDL sites are simultaneously present in the liquid at ambient conditions, consistent with the idea that the inherent potential energy surface of the liquid has two dominant conformations, with low and high density respectively.

Acknowledgements

I am fully appreciated to my thesis advisor Prof. Roberto Car for his long-term advising and generous support during the Ph.D period. It has been a great experience to perform computational physics research under his instruction. His sharp insight and serious attitude toward physics greatly influences on me. Besides, his enthusiasm in understanding physics motivates me to explore challenging problems and seek ingenious and beautiful solutions.

I want to thank Prof. Willam Happer and Prof. David Huse to be my thesis committee members. The questions arising from prethesis and FPO presentations show valuable suggestions from their viewpoints. I am feeling grateful to Prof. Salvatore Torquato to be my thesis reader. The physics department is a very nice place and I always take it as my home. Secretaries and staffs serving our department offer numerous assistances to orientate me in graduate student's life at Princeton.

I would want to thank my group members for the great collaborations, talks and others. I had a great time with Xifan when we collaborated to perform PBE0 simulation study. I am also working quite well with Rob in studying vdW effect. Other members includes Wei, Hongzhi, Liming, Lingzhu, Weiyi, Jia , Patrick and Federico, Uli, Santra. I still remember those happy days with John, Xinxin, Alex, Tibi, Richard and other graduate fellows to prepare prelims and take classes. I also want to thank Lifeng for useful discussions and suggestions.

My study and life at Princeton become more wonderful when I met with my wife, Jun Liu. Without her endless love and courage, this thesis would be impossible to publish. I want to say tons of thanks to my mum Meizhen Zheng and my dad Lesheng Li. Their hard working all over their lives always

impacts me deeply. Their decision to support my study from elementary school to college is a grand, grand love that I can feel and touch. I am also feeling appreciated to my uncle Weiguo Zheng, who is my early advisor of life and study.

To them I dedicate this thesis.

Contents

List of Figures

1 Introduction

Water and aqueous solutions are ubiquitous in Nature. However, while their properties have been extensively investigated experimentally, their fundamental structure is still poorly understood from a theoretical viewpoint. This is due to the difficulty of modeling accurately the atomic interactions, i.e. potential energy surface and the difficulty of having to deal with large quantum effects when classical simulations are insufficient. Moreover, the highly disordered nature of the liquid requires extensive sampling and therefore it takes long and expensive AIMD simulations to achieve statistically meaningful results. Experimentally the liquid structure is probed in diffraction experiments (x-ray and neutron) but sometimes structural information is also extracted from electron excitation spectroscopy, particularly core excitation spectroscopy. In that case there is an additional difficulty in interpreting the experimental data, because the experimentally measured cross sections require in this case a proper many-body theory of the excitation process in addition to a good structural model. Thus at present there are considerable controversies in the literature[1]. Nevertheless, advances in computational methods combined with modern high performance computational techniques have the potential to achieve major breakthroughs and hence achieve definitive results.

Despite the great importance of water in physics, chemistry and the life sciences, a number of uncertainties still remain regarding its internal structure. For instance, there is still considerable uncertainty in the first peak of the oxygen-oxygen (OO) radial distribution function (RDF) extracted from x-ray data due to the uncertainty in the form factor, i.e. the electronic charge distribution in the molecule in condensed phase, and the uncertainty in the

transformation from the experimental accessible structure factor to the real space RDF, as recently reported by A.Soper[2].

The structure of the liquid as obtained from first-principle simulations based on common DFT approximations, such as PBE or BLYP[3, 4], the two most used functionals adopted in liquid water simulations, does not yet agree with experiment to a high degree of accuracy[5, 6]. In particular, PBE/BLYP simulations of water consistently result in a liquid that is considerably over-structured compared to experiment and shows very sluggish diffusion at temperatures close to room temperature, consistent with a substantial underestimation of the broken H-bond fraction. These difficulties reflect the dependence of the H-bond strength on a delicate balance between covalent and electrostatic interactions. Moreover, a purely classical model of the nuclei is not good enough, as indicated by the importance of isotope effects in static equilibrium properties, such as the freezing temperature and even the boiling temperature[7]. The limits of classical physics are particularly evident in the momentum distribution of the protons in ice and water, which has been measured by deep inelastic neutron scattering [8, 9] and is very different from the classical Boltzmann distribution at the temperature of equilibrium of the system. The quantum character of the protons shows up in several properties and its structural effect cannot be neglected, as shown by recent work of Morrone and Car[10]. The difficulties encountered so far in first-principle studies of liquid water structure are closely related to the above issues.

Therefore, on one hand more accurate functionals (e.g. the hybrid functionals for nonlocal exchange, and including non-local correlations responsible for vdW interactions) are needed to better improve the description of the electronic ground state over semi-local GGA for exchange-correlation[3]. Hybrid functionals (e.g. PBE0[11]) are computationally expensive and have been

used so far only rarely in water simulations[12, 13]. In addition, non-local correlation functionals, do play a role. On the other hand, large anharmonic zero point motion effects should be properly included. As for the structural effect of quantum nuclei, in principle, PI AIMD simulation can address this issue without approximation, but these calculations are also very expensive and finding ways of including nuclear quantum effects in an economical, yet sufficiently accurate way is a very important goal.

This thesis concentrates on studying the effect of more accurate new exchange-correlation functionals on the water structure: more specifically, the hybrid functionals that mix exact exchange into GGA functional to reduce self-interaction delocalization error, and including non-local correlation responsible for vdW interactions. The expensive calculation of exact exchange in the hybrid functionals is carried out using a recently developed order N algorithm based on Maximally Localized Wannier functions[14]. The non-local correlation functional adopts the Tkatchenko-Scheffler (TS) pairwise interatomic interaction scheme[15], in which the C_6 coefficients and the damping functions are explicit functionals of the electron density. Nuclear quantum effect can be studied accurately with the PI AIMD methodology[10]. However, this calculation is quite expensive and has not been considered in the present study. Focusing on the O site distribution, the effect of quantum nuclei can be accounted for by an increase of approximately 30K in temperature, as suggested by previous PI AIMD simulation results[10]. Including all these effects (hybrid functionals, non-local correlation, and temperature increase to account approximately for quantum nuclei), the OO RDF and the structure factor resulting from the simulation show a dramatically improved agreement with the available experimental data, suggesting that we have captured the main physics relevant to liquid water structure. Our re-

search therefore constitutes an important step forward in the description of liquid water from first principles.

This thesis is organized as follows: Chapter 1 contains the introduction. Chapter 2 briefly reviews some properties water and mentions physical and chemical properties that are not yet fully understood experimentally or theoretically. This chapter also illustrates the shortcoming of GGA functionals in the description of water.

Chapter 3 is devoted to the development of an improved exchange-correlation functional including: (a) a hybrid functional scheme, i.e. the PBE0 functional that incorporates a fraction of exact Hartree Fock exchange into the PBE (GGA) functional; (b) the vdW non-local correlation that is missing in PBE0 functional but is important in liquid water. The implementation and optimization scheme is also discussed.

Chapter 4 discusses comparison between simulation and experiment, and the influence of various interactions and dynamic effects on the liquid water structure. The simulation with PBE0 hybrid functional and vdW correlation at 330K(to mimic quantum effects) leads to an excellent agreement with experiment at 300K (OO RDF and structure factor).

Chapter 5 contains a summary and the major conclusions of this study.

2 General background

2.1 A brief description of water

2.1.1 An important liquid with unique properties

Water, called the "molecule of life" or the "matrix of life", underlines its biological importance. Water acts as solvent and dispersing and lubricating medium. It is also a versatile reactant. Life and water are therefore inseparable. It is hardly surprising that living organisms are sensitively attuned to the properties of water.

The physical and chemical anomalies of water are well known and documented, but their significance in quantitative terms is not always realized[16]. Just to name a few, the large heat capacity of water reflects its ability to store considerable quantities of thermal energy. The ocean thus acts as vast thermostats from which the heat energy is carried to the cooler regions by currents. Another property connected with water movement is the density maximum near 4°C. This results in the freezing of rivers from the surface downward instead of upward, enabling fishes to survive in the cold weather.

Some notable physical properties are the following: negative volume of melting; numerous crystalline polymorphs; high dielectric constant; anomalously high melting, boiling and critical temperatures for a low-molecular-weight substance; increasing liquid fluidity with increasing pressure; high mobility transport for H^+ and OH^- ions. Other remarkable properties includes the almost universal solvent action of liquid water, making it rigorous purification extremely difficult. Besides, water is one of the most reactive chemicals, in that it readily interacts with ions and molecules.

Most of these anomalies are rationalized in terms of an underlying tetrahedral H-bonded network. For instance, water contracts on melting while most materials do expand on melting. The reason is that the structure of ice (e.g. hexagonal ice (Ih)) is open with a low packing efficiency where all the water molecules are involved in four straight tetrahedrally-oriented H-bonds. On melting, some of these ice (Ih) bonds break, others bend and the structure undergoes a partial collapse. This is different from what happens with most solids, where the extra movement available in the liquid phase requires more space and therefore melting is accompanied by expansion.

2.1.2 Liquid water structure

The knowledge about how a water molecule interacts with its neighbors in the liquid is the essential to understand the unique properties of liquid water. However, the liquid water structure is still not completely understood both experimentally and theoretically. The more we look, the more the problems accumulate[1]. As a proof of this, only a few years ago, the exact determination of the structure of water was listed among the most outstanding problem in science[17, 18].

The instantaneous local structure of liquid water has been modeled as an approximate tetrahedron connected via H-bonds, similar to that in crystalline phase. In the liquid, this tetrahedral structure not only distorts, but also breaks and reforms on a time scale of about 0.5 picosecond (ps)[19]. This repeated, constantly reorganizing unit defines a three-dimensional network extending throughout the liquid. This prevailing view comes largely from x-ray, neutron-scattering studies and computer simulations. This makes good sense in the light of the unambiguously tetrahedral arrangement of molecules

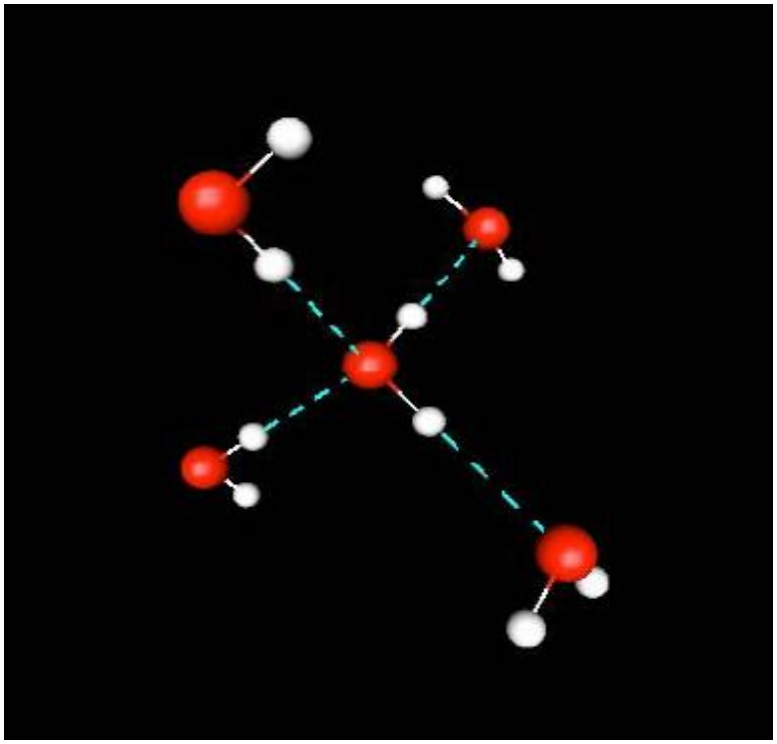

Figure 2.1: Each water molecule(center) is in local tetrahedral cage in the H-bond network. Red denotes oxygen atom and white denotes hydrogen. Each water molecule has two donors and two acceptors. The dash lines indicate H-bond connections among water molecules.

in ice crystals. The local structure of liquid water has been widely accepted to have this tetrahedron connection with H-bonds, as seen in Fig.[2.1] and Fig.[2.2].

How exactly large this tetrahedrality is in liquid water remains to be determined. Recent *ab initio* simulation[10] with H-bond definition from Ref.[10] finds that on average $\sim 7\%$ H-bonds are broken with classical nuclei in BLYP functional. Quantum nuclei effect increases this to $\sim 11\%$ and therefore leads to a reduction of tetrahedrality. Improved functionals compared to semi-local

GGAs increase the average broken H-bond fraction. Though no single experiment is capable of characterizing the molecular architecture of the liquid network in detail, a sizeable fraction around 13% of H-bonds are estimated to be broken, based on the fact that the heat of melting is 13% of the sublimation energy of the solid roughly assuming that the whole energy is used only to break H-bond[20]. A proposal of a different structural organization consisting of chains and large rings of H-bonded molecules[21] in which 80% H-bonds are broken is inconsistent with this point of view. Moreover, it is inconsistent with basic energetics and thermodynamics.

2.1.3 Hydrogen-bond network in water

The distinctive properties of liquid water can be attributed to H-bond[22, 23]. It is widely accepted that fleeting H-bond is one aspect of water's molecular structure setting it apart from most other liquids. These feeble bonds that link the molecules constantly break and form above water's melting point, yet still impose a degree of structure.

The concept of H-bond was first systematically studied by Latimer and Rodebush nearly one century ago[24]. Generally, a H-bond specifically exists between electronegative atoms (e.g. nitrogen, oxygen, fluorine and chlorine) and hydrogen (or deuterium) atoms, particularly when electropositive atoms are themselves chemically bonded directly to electronegative atoms. The H-bond is strongly *directional*. So the chemical bond containing the hydrogen atom that is donated to H-bond tends to point directly at the nucleus of the hydrogen-accepting electronegative atoms. The strength of the H-bond in water is roughly 0.2 eV and is normally intermediate between weak vdW interactions (~ 0.02 eV) and strong covalent chemical bonds (~ 2 eV). Fig.[2.3]

shows water molecule interaction via the H-bond in the water dimer. As water dimer contains only one H-bond, it is usually taken as accurate benchmark calculation to illustrate dissociation energy variation, thus providing evidence about H-bond strength.

Liquid water consists of a macroscopically connected, random network of H-bonds, with frequent strained and broken bonds, as seen in Fig.[2.2]. Anomalous properties of water arise largely due to H-bond. For instance, liquid water's high boiling point is due to the higher number of H-bonds each molecule can form. Water is unique because its oxygen atom has two lone pairs and two hydrogen atoms, so the total number of H-bonds per water molecule is two as each H-bond is shared by two water molecules. That is, each molecule will act as a H-bond donor through pointing its (positively charged) hydrogens at a lone-pair (negative) region of each of two neighboring waters. In the other two interactions, the central waters will act as a H-bond acceptor of two neighboring waters pointing their hydrogens towards the lone-pair regions of the central water molecule. Based on this simple model, the four-coordinated motif is built (see Fig.[2.1]). And due to the difficulty to break these bonds(H-bond strength is ~ 0.2eV and is one order magnitude larger than k_BT with ~ 0.02eV), water has a very high boiling point, melting point and viscosity compared to other molecular liquids that are not H-bonded (e.g. CH_4, SiH_4).

To interpret dynamic aspects of H-bonding theoretically and experimentally, one relies on the definition of H-bond in liquid water. There have been various widely adopted criteria to define H-bond based on inter-atomic distances and angles. For instance, in Ref.[26], two water molecules are chosen as being H-bonded only if their inter-oxygen distance is less than 3.5 Å, and simultaneously the $\mathrm{H}-\mathrm{O}\cdots\mathrm{O}$ angle is less than 30^o. In Ref.[10], a H-bond

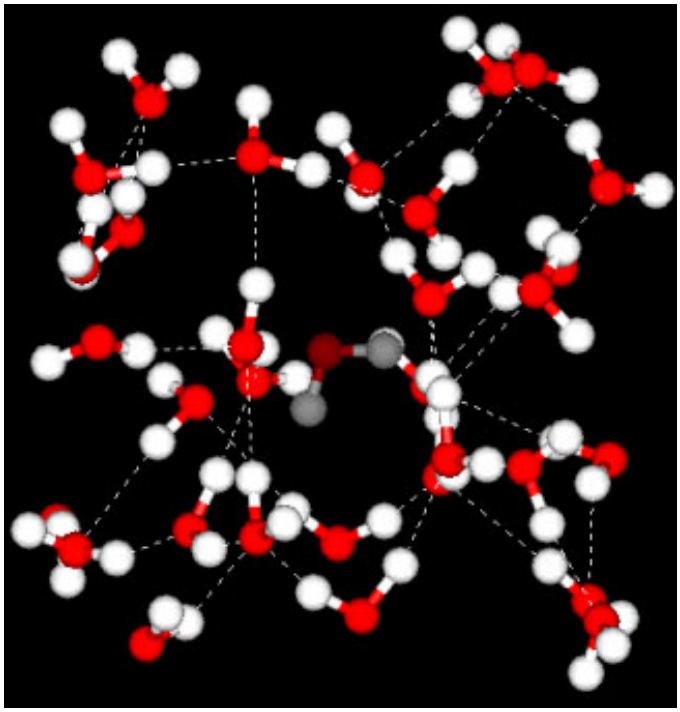

Figure 2.2: Snapshot of a computer simulation of liquid water. Oxygen atoms (red balls) and hydrogen atoms (white balls) form molecule via covalent bonding (solid lines). Neighboring molecules interact via H-bonding (dash lines). In this image, the center molecule is tetrahedrally coordinated, forming four H-bonds. Image taken from[25].

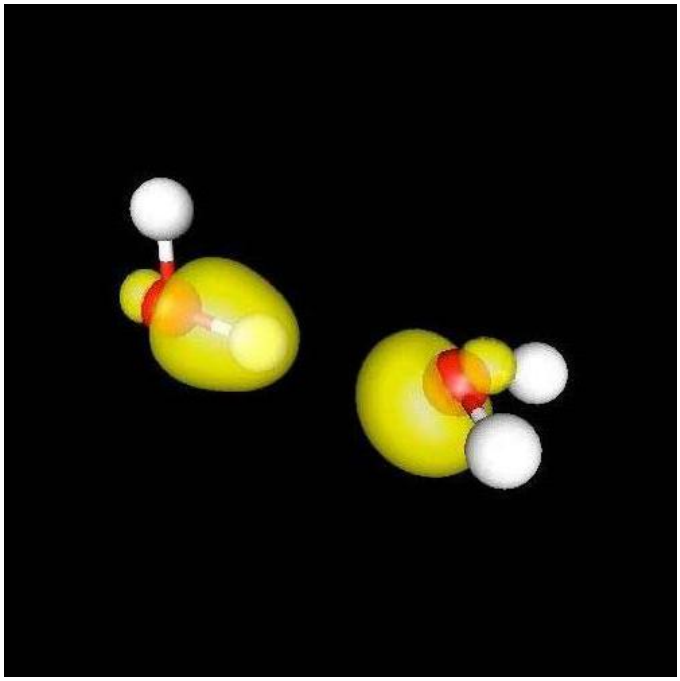

Figure 2.3: Water molecules interact via H-bond.

is defined in geometric terms by oxygen-to-oxygen and oxygen-to-hydrogen distance cutoffs that are equal to the minima of the H-bonding peaks of the RDF, and a H-bond angle greater than 140^o. Other H-bond definitions according to geometry criteria can be found in Refs.[27, 28].

In addition to geometry criteria, there are also definitions in terms of energetics cutoffs. For instance, within a given model, one can calculate the distribution of binding energies for all pairs of molecules in the liquid, and define the pairs within the first peak of this distribution to be H-bonded[20] . A detailed review and analysis of H-bond definition is presented in Ref.[29], where a criterion based on electronic structure is also suggested. We have to notice that these various H-bond definitions in terms of geometry or energy cutoff are subject to changes of density and temperature and therefore, some minor difference would appear for different approaches. However, the test for some H-bond definitions shows that the difference is small and the number of H-bonds per molecule in water simulations is typically between 3 and 4 at ambient conditions(since each H-bond is shared by a pair of molecules, the average number of H-bonds per molecule is 2 or less)[29], i.e. the average of broken H-bond is between 10% and 20%.

In the following chapters, we will see that H-bond statistics provides a useful and physically appealing tool to monitor the degree of liquid water overstructuring when different DFT approximations are adopted.

2.2 Experimental probes of water structure: x-ray and neutron scattering

The most important microscopic measurements to probe liquid water structure are radiation scattering experiments. Elastic scattering of neutrons or x-rays, in which the scattering crosssection is measured as a function of momentum transfer between the radiation and the sample, is the source of our experimental knowledge of the static structure of a liquid. Liquid water structure has been widely investigated experimentally through elastic x-ray scattering and neutron scattering[30, 31]

Ab initio simulations of liquid water structure usually focus on real space structural information (e.g. pair correlation function, or radial distribution function (RDF)) while experimental scatterings explore reciprocal space. In order to take experimental data as reference, the procedure of converting the reciprocal space diffraction data into real space correlations is required. However, this transformation is far from trivial. For instance, neutron scattering measures a combination of internuclear structure factors weighted by the scattering lengths of the constituted nuclei. In order to extract the individual pair correlations (they are oxygen-oxygen, oxygen-hydrogen, hydrogen-hydrogen pair correlation functions in liquid water), three different experiments are performed on samples of different isotopic composition and therefore different scattering lengths. The data analysis is complex and requires the applications of a variety of corrections to the raw data. In addition, the scattering is dominated by the hydrogen and the deuterium atoms, which leads to large uncertainties in oxygen-oxygen (OO) RDF. By contrast, x-ray scattering is dominated by the oxygen-oxygen correlations, since most of the charge density is localized around the oxygen atoms. However, the

procedure of extracting OO RDF from the experimental observed scattering intensity is not simple and involves a number of theoretical assumptions. Due to this, it is useful to recall the theory of elastic scattering in order to better understand the validity of various experimental available data.

Neutrons interact predominantly with atomic nuclei and the strength of the interaction is quantified by the atomic scattering length b, the value of which varies from atoms to atoms(according to the Fermi contact approximation, b is scalar here as nuclei are considered to be point like particles). The neutron scattering intensity or diffraction cross section per atom for a system containing N atoms is described by an expression of the form[32]

$$\begin{aligned} I(q) \equiv \frac{d\sigma}{d\Omega} &= \left\langle \left| \sum_{i=1}^{N} b_i \exp(-i\mathbf{q}\cdot\mathbf{r_i}) \right|^2 \right\rangle \\ &= \left\langle \sum_{i=1}^{N}\sum_{j=1}^{N} b_i b_j \exp(-i\mathbf{q}\cdot(\mathbf{r_i}-\mathbf{r_j})) \right\rangle \end{aligned} \tag{2.1}$$

where $\mathbf{q}$ is the wavevector transfer in the scattering experiment, and $\mathbf{r_i}$ is the position vector of atom i. The angle brackets represent an ensemble average over the system, and an average over the spin and isotope state of the nuclei, since the neutron scattering length can be dependent on both the spin and isotope state. Fortunately the spins and isotopes are not normally correlated with the atomic positions, unless the system is magnetic, so this average can be performed independently from the ensemble average.

Since for a homogeneous liquid(e.g. liquid water that we are interested in here), the separation vector between atoms is not correlated with the direction of q. Then the differential scattering cross-section can be simplified to the form

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)^{self} + \left(\frac{d\sigma}{d\Omega}\right)^{distinct} \tag{2.2}$$

where

$$I_{self}(q) \equiv \left(\frac{d\sigma}{d\Omega}\right)^{self} = \sum_{\alpha} c_\alpha \langle b_\alpha^2 \rangle \tag{2.3}$$

$$I_{distinct}(q) \equiv \left(\frac{d\sigma}{d\Omega}\right)^{distinct} = \sum_{\alpha,\beta \geq \alpha} (2 - \delta_{\alpha\beta}) c_\alpha c_\beta \langle b_\alpha \rangle \langle b_\beta \rangle S_{\alpha,\beta}(q) \tag{2.4}$$

where $\delta_{\alpha\beta}$ is the Kronecker δ function, $c_\alpha = \frac{N_\alpha}{N}$ is the atomic fraction of atomic species α with N_α the number of that species. The angular brackets represent spin and isotope average. $S_{\alpha,\beta}(q)$ is the partial structure factor between atom α and β defined by

$$S_{\alpha,\beta}(q) = 4\pi\rho \int_0^\infty r^2 [g_{\alpha,\beta}(r) - 1] \frac{\sin(qr)}{qr} dr \tag{2.5}$$

Assuming corrections for inelasticity effects have been made satisfactorily, then the self-scattering for neutrons in Eq.[2.3] is a constant and is independent of q.

For molecular systems, it usually splits the distinct term into two terms, one intramolecular and the other intermolecular. The former consists of pairs of atoms on the same molecule while the latter is composed of pairs of atoms on different molecules. It is possible to perform a direct Fourier inversion of $I_{inter}(q)$ to have appropriately weighted sum of site-site radial distribution functions(RDFs)

$$\frac{1}{2\pi^2\rho}\int q^2 I_{inter}(q)\frac{\sin qr}{qr}dq = \sum_{\alpha,\beta}(2-\delta_{\alpha\beta})c_\alpha c_\beta \langle b_\alpha\rangle\langle b_\beta\rangle(g_{\alpha,\beta}(r)-1) \quad (2.6)$$

For x-ray diffraction, the expressions for the differential scattering cross section are analogous to neutron case from Eqs[2.1-2.6], within the standard approximation that the electron form factor $F^2(q)$ for each atom is spherically isotropic. This electron form factor replaces the neutron scattering length b in Eqs[2.1-2.6] while the rest stay the same.

2.3 Structure factor and partial correlation function

2.3.1 Structure factor

As has been discussed, the pair correlation function or radial distribution function is not directly observable from experiment, but rather derived from Fourier transform of structure factor(see Eq.[2.17]). In x-ray experiment, the liquid water structure factor $S(q)$ in experiment is extracted from the observable scattering intensity $I(q)$ assuming that the x-ray are scattered by oxygen atoms. The form factor $F^2(q)$ can be calculated from a quantum mechanical calculation of an isolated water molecules in gas phase. Both assumptions are reasonable, but the one on the form factor should be verified. Given the assumption that the center of the electronic molecular charge coincides with the oxygen atom, which should be correct to very good extent, the structure factor has a particularly simple expression:

$$S(\mathbf{q}) = \langle \frac{1}{N} \rho_{\mathbf{q}} \rho_{\mathbf{-q}} \rangle \tag{2.7}$$

where ρ_q is a Fourier component of the microscopic density

$$\rho_q = \int \rho(\mathbf{R}) \exp(-i\mathbf{q} \cdot \mathbf{R}) d\mathbf{R} = \sum_{i=1}^{N} \exp(-i\mathbf{q} \cdot \mathbf{R}_i) \tag{2.8}$$

Then we have

$$S(q) = \frac{1}{N} \sum_{i,j} e^{i\mathbf{q} \cdot \mathbf{R_{i,j}}} \tag{2.9}$$

where $R_{i,j} \equiv R_i - R_j$ and the sum extends to the pairs of oxygen atoms only. According to the definition of the x-ray intensity, it is given by

$$I(q) = F^2(q)\ S(q) \tag{2.10}$$

Let's first focus on structure factor. For a macroscopically isotropic sample(e.g. liquid water), $S(q)$ depends only on the modulus of q. In the simulation N is the number of oxygen atoms in the cell and the pairs are only those contained in the cell adopting the minimum image convention. The q vectors that should be included are the reciprocal lattice vectors of the simulation cell, i.e. $q = \frac{2\pi}{a}(n_1, n_2, n_3)$ for a simple cubic cell. This means we will obtain a histogram representation of $S(q)$ which has better resolution as q becomes larger. Also notice that the minimum q accessible in a simulation is $q_{min} = \frac{2\pi}{a}$.

In calculating $S(q)$ a spherical average over the members of each shell of q vectors must be performed. Thus for a given reciprocal lattice vector modulus q we have:

$$S(q) = \frac{1}{N}\frac{1}{N_q}\sum_{i,j}\sum_{q} e^{i\mathbf{q}\cdot\,\mathbf{R_{i,j}}} \tag{2.11}$$

In which the sum over q includes only the N_q reciprocal lattice vectors that have modulus $|\mathbf{q}| = q$. According to this, we could obtain the structure factor of liquid water from the *ab initio* simulation results.

2.3.2 Form factor

For the form factor, the definition is as follows:

$$F^2(q) \equiv \langle |F(q)|^2 \rangle \tag{2.12}$$

where

$$F(q) \propto \int dr \rho(r) e^{i\mathbf{q}\cdot\mathbf{r}} \tag{2.13}$$

$\rho(r)$ is the electron density distribution of a water molecule and it is defined within a normalization constant. The angular brackets in the $F(q)$ imply a spherical and an ensemble average. The latter is only important for a molecule in a condensed environment.

The form factor for a molecule in gas phase is first calculated. In this case, we could use the 64-molecule cell size and include only one water molecule in the cell in its zero temperature equilibrium configuration. Then

$$F^2(q) = \frac{1}{N_q} \sum_q F(-q)F(q) \tag{2.14}$$

As in the case of the structure factor the average is made over the shell of reciprocal lattice vectors having modulus equal to q. Notice for a non-spherical charge distribution, form factor depends on the orientation of the molecule in the cell and in fact one should average over also the molecular orientations. However, the non-sphericity should be small and moreover it would disappear at larger q. The test shows that this is indeed the case by randomly rotating the molecule and averaging over the orientation. The next step is to check how well $F^2(q)$ calculated in gas phase corresponds to the ensemble average for the molecules in liquid. In order to check that we need to identify the $\rho(r)$ of an individual molecule in the liquid. The latter density

in not uniquely defined because the molecules in the liquid interact among themselves. An optimal way to identify the molecular charge in liquid water is via the Maximally Localized Wannier Functions (MLWF)[33]. There are 4 MLWFs per water molecule. The molecular densities constructed in this way integrate to the correct molecular electron charge and show a very small overlap among neighboring molecules. Then the ensemble average can be performed easily by averaging over molecules and over molecular dynamics snapshots in addition to averaging over the reciprocal vector shells. Test has shown that several snapshots are sufficient to compute the ensemble average.

We have to notice that, the experimental form factor involves the total (valence plus core) electronic charge, but in the current pseudopotential calculations we only access the valence pseudo-charge. This is the reason for having only 4 MLWFs per molecule, containing 2 electrons each for a total of 8 electrons, but the total number of electrons in a water molecule is 10. However the 1s oxygen core electrons have definitely a spherical charge distribution highly localized around the oxygen nucleus, while all the environmental effects we are after are associated to the valence electrons. In particular the difference between the average pseudocharge of a molecule in the liquid and the spherically averaged molecular pseudocharge of a gas phase molecule should provide an accurate estimate of the change of the form factor going from the gas to the condensed phase.

Fig.[2.4] shows the calculated x-ray scattering intensity $F^2(q)$ for a single water molecule in gas phase and condensed phase. In gas phase, the single water molecule is put at the center of a large supercell. In a pseudopotential framework, the core contribution results in a (pseudo) form factor which differs from the one obtained from an all electron calculation[34]. However, the difference from gas to liquid phase should be represented accurately by

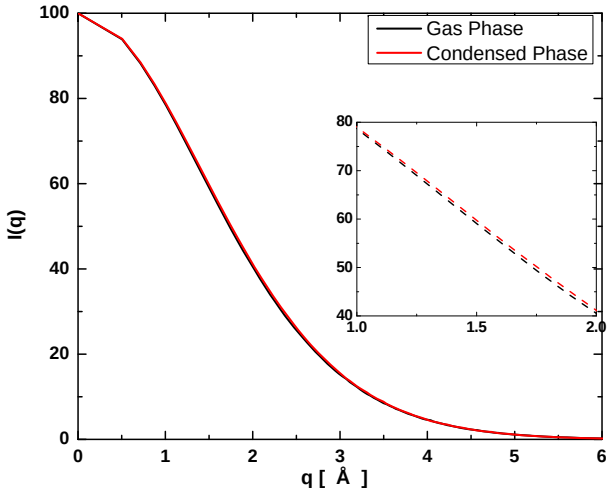

Figure 2.4: *Ab initio* calculated x-ray scattering intensities of a water molecule in the gas phase (black line) and condensed phase (red line). The inset depicts the zoom-in results to distinguish these two difference, where the variation of the two is quite small, within $\sim 2\%$. As the water molecule form factor in gas phase from all electron calculation is usually adopted, it leads to $\sim 2\%$ overestimation in the experimental structure factors due to this uncertainty.

the difference between the corresponding pseudocharge densities. Fig.[2.4] thus provides us sufficient information that the form factors in both phases are quite close and therefore, we can use the water molecule form factor in gas phase in dealing with structure factor and x-ray scattering intensities. The reason is that it is easy to do an accurate all electron calculation for the gas phase but it would be considerably more complicated to do it for liquid water.

2.3.3 Pair correlation function

The complete description of the structure of a liquid usually refers to particle densities and the closely related, equilibrium particle distribution functions[32]. The generalized quantity $\rho_N^{(n)}(r^n)dr^n$ yields the probability of finding n particles of N particles system with coordinates in the volume element dr^n, irrespective of the positions of the remaining particles and irrespective of all momenta. Usually, knowledge of the low-order particle distribution functions, in particular of the pair density $\rho_N^{(2)}(r_1, r_2)$ is often sufficient to calculate the equation of state and other thermodynamic properties of the system in gas phase. This two dimensional pair distribution functions measure the extent to which the structure of a liquid deviates from complete randomness. If the system is homogenous and isotropic(like liquid water), the pair distribution function is a function only of the separation $r_{12} = |r_2 - r_1|$. It is then usually called the radial distribution function (RDF) and written simply as $g(r)$.

The RDF between the atom A and B, $g_{AB}(r)$, is obtained from the average of the atom-atom pair correlation function over a spherical shell of width dr and number density ρ

$$4\pi r^2 dr \rho g_{AB}(r) = \int d\Omega \langle \delta(r - (r_A - r_B)) \rangle \tag{2.15}$$

where the integral is taken over the solid angle Ω , δ is the Dirac delta function, and the average is taken over all of phase space. $g_{AB}(r)$ evaluates the probability that atom A is at a distance r from atom B. The peaks of the RDF indicate features in the structure of the system of interest. When no correlations are presents (i.e. in an ideal gas or at long distances in liquid systems) the RDF approaches one.

RDF plays a key role in the physics of liquids. Based on the definition of $g_{AB}(r)$, another relevant quantity $4\pi r^2 \rho g_{AB}(r)dr$ is the average number atoms in a spherical shell of thickness dr, at a distance r from the central atom. This is informative if we plot $4\pi r^2 \rho g_{AB}(r)dr$ as a function of r, as the area under the curve between any two values of r is equal to the number of neighboring atoms within that range of distance. For instance, an important quantity called coordination number is employed to describe liquid water structure . This quantity is the integration of $4\pi r^2 \rho g_{AB}(r)dr$ till the first minima position in the OO RDF. A typical value between 4 and 5 is reported in various *ab initio* simulations and experiments, indicating that liquid water preserves much of its ice-like tetrahedral structuring but with differences in H-bonding patterns.

The early investigation of liquid water structure based on x-ray scattering was performed by Bernal and Fowler in the 1930s[35], who inferred that tetrahedral grouping of molecules occur frequently in the liquid water. This fundamental discovery, however, was made without determining the RDF. Instead, they calculated the diffraction patterns expected for various models of the liquid water structure and compared the results with the observed pattern. Narten, *et al.* (in the 1970s)[36, 37], on the basis of the experimental RDF at different temperatures, suggested that the tetrahedral coordination might bear some resemblance to ice I_h for water at 2^oC, and that the contraction of water during warming to 4^oC might represent the completion of a transition to a more compact form of tetrahedral coordination. The overall conclusion is that water molecules are predominantly coordinated to four neighboring molecules at room temperature and below, thus forming an extended, three dimensional network of molecules, but that this four-coordination breaks down when temperature raises. However, we have to notice that simply fo-

cus on coordination is not enough to establish tetrahedral coordination and angular correlations are very important too.

The RDF is garnered indirectly from scattering experiments[38, 39, 40]. Over the past decade, it was believed that for pure water the neutron[31, 41] and x-ray scattering[42, 43] data were in good agreement with each other. If this was the case, then the experimentally obtained RDF may be used as a benchmark for the study of the structure of water and the viability of any theoretical model. Particular importance is cast upon the first peak of the OO distribution. However, a recent study has cast doubt upon this claim[2], and has shown that there is significant variance in the peak height of the OO distribution. Much of the confusion originates from how precisely the RDFs are obtained through transform of the experimental derived structure factor. The experimental procedures of each method obtain liquid water structure factor, $S_{AB}(q)$, which is related to the RDF by the following equation[40]:

$$S(q) = 4\pi\rho \int_0^\infty r^2[g(r) - 1]\frac{\sin(qr)}{qr}dr \tag{2.16}$$

Through inverse FFT, we would obtain

$$g(r) = 1 + \frac{1}{2\pi^2 r\rho}\int_0^{q_{max}} qs(q)\sin(qr)dq \tag{2.17}$$

We can see that Eq.[2.17] is the same as Eq.[2.5] except the upper limit in the integral is q_{max} instead of infinity, which Eq.[2.17] engenders the difficulty to obtain accurate RDFs. For instance, it engenders non-physical error in the small OO distance in the RDF from x-ray derived structure factor. There are several techniques that have been introduced in order to extract the RDF from the scattering information[2, 40]. A number of these methods, including reverse Monte Carlo[44] and empirical structure potential refinement (ESPR)[31], are to fit the experimental data into a computer simulation of

water. However, these techniques usually bias the experimental observations. They show strong dependence upon the reference water models that are utilized. This problem is exacerbated by the fact that the structure factors do not appear to be very sensitive towards the first peak of the OO RDF. On the other hand, the OH distribution appears to have greater reliability .

We will discuss more about RDF in the coming chapters when it is used to study the liquid water structure.

2.4 Computer simulation studies: Molecular dynamics

There are a range of techniques of a quasi-experimental character, referred to collectively as computer simulation, the importance of which in the development of liquid state theory can hardly be overstated. Simulation provides what are essentially exact results for a given potential model. Its usefulness rests ultimately on the fact that a sample containing a few hundred or few thousand particles is in many cases sufficiently large to simulate the behavior of a macroscopic system.

There are two classical approaches: the Monte Carlo method and the method of molecular dynamics. There are many variants of each, but in broad terms a Monte Carlo calculation is designed to generate static configurations of the system of interest, while molecular dynamics involves the solution of the classical equations of motion of the particles. In the application to *ab initio* field, for instance, first principle quantum Monte Carlo and diffusion quantum Monte Carlo are highly accurate in computing electronic energies for geometries optimized with different approaches (e.g. quantum chemical approach, DFT)[45]. However, the application to liquid water is too expensive to perform.

Molecular dynamics has the advantage of allowing the study of time-dependent processes. It is based on forces (thus they collectively move all the molecules and nuclei) which is not what is done in simple MC methods. Molecular dynamics simulations have increased in complexity from simple pairwise interaction potentials to fully quantum mechanical descriptions. In particular, the efficient AIMD methodology has enabled simulations of liquid water entirely from first principles[46]. Combined with significant increase in available computing power, it has pushed the computational simulation of water to an

exciting stage.

In this light, it is of great importance to understand the accuracy and predictive power of AIMD simulation. Right here some brief introductions of key concepts are mentioned. More detailed explanations can refer to relevant appendix in the end of this thesis and some cited references.

Molecular dynamics is powerful to simulate classical many-body systems[47]. In order to simulate a specific material, one needs a realistic model of its potential energy $\Phi(R)$. There are mainly two directions in approaching the interactions of the systems. The first is to build a potential utilizing model functions based mostly upon physical intuition. The parameters are chosen such that the results from the model match either experiment or the results of accurate quantum chemical calculations. These model potentials are known as "force fields"[48, 49]. However, the basic difficulty lies in the case that *transferability* is limited in "force fields" approach. Force fields are typically unable to treat bond breaking and forming events, which are essential to describe chemical reactions.

The second approach, which is what we are concentrating on in this entire thesis, is to start from the quantum mechanical description of atomic interactions, and make necessary approximations in order to solve these many-body complex equations. This would be highlighted in the coming subsection.

2.4.1 Density functional theory

The fundamental laws necessary to the mathematical treatment of large parts of physics and the whole of chemistry are thus fully known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved.

P.A.M. Dirac (1929)[50]

The above statement is written more than eighty years ago. There is an oral tradition that, shortly after Schrödinger's equation for the electronic wave function Ψ had been put forward and spectacularly validated for small system like He and H_2. P.M.Dirac declared that chemistry had come to an end, as its content was entirely contained in that powerful equation. Too bad, he is said to have added, that in almost all cases, this equation was far too complex to allow solution.

R.P.Feynman reiterated this practical difficulty after 30 years as:

Don't forget that the reason a physicist can really calculate from first principles is that he chooses only simple problems. He never solves a problem with 42 or even 6 electrons in it. So far, he has been able to calculate reasonably accurately only the hydrogen atom and the helium atom.

R.P. Feynman(1969)[51]

These complex problems could not be solved with sufficient accuracy until the birth of modern density functionl theory (DFT)[52, 53]. DFT of electronic structure is widely and successfully applied in simulations throughout science and engineering. Kohn noted in this Nobel lecture that DFT "has been most useful for systems of very many electrons where wave function methods encounter and are stopped by the exponential wall"[54]. The beauty of DFT

is that its formalism is in principle exact yet efficient, with one determinant describing the electron density – all of the complexity is hidden in one term, the exchange correlation functional. While the exchange-correlation energy is a functional of the density, the explicit form of this functional in general is not known.

DFT is an exact reformulation of the ground-state non-relativistic quantum mechanics of many-body systems. It provides an unprecedented balance of accuracy and efficiency for electronic structure calculations in molecules, clusters, and solids. Our theoretical investigating tool of liquid water structure is the *ab initio* molecular dynamics. These key features of Kohn-Sham DFT, especially Car-Parrinello molecular dynamics that ingeniously combines the DFT and molecular dynamics are reviewed in Appendix A, B.

2.4.2 Exchange correlation functional

The ground-state energy contribution that needs practical approximations within KS-DFT approach is the exchange correlation functional. This term holds the key to the success of DFT. Exchange arises from antisymmetry due to the Pauli exclusion principle, and correlation accounts for the remaining complicated many-body effects that need many determinants to be fully described. However, the form of exchange-correlation in terms of the density remains unknown and it is necessary to use approximations. The choice of exchange-correlation functional is critical in generating accurate results from DFT methodology. In the following section, major triumph but also major failures of some explicit approximation would be reviewed, which is the motivation to develop advanced improved functional (e.g. PBE0 hybrid functional) and vdW dispersion interaction correction.

The most simple approximation is to assume that the exchange-correlation functional is solely dependent upon the local electronic density. This is well known as the local density approximation (LDA)[53]. In this approximation, it is assumed that at each point in space the exchange correlation energy per particle is given by its value for a homogeneous electron gas, namely,

$$E_{xc}^{LDA} = \int d\mathbf{r}\ n(\mathbf{r})\ \epsilon_{xc}(n(\mathbf{r})) \tag{2.18}$$

The LDA exchange-correlation functional is garnered from the homogenous electron gas and can be computed to a high accuracy via quantum Monte Carlo simulation[55], and the LDA exchange-correlation energy may be extracted from these results. Although the LDA approximation is successful in describing many systems, it does not perform well in molecular, H-bonded systems such as water, as density is clearly very far away uniform in these systems.

Many of quantitative failures of the LDA are remedied using the generalized gradient approximation (GGA). The GGA functional depends not only on the local electron density, but also on the local gradient of the density in order to take into account approximately of the inhomogeneities in the density.

$$E_{xc}^{GGA} = \int d\mathbf{r}\ n(\mathbf{r})\left(\epsilon_x(n(\mathbf{r}), \nabla n(\mathbf{r})) + \epsilon_c(n(\mathbf{r}), \nabla n(\mathbf{r}))\right) \tag{2.19}$$

where ϵ_x and ϵ_c are the exchange and correlation energy densities, respectively.

There are many different choices of GGA functionals. For instance, Perdew-Burke-Ernzerhof(PBE) functional[3] are derived from purely physical considerations, that is, they are not fitted to any specific system, and are instead made to satisfy some general physical properties[56]. Some make use of fitting parameters. For example, the exchange functional of Becke[57] contains one parameter that is fit to the exact exchange (from a Hartree-Fock calculation) of six noble gas atoms. The Lee-Yang-Parr (LYP) correlation functional[4] contains four parameters that originate from the investigation of helium. These two functionals are often employed together, denoted as BLYP functional.

Semi-local GGA corrects (often slightly overcorrects) the LDA overestimation of bonding strength and provides an absolute average error of molecular atomization energy of the order of 0.3 eV for 20 selected molecule systems[3]. This is still above the desired chemical accuracy, but much better than LDA[58, 59, 60]. GGA works well for the H-bond in the water dimer and also for the dipole moment of the water molecule. Thus one would have expected that it should work well for the liquid as well. It was only in the last decade that it became clear that this was not the case. The application of GGA functional to liquid water structure at ambient condition, although

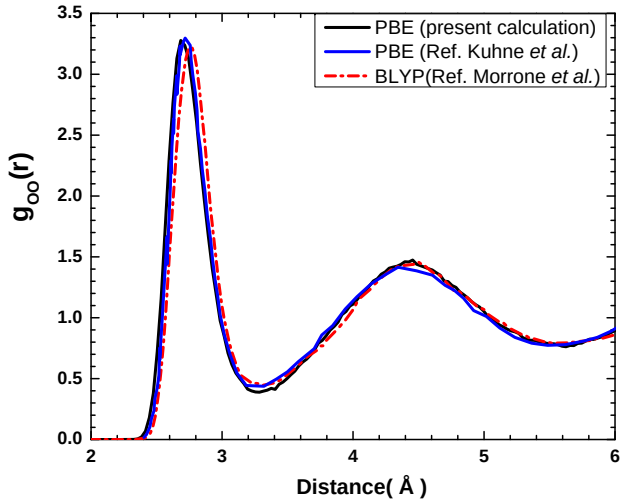

Figure 2.5: Benchmark calculation of OO RDF with PBE functional compared to reference, with PBE(Ref. Kuhne *et al.*)[6] and BLYP(Ref. Morrone *et al.*)[10].

often qualitatively correct, yields an over-structured oxygen-oxygen (OO) RDF than experiment[5]. Moreover, GGA functional is usually poor in the description of water density and pressure. This indicates that GGA functional is far from perfect in the correct description of liquid water and thus needs improvement both for exchange and correlation.

2.4.3 Liquid water structure with the PBE functional

For further investigation of liquid water structure with more accurate and improved exchange-correlation functional, it is necessary to have a benchmark calculation of liquid water structure with GGA functional. The RDFs obtained from AIMD simulation are always considered as a main tool on

monitoring liquid water structure. The liquid water structure with GGA functional has been widely investigated in the past few years[5, 6, 10, 61]. Typical AIMD simulation of liquid water uses 64 water molecules with periodic boundary conditions (refer to Appendix C to learn more about finite size effects). In our simulation, a larger supercell size with 128 water molecules has been used to have better statistical data and to better compare with the experimental structure factor. This will be illustrated when we discuss the simulation results in the later chapter.

Appendix C discusses how to perform a meaningful CPMD simulation of liquid water. Our CPMD simulation using Quantum-ESPRESSO code[62] with GGA functional (PBE here) yields a liquid water structure in terms of RDFs (g_{OO} and g_{OH}) quite close to other independent AIMD simulation results mentioned previously. Fig.[2.5] plots the OO RDF of AIMD simulation with GGA functional compared to two other reference data[6, 10]. We can observe that all three $g_{OO}(r)$ are close to each other. Moreover, two $g_{OO}(r)$ with PBE functional are even more close, with some tiny difference compared to the one with BLYP functional. The liquid water structure with various GGA functional would lead to some small difference as the derivations of different GGA functionals discussed in section[2.4.2] varies. Nevertheless, it is widely accepted that GGA functional leads to an overstructured liquid water structure. And our calculation with PBE is consistent with these independent reference calculation, indicating that the benchmark calculation with GGA functional is successful, i.e. we have performed a meaningful AIMD simulation and this provides a strong cornerstone for the coming simulations with more accurate and advanced functionals to see how they truly improve over PBE.

2.5 Improving *ab-initio* liquid water

2.5.1 The effect of quantum nuclei

The electronic structure calculations based on DFT have been widely used in the studies of molecular and condensed-matter systems. Most DFT applications to extended material systems adopts classical nuclei treatment and use the semilocal generalized gradient approximation (GGA) for exchange-correlation functional. The structure of the liquid as obtained from first-principle simulations based on DFT yields an over structured radial distribution function(RDF) for oxygen-oxygen(OO), oxygen-hydrogen(OH) and hydrogen-hydrogen(HH) compared to experiment[5, 6].

Part of this over-structuring arises from the classical treatment of nuclei, as nuclear quantum effect plays an important role as seen from isotope experimental evidence. The lighter the species, the larger quantum effects should be expected. The trend in the Fig.[2.6] indicates that nuclear quantum effect weakens the H-bond network in the water system. For instance, the melting and boiling point increases as does the enthalpy of formation and fusion from light water (H_2O) to heavy water (D_2O).

It has been recognized recently that a sizeable portion of the over structuring should be attributed to nuclear quantum mechanics, where the path-integral AIMD (PI AIMD) methodology is employed to include nuclear quantum effect[10]. The structural effect of quantum nuclei in water is illustrated in Fig.[2.7]. From the $g_{OO}(r)$ and $g_{OH}(r)$, we can see that the inclusion of nuclear quantum effect leads to significantly less structured RDFs than the corresponding AIMD simulation with classical nuclei. The second peak of the OH RDF, which corresponds to H-bonding interactions, becomes less

	Melting Point	Boiling Point	Enthalpy of formation [$X_2O_{(\ell)}$]	Enthalpy of fusion	Dielectric Constant
H_2O	0.0°C	100.0°C	-285.8 kJ/mol	6.01 kJ/mol	88.2 (T=0°C) 78.4 (T=25°C)
D_2O	3.8°C	101.4°C	-294.6 kJ/mol	6.28 kJ/mol	85.8 (T=5°C) 78.1 (T=25°C)
T_2O	4.5°C	101.5°C	—	—	—

Figure 2.6: Selected experimental properties of light, deuterated and tritiated water. Tables taken from thesis of J.Morrone[63].

sharper. The first peak of the OO RDF is a useful marker of the relative structuring of water. A recent analysis of the experimental data has revealed significant uncertainty in the peak height[2]. Two data sets from Ref.[2] are plotted in Fig.[2.7] alongside results from Ref.[31]. The resultant peak from the path-integral simulation is 2.84, which is closer to the latest experimental peak height (2.58)[64] than the value of 3.23 obtained from a standard AIMD simulation. This result is in agreement with increased average fraction of broken H-bonds observed in the path integral result. However, there are still some overstructuring in its first minima and the second peak compared to experimental $g_{OO}(r)$, where there is little controversies among various experimental data.

From Fig.[2.7] we can observe that $g_{OO}(r)$ with quantum nuclei at 300K (black line) is very close to $g_{OO}(r)$ with classical nuclei at 330K (red dot). This observation provides the rationale for mimicking quantum effects on the OO RDF by a temperature increase of ∼30K, and gives a justification for a procedure that we adopt later in our study to include nuclear quantum effect via 30K temperature increase with improved exchange-correlation functional for liquid water.

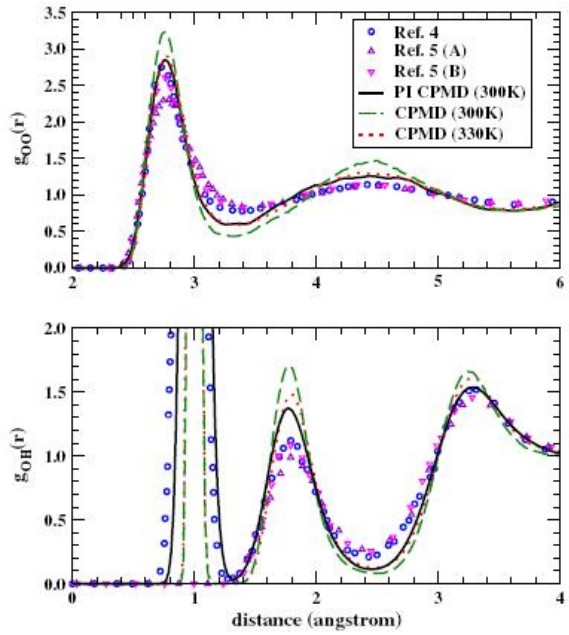

Figure 2.7: The OO (top) and OH (bottom) RDFs in liquid water from a PI CPMD simulation at 300 K (solid line) and standard CPMD simulations at 300 K (dashed line) and 330 K (dotted line) are reported with joint neutron/x-ray data. Plots taken from Ref.[10]

.

Based on these observations, we can conclude that the overstructuring present in standard AIMD simulations at room temperature is in part mitigated by the inclusion of nuclear quantum effects. Yet, despite the improvement, still there remains some degree of overstructuring.

2.5.2 The PBE0 hybrid functional

The remaining deviations from experiment in the calculated RDFs even after nuclear quantum effects are included bring up the issue of the DFT accuracy in the description of the Born-Oppenheimer potential energy surface of water. Though GGA functional often improves quantitatively over LDA, it is not a panacea. Both actually suffer from serious formal deficiencies. One important deficiency at the heart of many problems that occur with LDA and GGA functional is the presence of self-interaction. Another is the missing of non-local correlation, i.e. van der Waals (vdW) dispersion interaction that is the topic in the next subsection.

The total energy expression within KS-DFT in Eq.[A.6], is partitioned into a classical Coulomb part and an everything else exchange-correlation part. As a result, the KS equation has an explicit classical electron-electron repulsion potential $e^2 \int d\mathbf{r}' n(\mathbf{r}')/|\mathbf{r} - \mathbf{r}'|$ known as Hartree potential. This expression views the quantum charge distribution as a distribution of continuous classical charge. This, however, means that it is inherently in error, because each electron is repelled from the total charge in the system, including a spurious repulsion from itself. This effect is known as the *self-interaction error*, noted by Fermi in the context of Thomas-Fermi theory. This spurious repulsion is particularly easy to see in one-electron system, where clearly there should be no electron-electron repulsion at all, and yet the Hartree term is not zero. Because the electron-electron interaction is, in principle, handled exactly in DFT, whatever error we are making in the Hartree term must be completely canceled out by the exchange-correlation term. Unfortunately, only partial cancelation is obtained in either LDA or GGA. This seemingly innocuous issue has far-reaching consequences and is one of the

most fundamental problems in DFT.

The AIMD simulations using GGA functionals in the Fig.[2.7] are affected by self-interaction delocalization errors[65]. These errors can lead to sizable structural effects in water and H-bonded systems, because H-bonds are relatively weak bonds resulting from a delicate balance between covalent and electrostatic effects. The delocalization error tilts slightly the balance favoring proton donation and enhances the H-bond strength. As a result, it leads to an over structured liquid.

Due to this shortcoming of GGA, new exchange-correlation functionals have been developed. One promising area is orbital-dependent density functionals[66]. Typically, there are four classes of orbital-dependent functional:(1): Hybrid functionals. (2): Self-interaction correction. (3): The meta-generalized-gradient approximation. (4): Exact exchange with compatible correlation. In the following we are adopting hybrid functionals(vdW non-local correlation that is missing in GGA and hybrid functionals as well will be included as well and will be mentioned shortly). For more detailed information, one can refer to Ref.[66]. In this scheme, one utilizes the exact (Hartree-Fock) exchange term mixed with an exchange-correlation functional, yielding a so-called *hybrid functional* [67]. For most systems, such hybrid functionals are the most accurate option available, due to the presence of Hartree-Fock exchange term, although they are of greater computational cost than GGA.

The physics hybrid functionals are based on is best understood within the framework of a fundamental DFT concept: adiabatic connection theorem[68, 69]. Becke[70] realized that the adiabatic connection formalism can serve as a practical tool for functional construction. He suggested a mixture of Hartree-

Fock and LDA or GGA and dubbed this hybrid functionals, i.e.

$$E_{xc}^{hyb} = bE_x^{exact} + (1-b)E_x^{app} + E_c^{app} \tag{2.20}$$

where the superscripts *hyb* and *app* denote the hybrid functionals and the approximate (typically semilocal) functional from which it is constructed.

What value of mixing parameter b should one choose and what improvement over the LDA or GGA, if any, is obtained? Becke suggested practical usage of Eq.[2.20] by determining b empirically through least-squares fitting to atomization energies, ionization potentials, and proton affinities of a canonical set of mostly organic molecules, based on first and second row atoms, known as the G2 set. The optimal value of b is found to be 0.16 to 0.28, depending on the type of semiempirical correlation functional used for the fit. These values resulted in mean atomization errors of only 0.13 or 0.09 eV, respectively for the G2 set. This proves that Eq.[2.20] is in principle a useful form.

From Becke's results, we can see that whether or not $b = 0.25$ is a practically useful value has been found to be a strong function of the type of GGA used[11, 71, 72]. Fortunately, for the nonempirical PBE GGA, $b = 0.25$ is found to be an excellent value[11]. For example, for the largest G3 set, the mean error obtained for the enthalpies of formation with PBE GGA is 0.94 eV, whereas it is only 0.20 eV if Eq.[2.20] is used with the PBE exchange correlation functional and $b = 0.25$. The PBE0 hybrid functional we are adopting here is the 0.75 PBE exchange and 0.25 exact exchange, with PBE correlation.

Recently, there have been extensive studies of liquid water structure with hybrid functionals, due to the deficiency of semi-local GGA functional. The applications to molecular systems have shown that hybrid functionals are generally superior to GGA in the description of electronic and structure

properties[66]. For instance, Zhang *et al.*[73] performed first principles simulations of the infrared spectrum of liquid water using hybrid density functionls and observed a much better agreement with experimental results than GGA description. In particular, the measured stretching and bending mode of liquid water is quite good. This improvement originates from a more accurate account of the vibrational properties of water monomer and dimer, an underlying structural model of the liquid with a smaller number of H-bonds and oxygen coordination and a smaller effective molecular dipole than GGA functional.

When extended systems are studied, some approximation in the calculation of exact exchange is sometimes adopted to alleviate the computational burden[74]. However, this approximation might lead to some deviation from the true hybrid functionals description of electronic structure, thus resulting in a deviation of how hybrid functionals truly describe liquid water structure. For instance, Guidon *et al.*[13] reported negligible difference in the RDFs, indicating tiny changes in the number and character of H-bonds found in the system. In these simulations, some screening techniques like integral screening and density matrix screening are adopted to alleviate expensive computation cost even with massively parallel computing resources.

The calculation of exact exchange, if performed within delocalized basis sets, e.g. plane-wave basis set, is quite expensive and scales cubically with respect to system size. This is the main reason one is trying to reduce this cost by performing some approximation for this term. Taking 32 water molecules system for instance, in the simulation of Zhang *et al.*, the ratio of the CPU time per AIMD step using PBE0 versus PBE is about a factor of 25[73] even after efficient massive parallelization scheme is implemented, if adopting the regular calculation within delocalized plane wave basis sets. Due to this,

it would take several months to have a meaningful and converged AIMD simulation as 20 picosecond (ps) simulation is typically needed. As we have already noticed, 64 water system is currently the appropriate choice in order to eliminate the finite size effect in liquid water structure description. It is therefore impossible to run a practical AIMD simulation of 64 water supercell system in the traditional scheme without some algorithmic breakthrough.

Due to this, we are motivated to include PBE0 functional in liquid water simulation to see the role PBE0 indeed plays in the description of liquid water structure. An efficient and accurate calculation of exact exchange is performed, adopting a linear scaling algorithm developed by Wu *et al.*[14].

2.5.3 Non local correlation functional for vdW dispersion interactions: the TS approach

Another deficiency of GGA and hybrid functionals is the missing of the vdW non-local correlation. While DFT has been widely successful in describing the properties in dense materials and isolated molecules with LDA and GGA approximations, sparse systems, soft matter, molecular van der Waals complexes, biomolecules and the like can not be adequately described where the role of vdW effect is indispensable. The origin of vdW interaction between two non-chemically bonded fragments is the coupling of the electric field generated by fluctuations in the electron density of one fragment with the density of the other fragment. This is a dynamical correlation effect that local (LDA) and semi-local functional (GGA) can not capture, and is either not related to the exchange, so the hybrid functionals we were talking about do not help either. The issue of vdW dispersion interactions is a difficult benchmark in many-body theory, which is what correlation functional should address.

These difficulties resulting from long range interactions have been widely studied in the past decades[75]. Generally two approaches are usually adopted. One scheme is the so-called explicit density functional methods. The exchange-correlation energies are given as explicit functionals. This means that once the occupied Kohn-Sham (KS) orbitals and hence the density is obtained, the exchange-correlation energy is simply evaluated. This requirement guarantees that the scaling of the computational cost with system size would not destroy the cubic scaling by ordinary DFT.

The modern version of the nonempirical vdW density functional (vdW-DF) was introduced few years ago by Dion *et al.*[76]. The fully self-consistent

version came out recently[77]. The vdw-DF method has shown promise for various systems where long range vdW interactions are important[78]. Recent study with vdW-DF applied to liquid water simulation yields much less structured compared to GGA results, though there is a very large collapse in the second coordination shell, suggesting possible overestimation of the vdW effect in this scheme[79]. Thus this approach might not be suitable for liquid water structure description.

The second developing approach is based on C_6 coefficients and damping functions. This is based on London's original work yielding the C_6/R^6 asymptotic dissociative scaling between two atoms at distance R[80]. The total vdW energy is given by

$$\sum_{I<J} C_{6I} C_{6J} / R_{IJ}^6 \tag{2.21}$$

The form C_6/R^6 is only valid at large separation R between two atoms. It must be damped at shorter distance simply because it becomes incorrect (much before repulsion terms dominate). At shorter separation it is replaced by the common DFT approximations.

The asymptotic C_6/R^6 form is usually added as an empirical correction to DFT energies (see e.g. Grimme in Ref.[81]). In the vdW scheme by Tkatchenko and Scheffler(vdW-TS)[15], however, the C_6 coefficients are given as an explicit functional of the density, making the scheme largely a non-empirical DFT scheme like the Dion *et al.* scheme[76]. Application to the S22 molecular data set indicates that in this data set the vdW-TS scheme is significantly more accurate than Dion *et al.* and also more accurate than the empirical Grimme's scheme[81]. Therefore in the following section, we adopt the accurate vdW-TS scheme by Tkatchenko and Scheffler[15] that is implemented in the current project.

With vdW-TS approach to study vdW effect, Santra *et al.*[45] recently found that vdW has a substantial effect on the transition pressures between crystalline ice phases, the consequence of which is that the transition pressures from DFT neglecting vdW effect are greatly overestimated. The role of vdW becomes more important in the high density phase region. As a result, vdW plays a crucial role in determining the relative stabilities and phase transition pressures in ice. These observations are of relevance to understand intermolecular interactions of water in all of its condensed phases. The studies of vdW effect on liquid water structure with vdW-DF form from Wang *et al.*[79] also shows that the underestimated equilibrium density given by GGA functional increases dramatically together with observed higher diffusivity when vdW density functional (vdW-DF) of Dion *et al.*[76] is included. On the other hand, however, this vdW-DF functional overestimates the vdW effect and thus excessively weaken the H-bond network, resulting in a collapsing of the second coordination cell. Recently Jonchiere *et al.*[82] presented a study of the vdW effects on liquid water based on semi-empirical correction of Grimme *et al.*[83], showing that the structure and diffusivity of ambient-like liquid water are sensitive to the fifth neighbor position, thus highlighting the key role played by this neighbor and that the semi-empirical correction by Grimme *et al.* improves significantly both the structural and the dynamical description of ambient water and supercritical water.

It is noteworthy that the effects of the environment on both the C_6 coefficients and the damping functions are completely determined by the electron density in accordance with DFT principles. One empirical parameter, fitted to the S22 database of molecular binding energies[84] is used to define the onset with distance of the dispersion correction. The vdW correction improves substantially the description of weakly bonded systems within DFT. For example,

the correct ground-state structure of the water hexamer, missed by standard DFT calculations, is obtained in this way[85]. It makes this scheme quite appealing for *ab initio* simulations of water. Based on this, our group include vdW correlation corrections according to this scheme in AIMD simulations. The results obtained for water clusters (e.g. the hexamer) including vdW environmental effects are compared with calculations carried out at the so called EXX/RPA(EXX denotes exact exchange and RPA the random phase approximation[86]) level recently used to investigate dispersion interactions in weakly bonded systems, such as benzene, methane[87] and self-assembled monolayers on surfaces[88]. Such a comparison would permit to test the robustness of the environmental approach described above. We note that thanks to recent progress in evaluating dielectric matrices, EXX/RPA calculations for relatively large systems(e.g. containing tens of water molecules) are now becoming feasible. The study of above should give us good insight on the various physical effects that determine the microscopic structure of liquid water.

2.6 Summary

Hybrid functionals improve the description of vibrational properties both in water clusters and bulk water. vdW's contribution is important for the transition pressures among ice phases and therefore one should infer that it plays an important role in defining the equilibrium density both in ice and liquid water. Both hybrid functionals and vdW would correct the deficiency of GGA functional and lead to a better liquid water description than semi-local GGA functional.

Furthermore, nuclear quantum effects are also important to correctly describe liquid water structure. Thus the most accurate description of liquid water would require a PI-AIMD simulation based on a functional like PBE0+TS (PBE0 functional plus vdW-TS correlation correction). This however, would be very expensive because of the computational cost of the hybrid functionals combined with the need to simulate multiple system replicas in a PI simulation. In the present thesis we have therefore opted for a PBE0+TS simulation in which the nuclei are treated as classical particles and decided to focus the attention on the site distribution of the oxygen nuclei, which are significantly heavier than the hydrogens and are thus less prone to quantum behavior. As shown in a recent PI-AIMD study the effect of quantum dynamics on the oxygen-oxygen RDF is captured with surprising accuracy by a temperature increase of about 30K. We therefore perform simulations at 330K to compare the oxygen pair correlations and structure factor to experiment at 300K.

3 Numerical implementation of the PBE0+TS functional

In this chapter, we focus on the numerical implementation of the PBE0+TS functional, with emphasis on the simulation of liquid water.

3.1 Numerical implementation of the PBE0 functional

3.1.1 Maximally localized wannier function dynamics

We calculate Hartree Fock exchange in the PBE0 functional efficiently working under maximally localized Wannier function (MLWF) representation according to Wu *et al.* algorithm[14]. This is the footstone to perform the linear scaling method. In this section, an efficient algorithm of wannier dynamics that we are currently using is presented and illustrated.

In the AIMD simulation, MLWFs are updated at each MD time step. Following the scheme by Marzari and Vanderbilt[33], we define overlap matrices

$$M_{ij}^{l} = \langle w_i \mid e^{-i\mathbf{G}_l \cdot \mathbf{r}} \mid w_j \rangle \tag{3.1}$$

where $\mathbf{G}_l$ are the primitive reciprocal lattice vectors and w_i are Wannier functions(WF). One then defines Ω which is the functional of WFs given by

$$\Omega = \sum_{i=1}^{N} \sum_{l=1}^{3} |M_{ii}^{l}|^2 \tag{3.2}$$

where the index i runs over all the occupied bands and l runs over the three Cartesian components. The spread S of the WFs is related to Ω by

$$S = \left(\frac{L}{2\pi}\right)^2 (3N - \Omega) \tag{3.3}$$

where L is the linear dimension of the unit cell. The MLWFs are obtained by minimizing the functional S or, equivalently, by maximizing the functional Ω. In AIMD scheme, the electronic degrees of freedom are propagated with the Verlet algorithm[89, 90]. This implies that the WFs at time $t + \Delta$ are related to those at time t and $t - \Delta$ via

$$\bar{w}(t + \Delta) = 2w(t) - w(t - \Delta) + \frac{\Delta^2}{\mu}\left(-\frac{\delta E(t)}{\delta w*} + \Lambda w(t)\right) \tag{3.4}$$

where $E(t)$ is the KS potential energy and $\delta E(t)/\delta w^*$ is the "force" acting on orbital w_i and Λ are Lagrange multipliers that impose orthonormality constraints on the orbitals. We denote the predicted WFs at time $t + \Delta$ by $\bar{w}(t + \Delta)$ to indicate that this is not a MLWF any more here because it gets delocalized due to one time step propagation of AIMD. To transform $\bar{w}(t+\Delta)$ into a MLWF, we need to reminimize the spread functionals at time $t + \Delta$. In this way, we obtain the appropriate unitary transformation U that minimize S at time $t + \Delta$. Thus:

$$w(t + \Delta) = U\bar{w}(t + \Delta) \tag{3.5}$$

We minimize S (i.e. maximize Ω) by damped dynamics for U. The equation of motion is:

$$Q\ddot{U} = \frac{d\Omega}{dU} - \gamma\dot{U} + \lambda\frac{d}{dU}(U^T U - 1) \tag{3.6}$$

where Q is a fictitious mass parameter for U, $\frac{d\Omega}{dU}$ is the force, and γ is the damping coefficient. λ is the Lagrange multiplier that ensures the unitarity of U.

The Verlet scheme is used to compute $U(t + \Delta)$ from $U(t)$ and $U(t - \Delta)$. Thus, Eq.[3.6] numerically becomes

$$\begin{aligned} U_{i,j}(t + \Delta) = 2U_{i,j}(t) - U_{i,j}(t - \Delta) + \frac{\Delta^2}{Q}[(\frac{d\Omega(t)}{dU})_{i,j}] \\ -\frac{\gamma}{2\Delta}[U_{i,j}(t + \Delta) - U_{i,j}(t - \Delta)] + 2\Sigma_k U_{ik}\lambda_{kj} \end{aligned} \tag{3.7}$$

which on rearrangement yields

$$U_{i,j}(t+\Delta) = \bar{U}_{i,j}(t+\Delta) + \sum_k U_{ik} X_{kj} \tag{3.8}$$

with

$$X = \frac{\gamma}{\Delta}\lambda \tag{3.9}$$

$$\bar{U}_{i,j}(t+\Delta) = \frac{1}{1+f}\left(2U_{i,j}(t) - (1+f)U_{i,j}(t-\Delta) + \frac{\Delta^2}{Q}[(\frac{d\Omega(t)}{dt})_{i,j}]\right) \tag{3.10}$$

where $f = \frac{\gamma\Delta}{2}$ and the force $(\frac{d\Omega(t)}{dt})_{i,j}$ is given by

$$\frac{d\Omega}{dU_{i,j}} = \sum_l \left[\sum_k \left(B_{ik}U_{kj}M_{jj}^{l*} + U_{kj}B_{ki}M_{jj}^{l*}\right) + C.C.\right] \tag{3.11}$$

with

$$B_{ij} = \langle \varphi_i \mid e^{-i\vec{G}_l\cdot\vec{r}} \mid \varphi_j \rangle \tag{3.12}$$

where $C.C.$ denotes the conjugate complex part in the parenthesis.

To compute the lagrange multiplier X in Eq.[3.9], we impose the constraint of unitarity on $U(t+\Delta)$. This constraint leads to a quadratic expansion of X, which is resolved by an iterative procedure, similar to what is done in classical molecular dynamics simulation to impose rigidity constraints. In this case, the unitary condition of $U(t+\Delta)$ i.e. $\sum_k U_{ki}U_{kj} = \delta_{ij}$ results in the following expression

$$C + BX + X^T B^T + X^T X = I \tag{3.13}$$

with

$$C_{ij} = \sum_k \bar{U}_{ik}^T \bar{U}_{kj} \quad B_{ij} = \sum_k \bar{U}_{ik}^T U_{kj} \tag{3.14}$$

The solution is obtained by the following iterative procedure. The initial term is set to the zero-order approximation

$$X_0 = \frac{1}{2}(I - C) \tag{3.15}$$

and then the iterative cycle proceeds according to the recursive formula

$$X_{k+1} = \frac{1}{2}\left[(I - C) + X_k^T(I - B^T) + (I - B)X_k - X_k^T X_k\right] \tag{3.16}$$

This algorithm typically converges in less then ten steps in the iterative procedure. Once the matrix of the constraints X has been determined, the new unitary matrix is calculated according to Eq.[3.8].

This procedure is tested to be significantly more efficient than the alternate procedure implemented in the Quantum-ESPRESSO package[91].

This scheme can be parallelized quite easily and therefore can be used to cope with large systems when the matrix size grows big correspondingly. In this scheme, there are massive matrix multiplication manipulations whose computational cost grows cubically with increased number of electrons in the system. The parallelization technique implemented is called Cannon's algorithm in computer science, a distributed algorithm for matrix multiplication for two-dimensional meshes first described in 1969 by Lynn Elliot Cannon[92]. It is especially suitable for computers to lay out in a N by N mesh. It has been shown that Cannon's algorithm work well in homogeneous 2D grids. The main advantage of the algorithm is that its storage requirements remain constant and are independent of the number of processors.

Under the same condition, this damped dynamics with respect to U shows better efficiency, requiring less than one second on overage per AIMD time step in 64 water system, which is quite efficient compared to the coming expensive exact exchange calculation in PBE0 functional.

3.1.2 Calculation of exact exchange

We adopt the scheme proposed in Ref.[14] to calculate exact exchange. This approach has computational cost that scales linearly rather than cubically with system size. We take PBE0[93] as the exchange correlation energy

$$E_{xc}^{PBE0} = \frac{1}{4}E_x + \frac{3}{4}E_x^{PBE} + E_c^{PBE} \tag{3.17}$$

where E_x denotes exact exchange, E_x^{PBE} is the PBE exchange, and E_c^{PBE} is the PBE correlation functional. The exact exchange energy E_x has the Hartree-Fock expression in terms of the one-electron(pseudo-) orbitals:

$$E_x = -2\sum_{i,j}\int\int \frac{\varphi_i^*(\mathbf{r})\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r})\varphi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}d\mathbf{r}' \tag{3.18}$$

The corresponding Kohn-Sham(KS) Eq.[A.8] for occupied orbitals is

$$\left[\frac{-\hbar^2\bigtriangledown^2}{2m} + V_{ion}(\mathbf{r}) + V_H(\mathbf{r}) + V_c^{PBE}(\mathbf{r}) + \frac{3}{4}V_x^{PBE}(\mathbf{r})\right]\varphi_i(\mathbf{r}) + \frac{1}{4}\int V_x(\mathbf{r},\mathbf{r}')\varphi_i(\mathbf{r}')d\mathbf{r}' = \varepsilon_i\varphi_i(\mathbf{r}) \tag{3.19}$$

$V_c^{PBE}(\mathbf{r})$ and $V_x^{PBE}(\mathbf{r})$, the PBE exchange and correlation potential respectively, are functionals of density and its gradient. For the Hartree Fock exact exchange part, the action of $V_x(\mathbf{r},\mathbf{r}')$ on each orbital φ_i results in an orbital dependent term

$$D_x^i(\mathbf{r}) \equiv \frac{\delta E_x}{\delta\varphi_i^*} = -2e^2\sum_j^{N/2}\int d\mathbf{r}'\frac{\varphi_j^*(\mathbf{r}')\varphi_j(\mathbf{r})\varphi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \tag{3.20}$$

In typical implementation for extended systems, the calculation of $\int d\mathbf{r}'\frac{\varphi_j^*(\mathbf{r}')\varphi_i(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ in D_x^i can be performed using the Fast Fourier Transform (FFT) algorithm with cost proportional to $N_{FFT}\ln(N_{FFT})$ for a plane wave grid of size N_{FFT}. Since each Bloch function φ_i is delocalized throughout the entire

supercell, the integral above results in overall computation proportional to $N^2 \times N_{FFT} \ln(N_{FFT})$, which, neglecting the logarithmic dependence, amounts to cubic scaling with system size.

Instead of evaluating exact exchange from delocalized Bloch orbitals, we turn to the Wannier representation. The Wannier functions $w_i(r)$ are obtained from the occupied Kohn-Sham orbitals through a unitary transformation which leaves the ground-state energy invariant, i.e. $w_i = \sum_{j=1}^{N/2} U_{ij}\varphi_j$.

In terms of the w_i, $D_x^i(\mathbf{r})$ becomes

$$D_x^i(\mathbf{r}) = -2(v_{ii}(r)w_i(\mathbf{r}) + \sum_{i \neq j} v_{ij}(r)w_j(\mathbf{r})) \tag{3.21}$$

where

$$\nabla^2 v_{ii}(\mathbf{r}) = -4\pi\tilde{\rho}_{ii}(\mathbf{r}) \ , \ \nabla^2 v_{ij}(\mathbf{r}) = -4\pi\tilde{\rho}_{ij}(\mathbf{r}) \tag{3.22}$$

with the overlap matrix

$$\tilde{\rho}_{ij}(\mathbf{r}) = w_i(\mathbf{r})w_j^*(\mathbf{r}) \tag{3.23}$$

Then in real space, the exchange interaction is only present in the region where two orbitals overlap, i.e. where $\tilde{\rho}_{ij} \neq 0$, then the pair potential v_{ij} is efficiently computed by solving the Poisson equation in a spatial region that is much smaller than the supercell used in the simulation.

In this approach, the computational cost strongly depends on the localization of the Wannier orbitals. Therefore we are MLWFs, which exhibit exponential spatial decay in insulators[94]. Then for each orbital, the number of pair exchanges that needs to be included is finite and is independent of the system size. So the computation cost of an exact exchange calculation scales linearly with system size, as seen from Fig.[3.1], which shows a tagged Wannier orbital interacting with finite number of its neighboring Wannier orbitals in

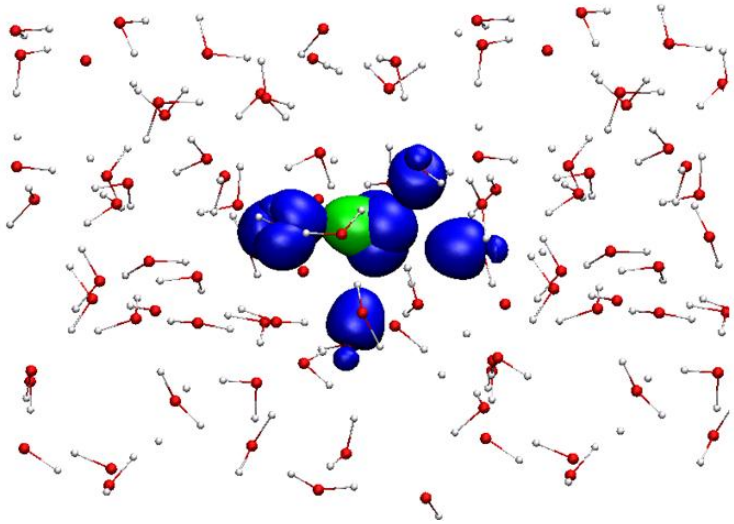

Figure 3.1: Overlap between a tagged Wannier orbital(green at the center) and its nearest neighboring Wannier orbitals(blue) in 64 water molecules supercell. Red dots denote oxygen atoms and white hydrogen atoms.

an AIMD simulation snapshot. Those Wannier orbitals far away from the tagged Wannier orbitals would have negligible overlap, therefore leading to negligible exchange interaction.

The practical solution of the above equations can be further improved. The density $\tilde{\rho}_{ij}$ in Eq.[3.23] is known for each pair of orbitals, we can associate with each pair of orbitals an orthorhombic box with sides(l_x,l_y,l_z) so that outside this box the $\tilde{\rho}_{ij}$ is smaller than a given cutoff value ρ_{cut}, the value of which is based on the testing results that there is negligible difference in the exact exchange energy and forces with even lower ρ_{cut}. Then Eq.[3.22] are solved in this greatly reduced grid box without loss of accuracy. For instance, in the 64 water molecules, the real space grid needed to compute the pair potential v_{ij} via Eq.[3.22] contains only 1/8 of the grid points of the simulation cell. Even fewer are only needed for pair potentials v_{ij} that are generated

by more distant orbitals. As the density $\tilde{\rho}_{ij}$ is vanishingly small when the distance between orbital i and j in terms of wannier center is sufficiently large, many pair interactions are negligibly small. In our simulation with 64 water molecules, the average number of pair exchange interaction is about 16 out of 255 orbital pairs. In Ref.[14], 64-atom Si supercell requires about 30 out of 127 pair exchange interactions. This makes sense as MLWFs in water molecules are more localized, associated with smaller wannier spread compared to Si. Therefore it requires fewer pair interactions among Wannier orbitals.

The Poisson equations (Eq.3.22) are solved numerically with Laplace operator ∇^2 discretized on 19 mesh points in three dimension (7 mesh points in one dimension). The resulting finite difference equation has the form of a linear matrix equation of the type $\mathbf{Ax} = \mathbf{b}$. The symmetric and positive-definite square matrix $\mathbf{A}$ is sparse with dimension n, where n is the number of mesh points inside the reduced box. The vector $\mathbf{x}$ corresponds to the unknown v_{ij} and the known $\mathbf{b}$ is the pair density $\tilde{\rho}_{ij}$. The value of v_{ij} at the boundary of the box are computed via multipole expansion.

$$v_{ij}(r) = 4\pi \sum_{l,m} \frac{1}{2l+1} q_{l,m} \frac{Y_{l,m}(\theta, \phi)}{r^{l+1}} \tag{3.24}$$

where the multipoles $q_{l,m}$ are given by the integrals

$$q_{l,m} = \int Y^*_{l,m}(\theta', \phi') \mathbf{r}'^l \tilde{\rho}_{ij}(\mathbf{r}') d\mathbf{r}' \tag{3.25}$$

$Y_{l,m}$ are spherical harmonics referred to the center of pair density. We found that choosing $l_{max} = 6$ is sufficient to achieve excellent accuracy convergence.

Solving the linearized Poisson equation $\mathbf{Ax} = \mathbf{b}$ is equivalent to finding the

vector $\mathbf{x}$ that minimizes the function

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x} + c \tag{3.26}$$

where c is an arbitrary constant. This minimization is efficiently performed with the conjugate gradient (CG) method[95]. The CG method is terminated with the residual of v_{ij} everywhere is smaller than 10^{-6} a.u. The calculation of v_{ij} that are out of reduced size box is performed according to multipole expansion according to Eq.[3.24].

With the calculated $D_x^i(\mathbf{r})$, the forces acting on the orbitals due to the exact exchange can be added in according to Eq.[3.19]. Finally, the exact exchange energy is given by the sum of the energies of the orbital pairs in the presence of the corresponding pair potential $v_{ij}(\mathbf{r})$,

$$E_x = -2\sum_{i,j}\int \tilde{\varphi}_i(\mathbf{r})\tilde{\varphi}_j(\mathbf{r})v_{ij}(\mathbf{r})d\mathbf{r} \tag{3.27}$$

3.2 Numerical implementation of the TS functional

In order to add the vdW correction self-consistently into AIMD methodology, the corresponding forces on wavefunctions and ions from vdW energy contribution need to be included into CP equation(B.3). Therefore the majority of vdW implementation is focused on the forces derivation, which is the content of the following. Fig.[3.2] illustrates basically how vdW correction is implemented.

The TS-vdW energy is given as a sum over unique pair energies between atoms A and B:

$$E_{vdW} = -\frac{1}{2}\sum_{A,B} E_{vdW}^{AB} = -\frac{1}{2}\sum_{A,B} f_{damp}^{AB} C_{6,AB}^{eff} R_{AB}^{-6} \tag{3.28}$$

in which f_{damp}^{AB} corresponds to a given vdW pair energy,

$$f_{damp}^{AB} = \frac{1}{1+\exp[-d(\frac{R_{AB}}{s_R R_{AB}^{0,eff}} - 1)]} \tag{3.29}$$

in an explicit function of the internuclear separation, R_{AB}, between atoms A and B,

$$R_{AB} = [(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2]^{1/2} \tag{3.30}$$

and the sum of their respective effective vdW radii, $R_{AB}^{0,eff}$,

$$R_{AB}^{0,eff} = R_A^{0,eff} + R_B^{0,eff} \tag{3.31}$$

In this expression for f_{damp}^{AB}, both d and s_R are empirical parameters.

The effective C_6 coefficient corresponding to atoms A and B, $C_{6,AB}^{eff}$, is given by the following combination rule:

$$C_{6,AB}^{eff} = \frac{2C_{6,AA}^{eff} C_{6,BB}^{eff}}{\left(\frac{\alpha_B^{0,eff}}{\alpha_A^{0,eff}}\right) C_{6,AA}^{eff} + \left(\frac{\alpha_A^{0,eff}}{\alpha_B^{0,eff}}\right) C_{6,BB}^{eff}} \tag{3.32}$$

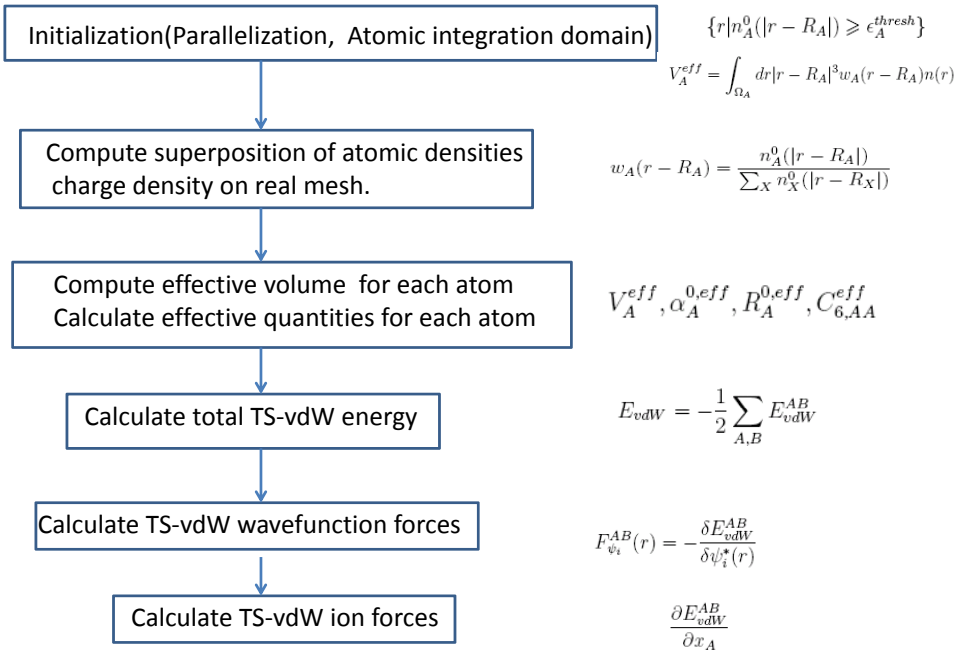

Figure 3.2: Flowchart of numerical implementation of vdW-TS scheme.

in terms of the effective homonuclear C_6 coefficients, $C_{6,AA}^{eff}$ and $C_{6,BB}^{eff}$, and the effective static dipole polarizability, $\alpha_A^{0,eff}$ and $\alpha_B^{0,eff}$.

Plug in $\alpha_A^{0,eff}$ and $\alpha_B^{0,eff}$ as a function of effective volume, we have

$$C_{6,AB}^{eff} = \left(\frac{V_A^{eff}}{V_A^{free}}\right)\left(\frac{V_B^{eff}}{V_B^{free}}\right) C_{6,AB}^{free} \tag{3.33}$$

3.2.1 Forces on wavefunctions

To derive the dispersion forces acting on a given electronic wavefunction, we need to consider the functional derivative of the TS-vdW pair energy with

respect to its complex conjugate, *i.e.*,

$$F_{\psi_i}^{AB}(r) = -\frac{\delta E_{vdW}^{AB}}{\delta \psi_i^*(r)} \tag{3.34}$$

which can be re-expressed as a functional derivative with respect to the charge density via:

$$\begin{aligned} F_{\psi_i}^{AB}(r) &= -\frac{\delta E_{vdW}^{AB}}{\delta \psi_i^*(r)} \\ &= -\int dr' \frac{\delta E_{vdW}^{AB}}{\delta n(r')} \frac{\delta n(r')}{\delta \psi_i^*(r)} \\ &= -\int dr' \frac{\delta E_{vdW}^{AB}}{\delta n(r')} \psi_i(r') \delta(r - r') \\ &= -\frac{\delta E_{vdW}^{AB}}{\delta n(r)} \psi_i(r) \end{aligned} \tag{3.35}$$

One can now define $U_{vdW}^{AB}(r) \equiv \frac{\delta E_{vdW}^{AB}}{\delta n(r)}$ as the dispersion potential arising from a given TS-vdW pair energy acting on each electronic wavefunction, yielding:

$$F_{\psi_i}^{AB}(r) = -U_{vdW}^{AB}(r)\psi_i(r) \tag{3.36}$$

The detailed derivative of vdW energy with respect to electron density is discussed in Ref.[96]. Fig.[3.3] illustrates how the derivation is generally performed.

3.2.2 Forces on ions

To derive the nuclear gradient of the vdW pair energy expression above, we will need to consider nuclear displacements of atoms A and B(which are explicitly included in the pair energy) and of a given atom K(which is not included in the pair energy)

Recalling that the vdW pair energy for atoms A and B is given by

$$E_{vdW}^{AB} = f_{damp}^{AB} C_{6,AB}^{eff} R_{AB}^{-6} \tag{3.37}$$

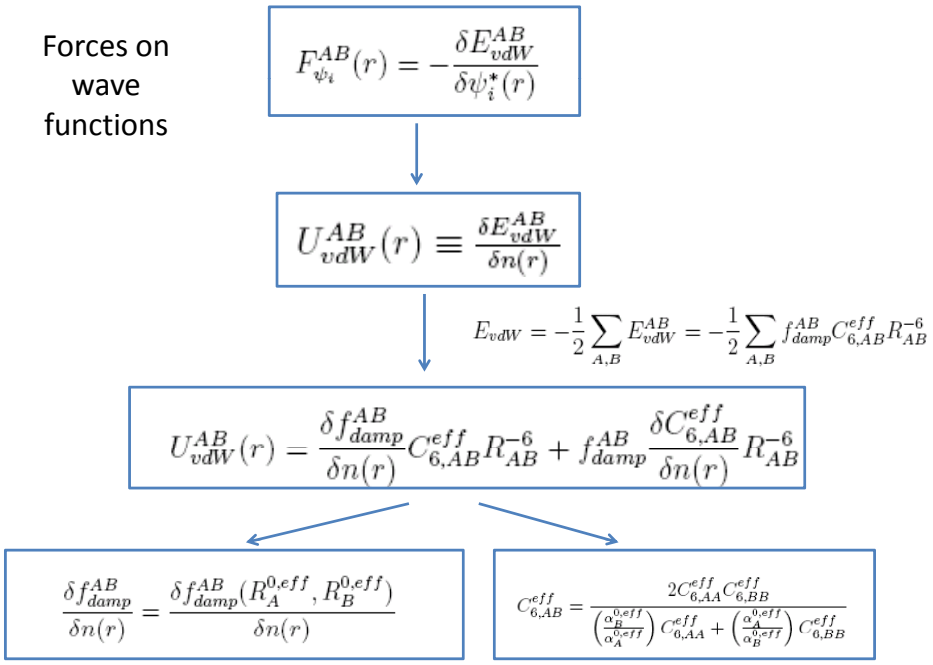

Figure 3.3: Flowchart of forces on wavefunctions in the numerical implementation of vdW-TS scheme.

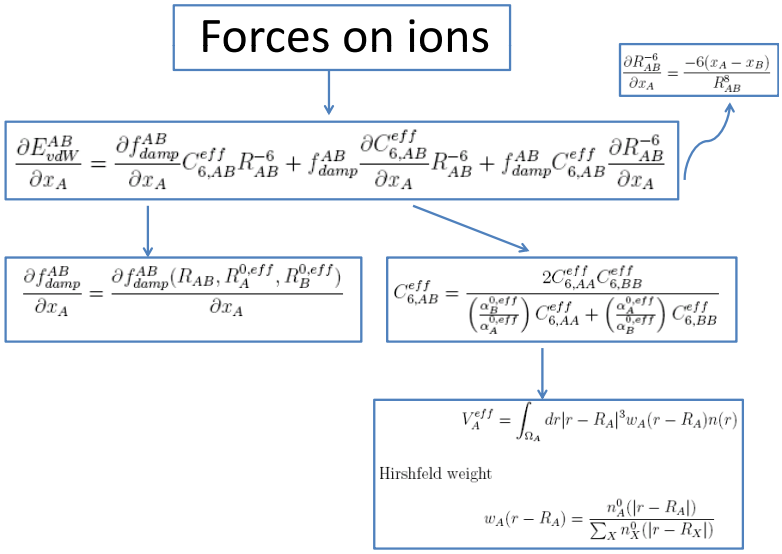

Figure 3.4: Flowchart of forces on ions in the numerical implementation of vdW-TS scheme.

Let's consider the derivative of E_{vdW}^{AB} with respect to a nuclear displacement of atom A along the x-direction:

$$\frac{\partial E_{vdW}^{AB}}{\partial x_A} = \frac{\partial f_{damp}^{AB}}{\partial x_A} C_{6,AB}^{eff} R_{AB}^{-6} + f_{damp}^{AB} \frac{\partial C_{6,AB}^{eff}}{\partial x_A} R_{AB}^{-6} + f_{damp}^{AB} C_{6,AB}^{eff} \frac{\partial R_{AB}^{-6}}{\partial x_A} \quad (3.38)$$

The derivative of vdW energy with respect to ions position is straightforward, though complicated. The detailed derivation is discussed in Ref.[96]. Fig.[3.4] is the flowchart showing the general derivation procedure.

3.3 Parallelization issues

3.3.1 Exact exchange calculation

The exact exchange calculation requires massive orbital information due to its non-local characteristics, as seen from Eq.(3.18). In practice, there is a tradeoff between improving computing efficiency and memory allocation when implementing the exact exchange calculation in PBE0 functional. For instance, in 64 water molecule system with cell size corresponding to experimental liquid water equilibrium density at ambient condition, the FFT mesh size is 128×128×128 and therefore the size of each orbital is roughly 16 megabytes(MB). If each processor is to hold 256 orbitals, that will be 4 gigabytes(GB). In reality, massively parallel computers that we are working on for simulation is only 2GB per core.

In order to solve this memory bottleneck, we utilize the excellent interprocessor communication among massively parallel computers. The National Energy Research Scientific Computing Center (NERSC)[97] provides amazing simulation environment where thousands of computing facilities are available, without which our calculation would not have been possible. Under this powerful computing tool, we can distribute the orbitals information among different processors rather than one processor holding all orbitals. For simplicity, each processor stores only one orbital information if we are taking 256 processors to deal with 64 water system simulation (which has 256 orbitals totally). Those orbitals needed for exact exchange calculation with a given orbital can be duplicated and transferred to corresponding processors at each MD time step. In liquid water, only 15 to 20 orbitals rather than 256 orbitals[98] are required for each orbital's pair exchange calculation. Correspondingly, the memory to hold these orbitals becomes extremely smaller

compared to that to hold the whole orbitals for each processor. The remaining exact exchange contributions due to other distant MLWFs are negligible due to the tiny overlap among MLWFs, as can be schematically seen from Fig.[3.1].

There is a huge advantage of calculating exact change in MLWF representations, as MLWFs are exponentially localized. The exchange interaction between two orbitals is restricted to the spatial region of orbital overlap, the amplitude of the exchange interaction between two MLWFs decays rapidly with the distance between their centers. Thus, typically each Wannier orbital exchanges with only a finite number of neighboring orbitals and the number of pair interactions per orbital is independent of system size. As a result, our procedure to compute exact exchange is order N, *i.e.*, its computational cost scales linearly with system size. Due to this, the scheme works also well in a larger system like 128 water molecules in the supercell with only double computation cost.

In addition, there are other useful parallelization strategies implemented to accelerate simulation. For instance, a so-called Task Group hierarchical parallelization strategy for three dimensional FFTs is also adopted[99]. The basic algorithm is shown in Fig.[3.5]. This strategy tries to utilize more processors when performing FFT or inverse FFT calculations to save computation time. In this case, not only the regular Z components of FFT for each electronic state is parallelized, different electronic states that constitute the total charge density are also parallelized, due to which the scalability performs well even when more processors are in use.

The calculation associated with exact exchange also implemented this strategy in order to be consistent with the code package expandability. This im-

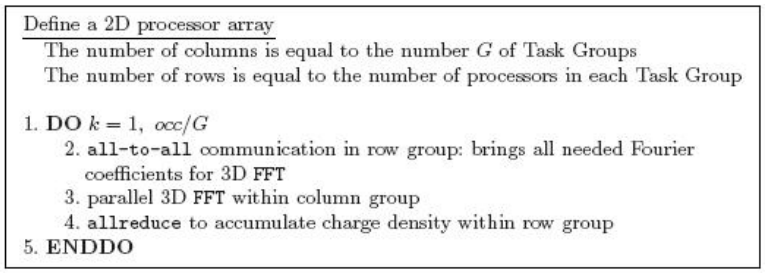

Figure 3.5: The Task Group 3D FFT scheme for the calculation of charge density $\rho(r)$. The G in the figure denotes the number of task group.

proves additional efficiency along with other parallelization strategies. Within these schemes, the calculation cost of exact exchange is about ∼30 seconds per AIMD step with 256 processors. Additional parallelization in which two processors handle one Wannier orbital pair exchange calculation almost double the efficiency and leads to ∼ 15 seconds per step with 512 processors. This has enabled us to perform a longer AIMD simulation (e.g. 20ps) within several weeks, an excellent achievement in solving this complex problem.

3.3.2 vdW calculation

The parallelization of vdW-TS scheme is pretty more straightforward than the complex exact exchange calculation explained above, due to two main advantages. First, the required memory allocation related with vdW becomes smaller as the C_6 coefficient and damping function depend on ground-state electron density instead of expensive wavefunction information. Therefore each processor is able to hold the superposition of charge density and molecular charge density simultaneously without memory relocation problem. Second, an efficient parallelization scheme is to distribute the calculation task along all the atoms in the system to different processors (i.e. processor of

index i computes the contribution of vdW from atom i), so that the calculation of vdW energy, forces on ions and wavefunctions attributed to each atom can be assigned to distributed processors. This calculation of each processor is independent with each other and does not require complex and expensive communication among processors all the time. The total energies and forces can then be gathered only *once* after the corresponding individual calculation on each processor is successfully performed.

Moreover, due to the linear scaling calculation in this efficient algorithm, one can easily perform a larger supercell simulation with only a reasonable increase of computation time. This enables us to choose 128 water molecule system instead of most-widely adopted 64 water systems in the later chapter in order to better simulate liquid water.

4 Comparison to experiment and analysis of short and intermediate range order in liquid water

In this chapter, first we investigate how well AIMD water structure compares to experiment when PBE0, vdW and quantum nuclei effects are included. As quantum nuclei are mimicked via a 30K increase in temperature based on the work by Morrone *et al.*[10], only the oxygen-oxygen RDF and the corresponding structure factor are considered. The agreement between theory and experiment is found to be quite good, indicating that we have obtained a realistic liquid water structure. Therefore, we are motivated to study how these various interactions and dynamic effects influence water structure. It is found that PBE0 and quantum nuclei are vital in first coordination shell because they control the fraction of broken H-bond. vdW on the other hand, plays a more important role in the second coordination shell by promoting penetration of interstitial molecules into first coordination shell. The analysis is greatly facilitated by using the so-called local structure index, an order parameter that allows to identify low- and high-density liquid environments. We find that both kinds of local environments coexist in the liquid at ambient pressure and density, consistent with the so-called two state model of water[100].

4.1 Comparison with experiment

4.1.1 PBE0+TS simulation at T=330K with 128 water molecules

A full PI-AIMD simulation of water at the PBE0+TS level would be extremely time consuming due to the need to simulate a large number (32 in Ref.[10]) of system replicas in a PI simulation. Some recently developed techniques based on colored noise thermostats could be used to accelerate PI simulations. The computational cost of this approach would still be heavy in spite of a substantially reduced number of replicas needed to converge the Feynman paths (only 4-8 replicas were found to be necessary to model ice close to melting with the colored noise approach in Ref.[101]). Instead of fully including nuclear quantum effects we use here a temperature 30K higher than experiment and focus on the O pair correlations only, as already discussed several times. In order to better compare the theory to the experimental structure factor we have performed an AIMD simulation on a sample of 128 water molecules in a cubic box with volume fixed at the experimental density at ambient temperature and pressure. A 128 molecule cell is necessary for a decent resolution in the structure factor in the q-range of the first diffraction peak. This is a rather expensive calculation: it took about 3 months to obtain an equilibrated trajectory lasting 10 ps.

4.1.2 Some comments on the effects of quantum nuclei

The OO RDF in Fig.[2.7], which reported both classical (AIMD) and quantum (PI-AIMD) data, showed that the effect of quantum fluctuations on the OO RDF could be mimicked with good accuracy by a classical simulation at an artificially higher temperature (e.g. 330K instead of 300K). This observa-

tion leads some support to the practice, often adopted, of performing AIMD simulations of water at higher effective temperature than the reference equilibrium temperature. This practice is however just an empirical cure that can hardly be justified on general grounds. One sees, for example, in Fig.[2.7] that a temperature of 330K, which is suitable for OO correlations, is not so for OH correlations. The inadequacy of a uniform temperature increase to properly account for the effects of quantum fluctuations is made dramatically evident by the momentum distribution of the protons. This quantity can be extracted from deep inelastic neutron scattering experiments[8, 9] and can be accurately calculated with PI-AIMD simulations[10]. The proton momentum distribution is very different from the classical equilibrium distribution: for example in ice Ih at T=269K, i.e. close to melting, the classical distribution that most closely reproduce the PI-CPMD data has a temperature of 869K[105]. Thus the temperature used to mimic quantum effects should be different for each physical property, suggesting that different effective temperatures should be assigned to particles with different masses. The second observation is that a Boltzmann distribution with a higher effective temperature does not capture important details of the actual distribution. In ice proton dynamics is quasi-harmonic and the momentum distribution is well represented by a multivariate Gaussian distribution, with the three principle frequencies associated respectively to stretching, and combinations of bending and libration[105]. Similar effects occur in liquid water[10]. These effects can be characterized as large zero point energy. They are likely to play an important structural role in all H-bonded systems but are typically neglected in MD simulations, both empirical and *ab initio*.

4.1.3 Structure factor comparison with experiment

X-ray diffraction experiment is the most direct way to probe the structure of the oxygen site network in liquid water experimentally, as x-rays are sensitive to electrons and therefore provides major information about OO structure distribution. However, neutrons are insensitive to OO correlations since hydrogen scatters neutrons more strongly than oxygen, which on the other way, makes neutron diffraction the best source of information on the oxygen-hydrogen and hydrogen-hydrogen correlations.

The x-ray diffraction intensity measured in experiments, is given by the product of the form factor $F^2(q)$ (see sect.2.3.2) and the structure factor $S(q)$ (see sect.2.3.1). In experiment, one relies on the form factor of water molecule in gas phase to approximate that in condense phase, as the latter is more complicated to compute in liquid water. In our pseudopotential calculation, we found that the form factor in gas phase and condense phase is quite close to each other, with only at most $\sim 2\%$ difference. Though the form factor is different between pseudopotential and all electron calculation, the difference from gas to liquid phase should be represented accurately by the difference between the corresponding pseudocharge densities. Therefore it is reasonable to use the water molecule form factor in gas phase in dealing with structure factor from x-ray scattering intensities.

Thus it is preferable to compare theoretical and experimental data for the structure factor $S(q)$, as it is more directly obtained from experiment and does not require a Fourier inversion. *Ab initio* simulation of liquid water can also calculate the structure factor $S(q)$ directly. Fig.[4.1] compares the calculated structure factor $S(q)$ using the PBE0+TS functional and two $S(q)$ from x-ray diffraction. The calculated $S(q)$ is discrete and therefore there is

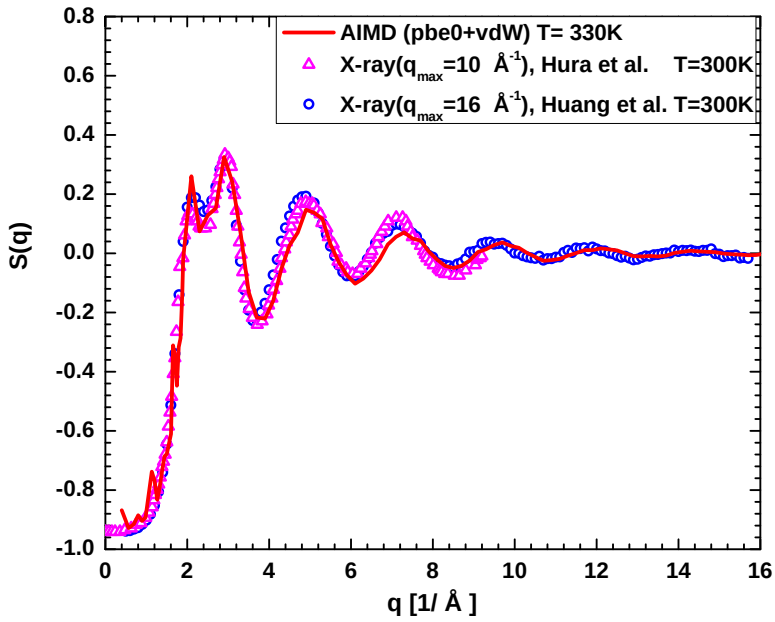

Figure 4.1: *Ab initio* calculated structure factor $S(q)$ with PBE0+TS at 330K compared to two experimental data, with Hura from Ref.[43] and Huang from Ref.[102].

some discontinuity particularly at small q. The experimental $S(q)$ shown here gives very consistent results in the two reported experiments. The calculated $S(q)$ is quite close to the two experimental $S(q)$ throughout the entire q range. Only some tiny right shift of the peak positions is observed, consistent with the OO RDF plot which is slightly left shifted compared to the experimental OO RDFs and this will be shown shortly.

4.1.4 OO RDF comparison with experiment

RDF is widely used in investigating liquid water structure, as it shows plenty of structure information in real space. *Ab initio* simulation can calculate the RDF directly as trajectories of all nuclei are recorded. X-ray/neutron diffraction experiments measures the scattering intensity in reciprocal space and can only calculate structure factor directly. To extract RDF, a transformation technique from reciprocal space to real space is required, which is not easy and can lead to uncertainties as to be discussed below.

A comparison between theory and experiment in real space is provided in Fig.[4.2], which shows the OO RDFs from two x-ray diffraction experiments, one joint x-ray/neutron diffraction experiment, and the one from our simulation. The agreement between theory and experiment is pretty good in the range from the first minimum of the RDF up to half the size of the supercell ($\sim$ 7.8 Å). In this range the three experiments agree very well with each other. It is more difficult to quantify the accuracy of our simulation in the region of the first peak, due to the large uncertainty in the experimental RDF extracted from two different experiments in that region. The two x-ray experiments had different resolutions (q_{max} = 10 Å^{-1} and q_{max} = 16 Å^{-1}, respectively). Recently, new high q-data with q_{max} = 23 Å^{-1} together with elimination of the OH contribution to the OO RDFs gives a first peak height of 2.58 at OO distance of 2.80Å[64], indicated by the brown star in Fig.[4.2]. This new x-ray experimental result is very close to the simulated peak height of 2.57 at a distance of 2.76 Å, providing strong support to the accuracy of our adopted functional approximation. We have shown in the Fig.[4.1] that the discrepancy between the two x-ray experiments in Fig.[4.2] is absent from the corresponding structure factors, indicating that the difference between

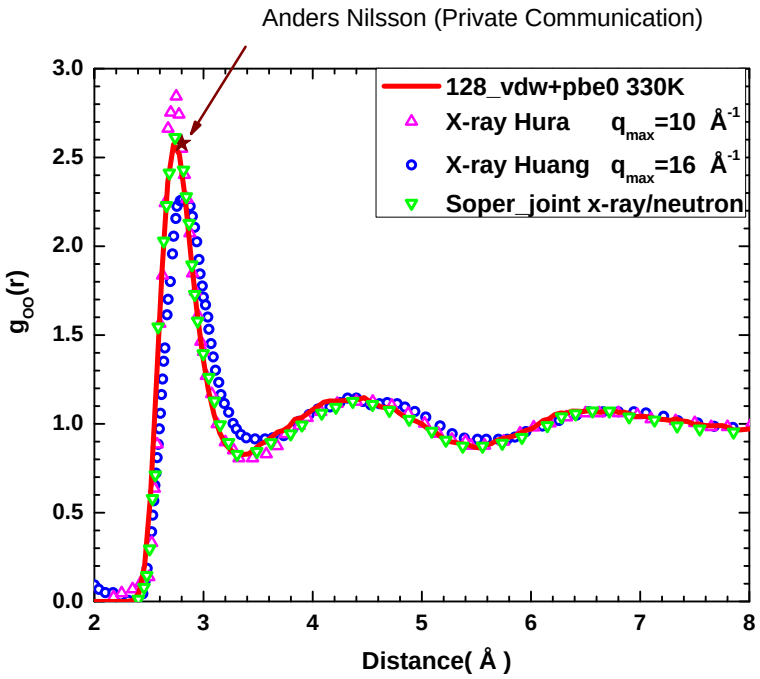

Figure 4.2: Calculated OO RDF with PBE0+TS at 330K from CPMD simulation, compared to two RDFs derived from x-ray diffraction experiment (x-ray Hura from Ref.[43], x-ray Huang from Ref.[102].) and one from joint x-ray/neutron diffraction experiment from Ref.[103].

the two x-ray experimental RDF's in the first peak of Fig.[4.2] should be associated to the procedure adopted to transform into real space data set extending only up to q_{max} in reciprocal space. The experimental OO RDF from joint x-ray/neutron diffraction experiment on the other hand agrees very well with the simulation also in the first peak region, in which a different RDF extraction technique call empirical potential structure refinement (EPSR)[104] is used and would be mentioned shortly.

4.1.5 Extracting the RDF from experiment

We have seen that two experimental derived OO RDFs show large uncertainty in the first peak region, but the corresponding structure factors have less difference (see Fig.[4.1] and Fig.[4.2]). Thus much of the experimental uncertainty derives from the transformation of the structure factor, the knowledge of which is limited by the experimental resolution. For instance, one experimental derived OO RDF from Ref.[102] in the Fig.[4.2] is obtained from structure factor $S(q)$ through the following relationship:

$$g(r) = 1 + \frac{1}{2\pi^2 r\rho_0}\int_0^{q_{max}} e^{-\alpha q^2} qS(q)\sin(qr)dq \tag{4.1}$$

where ρ_0 is the average molecular density of water. q_{max} is the maximal momentum transfer that can be achieved in experiment. The window function $e^{-\alpha q^2}$ is adopted to decrease the magnitude of spurious ripples resulting from the truncation errors associated with the Fourier transform[102].

This Fourier inversion in Eq.[4.1] is limited by q_{max}, which should be large enough to obtain a converged result. For instance in the experiment of Ref.[102], q_{max} was equal to $\sim$ 16 Å^{-1}. Though this is sufficient to capture the main information of the structure factor, an accurate determination of its first peak of the RDF and a complete elimination of spurious ripples would require higher q_{max}.

Fig.[4.3] plots the OO RDF extracted from $S(q)$ according to Eq.[4.1] without the usage of window function for smoothing, i.e. setting $\alpha = 0$. We observe that spurious ripples exist throughout the r range. Moreover, it is found that as q_{max} increases, the first peak region of OO RDF, usually an indicator of liquid water structuring, shows some small but non-negligible shift. As q_{max} increases from 9, to 12 and 16 Å^{-1}, the first peak position decreases from

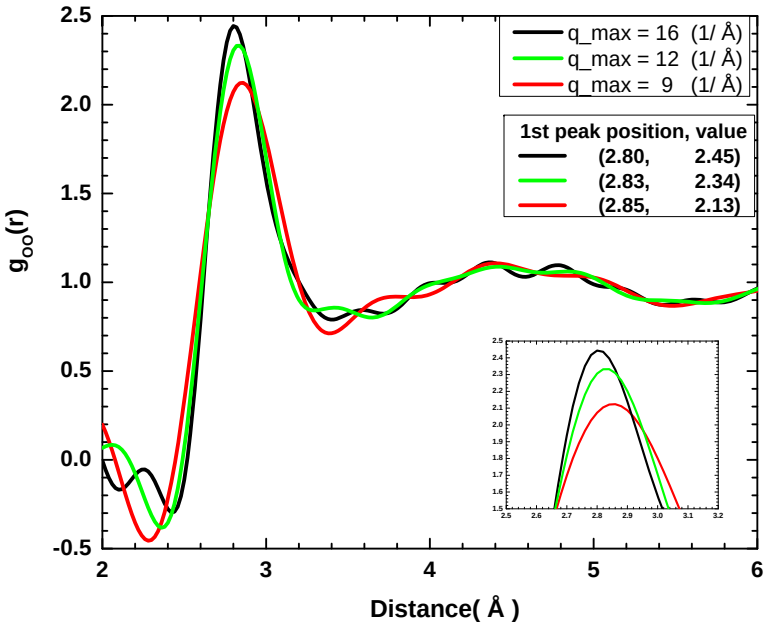

Figure 4.3: Calculated OO RDFs via Fourier transformation of structure factor $S(q)$ with several q_{max} cutoffs. The inset displays the zoom-in region of OO first peak, indicating a shift of both first peak position and value depending upon q_{max}.

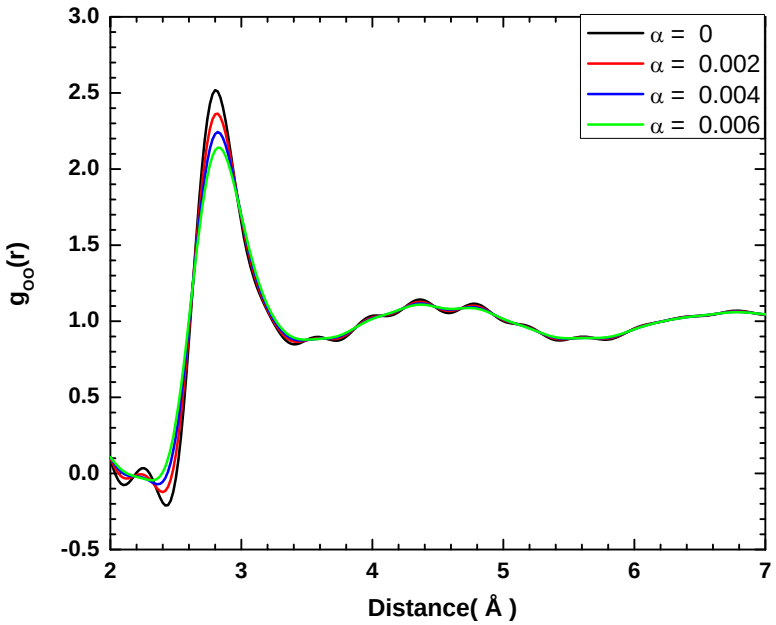

Figure 4.4: Obtained OO RDFs via Fourier transformation of structure factor $S(q)$ with several values of α in the window function.

2.85 to 2.83 and 2.80, while the peak intensity increases from 2.13 to 2.34 and 2.45 Å, respectively. With this trend in mind, we may expect that the converged peak value should be larger than 2.45 at a position smaller than 2.80 Å when more accurate high q experimental data would be available. This is indeed confirmed by a very recent experiment with $q_{max} = 23$ Å^{-1}, which reports a peak value of 2.58 with first peak position at 2.80 Å[64].

In practice, in order to eliminate spurious ripples due to truncation error, an exponentially decaying window function $e^{-\alpha q^2}$ is adopted in Eq.[4.1]. The Fig.[4.4] plots the OO RDF curve through Fourier transform of $S(q)$ with window function. It is observed that relatively strong spurious oscillations are present when the window function is not adopted (black line with $\alpha = 0$), which is the case in the early Narten's approach[38]. The spurious oscillations are gradually eliminated with increasing α. However, the first peak value in OO RDF is observed to be quite sensitive to small α variation. Therefore though window function can smoothen the RDFs, it leads the first peak OO RDF unreliable. This technique is adopted in Ref.[102] with the blue curve seen in Fig.[4.2] and therefore can explain the deviation of this experimental derived OO RDF from simulation.

Another experimental derived OO RDF from Ref.[43] in the Fig.[4.2] adopts another transformation technique. Instead of extracting the RDF from the structure factor $S(q)$ via a direct Fourier transformation, Hura, *et al.*[43], started in real space to construct a linear combination of RDFs chosen from various theoretical and experimental curves, as shown below

$$g_{OO}(r) = \sum_{i=l}^{N} a_i g_{OO}^{i}(r) \tag{4.2}$$

where $g_{OO}^{i}(r)$ are the basis functions from simulations of various potentials[106, 107] and various experimental curves[38, 108, 109]. The a_i are calculated

according to a linear least-squares optimization scheme performed on the entirely q-space, using linear superpositions of the corresponding $S^i_{OO}(q)$ (a modified atomic form factor is used in their approach to obtain structure factor from scattering intensity)[43]. The advantage of this fitting is that, it does not show any nonphysical features, because the basis RDFs do not show any nonphysical features for small distance r. The limit of this approach is that it does not rely entirely on experiment and/or data of high accuracy (such as the calculated form factor). Therefore $g(r)$ is biased by theoretical empirical potential simulations that lack direct independent experimental validation. This technique is adopted in Ref.[43] with the pink curve seen in Fig.[4.2] and therefore might explain the deviation of this experimental derived OO RDF from simulation.

Another approach widely used for exploration of the geometries of the intermolecular structures consistent with the measured radial distribution functions is the empirical potential structure refinement (EPSR) procedure[104]. This process performs a Monte Carlo computer simulation of the system under study to generate ensembles of water molecules whose structure is consistent with measured diffraction data. The water molecule ensembles so obtained are used to calculate structural quantities such as site-site radial distribution functions and spatial density functions. All of the tests that have been done to date using this procedure indicate strongly that forcing the simulated molecular ensembles to reproduce the measured radial distribution functions is a substantial constraint on the three-body and high-order correlation functions, and will almost certainly capture the essential topology of the local order. One example, where the robustness of the structures determined can be seen, is a comparison of the structures determined for a dilute alcohol water solution: two independent and differently weighted sets

of isotopic substitution data on the same chemical system were both analyzed by the EPSR procedure. The results are found to display the same local order. This technique is adopted in Ref.[103] to derive OO RDF with the green curve seen in Fig.[4.2]. The experiment and theory agree very well with each other, suggesting the advantage of this procedure and more importantly: a realistic water structure is obtained from *ab initio* simulation.

4.2 Effect of the functional approximation and quantum nuclear dynamics on the local order of liquid water

From the *ab initio* perspective it is important to understand how the various effects discussed above correlate with the molecular forces and dynamics.

4.2.1 Mean square displacement

For that purpose we consider the diffusion coefficient, which can be estimated from the mean square displacement(MSD) through the Einstein relation:

$$6D = lim_{t\to\infty}\frac{d}{dt}\langle|\mathbf{r}_i(t) - \mathbf{r}_i(0)|^2\rangle \tag{4.3}$$

where "D" is the diffusion coefficient, $\mathbf{r}_i(t)$ is the ith atom's position at time t. The MSD of the oxygen atoms was determined over the trajectories. The dynamic properties of liquid water in the simulation provide us useful information related with liquid water structuring. The MSD allows us to monitor the fluidity of a liquid. Our estimated D is affected by large uncertainty, as our trajectories are too short (typically ∼100 ps would be necessary for an accurate determination of D). Moreover in our simulation we use massive Nose-Hoover chain thermostats, which likely affects the dynamics. However, the diffusion coefficient extracted from the simulation still provides valuable information: an over-structured RDF would lead to a smaller diffusion coefficient, due to the overestimated H-bond strength.

Fig.[4.5] shows the MSD of liquid water with different exchange-correlation approximations. PBE0 increases fluidity over PBE as the H-bond strength is reduced in PBE0 due to a reduction of self-interaction error. vdW interactions also increase diffusion as they also tend to lead to more broken H-bonds.

Obviously a temperature increase of 30K increases diffusion compared to a simulation at 300K.

The diffusion coefficient that we obtain from the simulation on a 128 molecule cell with PBE0+TS at 330 K is $1.4 \times 10^{-5}\mathrm{cm}^2/\mathrm{s}$. As has been discussed in Ref.[6], there is size dependence of the translational self-diffusion D which arises from the fact that a diffusing particle sets up a hydrodynamic flow that decays as slowly as $1/r$. In a periodically repeated system this leads to an interference between a diffusing particle and its periodic images. This effect has been analyzed by Dunweg and Kremer[110, 111], who have established the following relation for the diffusion coefficient under periodic boundary conditions(PBC) as a function of simulation box length L.

$$D_\infty = D_{\mathrm{PBC}}(L) + \frac{k_B T \zeta}{6\pi\eta L} \tag{4.4}$$

where D_∞ is the diffusion coefficient for an infinite system, $\zeta = 2.837$ a numerical coefficient similar to the Madelung constants which results from an infinite summation over all replicas, and η the translational shear viscosity. If in Eq.[4.4], we use our calculated D_{PBC} and take the experimental value of $\eta = 8.92 \times 10^{-4}\mathrm{Pa} \cdot \mathrm{s}$, we extrapolate a value of $2.0 \times 10^{-5}\mathrm{cm}^2/\mathrm{s}$ for D_∞, which is quite close to the experimental value of $2.2 \times 10^{-5}\mathrm{cm}^2/\mathrm{s}$ at room temperature. Clearly one should not weight too strongly this comparison, in view of the short simulation trajectory, the adoption of massive thermostats for thermal equilibration and the crude approach used to mimic nuclear quantum effects. It is interesting, however, that the same accurate functional that gives a very good water structure, also gives diffusion properties close to experiment.

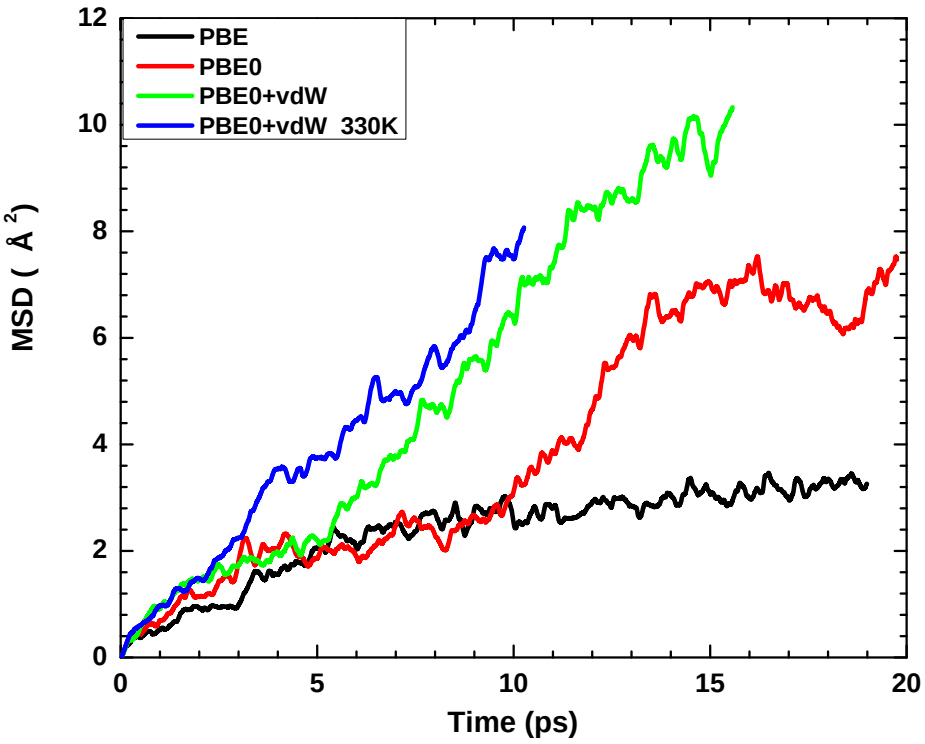

Figure 4.5: The MSD evolution of simulation time, with PBE(black), PBE0(red), PBE0+TS at 300K(green), PBE0+TS at 330K(blue). Note that the first three effects are all performed at 300K. In what follows, this holds the same meaning in the remaining plots.

4.2.2 Oxygen-oxygen radial distribution function

It is of interest to analyze how the various physical effects contribute to the local order in water. Fig.[4.6] shows the OO RDF with PBE, PBE0, PBE0+TS and PBE0+TS at 330K. It is found that PBE0, vdW and the temperature raise mimicking quantum effects all lead to a softened structure. Quantitatively, these effects have more or less the same contribution in the first peak region. On the other hand in the first minimum and second peak region, the vdW contribution is larger indicating that the primary role of vdW interactions should be associated to the penetration of more molecules from the second coordination shell to the first shell. This behavior becomes more clear when the RDF is decomposed into bonded and non-bonded subsets in HDL- and LDL-like sites, as will be discussed shortly.

The observed less structured RDF from PBE to PBE0 is qualitatively consistent with Ref.[12]. This result is also consistent with recently reported RDF results with the PBE0 functional by Zhang *et al.*[73]. The better accuracy of PBE0 over PBE in the description of the H-bond in water is well documented by detailed studies of the water dimer in Appendix D. Adding vdW interactions to PBE0 further improves the description of liquid water, indicating that both H-bonding and vdW interactions are crucial in liquid water. Quantum nuclei effects, mimicked here by a temperature raise, further soften the OO RDF. In absence of a full quantum mechanical treatment by PI-AIMD, we limit our discussion here to the OO RDF, a complete study of OO, OH and HH correlations will have to wait for a full PI-AIMD simulation based on PBE0+TS.

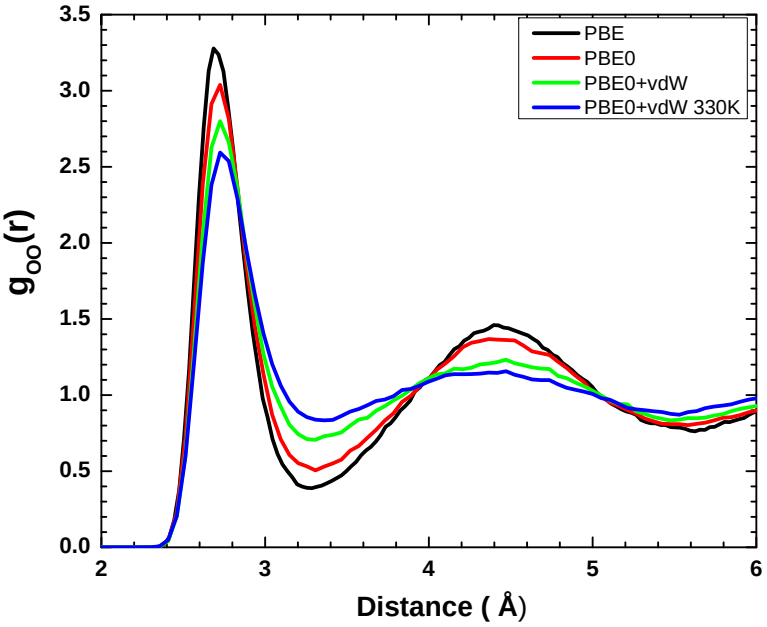

Figure 4.6: Calculated OO RDFs with PBE, PBE0, PBE0+TS at 300K, PBE0+TS at 330K.

4.2.3 Triplet correlation function

To better quantify the local order, higher order correlations than two-body ones would be important. Unfortunately they are not directly accessible from diffraction experiment. A very important three-body correlation function is the triplet distribution $p(\theta_{OOO})$. This quantity has been extracted from experiments recently[103] .

In this analysis, three water molecules are regarded as a triplet if two of them lie within 3.18 Å of the third: this is the distance at which OO coordination number reaches $\sim$ 4.0. The angle calculated is the included angle made by these two molecules with the third to which they are "bonded". Fig[4.7] plots the oxygen-oxygen-oxygen(OOO) three body triplet distribution with PBE, PBE0, PBE0+TS at 300K, PBE0+TS at 330K. This distribution shows a peak value with position close to the perfect tetrahedral angle of 109.47^o. The tetrahedrality is more pronounced in PBE and decreases when PBE0, vdW and approximate quantum nuclear are included.

In order to quantify the tetrahedrality of the liquid[112, 113], one uses a local order parameter "q", which is defined as:

$$\langle q \rangle = 1 - \langle (cos(\theta_{OOO}) + \frac{1}{3})^2 \rangle \tag{4.5}$$

the average is taken over all appropriate triplets in the liquids.

Fig[4.8] illustrates how tetrahedrality of liquid water evolves with different functionals. $\langle q \rangle$ systematically decreases when PBE0, vdW and nuclear quantum effect are included. The $\langle q \rangle$ is 0.65 when all these corrections are considered, which is quite close to experimental derived value of 0.58 for light water.

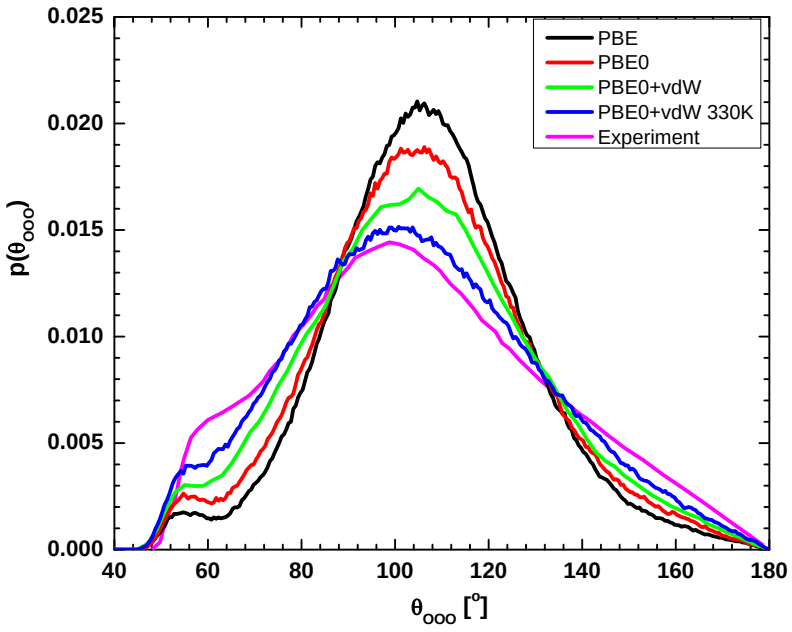

Figure 4.7: The distribution of triplet angles for triplets of water molecules in the CPMD simulations (black for PBE, red for PBE0, green for PBE0+TS, blue for PBE0+TS at 330K.)

PBE	**0.79**
PBE0	**0.75**
PBE0+vdW 300K	**0.70**
PBE0+vdW 330K	**0.65**
Experiment	**0.58**

Figure 4.8: Tetrahedrality of liquid water $\langle q \rangle$.

4.2.4 Broken H-bond statistics

Further important information comes from the broken H-bond statistics. We here adopt the geometrical definition of H-bond suggested in Ref.[10], i.e. a H-bond is defined in geometric terms by oxygen-to-oxygen and oxygen-to-hydrogen distance cutoffs that are equal to the minima of the H-bonding peaks of the RDFs, and a H-bond angle greater than 140^o. The fraction of broken hydrogen, as seen in Fig.[4.9], increases approximately from 7.5% with PBE to 10.5% with PBE0, indicating increased fluidity. This behavior is consistent with corresponding trend in the RDFs. The fraction of broken H-bond increases up to 14% when the vdW correction is added to PBE0. The fraction of final broken H-bond is about 18% when nuclear quantum effects are also approximately taken into account. This fraction of broken H-bond is consistent with most estimates and experimental data. It is definitely inconsistent with the proposal that $\sim$ 80 percent of the H-bonds are

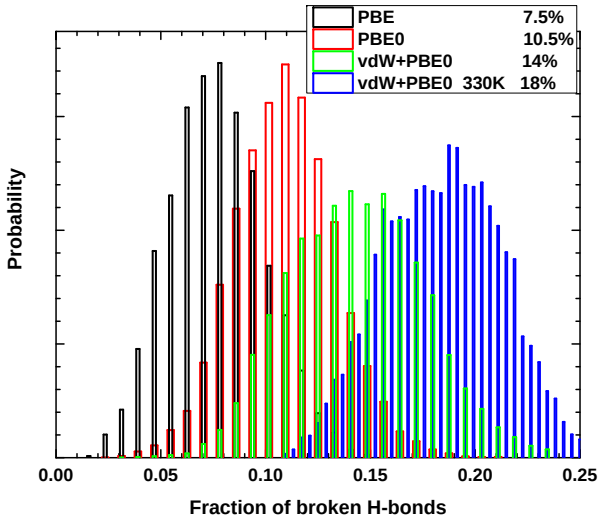

Figure 4.9: The distribution of the fraction of broken H-bonds in the CPMD simulations (black for PBE, red for PBE0, green for PBE0+TS, blue for PBE0+TS at 330K.)

broken in a liquid made by chains and rings of molecules. This conclusion is independent of the criteria adopted for the H-bond definition.

4.3 Local structure index analysis: high- and low-density sites

Some anomalies in water have been tentatively associated with the presence of two competing preferential local structures associated to local environments of high and low density [100], in a picture inspired by the existence of at least two different forms of amorphous glass states, namely low-density amorphous ice (LDA) and high-density amorphous ice (HDA)[114].

In liquid water, high density liquid (HDL)-like sites are characterized by the presence of interstitial molecules between the first and second coordination shells, leading to perturbed H-bonding and higher density. Low density liquid (LDL)-like sites exhibit more tetrahedral-like local order with well-separated first and second shells, and thus lower density close to that of hexagonal ice. The structures of HDL and LDL have been characterized experimentally in undercooled water by A.Soper *et al.*[115]. The water structure was measured as a function of pressure using neutron diffraction at a temperature of 268 K. The measured structure functions imply a continuous transformation with increasing pressure from a LDL with an open, H-bonded tetrahedral structure, to a HDL with a collapsed second coordination shell,

To identify these "structured" and "unstructured" molecules, a more elaborate parameter called the Shiratani-Sasai local-structure-index (LSI) has been used[116, 117]. This LSI has been successfully applied to the inherent structure in force field simulations of supercooled and ambient water and reveals a structural bimodality in the underlying potential energy surface on which the thermal motion evolves [118, 119, 120].

The Shiratani-Sasai LSI distinguishes molecules with well separated first and

second coordination shells from molecules with a more disordered environment that contains neighboring molecules in interstitial positions. The index $I(i,t)$ is defined for molecule i at time t. For each molecule i one orders the rest of the molecules depending on the radial distance r_j between the oxygen of the molecule i and the oxygen of molecule $j : r_1 < r_2 < r_j < r_{j+1} < ... < r_{n(i,t)} < r_{n(i,t)+1}, n(i,t)$ is chosen so that $r_{n(i,t)} < r_{th} = 3.7\text{Å} < r_{n(i,t)+1}$. Then $I(i,t)$ is defined as

$$I(i,t) = \frac{1}{n(i,t)} \sum_{j=1}^{n(i,t)} [\Delta(j;i,t) - \bar{\Delta}(i,t)]^2 \tag{4.6}$$

where $\Delta(j;i,t) = r_{j+1} - r_j$ and $\bar{\Delta}(i,t)$ is the average over all molecules of $\Delta(j;i,t)$. Thus, $I(i,t)$ expresses the inhomogeneity in the radial distribution within a sphere of radius around 3.7Å. The choice of 3.7Å comes from the observation of the existence of certain molecules which show an unoccupied gap between 3.2Å and 3.8Å in their radial neighbor distribution for certain periods of time[116, 117]. A low LSI value corresponds to a disordered local environment with high-local density while a high LSI indicates a highly structured, tetrahedral coordination with low local density.

We apply the LSI order parameter directly to the real thermally disordered structure in the simulations. A LSI cutoff value is selected to separate the molecules with an approximately 1:1 ratio for AIMD liquid water with PBE0+TS at 330K($I_c = 0.025$Å). The LSI distributions in PBE, PBE0, PBE0+TS at 300K, PBE0+TS at 330K are all studied to understand how the various effects (PBE0, vdW and quantum nuclei) correlate with the molecular forces and dynamics. Fig.[4.10] plots the LSI distribution in these various environments, reflecting the changing population of the two sets in different samples. With the chosen LSI cutoff, the fraction of HDL-like sites is 23%

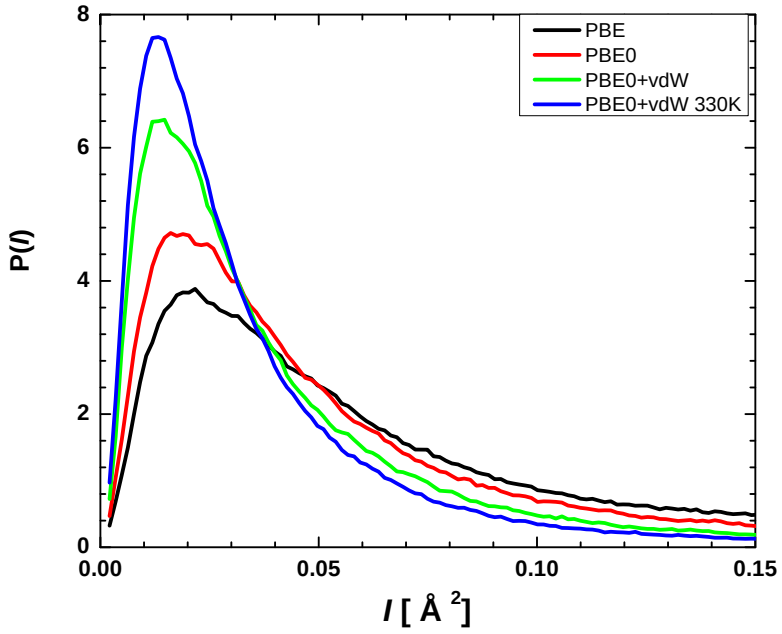

Figure 4.10: Probability density of finding a molecule with local structure index I from PBE , PBE0, PBE0+TS, PBE0+TS at 330K.

with PBE, 30% with PBE0, 43% with PBE0+TS, 50% with PBE0+TS at 330K. This increase of HDL-like sites due to PBE0 and quantum nuclei are more or less the same ($\sim$7% increase for each effect), while the vdW interaction has a larger contribution in increasing the population of HDL-like sites ($\sim$13% increase). This indicates that vdW interactions play an more important role in HDL-like sites formation, which is largely via the penetration of more interstitial molecules into first coordination shell.

Fig.[4.11] illustrates the calculated OO RDFs for LDL- and HDL-like sites compared to those derived from experiment. The RDF of LDL-like exhibits a highly structured distribution, while the first minimum of HDL-like RDF

is difficult to locate due to the greater population of interstitial molecules in HDL-like sites leading to a collapse of the second coordination shell. Both LDL- and HDL-like RDFs show similarity compared to experimentally derived RDFs for LDL and HDL water.

HDL- and LDL-like sites have distinct structural features. The high-LSI sites (LDL like) are characterized by very well defined first and second coordination shells and a deep first minimum similar to the experimentally derived RDF of LDL water. This is in sharp contrast to the low-LSI sites (HDL like) which feature a collapsed second coordination shell and a pronounced shoulder at interstitial distances around 3.5 Å, similar to what is observed for high density amorphous (HDA) ice.

The coordination number that is associated with OO RDFs, is studied to show the average number of neighboring molecules of each molecule. The calculated value is $\sim$4.66 in the simulation of liquid water, and is $\sim$ 4.83 and $\sim$ 4.43 for HDL- and LDL-like sites, respectively. The higher coordination number in HDL-like is due to more population of interstitial molecules between the first and second coordination shell.

The H-bond statistics is also interesting to investigate on LDL- and HDL-like sites. With our adopted H-bond definition, the average fraction of broken H-bond is 10% in LDL-like and 19% in HDL-like sites. Connected largely by a tetrahedral structure, water molecules in LDL-like sites are mainly connected with 4 H-bonds with some degree of distortion and breaking of H-bonds. The HDL-like sites have a higher broken H-bond fraction, combined with a higher fraction of interstitial molecules. However, more than three out of four H-bonds are still connected, indicating that the differences in these two sets are largely due to interstitial molecules's role instead of the H-bonded peak. To

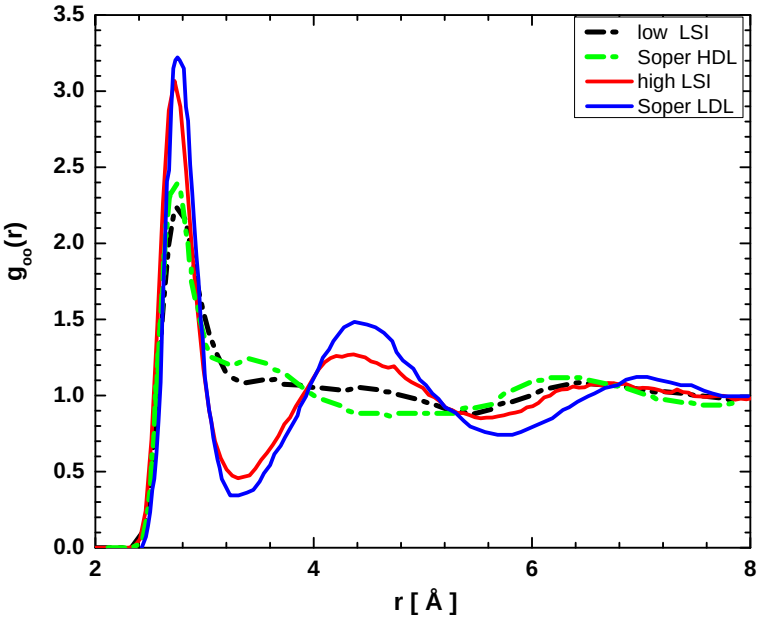

Figure 4.11: Comparison of experimentally derived HDL and LDL OO RDF [115] with that of sites in PBE0+TS at 330K simulation that distinguish HDL- (low LSI) and LDL-like (high LSI) molecules according to one given cutoff of LSI ($I_c = 0.025$Å).

fully understand this behavior, a detailed quantitative analysis of H-bonded and non H-bonded subsets is required.

Fig.[4.12] and Fig.[4.13] plot the OO RDFs for H-bonded and non H-bonded molecules in HDL-like, LDL-like sites respectively. In the H-bonded region of the first coordination shell, PBE0 and the quantum nuclei have a larger effect in softening HDL-like structure. In the non H-bonded region, vdW has a more important role by pulling interstitial molecules into first coordination shell. This effect is pretty evident in the non H-bonded HDL-like OO RDF from Fig.[4.12]. Hybrid functionals and the quantum nuclei soften the liquid structure but do not change significantly the average network topology by favoring more high-density configurations (i.e. configurations with more interstitial molecules than simply allowed by the broken H-bonds). The increase of the non-bonded fraction in excess to what is made possible by the broken H-bonds is due to vdW interactions, which increase the number of interstitial molecules by displacing molecules from the second to the first coordination shell. Since the interstitial molecules are not H-bonded to the central molecule, the only attractive interaction that may bring them closer to the central molecule is the vdW interaction. On the other hand, in the low density local structure LDL-like sites having a network topology closer to that of LDA ice, the fraction of interstitial molecules is much less and the contribution of vdW is weaker and leads to structural changes similar in magnitude to those generated by the hybrid functionals and the quantum nuclei. Interestingly, many more interstitial molecules are present in the neighborhood of HDL-like sites than in that of LDL-like sites, but the H-bonded fraction changes by a lesser amount. This is due to a change of network topology between the two sets. This explanation is supported by the coordination number analysis in the HDL- and LDL-like sites that is decom-

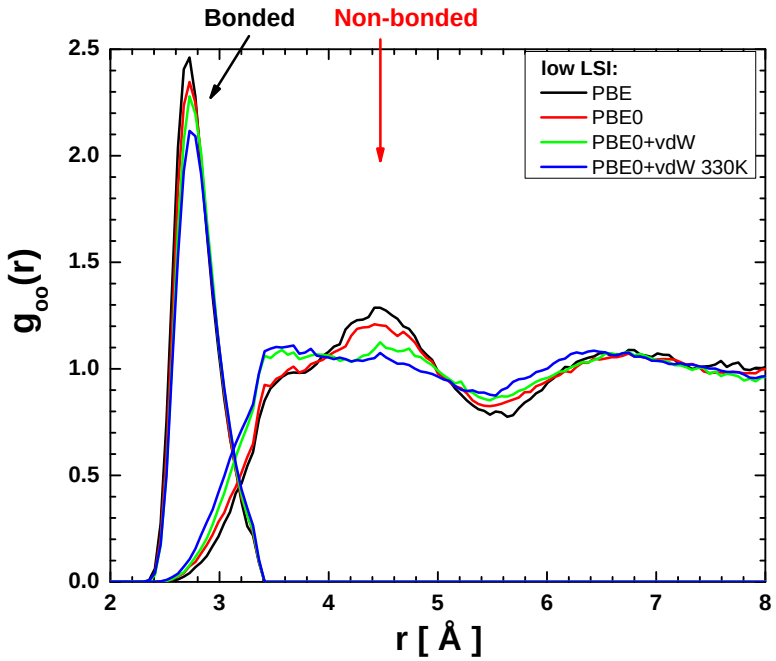

Figure 4.12: OO RDFs of H-bonded and non H-bonded HDL-like species with PBE, PBE0, PBE0+TS, PBE0+TS at 330K. In H-bonded region of first coordination shell, PBE0 and quantum nuclei (mimicking) have a larger effect in softening HDL-like structure. In the non H-bonded region, vdW have a more important role by pulling interstitial molecules into first coordination shell.

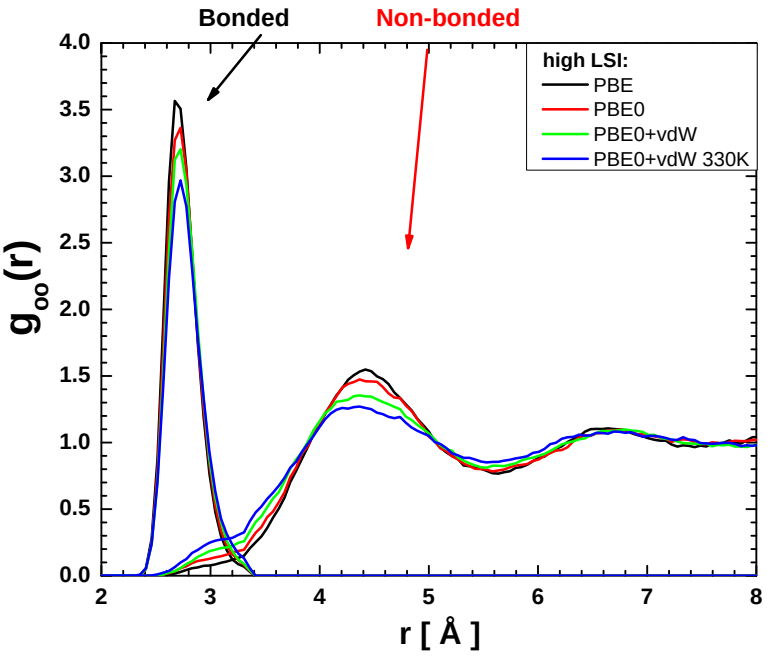

Figure 4.13: OO RDFs of H-bonded and non H-bonded LDL-like sites with PBE, PBE0, PBE0+TS, PBE0+TS at 330K. Different from HDL-like species, interstitial molecules are fewer in LDL-like species as seen from Fig.[4.11]. PBE0, vdW and quantum nuclei (mimicking) show similar effect in both H-bonded and non H-bonded regions.

	PBE	PBE0	PBE0+vdW	PBE0+vdW 330K
Low -LSI H-bonded	3.61	3.52	3.44	3.34
Low-LSI non H-bonded	0.92	1.06	1.31	1.46
High-LSI H-bonded	3.95	3.87	3.81	3.68
High-LSI non H-bonded	0.27	0.41	0.53	0.70

Figure 4.14: The coordination number in H-bonded and non H-bonded HDL-like, LDL-like sites with PBE, PBE0, PBE0+TS, PBE0+TS at 330K.

posed into H-bonded and non H-bonded contribution, as seen in Fig.[4.14]. The relative change in the bonded fraction between the two sets is smaller than the corresponding change in the non-bonded fraction, i.e., the H-bonded fraction changes less.

In the H-bonded region, on the other hand, the local structure of HDL- and LDL-like sites shows similar behavior, being almost equally tetrahedral in both cases. The Fig.[4.15] plots the three body oxygen-oxygen-oxygen (OOO) angle distribution. In a perfect tetrahedral structure, this angle distribution is a delta function centered at 109.47^o. In both H-bonded HDL- and LDL-like sites, this tetrahedral structure is primarily maintained, with angle distribution close to gaussian distributions with average 107.4^o, 107.6^o and standard deviation 22.3^o, 20.1^o in HDL-like, LDL-like respectively. This indicates that the local bonding environment of HDL- and LDL- like are generally the same, with only slight difference.

Last but not least, it is interesting to investigate how HDL- and LDL-like sites organize in liquid water, i.e., what's the local ordering of these two

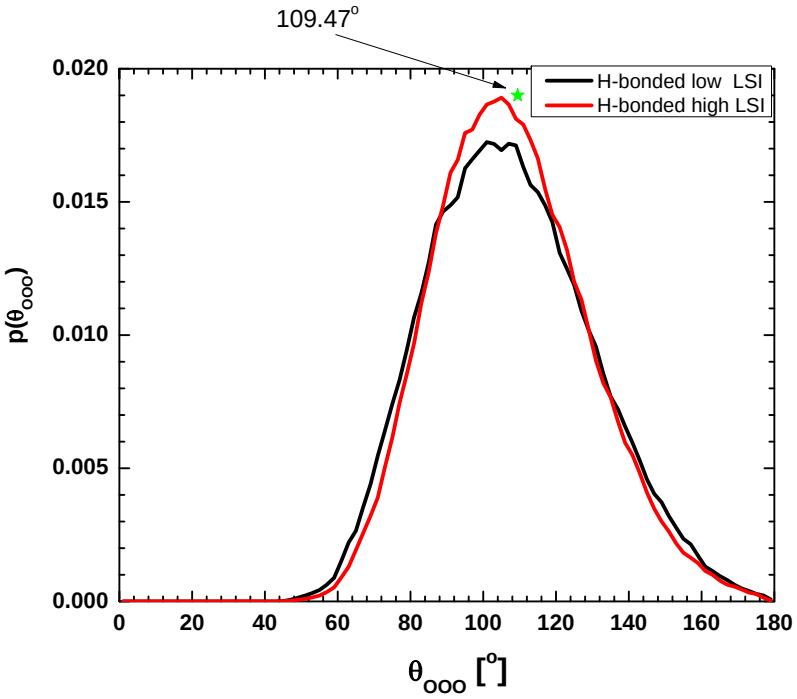

Figure 4.15: Three body oxygen-oxygen-oxygen (OOO) angle distribution of H-bonded HDL-like, LDL-like sites with PBE0+TS at 330K.

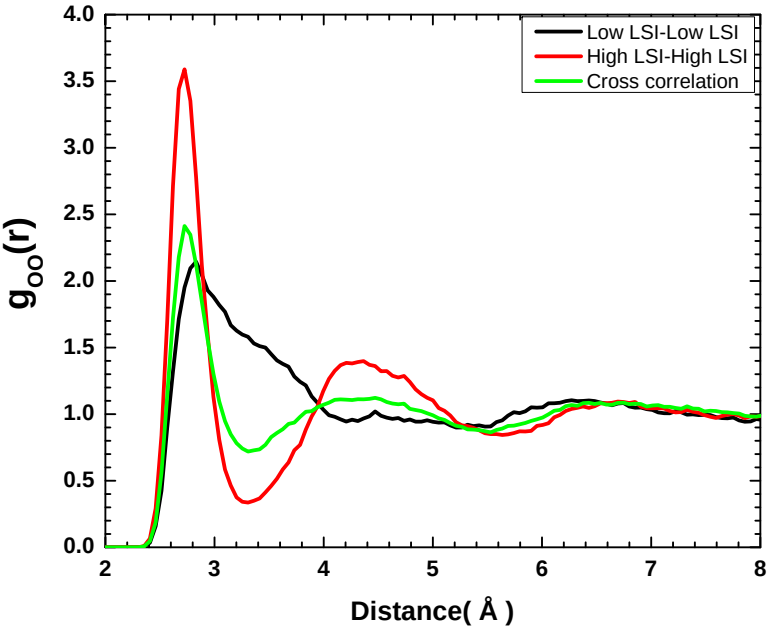

Figure 4.16: OO RDFs of HDL-HDL, LDL-LDL and their cross correlation with PBE0+TS at 330K.

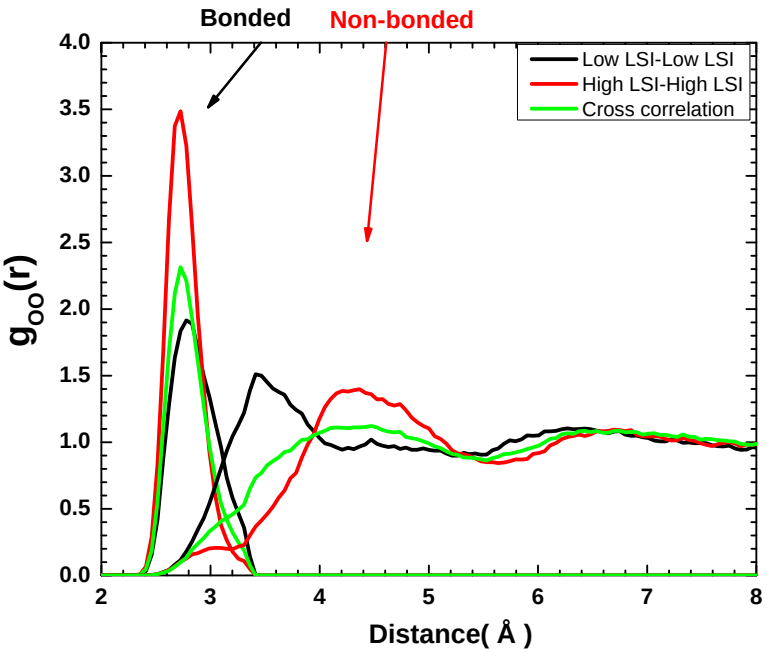

Figure 4.17: OO RDFs of H-bonded and non H-bonded in the subsets of HDL-HDL, LDL-LDL and their cross correlation with PBE0+TS at 330K.

sites? Are they mixed or clustered? In order to answer this, an analysis of cross correlation between HDL and LDL and relevant coordination number is required. Fig.[4.16] plots the OO RDFs of HDL-HDL, LDL-LDL and their cross correlation as resulting from our simulation with PBE0+TS at 330K. In this plot, RDF of HDL-HDL shows more interstitial molecules and RDF of LDL-LDL is more or less a tetrahedral structure. This suggests that in the short range, LDL-like sites have some degree of clustering as seen in the first peak region. On the other hand, HDL-like sites correlation is more pronounced in the interstitial range, resulting in a collapse of second coordination shell. Both HDL-HDL and LDL-LDL correlations are observed in the first and second coordination shell, but become weaker after the third coordination shell. Fig.[4.17] decomposes these correlations into H-bonded and non H-bonded subsets. The distinction between low LSI-low LSI and high LSI-high LSI becomes more obvious. Low LSI-low LSI which corresponds to high local density, show a more obvious peak around 3.5 Å and this value is roughly the first minimum of OO RDF of liquid water. This indicates that more interstitial molecules are in between first and second coordination shell. High LSI-high LSI shows a peak around 4.3 Å which is roughly the position of second peak of OO RDF of liquid water. Therefore, this means a normal tetrahedral structure for high LSI-high LSI correlation.

Fig.[4.18] shows the coordination number of H-bonded and non H-bonded in the subsets of low LSI-low LSI, high LSI-high LSI and cross correlation from simulation with PBE0+TS at 330K. In the H-bonded region, the coordination number of low LSI-low LSI correlation is 1.73, which is close to that in the cross correlation (1.62). Both are a little smaller than the coordination number of high LSI-high LSI sites (2.05). This indicates that in H-bonded region, LDL-like sites have some degree of clustering but not much, while

	Low LSI-Low LSI	High LSI-High LSI	Cross correlation
H-bonded	1.73	2.05	1.62
Non H-bonded	1.01	0.28	0.51

Figure 4.18: Coordination number of H-bonded and non H-bonded in the subsets of Low LSI-low LSI, High LSI-High LSI and cross correlation with PBE0+TS at 330K

HDL-like sites have little clustering. This analysis is consistent with that from OO RDF. In the non H-bonded region, the coordination number of HDL-HDL is 1.01. This value is higher than the number from cross correlation (0.51) and the number from LDL-LDL (0.28). Therefore HDL-sites are somewhat clustered in interstitial range. The basic message that one can extract from that is: at short range (nearest neighbor distance) some clustering is present but mainly in high LSI sites and negligibly in low LSI sites. At the interstitial range (around 3.5 Å), low LSI sites show some clustering.

With this, we can conclude that the analysis of HDL- and LDL-like sites in *ab initio* water is consistent with the two state model of water.

5 Conclusion

In this thesis, we focus on ambient water structure from AIMD simulation. We found that in order to achieve an excellent description of liquid water structure, three improvements are required to cure the deficiency of GGA functional with classical nuclei simulation:

1: Quantum treatment of hydrogen instead of classically as we should, leads to less overstructuring in liquid water[10];

2: Inclusion of PBE0 hybrid functional over semi-local GGA functional(PBE) corrects the spurious self-interaction error and this leads to an improved liquid water description;

3: The vdW non-local correlation that is missing in GGA and hybrid functionals is also found to be important.

We observed that inclusion of all three important effects could achieve dramatically great results by referring to experiment data. The OO RDF, structure factor $S(q)$, average broken H-bond statistics, the tetrahedrality of liquid water $\langle q \rangle$ support this conclusion. Though currently quantum nuclei effect is mimicked by a temperature increase due to the current computing limitation, a complete simulation accounting for the above three effects will be possible in the coming future with the higher-speed computing power and more ingenious algorithm development. Moreover, the analysis of HDL- and LDL-like sites in liquid water shows more clear the true role that PBE0, vdW and quantum nuclei play. And more importantly, the current study provides evidence that sites having LDL like and HDL like character coexist in liquid water at ambient conditions. The coexistence of these sites with distinct characteristics is definitely related to the existence of two amorphous

ice forms in water. Whether or not this is related to the presence of a first order transition in the metastable liquid regime, as hypothesized in Ref.[100], is beyond our study.

Improving the modeling of water to have a better physics understanding of liquid water structure not only enhances the predictive power of computational methodologies based on first principle electronic structure theory, but also provides novel insight for experimental observations. Our detailed investigations of the microscopic properties of bulk water will also be a major step forward to better understand water's role in different environments, such as interfacial water in materials and water at biological interfaces. Another important issue is a better understanding of water as a solvating medium and provide insight about how water rearranges around small solutes (both hydrophilic and hydrophobic), and what signatures of important water defects(hydronium, hydroxyl ions, the hydroxy radical etc.) can be detected with spectroscopic means.

It should be stressed that the connection between theory and experiment works in both directions. On one hand experimental data is essential to validate theory, on the other hand theory helps in interpreting the experiments and can even lead to extract more accurate experimental information. For instance, the experimental uncertainty in the first peak of the OO RDF comes from the incomplete knowledge of the x-ray scattering form factor for oxygen in condensed water phases and the finite q_{max} scattering density. In theory, we could contribute to further solving the uncertainty in this OO RDF by extracting the form factor from accurate DFT calculation. Our simulated ambient water gives a first peak value of OO RDF ~ 2.57 around the position of 2.76 Å, in pretty good agreement with latest experiment determination derived from the highest q_{max} data currently available. This can

shed light on the validation and verification of approximations used in the experimental data processing.

Lastly, computer simulation has been playing a vital role in both the quantitative characterization of liquid water and the advancement of our qualitative understanding of water and its anomalies.

Appendices

A Kohn Sham density functional theory

Instead of empirical potential energy, a more fundamental approach without empirical fitting is to derive the potential energy surface directly from the elementary interaction. The atom is consisted of electrons and nuclei in the atomic scale. While it is often efficient to treat nuclei classically, the electrons need to be treated quantum mechanically. And to a good approximation, the light electrons follow adiabatically the heavy nuclei, which is the essence of *Born-Oppenheimer* adiabatic approximation. According to this approximation, the electrons that are initially in the ground state will remain in the ground state corresponding to the nuclear configuration visited at that particular instant. The great approximation would lead to the separation of electronic and nuclei dynamics. Therefore, the nuclear parameters R can be treated as external parameters in the many-body electronic Hamiltonian $\hat{H}$:

$$\hat{H} = \frac{-\hbar^2}{2m}\sum_i \nabla_i^2 + \sum_{i,I}\frac{-Z_I e^2}{|r_i - R_I|} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|r_i - r_j|} \tag{A.1}$$

In this equation, upper case indices refer to nuclei and lower case indicates to electrons. Z_I is the atomic number, e is the absolute value of the electron charge, m the electron mass and the sum run over nuclei and electrons.

The ground state of electrons is found by minimization:

$$E_{GS}(R) = Min_{\Psi}\langle\Psi|\hat{H}|\Psi\rangle \tag{A.2}$$

Here $\Psi(r) \equiv \Psi(r_1, r_2, r_3, ..., r_{Ne})$ is a normalized many-electron wavefunction and N_e is the total number of electrons(For simplicity, spin variables are

omitted here). Then the nuclear potential energy surface is given by

$$\Phi(R) = E_{GS}(R) + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J e^2}{|R_I - R_J|} \tag{A.3}$$

The classical nuclear trajectories can be computed without empirical fitting of potential. However, solving this Eq.[A.2] is not easy as this is a formidable quantum many-body problem that requires further approximation.

To simplify this problem , we are adopting Kohn-Sham DFT(KS-DFT), a formally exact scheme to map a system of interacting electrons into a fictitious system of non-interacting electrons with the same density. KS-DFT is an approach to the many-electron problem in which electron density, rather than the many-electron wave functional, plays the central role[54, 121]. It makes use of the Hohenberg-Kohn theorem, which states that there is a one-to-one relationship between the ground state energy and the electronic density, *n(r)*. Thus, it is possible to express the energy as a functional of the electron density. This greatly simplifies the problem due to the fact that density is only a function of single three-fold vector, whereas the wavefunction is a function with three dimensional position vectors times the number of states.

In the recent years, DFT has become the method of choice for electronic-structure calculations across an wide variety of fields, from organic chemistry to condensed matter physics[56]. There are two main reasons for the success of DFT: First, DFT offers the only currently known practical method of fully quantum mechanical calculations for systems with many hundreds or even thousands of electrons. Second, it enhances our understanding by relying on relatively simple, physically accessible quantities that are easily visualized even for large systems[122].

According to DFT, the ground state energy of the interacting system with given nuclear configuration R, can be obtained by minimizing a functional of the electron density.

$$E_{GS}(R) = \mathrm{Min}_{n(r)} E_V[n] \tag{A.4}$$

Here $V \equiv V(r) = \sum_I \frac{-Z_I e^2}{|r-R_I|}$ is the external potential of the nuclei acting on the electrons. And the potential V depends parametrically on the nuclear configuration R. $E_V[n]$ is a functional of the orbital, with $n(r) = 2\sum_{i=1}^{N_e/2} |\psi_i(r)|^2$ for a closed shell system(For other systems, it is straightforward to write down the n(r) expression as well). In this case, the minimization problem is equal to the case with respect to the orbitals

$$E_{GS}(R) = \mathrm{Min}_{\psi^*} E_V[\psi, \psi^*] \tag{A.5}$$

Eq.[A.5] now becomes much simpler compared to Eq.[A.2], as Eq.[A.2] has variables that is a many-body wavefunction and is growing exponentially with the number of electrons. However, the Eq.[A.5] has $N_e/2$ independent variables with only orthonormality constraint, i.e. $\langle\psi_i|\psi_j\rangle = \delta_{ij}$. With this, the energy functional in Eq.[A.5] becomes:

$$\begin{aligned} E_V[\psi^*, \psi] = 2\sum_{i=1}^{N_e/2} \langle\psi_i| \frac{-\hbar^2\nabla^2}{2m} |\psi_i\rangle + \int V(r) n(r) dr \\ + \frac{1}{2} \int\int dr dr' \frac{n(r)n(r')e^2}{|r-r'|} + E_{XC}[n] \end{aligned} \tag{A.6}$$

In this equation, the first term is the kinetic energy associated to the single particle orbitals. The second term is the potential energy of the electrons in the field of the nuclei and the third is the average Coulomb energy of the electrons. $E_{XC}[n]$ accounts for the exchange and correlation energy, which is the remaining contribution to E_{GS}.

$E_{XC}[n]$ is an unknown universal functional of the density. The choice of exchange correlation functional is critical in generating accurate results from DFT methodology. And this thesis is dedicated to improve this important term in order to better describe the microstructure of liquid water .

The minimization problem in Eq.[A.5] then leads to Euler-Lagrange equations

$$\frac{\delta E_V[\psi^*, \psi]}{\delta \psi_i^*(r)} - \epsilon_i \psi_i(r) = 0 \tag{A.7}$$

where ϵ_i are lagrange multipliers that keep the norm of the orbitals unitary. Expanding the functional derivative would give

$$(\frac{-\hbar^2 \nabla^2}{2m} + V(r) + \int \frac{n(r')e^2}{|r - r'|} dr' + V_{XC}(r))\psi_i(r) = \epsilon_i \psi_i(r) \tag{A.8}$$

The left term in the bracket is the Kohn-Sham Hamiltonian and $V_{XC}(r) = \frac{\delta E_{XC}[n]}{\delta n(r)}$ is the exchange-correlation potential. Eq.[A.8] is well known as the Kohn-Sham equations.

If the form of this functional was known exactly, then the density of the auxiliary system would be exactly that of the real system, and we would be able to solve any problem within the limit of numerical accuracy. Unfortunately, this exact form is unknown. In practice, approximations are necessary to write explicit expressions for the exchange-correlation energy and potential as functionals of the density. Commonly used approximations are the local density approximation(LDA)[53] or the generalized gradient approximation(GGA)[54, 123].

B Car-Parrinello molecular dynamics

The emerging of Car-Parrinello molecular dynamics(CPMD) (1985) originates from the limitations suffered by both DFT and classical molecular dynamics, which effectively restricted their application to the realistic simulation of condensed matter at finite temperature to specific cases. Pure DFT was mainly applicable to the electronic structures of ordered and homogenous systems. On the other hand, the forces between atoms sued in molecular dynamics did not take into account the fact that the electronic potentials varied with the atomic movement during the progress of a simulation. By using DFT to calculate the potential 'felt' by atoms and letting such potential evolve with each step of the simulations, the Car-Parrinello method allowed a much wider range of disordered and therefore more realistic materials systems to be studied.

The main obstacle that prevents the use of potential energy surfaces derived from DFT in molecular dynamics simulations is the high cost of the quantum mechanical calculation of the electronic energy and electronic forces on the nuclei. In a molecular dynamics simulation, the Kohn-Sham equation need to solved self-consistently at all the nuclear configurations visited in a trajectory. In order to compute meaningful statistical averages the number of nuclear configurations in a numerical trajectory must be large, i.e. long time simulation is always needed to obtain converged result. And this cost exists until the formulation of the Car-Parrinello approach come out.

So right here we are giving a brief introduction about CPMD approach, as this is the computation tool we are using to investigate liquid water structure. The Car-Parrinello approach[46] is an extended Lagrangian formulation, in which both nuclear and electronic degree of freedom act as dynamic variables.

The dynamics derives from the Lagragian postulated by Car and Parrinello:

$$L_{CP} = \frac{1}{2}\sum_I M_I \dot{R}_I^2 + 2\mu \sum_i \langle \dot{\psi}|\dot{\psi}\rangle - \Phi[R,\psi^*,\psi] + 2\sum_{i,j}\lambda_{i,j}(\langle\psi_i|\psi_j\rangle - \delta_{ij}) \quad \text{(B.1)}$$

Here μ is the mass parameter with unit of a mass time a length squared. The mass parameter is used to control the dynamical response of the electronic orbital $\psi_i(r,t)$. Note that in order to keep the derivation as transparent as possible, the more convenient atomic units(a.u.) is used here. And the $\dot{\psi}_i(r,t) \equiv \partial\psi_i(r,t)/\partial t$ describes the rate of change of the orbitals with respect to time. λ_{ij} is the lagrange multipliers that impose orthonormality among the orbitals. The second term in the equation gives the kinetic energy associated with the time evolution of the orbitals. If we combine the nuclear and electronic parameter space together, we would have potential energy surface

$$\Phi_{CP}[R,\psi^*,\psi] = E_{V(R)}[\psi^*,\psi] + \frac{1}{2}\sum_{I\neq J}\frac{Z_I^v Z_J^v}{|R_I - R_J|} \quad \text{(B.2)}$$

The first term is the Kohn-Sham energy functional that has been discussed above in Eq.[A.6].

From the Lagrangian equation, we obtain the equation of motions for orbitals and nuclei:

$$\begin{aligned} M_I\ddot{R}_I &= -\frac{\partial\Phi_{CP}[R,\psi^*,\psi]}{\partial R_I} \\ \mu|\ddot{\psi}_i\rangle &= -H_{KS}|\psi_i\rangle + \sum_j |\psi_j\rangle\lambda_{ji} \end{aligned} \quad \text{(B.3)}$$

where $H_{KS} = \frac{\delta\Phi_{CP}[R,\psi^*,\psi]}{2\delta\langle\psi_i|}$.

These two equations are usually called Car-Parrinello equations. They generate trajectories in the extended parameter space of nuclear and electronic degrees of freedom. And the Lagrange multipliers that ensure the orthogonality of Kohn-Sham orbitals are given by

$$\lambda_{ij}(t) = \langle\psi_i(t)|\hat{H}_{KS}(t)|\psi_j(t)\rangle - \mu\langle\dot{\psi}_i(t)|\dot{\psi}_j(t)\rangle \quad \text{(B.4)}$$

Eqs.[B.3] has conserved total energy in extended parameter space:

$$H_{CP} = K + K_e + \Phi_{CP} \tag{B.5}$$

Here $K(\dot{R}) = \frac{1}{2}\sum_I M_I \dot{R}_I^2$ is the kinetic energy of the nuclei. The $K_e = 2\mu \sum_i \langle \dot{\psi}_i(t) | \dot{\psi}_i(t) \rangle$ is the fictitious kinetic energy of the electronic orbitals under the dynamics generated by Eq.[B.3].

The essence of CPMD is that owing to the mass separation between the electrons and the ions, the electrons follow the ionic motion adiabatically. The electronic evolution may be modeled by means of a fictitious dynamical scheme. In this approach, the first principle molecular dynamics trajectory is generated without the need for re-minimization of the electron wavefunction at each time step, that is, it is possible to generate one step of electronic motion per step of nuclear evolution. They presented a unified dynamical framework by which the electrons and ions evolve adiabatically. A CPMD time step requires evaluating the action of the Hamiltonian H_{KS} on the electronic orbitals, which is much simpler and less time consuming compared to a self-consistent diagonalization of Kohn-Sham Hamiltonian. And the nuclear dynamics generated by Eq.[B.3] is an excellent approximation of the Born-Oppenheimer time evolution when orbital dynamics is fast and follows adiabatically nuclear motion. Therefore CPMD dramatically decreases the amount of computational effort necessary and has paved the way for studies of a large variety of systems that would not be possible otherwise.

The remaining task is to separate in practice nuclear and electronic motion such that the fast electronic subsystem stay cold also for long times but still follows the slow nuclear motion adiabatically. And adiabaticity is controlled by the fictitious mass parameter μ. Simultaneously, the nuclei must nevertheless be kept at a much higher temperature. This can be achieved in

nonlinear classical dynamics via decoupling of the two subsystems(fast electrons and slow ions) and adiabatic time evolution. This is possible if the power spectra stemming from the dynamics de not have substantial overlap of their respective vibrational density of states, so that the energy transfer from the “hot nuclei” to the ”cold electrons” becomes practically impossible on the relevant time scales.

C Performing meaning CPMD simulations

CPMD is a so useful theoretical tool that it would be necessary to have a few words on how to perform a meaningful and correct simulation. First a general procedure to perform CPMD simulation is recalled, followed by several key quantities to monitor to make sure a meaningful CPMD is run. After that, the issue about finite size effect is talked about. In CPMD simulation with NVT ensemble, an effective temperature control technique is required and thus Nose-Hoover thermostat that is adopted in the entire simulation is reviewed.

C.1 The procedure of running CPMD simulation

In order to achieve an excellent and correct CPMD simulation, typically we need a good initial configuration which is often taken from empirical force field simulation, as *ab-initio* simulation is much more expensive.

After the initial configuration is successfully obtained, the electronic wavefunctions would be optimized to ground state with damping CPMD scheme[47], which has proved efficient in optimizing electrons wave functions. In order to run a statistically meaningful CPMD, the configuration would need to modified around their equilibrium positions by adding in a random displacement for each atom, so that ions would deviate from their equilibrium and perform statistical sampling. Otherwise, the ions would move little in molecular dynamics and lead to poor statistical sampling. A simple check whether the configuration is appropriate can be monitored by looking at the forces acting on ions. Typically an order of 10^{-2} a.u. is suitable for CPMD simulation. If the forces is quite smaller like 10^{-4} a.u., then a larger random placement for

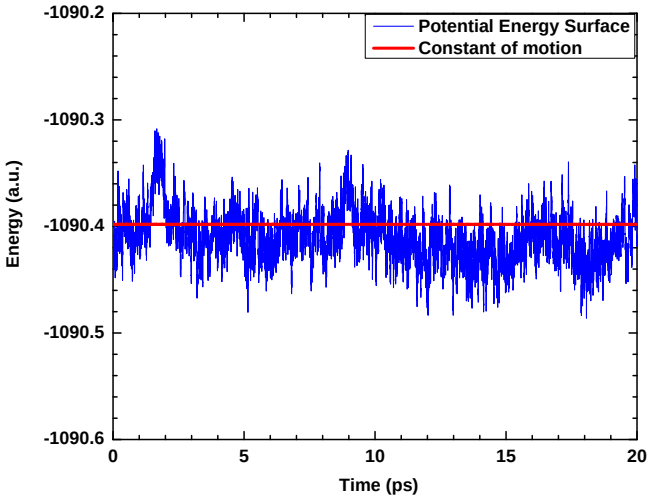

Figure C.1: The time evolution of potential energy surface and constant of motion in a typical CPMD simulation.

the configuration needs to be added in.

There are typically four key quantities to keep an eye on in CPMD simulation:

(1): The potential energy surface fluctuates as a function of simulation time and does not show any drift, otherwise the system is still evolving into desired equilibrium state.

(2): The constant of motion, a quantity that is a constant in the CPMD methodology should not show any drift, according to the Eq.[B.5] in micro-canonical ensemble or Eq.[C.8] in canonical ensemble, otherwise some numerical integration or other related issues may have something wrong.

(3): The average temperature of oxygen and hydrogen is the desired temperature under canonical ensemble.

(4): The ratio of ions kinetic energy versus the fictitious electron kinetic

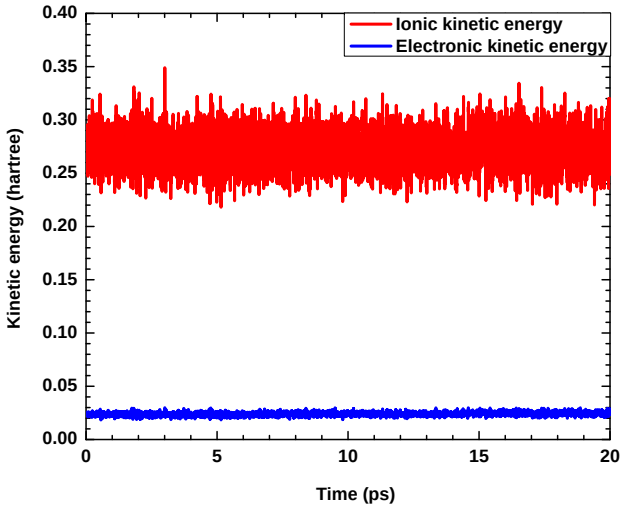

Figure C.2: The fictitious Car-Parrinello electron kinetic energy(blue line) and the kinetic energy of the ions(red line) are plotted as a function of time in a typical CPMD simulation.

energy should be large enough to satisfy the adiabatical conditions between ions and electrons, i.e. electrons follows adiabatically the motion of ions.

A typical simulation result with PBE functional is shown for better illustration. Fig.[C.1] shows the potential energy surface and the constant of motion over a typical CPMD running. There is no drift in the potential energy indicating that the system is in equilibrium state. The constant of motion is a straight line, meaning that the integration over CPMD is well performed.

Fig.[C.2] shows the evolution of electron fictitious kinetic energy and ions kinetic energy in CPMD simulation. The simulation is done under NVT ensemble with T=300K. The plotted energy is converted to atomic unit with the relation $E_{ions} = \frac{3}{2}Nk_BT$. There is no drift in both electronic fictitious

kinetic energy and ionic kinetic energy and the ratio of ions over electrons is around 10. This indicates that the adiabatic condition between ions and electrons are satisfied quite well, i.e. there are no energy transfer from ions to electrons in the simulation.

C.2 Finite size effect

In AIMD simulation, it is preferred to work in the supercell with periodic boundary conditions, in which water molecules are put into the box with system size that corresponds to experimental density. In periodic boundary conditions that is first developed by Born and von Karman in 1912, the cubic box is replicated throughout space to form an infinite lattice. Each molecule interacts with a series of periodic images of the system. This greatly simplifies the problem of simulating large system with almost infinite atoms in real world.

However, it is important to ask if the properties of a small, infinitely periodic system can represent the macroscopic system. This so-called finite size effect exists in expensive *ab initio* simulation and needs to be minimized as much as possible. The computing transport properties such as the diffusion constant has been shown to depend on system size[6]. Furthermore, since the system is periodic on a small interval, the RDF is only valid up to a distance as long as only the box size allows.

Fig.[C.3] taken from Kuhne *et al.*[6] shows the OO RDF with semi-local GGA functional in three supercells(32, 64, 128 water molecules respectively). The difference is quite huge between 32 and 64 water system, while it is almost negligible between 64 and 128 water system. Therefore at PBE functional level the finite size effect is not negligible in 32 water system, that is, the

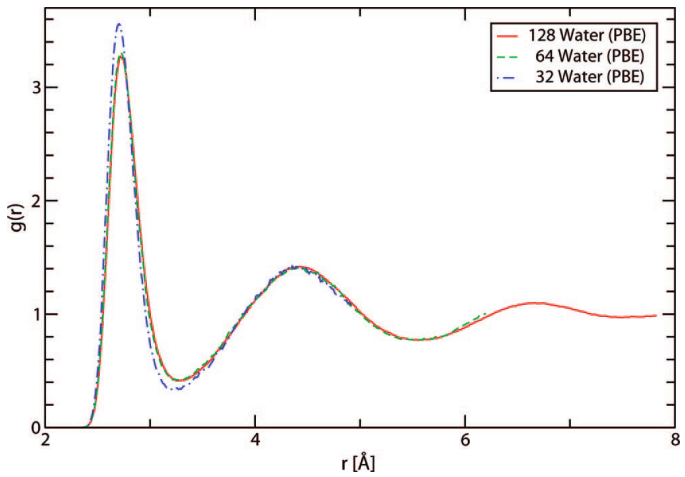

Figure C.3: Comparison of the oxygen-oxygen(OO) RDFs, obtained from AIMD simulations with 32, 64, and 128 water molecules in the supercell, plot taken from Ref.[6]

liquid water properties at ambient conditions can not be represented well in this small system. In the current CPMD code package[62] , the computation cost typically scales cubically with system size in planewave basis set representation and therefore there is a tradeoff between accuracy and efficiency. Due to this, we need to adopt at least 64 water system instead. In order to calculate In what follows, all the *ab initio* simulations are performed in the supercell with 64 water molecules, unless pointed out specifically.

C.3 Nose-Hoover thermostat

To study liquid water at ambient temperature within AIMD methodology, it is preferable to work under NVT canonical ensemble, i.e. the number of atoms, volume and temperature are kept fixed in the simulation, though MD simulations are most easily carried out in microcanonical ensemble(NVE

ensemble). In investigating liquid water properties, there are advantages to perform simulations in the canonical ensemble. For instance, one main objective of performing simulation is to calculate experimentally measurable quantities. The experiments with which we want to make contact are usually carried out under the conditions of constant particle number and temperature and either constant pressure or volume. Therefore, temperature control in molecular dynamics is essential as good scheme would be able to equilibrate the system quite efficiently, so that the ergodicity can be achieved successfully and the molecular dynamics statistical results will be meaningful.

This issue arises more serious in liquid water simulation with widely adopted Generalized Gradient approximation(GGA) for exchange-correlation functional[5], within which the H-bond strength is over estimated, as indicated from the overstructured RDFs compared to experimental data. This means under ambient condition, the water would have glassy behavior instead of liquid property within GGA approach. The mean square displacement(MSD), an indicator of liquid fluidity, shows smaller diffusion coefficient compared to experiment. Due to this, it would take much longer time in the expensive *ab initio* simulation to achieve converged statistical results. Furthermore, a good understanding of thermostat applied to canonical ensemble(NVT) would be easily transferred to other widely used ensemble(e.g. isobaric-isothermal ensemble(NPT)), in which the underlying physics is the same. Therefore, it is necessary to have a brief introduction and investigation of thermostats that are frequently used in CPMD simulation in order to harness this technical tool correctly.

One class of constant temperature simulation methods that has numerous applications to a wide range of problems is the so-called "extended system"(ES)[124]. The ES method, originally proposed by Nose[125], is rewritten by Noover[126]

in a more convenient form. We are briefly review the equation of motion in the Nose-Hoover chain dynamics. We consider the case of n chains where the jth chain is coupled to g_j degrees of freedom, where

$$\ddot{q}_{ij} = \frac{F_{ij}}{m_{ij}} - q_{ij}\dot{\eta}_{1j} \tag{C.1}$$

$$\ddot{\eta}_{1j} = \frac{\sum_{i=1}^{N} g_i m_{ij}\dot{q}_{ij} - g_j k_B T_j}{Q_{1j}} - \dot{\eta}_{1j}\dot{\eta}_{2j} \tag{C.2}$$

$$\ddot{\eta}_{kj} = \frac{Q_{k-1,j}\dot{\eta}^2_{k-1,j} - k_B T_j}{Q_{kj}} - \dot{\eta}_{kj}\dot{\eta}_{k+1,j} \tag{C.3}$$

$$\ddot{\eta}_{M_j j} = \frac{Q_{M_j-1,j}\dot{\eta}^2_{M_j-1,j} - k_B T_j}{Q_{M_j j}} \tag{C.4}$$

It can be shown that the extended system has the conserved quantity.

$$H' = \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{g_j} m_{ij}\dot{q}^2_{ij} + \frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{M_j} Q_{ij}\dot{\eta}^2_{ij} + U(q_{ij}) + \sum_{j=1}^{n} g_j k_B T_j \eta_{1j} + \sum_{j=1}^{n}\sum_{i=2}^{M_j} k_B T_j \eta_{ij} \tag{C.5}$$

The Nose-Hoover chain thermostats are attractive because they produce continuous dynamics with a well defined, conserved quantity. Assuming the dynamics is ergodic, they generate canonically distributed positions and momenta.

Combining this thermostat equation with Car-Parrinello approach expressed in Eq.[B.3], the equation of motion for ions coupled with thermostat is:

$$M_I \ddot{R}_I = -\frac{\partial \Phi_{CP}[R, \psi^*, \psi]}{\partial R_I} - \epsilon_R M_I \dot{R}_I \tag{C.6}$$

where the ions are coupled to independent thermostats, with ϵ_R the thermostat variables that act as dynamical friction coefficients on the nuclei

Correspondingly, the dynamics of the thermostats is governed by the following equation:

$$Q_R \dot{\epsilon}_R = \left(\sum_I M_I \dot{R}_I^2 - g k_B T \right) \tag{C.7}$$

Here Q_R is the thermostat "masses", g the number of independent internal nuclear degrees of freedom(g= $3N-3$ in molecular dynamics simulations of extended systems with periodic boundary conditions). The mass Q_R controls the dynamical response of the thermostat. The values are chosen to ensure good dynamical coupling between nuclei and the corresponding thermostats. Though Eq.[C.7] are not equivalent to Hamiltonian equations of motion, there is still a conserved quantity associated to this non-Hamiltonian dynamics , namely:

$$H_{CP}^{NH} = K_e + K + \Phi_{CP} + \frac{Q_R}{2}\epsilon_R^2 + g k_B T \int_0^t dt' \epsilon_R(t') \tag{C.8}$$

Eq.[C.7] eliminates the systematic drift of ions kinetic energy K in the simulations. In practical simulation, the conserved quantity(constant of motion) can be monitored to check whether there is something wrong, algorithmically or numerically in the implementation.

To numerically integrate these equations of motion is nontrivial especially Eq.[C.1], in which there are velocity dependent forces in the equation of mo-

tion. The original implementation adopted $\dot{\eta}_{1j} = (\eta_{1j}(t) - \eta_{1j}(t - \Delta t))/\Delta t$ to approximate Eq.[C.9], which is accurate only to the order of Δt and therefore leads to a nonnegligible drift in the constant of motion quantity. This drift accumulates as time evolves. If not handled properly, the simulation would lead to the deviation of correct ions dynamics and cause some underlying non-physical description.

If we apply the Verlet algorithm[89, 90] to the quantities q_{ij} and η_{1j} that have the accuracy to the order of Δt^2, the drift of constant of motion would be eliminated. Eq.[C.1] then numerically becomes

$$\frac{q_{ij}(t + \Delta t) + q_{ij}(t - \Delta t) - 2q_{ij}(t)}{(\Delta t)^2} = \frac{F_{ij}(t)}{m_{ij}} - q_{ij}(t)\frac{\eta_{1j}(t + \Delta t) - \eta_{1j}(t - \Delta t)}{2\Delta t} \quad \text{(C.9)}$$

If we wants to compute the atomic positions at time $t + \Delta t$ based on known quantities at time t, $\eta_{1j}(t + \Delta t)$ needs to be solved first. However this is impossible, as seen from Eq.[C.2], $\eta_{1j}(t + \Delta t)$ can only be solved with known $q_{ij}(t + \Delta t)$. Therefore Eqs.[C.1, C.2]need to be solved self-consistently, otherwise there would be a drift in the constant of motion due to the numerical integration error. A working alternative is to start with $\eta_{i,j}(t)$ as the initial guess to approximate $\eta_{i,j}(t + \Delta t)$, then the whole equation of motion quantities at $t + \Delta t$ are calculated. These calculated $\eta_{i,j}(t + \Delta t)$ can be taken as the new input to solve Eqs.[C.1, C.2] again. This algorithm is proved to worked well and do not require too many modifications in the original code within standard Verlet algorithm scheme.

The code package we are working on relies largely on Verlet algorithm for numerical integration. Although simple in form, the original Verlet algorithm and its modifications are at least as satisfactory as higher-order schemes that make use of beyond second derivatives of the particle coordinates. It may be less accurate than others at short times but, more importantly, it conserves

energy well even over long times. It is also time reversible, as it should be for consistency with the equations of motion. Some understanding of the reasons for the stability of the algorithms can be obtained in the reference[127].

Another non-trivial issue when using the Nose-Hoover algorithm is the set up of parameters, like the $Q_{i,j}$. The appropriate value for this parameter is inversely proportional to the characteristic frequency of the nuclear system to which it is coupled. In order to equilibrate the ions system as fast as possible to the desired temperature, we should turn to the ions spectrum of liquid water and choose the frequency that corresponds to the main vibrational spectrum. Moreover, we need to keep in mind that, in order to keep electrons to ground state, there needs to be no spectrum overlap between electrons and ions, or electrons and thermostats. Last but not the least, since oxygen and hydrogen has various spectrum ranges due to the mass, different Q_{ij} should be set according to the relations $Q_{kj} = (m_j/m_H)k_BT\tau$ suggested by Martyna *et al.*[128]. Where τ is the characteristic time scale of the system, m_H is the hydrogen mass and m_j is the mass of particles that are coupled to the thermostat. This enables both the hydrogen and oxygen to evolve to the desired temperature more quickly.

D Water dimer calculation with PBE0

To see how better hybrid functionals perform over GGA, it would be necessary to first look into the geometry properties of small water cluster system which has been widely studied and the accurate reference data are available. In small water clusters, hybrid functional, e.g. Perdew-Burke-Ernzerhof (PBE0) functional is found to give a more accurate description over PBE functional, by referring to accurate quantum chemistry calculation[129] in which the global minimum energy structures of small water clusters from dimer to pentamer is computed with reference to second order Moller-Plesset perturbation theory(MP2)[130]. From the Fig.[D.1], we can see that binding energy of water dimer decreases from 227.8 eV with PBE by 6.0 eV to 221.8 eV with PBE0. This decrease is fairly agreement with reference calculation[129] where 220.1 eV with PBE dropped by 5.6 eV to 214.5 eV. This decrease indicates that H-bond strength becomes weaker from PBE to PBE0, as the spurious delocalization error in PBE is cured by the introducing of exact exchange in PBE0.

The equilibrium geometric properties of water dimer are also fully investigated, as seen from Fig.[D.3]. The definition of these parameters can refer to Fig.[D.2]. Regarding the structural properties of the dimer, it shows the correct structural trend from PBE to PBE0. This follows the trends in the PBEPBE and PBE1PBE of Ref.[131] , which refers to PBE and PBE0 in our simulation, respectively. The PBE0 numbers are actually quite close to what is called best *ab initio*. The error of $\sim$ 1 percent in the equilibrium distance R is very good and within the accuracy of DFT. Also the fact that PBE and PBE0 give quite similar results is what we expect given that PBE already works quite well. The residual differences between Xu *et al.*[131] and

Binding Energy(Mev/ H bond)	Our calculation	Reference*
PBE	227.8	220.1
PBE0	221.8	214.5
PBE-PBE0	6.0	5.6

Figure D.1: Comparison of binding energy in water dimer with PBE and PBE0 functional. Reference data are taken from Ref.[129]

ours should be due to numerical effects (different basis sets, etc).

The most important effect is the fact that, while R is the same between PBE and PBE0 (within 0.2 percent), r_d is shorter by ~ 1.2 percent in PBE0. This indicates that the H-bond is going to be slightly less stronger with PBE0. Indeed the binding energy that we find is 5.67 kcal/mole with PBE and 5.54 kcal/mole with PBE0(1 kcal/mole = 43.3641 meV/mole). This difference between PBE and PBE0 functional is exactly the same as in Ref.[131], where for PBE it is 5.11 kcal/mole and 4.98 kcal/mole with PBE0.

If we take the reduction of binding strength as an indicator of the difference that we should expect in a bulk calculation (this is a very rough estimate because we neglect cooperative effects, i.e. we assume them to be the same in PBE and PBE0) a different in H-bond strength of ~ 2 percent which is equivalent to an effect of $\sim$ 30-40 K in temperature units, which is precisely the kind of effect we are looking for.

Based on these observations, PBE0 functional is found to perform better than PBE functional in small water cluster. This leads to the expectation

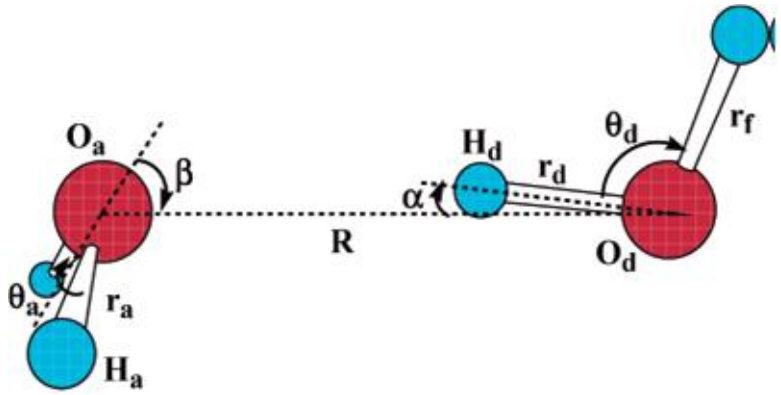

Figure D.2: Bonding geometry of the ground state water dimer, adopted from Ref.[131]

	R	r_d	r_f	r_a	θ_d	θ_a	α	β
PBE	2.890	0.980	0.968	0.970	104.39	104.85	3.33	56.4
PBE0	2.885	0.968	0.957	0.959	104.96	105.29	4.00	55.2
Best ab.initio	2.912	0.964	0.957	0.958	104.8	104.9	5.5	56.6

Figure D.3: Geometric properties(Å , deg) of water dimer. The *ab initio* data are taken from Ref.[131].

that PBE0 functional would have a better liquid water structure description compared to PBE functional when we turn to liquid water in condensed phase.

References

[1] P.Ball, Nature, **452**, 291(2008).

[2] A.Soper, J.Phys.Condens.Matt. **19**, 335206 (2007).

[3] J.P.Perdew, K.Burke, and M.Ernzerhof , Phys.Rev.Lett. **77**, 3865(1996).

[4] C.Lee, W.Yang and R.Parr, Phys.Rev.B **37**, 785(1988).

[5] J.C.Grossman, E.Schwegler, E.W.Draeger, F.Gygi and G.Galli, J.Chem.Phys. **120**, 300(2004).

[6] T.D.Kuhne, M.Krack, and M.Parrinello, J.Chem.Theory Comput. **5** , 235 (2009).

[7] N.Greenwood and A.Earnshaw, *Chemistry of the Elements* $2^{nd}ed.$,(Butterworth Heineman, Oxford, 1997).

[8] C.Andreani, D.Colognesi, J.Mayers, G.F.Reiter and R.Senesi, Adv.Phys. **54**, 377(2005).

[9] G.F.Reiter, J.C.Li, J.Mayers, et al., Braz J Phys, **34**, 142(2004).

[10] J.A.Morrone and R.Car, Phys. Rev. Lett. **101**, 017801(2008).

[11] C.Adamo and V.Barone, J.Chem.Phys. **110**,6158(1999).

[12] T.Todorova, A.P.Seitsonen, J.Hutter, I.W.Kuo and C.J.Mundy, J.Phys.Chem.B. **110**, 3685(2006).

[13] M.Guidon, F.Schiffmann, J.Hutter, J.VandeVondele, J. Chem. Phys. **128**, 214104 (2008).

[14] X.Wu, A.Selloni and R.Car, Phys.Rev.B. **79**, 085102 (2009).

[15] A.Tkatchenko and M.Scheffler, Phys.Rev.Lett. **102**, 073005 (2009).

[16] F.Franks, *Water: a comprehensive treatis. Vol 1, The physics and physical chemistry of water* (Plenum Press, New York 1972).

[17] D.Kennedy and C.Norman, Science, **309**, 75,(2005).

[18] T.Head-Gordon and M.E.Johnson, Proc.Natl.Acad.Sci. **103**, 7973 (2006).

[19] S.Woutersen, U.Emmerichs, H.K.Nienhuys and H.J.Bakker, Phys.Rev.Lett. **81**,1105(1998).

[20] F.H.Stillinger, Science. **209**, 4455 (1980).

[21] P.Wernet *et al.* Science, **304**, 595(2004).

[22] G.A.Jeffrey, *Introduction to Hydrogen Bonding* (Oxford University Press, New York, 1997).

[23] A.D.Buckingham, J.E.D.Bene, and S.A.C.McDowell, Chem.PhysLett. **463**, 1-10 (2008).

[24] W.M.Latimer and W.H.Rodebush, J.Am.Chem.Soc. **42**, 1419(1920).

[25] See http://ssrl.slac.stanford.edu/nilssongroup/index.html

[26] A.Luzar and D.Chandler. Phys.Rev.Lett. **76**,928 (1996).

[27] B.Hetsenyi, F.D.Angelis, P.Giannozzi, and R.Car, J.Chem.Phys. **120**, 8632(2004).

[28] D.Prendergast and G.Galli, Phys.Rev.Lett. **96**, 215502 (2006).

[29] R.Kumar, J.R.Schmidt, and J.L.Skinner, J.Chem.Phys. **126**, 204107 (2007).

[30] G.Hura *et al.*, Phys.Chem.Chem.Phys. **5**, 1981 (2003).

[31] A.Soper, Chemical Physics, **258**, 121(2000).

[32] J.P.Hansen and I.R.McDonald, *Theory of simple liquids 2nd edition*(London:Academic)(1986).

[33] N. Marzari and D. Vanderbilt, Phys. Rev. B. **56**, 12847(1997).

[34] M.Krack, A.Gamirasio and M. Parrinello, J.Chem.Phys., **117**, 9409(2002).

[35] J.D.Bermal and R.H.Fowler, J.Chem.Phys. **1**, 515 (1933).

[36] A.H.Narten, M.D.Danford and H.A.Levy, *Oak Ridge National Laboratory Report ORNL-3997* (1966).

[37] A.H.Narten, M.D.Danford and H.A.Levy, Discuss.Faraday Soc, **43**, 97 (1967).

[38] A.Narten and H.Levy, J.Chem.Phys. **55**, 2263 (1971).

[39] A.Soper and M.Philips, Chemical Physics **107**, 47(1986).

[40] G.Hura and T.Head-Gordon, Chemical Review **102**, 2651(2002).

[41] A.Soper, F.Bruni and M.Ricci, J.Chem.Phys. **106**, 247(1997).

[42] G.Hura, J.Sorenson, R.Glaser and T.Head-Gordon, J.Chem.Phys. **113**, 9140 (2000).

[43] J.Sorenon, G.Hura, R.Glaser and T.Head-Gordon, J.Chem.Phys.**113**, 9149 (2000).

[44] R.McGreevy and L.Puszati, Mol.Simul. **1**, 359(1988).

[45] B.Santra, J.Klimes, D.Alfe, A.Tkatchenko, B.Slater, A.Michaelides, R.Car and M.Scheffler. Phys.Rev.Lett. **107**, 185701(2011).

[46] R.Car and M.Parrinello, Phys.Rev.Lett. **55**, 2471 (1985).

[47] R.Car, in *Conceptual Foundations of Materials: A Standard Model for Ground- and Excited-Stated Properties, Contemporay Concepts of Condensed Matter Science*, edited by S.G.Louie and M.L.Cohen(Elsevier, Amsterdam, 2006).

[48] H.J.C.Berendson, J.P.M.Postma, W.F.van Gunsteren and J.Hermans, in *Intermolecular Forces*, edited by B.Pullman (Reidel, Dordrecht, ADDRESS, 1981).

[49] F.Stillinger and A.Rahman, J.Chem.Phys. **60**, 1545(1974).

[50] P.A.M Dirac, Proc. Roy. Soc. A **123**, 714 (1929).

[51] R.P. Feynman, *The Feynman lectures on physics* , 3^{rd} printing (1969).

[52] P.Hohenberg and W.Kohn, Phys.Rev. **136**, B864 (1964).

[53] W.Kohn and L.J.Sham, Phys.Rev. **140** , A1133(1965).

[54] W.Kohn, Rev. Mod. Phys. **71** , 1253(1999).

[55] D.Ceperley and B.Alder, Phys.Rev.Lett. **45**, 566 (1980).

[56] R.M.Marin, *Electronic structure:Basic Theory and Practical Methods*(Cambridge University Press, Cambridge, 2004).

[57] A.Becke, Phys.Rev.A **38**, 3098(1988).

[58] A.D.Becke, J.Chem.Phys. **96** , 2155 (1992).

[59] A.D.Becke, J.Chem.Phys. **97** , 9173 (1992).

[60] J.P.Perdew and S.Kruth, 2003, in *A Primer in Density Functional Theory*, edited by C.Fiolhais, F.Nogueira and M.A.L.Marques(Springer, Berlin).

[61] E.Schwegler, J.C.Grossman, F.Gygi and G.Galli, J.Chem.Phys. **121**, 5400(2004).

[62] P. Giannozzi *et al.* J. Phys. Condens. Matter **21**, 395502 (2009).

[63] J.A.Morrone, *Nuclear quantum effects in water: A study in position and momentum space*, Princeton University, 2008.

[64] Private communicate with Prof.Anders Nilsson from Stanford university.

[65] A.J.Cohen, P.M.Sanchez and W.Yang, Science, **321**, 792 (2008).

[66] S. K*ü*mmel and L. Kronik, Rev.Mod.Phys. 80, 3 (2008).

[67] A.D.Becke, J.Chem.Phys. **98**, 1372(1993).

[68] J.Harris and R.O.Jones, J.Phys.F:Met.Phys. **4**, 1170 (1974).

[69] D.C.Langreth and J.P.Perdew, Phys.Rev.B **15**, 2884(1977).

[70] A.D.Becke, J.Chem.Phys. **98** , 1372 (1992).

[71] C.Adamo and V.Barone, Chem.Phys.Lett. **274**,242(1997).

[72] C.Adamo and V.Barone, Chem.Phys.Lett. **298**,113(1998).

[73] C.Zhang, D.Donadio, F.Gygi and G. Galli, J.Chem.Theory Comput. **7** , 1443(2011).

[74] J.Heyd and G.E.Scuseria, J.Chem.Phys. **118**, 8207 (2003).

[75] R.H.French *et al.* Rev.Mod.Phys. **82**,1887(2010).

[76] M.Dion, H.Rydberg, E.Schröder, D.C.Langreth and B.J.Lundqvist, Phys.Rev.Lett. **92**, 246401 (2004).

[77] T.Thonhauer, V.R.Cooper, S.Li, A.Puzder, P.Hyldgaard and D.C.Langreth, Phys.Rev.B **76**, 125112(2007).

[78] J.Kleis, B.I.Lundqvist, D.C.Langreth and E.Schröder, Phys.Rev.B **76**,100201 (2007).

[79] J.Wang, G.R.Perez, J.M.Soler, E.Artacho and M.V.F.Serra, J.Chem.Phys. **134**, 024516 (2011).

[80] W.Heitler and F.London. Z.Phys. **44**, 455 (1927).

[81] S. Grimme, J. Comput. Chem. **25**, 1463 (2004).

[82] R.Johchiere, A.P.Seitsonen, G.Ferlat, A.M.Saitta and R. Vuilleumier, J.Chem.Phys. **135**,154503(2011).

[83] S.Grimme, J.Antony, S.Ehrlich and H.Krieg, J. Chem. Phys. **132**, 154104 (2010).

[84] P.Jurecka, J.Sopner, J.Cerny, *et al.*, Phys.Chem Chem Phys, **8**, 1985(2006).

[85] B.Santra, A.Michaelides, M.Fuchs, *et al.*, J.Chem.Phys. **129**, 194111(2008).

[86] D.C.Langreth and J.P.Perdew, Phys.Rev.B., **15**, 2884(1977).

[87] D.Lu, F.Gygi and G.Galli, Phys.Rev.Lett. **100**, 147601 (2008).

[88] Y.Li, D.Lu, H-V.Nguyen and G.Galli, J.Phys.Chem A, **114**, 1944(2010).

[89] M.P.Allen and D.J.Tildesley, *Computer simulation of Liquids*, (Claredon Press, Oxford, 1991).

[90] D.Fenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications.* (Academic Press, New York, 2002).

[91] M.Sharma, Y.Wu and R.Car, Int.J.Quantum.Chem. **95**,821(2003).

[92] Lynn Elliot Cannon, *A cellular computer to implement the Kalman Filter Algorithm, Technical report, Ph.D. Thesis, Montana State University*, 14 July 1969.

[93] J.P.Perdew, M.Ernzerhof and K.Burke, J.Chem.Phys. **105**,9982(1996).

[94] L.X.He and D.Vanderbilt, Phys.Rev.Lett. **86**, 5341(2001).

[95] W.H.Press, S.A.Teukolsky, W.T.Vetterling and B.P.Flanery, *Numerical Recipes*(Cambridge University Press, Cambridge, 1992).

[96] R.A.DiStasio, Z.Li and R.Car, to be published(2012).

[97] See http://www.nersc.gov/ .

[98] Within pseupotential framework, one water molecule has 8 valence electrons: 2 from two hydrogen atoms an 6 from one oxygen atom. Totally there are 512 valence electrons, corresponding to 256 electron orbitals due to the spin degeneracy.

[99] C.Bekas , A.Curioni, W. Andreoni, *New Scalability Frontiers in Electronic Structure Calculations Using the BG/L Supercomputer*, In PARA 1026-1035(2006).

[100] O.Mishima, H.E.Stanley, Nature **396**, 329(1998).

[101] M.Ceriotti, G.Bussi and M.Parrinello, Phys.Rev.Lett. **103**, 030603(2009).

[102] C.Huang *et al.*, Phys.Chem.Chem.Phys., **13**, 19997(2011).

[103] A.K.Soper and C.J.Benmore, Phys.Rev.Lett. **101**, 065502(2008).

[104] A.K.Soper, Chem.Phys. **202**, 295(1996).

[105] L.Lin, J.A.Morrone, R.Car and M.Parrinello, Phys.Rev.B, **83**, 220302(2011).

[106] T.Head-Gordon and F.H.Stillinger, J.Chem.Phys. **98**, 3313(1993).

[107] G.C.Lie, E.Clementi and M.Yoshimine, J.Chem.Phys. **64**, 2314(1976).

[108] A.K.Soper, F.Brune and M.A.Ricci, J.Chem.PHys. **106**, 247(1997).

[109] Y.E.Gorbaty and Y.N.Demianets, Mol.Phys. **55**, 571(1985).

[110] B.Dunweg and K.Kremer, J.Chem.Phys., **99**, 6983(1993).

[111] I.C.Yeh and G.Hummer, J.Phys.Chem.B., **108**, 15873(2004).

[112] P.L.Chau and A.J.Hardwick, Mol.Phys. **93**, 511 (1998).

[113] J.R.Errington and P.G.Debenedetti, Nature (London) **409**, 318 (2001).

[114] O.Mishima, L.D.Calvert, E.Whalley, Nature **310**, 393(1984).

[115] A.K.Soper and M.A.Ricci, Phys.Rev.Lett.,**84**, 2881(2000).

[116] E.Shiratani and M.Sasai, J.Chem.Phys., **104**, 7671(1996).

[117] E.Shiratani and M.Sasai, J.Chem.Phys., **108**, 3264(1998).

[118] S.R.Accordino, J.A.Rodriguez Friz, F.Sciortino and G.A.Appignanesi, Eur. Phys. J.E, **34**, 48(2011).

[119] S.R.Accordino, J.A.Rodriguez Friz, F.Sciortino and G.A.Appignanesi, Eur. Phys. J.E, **29**, 305(2009)

[120] K.T.Wikfeldt, A.Nilsson and L.G.M.Pettersson, Phys.Chem.Chem.Phys.,**13**, 19918(2011).

[121] Parr,R.G. and W.Yang, *Density Functional Theory of Atoms and Molecules*(Oxford University Press,Oxford)(1989).

[122] P. Geerlings, F.de Proft and W.Langenaeker, Chem.Rev. **103**, 1793 (2003).

[123] R.M.Dreizler and E.K.U.Gross Springer,Berlin,(1990).

[124] S.Nose. J.Chem.Phys. **81**, 511(1984).

[125] S.Nose. Prog.Theor.Phys.Suppl. **103**, 1(1991).

[126] W.G.Hoover, Phys.Rev.A. **31**, 1695(1985).

[127] M.Tuckerman, B.J.Berne and G.Martyna, J.Chem.Phys. **97** , 1990(1992).

[128] G.J.Martyna, M.L.Klein and M.Tuckerman, J.Chem.Phys. **97**, 2635(1992).

[129] B.Santra, A.Michaelides and M.Scheffler, J.Chem.Phys. **127**, 184104 (2007).

[130] C.Moller and M.S.Plesset, Phys.Rev. **46**, 618 (1934).

[131] X.Xu and W.A.Goddard, J.Phys.Chem.A **108**, 2305(2004).