

MODELING, SIMULATION AND ANALYSIS OF THE REACTION FIELD FOR
ELECTROSTATIC INTERACTIONS IN AQUEOUS SOLUTION

by

Wei Song

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2013

Approved by:

Dr. Donald J. Jacobs

Dr. Dennis R. Livesay

Dr. Jun-Tao Guo

Dr. Irina Nesmelova

Dr. Shao-Zhong Deng

©2013
Wei Song
ALL RIGHTS RESERVED

ABSTRACT

WEI SONG. Modeling, simulation and analysis of the reaction field for electrostatic interactions in aqueous solution. (Under the direction of DR.DONALD J. JACOBS)

How to deal with the long-range electrostatic interactions theoretically and computationally has been well studied due to their importance in biological processes and time consuming summations in computer simulations. The main focus of our research has been on the design and application of a new type of hybrid model that combines both the explicit and implicit solvent models using a reaction field (RF) approach, for accurate and efficient electrostatic calculations. This hybrid model, named as Image Charge Solvation Model (ICSM), replaces an infinite Coulomb summation by two finite sums over direct interactions plus image charges for RF. To characterize the ICSM, the electrostatic torques and forces using different model parameters are compared through various histogram distributions. The contributions of RF are 20% and 2% of the total electrostatic torques and forces, respectively, suggesting that the main effect of RF is to maintain the orientation of water dipoles in the solution. Considering systematic artifacts of the discontinuous dielectric constant at the edge of the cavity, we modified the image charge formula in an optimal way to better account for the continuously changing dielectric profile near the boundary, which provides a computational procedure to determine the most accurate RF possible for a specified water model. The Periodic Boundary Conditions (PBC) in ICSM reduces the size of the productive region and introduces unphysical correlations between ions in ionic solution. With combination of finite boundary conditions, mean field theory for short-range forces and multiple constraint forces applied to water molecules in a buffer layer, bulk water properties are maintained

without problems from imaged ions in a much bigger usable region than before. To summarize, the results presented in this work provide a complete characterization, optimization and improvement of the ICSM for electrostatic calculations.

DEDICATION

For my parents and teachers who introduced me into science.

ACKNOWLEDGMENTS

I would like to give my greatest admiration and appreciation to Dr. Jacobs for his most valuable guidance and extreme patience on my research in the past five years. He is very clever, fast thinking and so knowledgeable scientist in mathematical, physical and biological fields, who not only teaches me how to do scientific research but also gives me instructions in daily life as a foreign student. I sincerely thank my committee members Dr. Dennis Livesay, Dr. Jun-Tao Guo, Dr. Shao-Zhong Deng, Dr. Irina Nesmelova, and Dr. Andrij Baumketner, for their time and kindly guidance both on classes and researches. I would like to thank the Department of Bioinformatics and Genomics, Department of Physics and Optical Science, and Graduate School in UNC Charlotte for providing me financial support in these years. I would also like to express my appreciation to all the staff of ISSO for their friendly help. Finally, I would like to give my loves and thanks to my family members, my parents and my fiancé Miss Shan Li, for their unconditional love and support on me, who bring more happiness and hopes in my life.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1 Introduction to Electrostatic Interactions	1
1.2 Electrostatic Interactions in Computer Simulations	3
1.3 Solvent Model and Boundary Conditions	5
1.4 Evaluation of Electrostatic Interactions	8
1.5 Hybrid Model for Electrostatic Calculations	9
1.6 Dissertation Objectives and Layout	11
CHAPTER 2: IMAGE CHARGE SOLVATION MODEL (ICSM)	14
2.1 Reaction Field and Multiple Image Charge Method	14
2.2 Image Charge Solvation Model (ICSM)	18
2.3 Boundary Conditions for Water Molecules	23
CHAPTER 3: CHARACTERIZATION OF REACTION FIELD	26
3.1 Introduction	26
3.2 Methods	27
3.3 Results and Discussions	29
3.4 Conclusions	45
CHAPTER 4: OPTIMIZATION OF IMAGE CHARGES	47
4.1 Introduction	47
4.2 Methods	48
4.3 Results and Discussions	59
4.4 Conclusions	80
CHAPTER 5: OPTIMIZATION OF BOUNDARY CONDITIONS IN ICSM	81

	viii
5.1 Introduction	81
5.2 Methods	84
5.2.1 Computational Details	84
5.2.2 Model Details	85
5.3 Results and Discussions	102
5.4 Conclusions	111
CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS	114
REFERENCES	118

CHAPTER 1: INTRODUCTION

1.1 Introduction to Electrostatic Interactions

Electrostatics, which accurately deals with the interactions of electric charges that have slow motions, underpins many important phenomena that we experience in our daily lives ranging from lightening during thunderstorms to the effects of electrolytes in our bodies. Electrostatic forces can be described by a simple formula called Coulomb's law, which was empirically discovered by Charles Augustin de Coulomb in 1785. Consider a system with two point charges as shown in Fig. 1.1 A. The force between these two charges is along the straight line that connects them, given by Coulomb's law:

$$F = \frac{1}{4\pi\epsilon_0\epsilon} \frac{Q_1Q_2}{R_{12}^2} \quad (\text{Eq. 1.1})$$

where ϵ_0 is the dielectric permittivity in free space, ϵ is the relative permittivity of the material containing charges, π is the circumference ratio, Q_1 and Q_2 are the values of charge, and R_{12} is the distance between the two charges. Fig 1.1 B shows the Coulomb force follows the $F \propto 1/R^2$ rule.

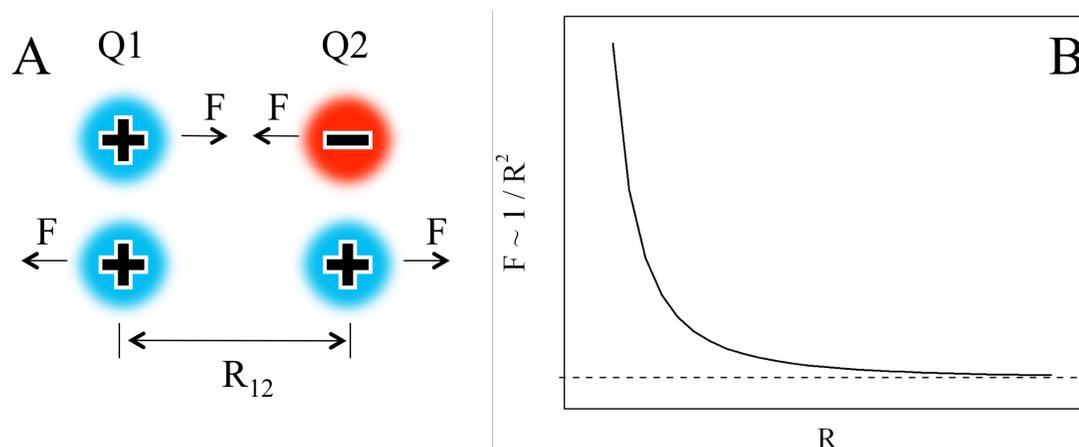


Figure 1.1: A) Two charges of opposite sign have attractive forces on each other and two charges of same sign have repulsive forces. B) Schematic plot for the Coulomb force verses the distance between two charges.

In life science, the electrostatic interactions play a critical role in biological processes. Here we take the hydrogen bond (H-bond) as an example. The electrostatic forces between polar molecules that involve the partial charges assigned to a donor atom, its hydrogen atom and an acceptor atom taken together model the H-bond. The H-bond is important in forming and stabilizing the structures of protein and DNA as shown in Fig. 1.2. Other important roles of electrostatic interactions in biological systems include signal transmission, ion-transport, molecular recognition, structure, stability and function of biopolymers such as DNA, RNA and proteins. Underlying all these illustrative examples is the importance of water, which is an essential substance that mediates biological processes. Moreover, electrostatic interactions are critical in determining the rich behavior that bulk water exhibits.

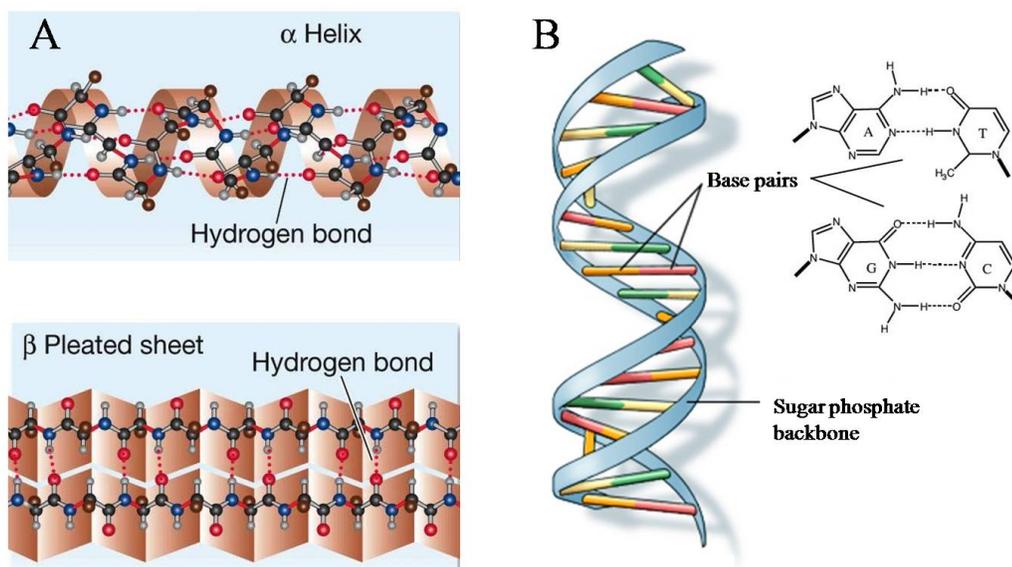


Figure 1.2: A) The hydrogen bonds formed between amino acid residues in α -helix and β -sheets help to stable the secondary structures in protein conformation. B) The hydrogen bonds between the A-T and C-G base pairs stabilize the double helix structure of DNA.

1.2 Electrostatic Interactions in Computer Simulations

The explosive development and continuing growth of the power of super computers make it possible to simulate and analyze biomolecular systems. Simulation has become an essential method following the traditional theoretical and experimental approaches in scientific research. The common indispensable tools in computational biology are Molecular Dynamics (MD) and Monte Carlo (MC) methods that provide ways for detailed studies on biopolymers to characterize their properties and to help understand how they perform their biological functions. Better algorithms, faster distributed computing and more user-friendly software have been applied to multi-component solvent (such as pure water, ionic solvent and other polar or non-polar solvent) simulations, which are normally encountered in experiments [1].

To model the biological processes precisely in simulations, a force field for biomolecules and solvent has to be developed to model all possible interactions in the system, which consists of bonded (covalent interactions such as bond-stretching, bond angle bending, bond twisting or dihedral angle torsion) and non-bonded interactions (such as van der Waals, electrostatic). A force field is calculated by:

$$\begin{aligned}
 V = & \sum_{bonds} \frac{K_l}{2} (l - l_0)^2 + \sum_{angles} \frac{K_\theta}{2} (\theta - \theta_0)^2 \\
 & + \sum_{torsions} \frac{V_n}{2} \{1 - \cos(n\omega - \gamma)\} \\
 & + \sum_i \sum_{j>i} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
 & + \sum_i \sum_{j>i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}
 \end{aligned} \tag{Eq. 1.2}$$

The last two non-bonded interactions are nominally computationally expensive because they are $O(N^2)$ summation with $N(N-1)/2$ pairs, which is prohibitive for large N . The van der Waals force is short-ranged, falling off to a negligible level after a distance of about 1.5 nm. However, the long-range electrostatic interactions decrease slowly as the distance r between molecules increases. The electrostatic energy between two charged molecules is proportional to r^{-1} , for a neutral molecule with a dipole moment and a charged molecule is proportional to r^{-2} , for two neutral molecules with dipole moments is proportional to r^{-3} . Two neutral molecules with quadrupole moments is proportional to r^{-5} , which is still longer ranged than van der Waals interactions that come from fluctuating induced dipole-dipole interactions (proportional to r^{-6}) [2]. For the electrostatic energy of a single charge with all other charges and multipoles surrounding it, the electrostatic interactions should be summed in full without applying cutoff distances. The results of simulations do vary according to how the electrostatic energy and interactions are

approximated, and simple cutoffs tend not to work well for high accuracy simulations [3, 4]. Therefore, in this thesis, a sum over all pairs of charges in the system is used.

1.3 Solvent Model and Boundary Conditions

The biggest part of the electrostatic calculations comes from the solvent molecules around the solutes and there are several main types of solvents commonly used in simulations, as shown in Fig. 1.3. The first approach is all-atom explicit solvent model, which considers the atomic details of all the atoms for both solute and solvent. Considering all the detailed information makes the explicit solvent model accurate but costs large amount of computing time to simulate large systems on time scales that are biologically relevant. Another approach is called coarse-grained (CG) modeling. One way is to use one or more beads to represent the solute and solvent molecules, which reduces the total numbers of degrees of freedom by eliminating fine details about inter-particle interactions. The simulation with CG can be substantially faster than the all-atom representation. Errors might be introduced due to the lack of complete molecular information. Modeling solvent implicitly as a dielectric continuum reduces degrees of freedom further. Modeling solute in atomic detail and solvent implicitly makes the simulation runs much faster. However, the molecular interactions near the implicit solvent interface are poorly modeled with strong surface effects, and these interactions are often critical to biological function.

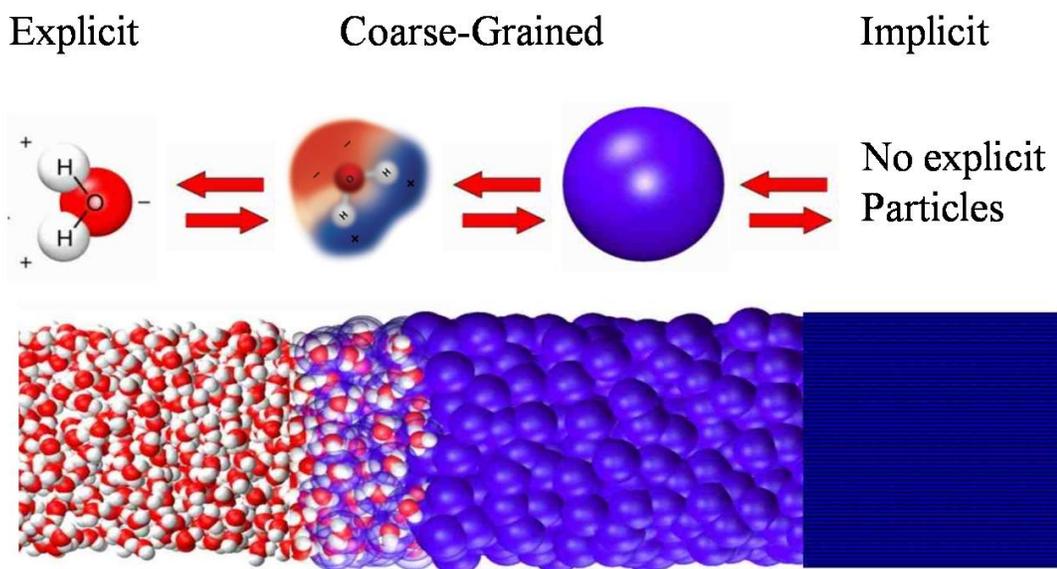


Figure 1.3: Three types of solvent models using water molecule as an example. Picture from the work of Delgado-Buscalioni *et al.*[5].

How to treat the infinitely long-ranged interactions in a finite size system is a long-standing problem. Based on different types of solvent models, there are many approaches for correct treatment of the electrostatic interactions in simulation. Historically, when there was much more limited computing power, the long-range electrostatic forces were truncated by cutoffs, which introduced severe artifacts in simulating peptides and nucleic acids [3, 4]. Because cutoffs do not work well, the finite size effects from the long-range electrostatic interactions must be treated carefully through boundary conditions.

Two predominant kinds of boundaries are periodic boundary conditions (PBC) and non-periodic boundary conditions (NPBC) as shown in Fig. 1.4. The periodic boundary conditions replicate the unit simulation box to completely fill the space, thereby reducing surface effects. The PBC removes any identifiable boundary of the system and yields essentially exact bulk properties of solutions. However, artifacts due to correlations caused by imaging the unit cell occur due to the long-range nature of electrostatic forces

when the box size is small. The effects of periodicity-induced artifacts is not fully understood at present [6-8]. A simple way to avoid this problem is to generously increase the unit cell size, but this would not follow common practice because doing so will severely increase the computational cost, perhaps to a point that is intractable. Artifacts are likely to appear using PBC whenever any of the following three aspects occur; (i) the solvent has low dielectric permittivity; (ii) the solute has non-negligible size compared to the size of the simulation box, and, (iii) the system has a net charge [9].

The non-periodic boundary conditions (NPBC) use a continuum dielectric medium start from dielectric cavity R_{RF} to infinity and Poisson-Boltzmann (PB) equations are solved to model the effects of solvent on solute [10-13]. The NPBC has no periodic artifacts but introduces the surface effects at the boundary between two different solvent models. The solution of three-dimensional differential equation for arbitrary shape is still a challenging computational problem. However, the spherical shape makes the mathematical problem completely tractable and computationally efficient, as will be described below in detail.

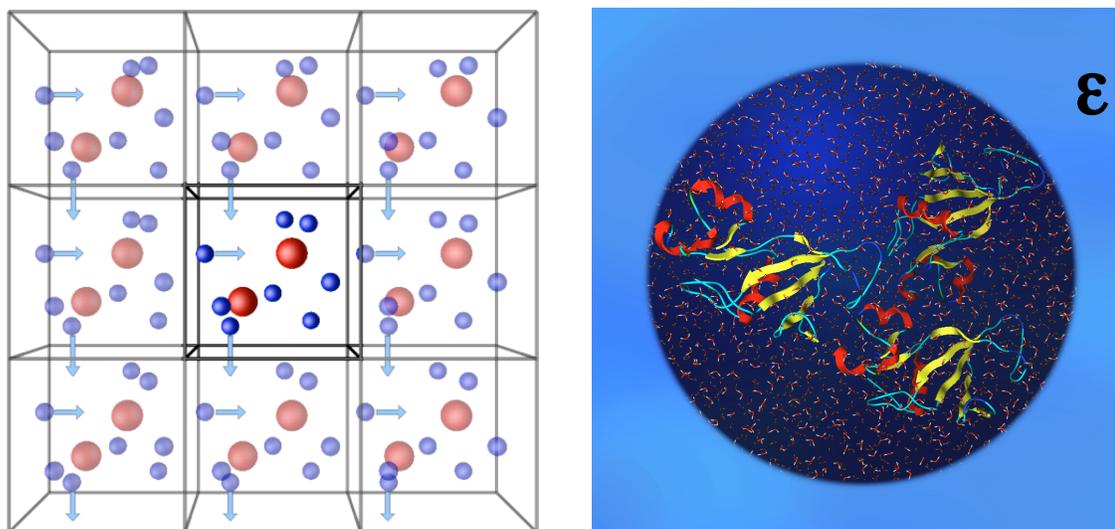


Figure 1.4: Periodic and non-periodic boundary conditions.

1.4 Evaluation of Electrostatic Interactions

Two main kinds of techniques have been developed to evaluate the long-range electrostatic interactions. The first type of technique is lattice summation, including the Ewald summation [14, 15], Particle Mesh Ewald (PME) [16, 17] and particle-particle particle-mesh [18, 19] techniques. All these methods rely on PBC, and include in the sum the electrostatic contributions from all partial charges in the system, together with all of their periodic images (an infinite number of them). While the Ewald summation is the most accurate of all these methods, it performs as $O(N^{3/2})$, which gives a windfall in efficiency compared to the original $O(N^2)$ problem. Unfortunately, $O(N^{3/2})$ is insufficient to conduct large-scale molecular dynamics simulations. The other methods split the electrostatic interactions into two parts, a short-range term that is calculated exactly in a direct sum, plus a smoothly varying long-range term that is handled approximately in a reciprocal sum by fast Fourier transform methods with a performance characteristic of $O(N \log N)$. These methods are commonly used, and have been proven to be useful, despite the potential pitfalls of using too small unit cell that is not justifiable. The second type of technique is the Fast Multipole Method (FMM) [20-22], which treats the electrostatic interactions exactly for the particles in the same simulation box or the nearby subcells, and calculates the potential of distant charges using multipolar expansions. FMM is useful when employing NPBC. Based on certain hierarchical tree construction, the initial simulation box is divided into self-similar subcells. Take the cubic box for example, the initial level zero ($l=0$) is only one box, the level one ($l=1$) has 8 subcells and level two has 64 subcells. The refinement process stops at $\log_3 N$ number of levels. For each cell i in level l , a multipole expansion is performed about the cell center to represent the

electrostatic field produced by the particles in that cell [23]. The FMM performs as $O(N)$ and can be used for both NPBC and PBC [24].

1.5 Hybrid Model for Electrostatic Calculations

Considering the advantages of the implicit and explicit approaches, combining the elements from both methods into one might be a good choice to design a better hybrid solvation model [12, 25], as shown in Fig.1.5. As I will show below numerous disadvantages emerge in a hybrid model that has to do with surface effects at the boundary between explicit and implicit representations of solvent. It suffices to say at this point that the disadvantages were first characterized and then surgically removed by designing the model carefully. A hybrid model considers all molecules inside the cavity explicitly, containing all atomic interactions among the solute and solvent molecules. The cavity is then embedded in a dielectric continuum which is represented by implicit solvent [26]. A buffer layer of explicit water molecules is inserted between the two solvents but experience different forces from those in the central cavity. The electrostatic forces on the explicit solvent molecules inside the cavity consist of two parts. First, there is a direct Coulomb's force between molecules. Second, a reaction field (RF) is created. The RF consists of the forces generated by the polarization of the continuum medium, which is induced by explicit charges within the spherical cavity. Differences among various hybrid models distill down to how they calculate the RF. A brute force way is to numerically solve the PB equation directly [27] or using an approximate theory for the PB equation [28, 29], although arbitrary geometries of the cavity will require high computational costs. Another approach [30] is to develop exact solutions for a certain convenient geometrical setup, such as for a plane or sphere [31].

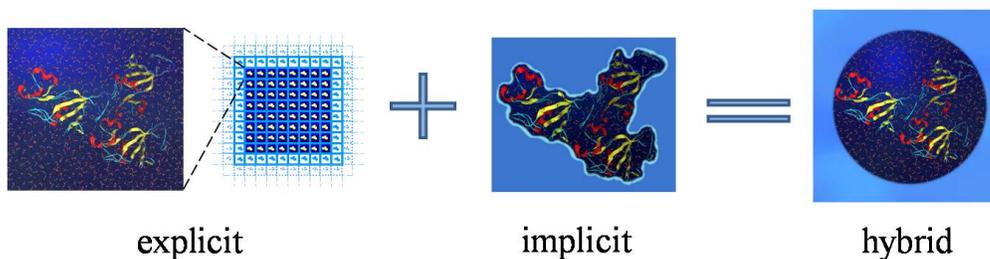


Figure 1.5: Different schemes for explicit, implicit and hybrid models. Explicit: All-atom model which use periodic boundary conditions, accurate but slow. Implicit: Continuum solvent model that treats solvent as a dielectric medium, faster but less accurate due to loss of atomic information. Hybrid: The protein and solvent inside the sphere are treated explicit. Outside is treated as a continuum dielectric medium.

For hybrid models [32-34] that invoke a spherical cavity, a series expansion that was first developed by Kirkwood [26, 35] can be used for the RF. Although arbitrary accuracy can be achieved by this expansion, it is slow to converge, especially near the cavity boundary. An image charge approach can be used alternatively [36]. Friedman [37] was first to apply an image charge method for the solvation problem. In his work, only one image charge for each source charge in a cavity is used to construct the RF, reaching the accuracy of $O(1/\epsilon)$ when the dielectric constant ϵ is high. Later a more accurate approximation was implemented by Abagyan and Totrov [15] that involved the classical Kelvin image charge for a perfect conductor [38]. The weaknesses of current hybrid models are 1) the accuracy of the Friedman expression loses accuracy as the dielectric constant decreases, and, 2) surface effects were not considered, which means non-negligible artifacts are plaguing simulations. The hybrid model I have been working on is called *Image Charge Solvation Model* (ICSM) [39] where the RF can be calculated to arbitrary high accuracy using the multiple image charge method [40-42]. The model is successful in dealing with pure water and extreme dilute ionic solutions [9, 39]. However,

all computational models/methods have advantages and disadvantages, and my interests are to maximize the former and minimize the latter.

1.6 Dissertation Objectives and Layout

The main objective of this dissertation is to characterize the ICSM and define its weaknesses, and based on that, optimize and improve the model for better performance. Performance means maintain accuracy, maximize the volume of the central part of the spherical cavity that can be used to collect data (called the productive region), and to get the results in the shortest possible time by using very efficient algorithms like FMM. To demonstrate this objective clearly, we need to answer the following questions: a) Why is the reaction field essential for correct dielectric properties? b) How does the imaged water in the buffer layer protect the productive region? c) Will the productive region be affected by varying the accuracy in the RF, or with different boundary conditions? d) How can we optimize the RF to maintain accuracy in the productive region, but not spend more time than needed to have greater accuracy than needed? e) Can we replace the discontinuous dielectric profile with a continuous one? f) Is it possible to increase the productive region by reducing the buffer layer using an optimized RF? g) How to avoid the unphysical correlation between ions and their periodic images in ICSM due to PBC? h) How to account for the artifacts of the discontinuous boundary conditions in the case of NPBC?

Because the ICSM used a multiple image charge method for the RF and used PBC for short-range forces inside a spherical cavity with a buffer layer for long-range forces, which are creative and novel elements to modeling, the objective of my first project is to characterize the ICSM to further explore the key role of RF, the function of buffer layer

and the effects of molecular rules for water. Based on this brand new ICSM when I started my PhD studies, various histogram distributions of electrostatic forces and torques on water molecules throughout the cavity was used to assess the model by quantifying relative contributions from the RF to total forces and torques. The results suggest that the main effect of the RF is to affect the orientation of water dipoles. The role of buffer layer is essential to reduce surface effects and to achieve high accuracy in ICSM by absorbing artifacts caused by the discontinuous boundary at the cavity wall in the non-productive region. Boundary conditions classified as: atom-, group- and ODL-based rules have little influence on the performance of ICSM. This work has been published in *Communication in Computational Physics* [43].

Consequently, a buffer layer is needed to absorb errors due to the discontinuity of the dielectric constant in ICSM. The objective of the second project is to reduce the artifacts by generating a different type of RF using optimized image charges, so that the thickness of the buffer layer can be decreased. The idea is to modify the image charge formula to better account for the a priori unknown continuously changing dielectric profile near the boundary. A novel inverse method was developed, and I showed that a notable improvement between the RF using optimal image charges to the PME results was obtained. Moreover, with this new method, the dielectric permittivity is self-consistently determined based on the set of optimized image charges. A pipeline of computationally determining the most accurate RF possible for a specified water model has been constructed. Although the electrostatic interactions have been improved, the short-range forces prevented the buffer layer to be reduced in size, which means there was no performance gain.

The ICSM experiences low efficiency and unphysical electrostatic interactions when it is used to simulate high ionic concentration solutions. The objective of my third project is to avoid the periodicity-induced problem. The PBC in ICSM is replaced by NPBC with multiple constraint forces and thermostats introduced into the buffer layer. We employed a novel mean field approach to determine the van der Waal interactions from continuum dielectric medium. Not only could the unphysical correlations due to imaged ions be removed, a much larger usable region than the original ICSM is achieved with correct bulk water properties. The manuscript of this work is under preparation and will be submitted to Journal of Chemical Physics B.

CHAPTER 2: IMAGE CHARGE SOLVATION MODEL (ICSM)

2.1 Reaction Field and Multiple Image Charge Methods

In the ICSM, a spherical cavity of radius R_c contains a system of molecules that are simulated. The interactions among the molecules inside the cavity are treated explicitly at the all-atom level. Outside the cavity the solvent is treated as a homogeneous dielectric medium of uniform dielectric constant ϵ_s , as shown in Fig. 2.1. The explicit charges in the cavity induce the polarization in the medium, which in turn create a RF back on the solvent inside. The electrostatic field due to the surrounding dielectric is given by:

$$\bar{E}_i = \frac{2(\epsilon_s - 1)}{2\epsilon_s + 1} \left(\frac{1}{r_c^3} \right) \sum_{j; r_{ij} \leq r_c} \bar{\mu}_j \quad (\text{Eq. 2. 1})$$

where μ_j are the dipoles of the neighboring molecules that are within the cutoff distance r_c of the molecule i . The interaction between the molecule i , and the RF, equals to $\bar{E}_i \cdot \bar{\mu}_i$. One of the problems of the RF is the discontinuities in the energy or force when the molecules enter or leave the cutoff radius r_c . This can be avoided by employing a switching function for molecules near the RF boundary. Another problem is the dielectric constant should be known a priori. However, most of the time the dynamic properties are insensitive to the ϵ_s . Instead of a spherical boundary, more complicated shape of boundary can be applied for better approximation of the true molecular surface.

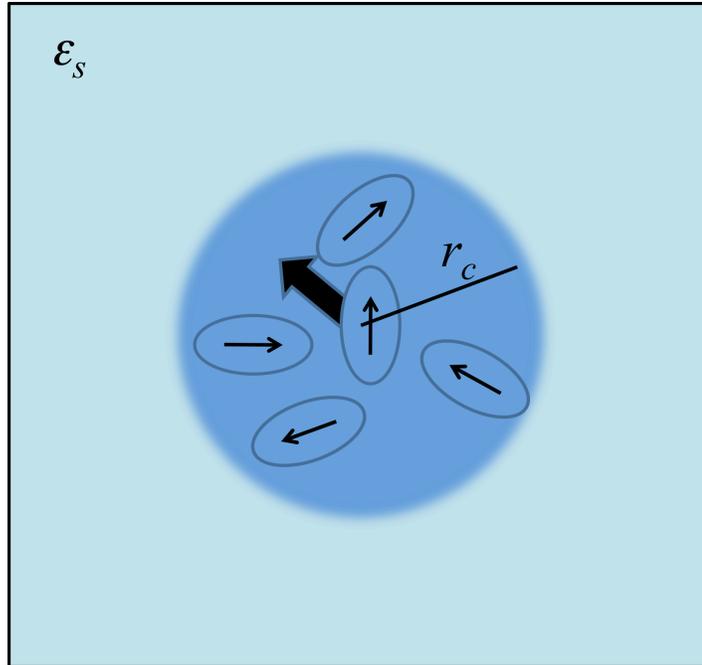


Figure 2.1: The RF method. The narrow arrows are the dipole moment in the spherical cavity of cutoff radius r_c . The shaded arrow represents the sum of all dipoles in r_c

There are many ways to calculate the RF within the spherical boundary. Common methods solve the linearized Poisson-Boltzmann (LPB) equation [27], employ approximate theories [28, 29], apply an exact series expansion [26, 35] for spherical cavities, or use a single image charge method [37]. The image charge method uses a spherical boundary and the RF due to the explicit charge inside cavity is generated by the so-called image charge located in the continuous dielectric medium beyond the sphere [37]. If the position of the explicit charge q_i is \vec{r}_i , then the location and magnitude of the image charge are:

$$\begin{aligned} r_{im} &= (R/r_i)^2 \vec{r}_i \\ q_{im} &= -\frac{\epsilon_s - \epsilon_r}{\epsilon_s + \epsilon_r} \frac{q_i R}{r_i} \end{aligned} \quad (\text{Eq. 2.2})$$

where R is the radius of the boundary sphere, ϵ_s and ϵ_r are the dielectric constants of the explicit cavity and the continuous dielectric medium, respectively. The expression is

accurate if the dielectric constant beyond the boundary is much larger than that inside ($\epsilon_s \gg \epsilon_r$). A problem of this method is that as the explicit charge approaches the boundary, so does its image charge with opposite sign and a singularity will occur.

In this dissertation, a more complete multiple image charge method [40, 44] has been used to calculate the RF. The method is based on the work of Neumann [45], who gave the electrostatic potential in terms of Kelvin image charge plus a line charge that falls off as a power law, starting from the Kelvin charge and extending to infinity [40, 44]. However, this method is not useful until it is converted into a numerical algorithm for calculating the potential field. The expression [40] below uses Gauss-Radau quadrature to reproduce the RF of line image charges by constructing a set of discrete image charges along the line, which makes the numerical calculate more efficient. Multiple image charges placed in the continuum medium are used to calculate the RF inside the sphere.

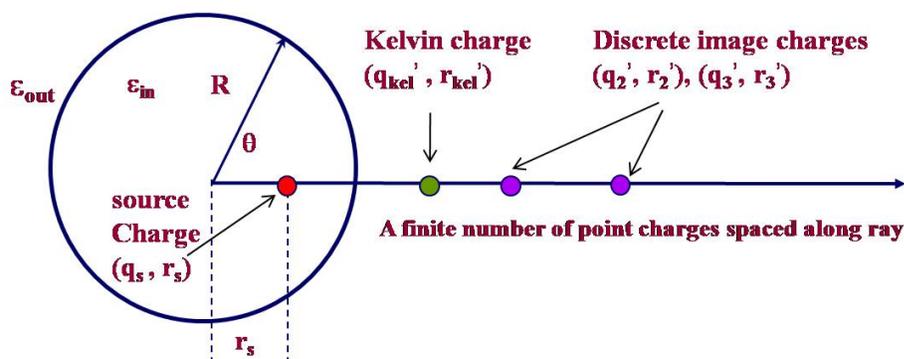


Figure 2.2: An illustration of how the multiple-image method is applied to compute RF in a spherical cavity with dielectric constant of ϵ_{in} embedded in the solvent bath with constant of ϵ_{out} . The polarization of the solvent inside the cavity by the source charge q at position r_s results in the RF $\Phi_{RF}(r, \theta)$ that is approximated by the potential created by auxiliary charges, referred to as image charges, $q_{Kel}, q_i, i \geq 2$, located at positions $r_{Kel}, r_i, i \geq 2$, respectively.

Then the total electrostatic field inside the cavity will consist of two parts: one from the coulomb's law among the source charges (Φ_{source}) and one from the image charge outside the sphere (Φ_{RF}).

$$\Phi_{total} = \Phi_{source} + \Phi_{RF} \quad (\text{Eq. 2.3})$$

As shown in Fig. 2.2, suppose there is one source charge (q_s, r_s) located inside the spherical cavity with dielectric constant ϵ_{in} , which is embedded in the infinite solvent with dielectric constant ϵ_{out} , it will generate an image charge outside the sphere along the radial direction, called the Kelvin image charge. Increasing accuracy can be obtained by systematically increasing the number of image charges, which represent the line charge of continuous charge density for each source charge. As further discussed later, small finite number of image charges (1 to 3) can be placed along a ray in the radial direction outside the sphere of radius R [40-42] to obtain more than enough accuracy to support MD studies.

The numerical approximation for the electrostatic potential of the RF is given in Eq. 2.4. Notice the superposition of M point image charges has a physically intuitive form, where the locations along the ray and corresponding charge for each image are given in Eq. 2.5.

$$\Phi_{RF}(\vec{r}) \approx \frac{q'_K}{4\pi\epsilon_i|\vec{r} - \vec{r}_K|} + \sum_{m=2}^M \frac{q'_m}{4\pi\epsilon_i|\vec{r} - \vec{r}_m|} \quad (\text{Eq. 2.4})$$

$$r_m = r_K \left(\frac{2}{1 - s_m} \right)^{1 + \frac{\epsilon_i}{\epsilon_o}}, \quad q'_K = \left(1 + \frac{\omega_1 \epsilon_i}{2\epsilon_o} \right) q_K, \quad q'_m = \frac{\gamma \epsilon_i}{2\epsilon_o} \frac{\omega_m r_m}{R} q \quad (\text{Eq. 2.5})$$

Here, $r_K = R^2/r_s$ is the distance from the origin to the Kelvin image of charge $q_K = \gamma R q / r_s$ when the source charge is r_s distance from the origin and has charge q . The

discontinuous mismatch in dielectric constants is reflected in the scale factor, $\gamma = (\epsilon_i - \epsilon_o) / (\epsilon_i + \epsilon_o)$. By employing the Gauss–Radau quadrature, where $\{s_m, \omega_m\}$ for $1 \leq m \leq M$ are the locations and weights [40] the Kelvin image charge is modified, which is denoted as q'_K . Higher accuracy can be achieved by introducing more image charges. If $\epsilon_{out} \rightarrow 0$, $q'_{kel} = q_{kel}$ at the location r'_{kel} , is same as the classical Kelvin image for a conductor in the electrostatic theory. The image charge formula is applied to all source charges within the sphere, and the total electrostatic potential or forces are calculated based on the linear supposition principle.

The advantages of the RF and image charge methods are conceptually simple, relatively easy to implement and computational efficient. They assume that the molecules beyond the cutoff sphere can be modeled as a continuum dielectric, but this is a reasonable assumption for homogeneous fluids. The dielectric constant of the surrounding continuum should also be specified, which can be taken from experimental data.

2.2 Image Charge Solvation Model (ICSM)

The ICSM is a hybrid model to deal with electrostatic calculations accurately and efficiently. It has been constructed by many researchers over the last 5 years at UNC Charlotte, and many of the technical details have been published [39-42]. Dr. Cai and Dr. Deng worked out the mathematical underpinnings for the image charge method and the adaptive FMM for all the charges in the system (images and source charges). Dr. Baumketner and Dr. Jacobs worked on the simulation details, and Dr. Baumketner developed the first prototype of the ICSM and wrote the original code. Under the guidance of Dr. Baumketner, Dr. Lin implemented the geometry of the simulation cell

using a Truncated Octahedron with periodic boundary conditions (referred to as TO-box), implemented a variety of subroutines to facilitate analysis of the simulation trajectories, and carried out simulations on bulk water and for investigating solvation free energy of single ions and ion pairs. I inherited the FORTRAN code from Dr. Lin, and have modified it as needed to carry out my work. Over the last few years, I have acquired first-hand knowledge about the details of the ICSM and the FORTRAN code, and have further extended it. I now summarize the ICSM, where the essential parts of the model are described.

For the electrostatics part of the problem, the ICSM is based on a spherical cavity filled with molecules representing solute and solvent. The model is shown in Fig. 2.3. Inside the sphere the dielectric permittivity is given as ϵ_{in} taken as 1 for vacuum, while outside the sphere is implicit solvent whose electrostatic effect is represented by the continuum media with a dielectric ϵ_{out} . The total electrostatic potential $\Phi(r)$ satisfies the LPB equations:

$$\epsilon_i \nabla^2 \Phi(r) = -\rho_{in}(r), \quad r \in V_{in}, \quad (\text{Eq. 2.6 a})$$

$$[\nabla^2 - \lambda^2] \Phi(r) = 0, \quad r \in V_{out}, \quad (\text{Eq. 2.6 b})$$

where the charge distribution $\rho_{in}(r)$ contains all the explicit charges from solute and solvent molecules. λ is the inverse Debye-Huckel screening length. For boundary conditions near the surface,

$$\Phi_{in} = \Phi_{out}, \quad \text{and} \quad \epsilon_i \frac{\partial \Phi_{in}(r)}{\partial n} = \epsilon_o \frac{\partial \Phi_{out}(r)}{\partial n}, \quad (\text{Eq. 2.7})$$

where n is the outward normal of spherical surface.

PBC are applied to the simulation cell for the short-range molecular forces between atoms. The PBC make the simulation protocol run at a fixed particle density and reduce finite size effects. The periodic boundaries are incorporated into a sphere, which can be done using a cube inscribed within the sphere. However, it is obvious that a cube and sphere do not occupy the same space, and given that the cube does not fit snugly inside the sphere, there will be a region of space that is defined outside of the cube and inside the sphere. This region is important, and will be described below in more detail. Most of this space is wasted using a cube, so the idea is to minimize the wasted space. Therefore, the cube is replaced with a TO-box as schematically illustrated in Fig. 2.3. The TO-box, Λ , is generated from a cube of length L by cutting its eight corners at a distance $(1/4\sqrt{2})L$ from the center. The TO-box has 8 hexagonal and 6 square faces. The distance from the origin to the square face is $(1/2)L$, to the hexagonal face $(\sqrt{3}/4)L$, and to the corners $R_c = (\sqrt{5}/4)L$. The short-range forces and local structure of water are modeled well using periodic boundaries because surface effects are minimized. Most other hybrid models use finite boundary conditions [28, 33, 46], which makes reproducing bulk properties of water near the edge very difficult if not impossible. This problem I will address later in Chapter 5. The ICSM encompasses the TO-box in the spherical cavity, and the TO-box is surrounded by a buffer layer that extends to the spherical cavity wall, of radius given by $R = R_c + \tau$. The spherical cavity separates two types of dielectric medium where the dielectric constants on the inside, ϵ_i , and outside, ϵ_o , of the sphere are respectively set to values for vacuum and bulk solvent, the latter being water in this case.

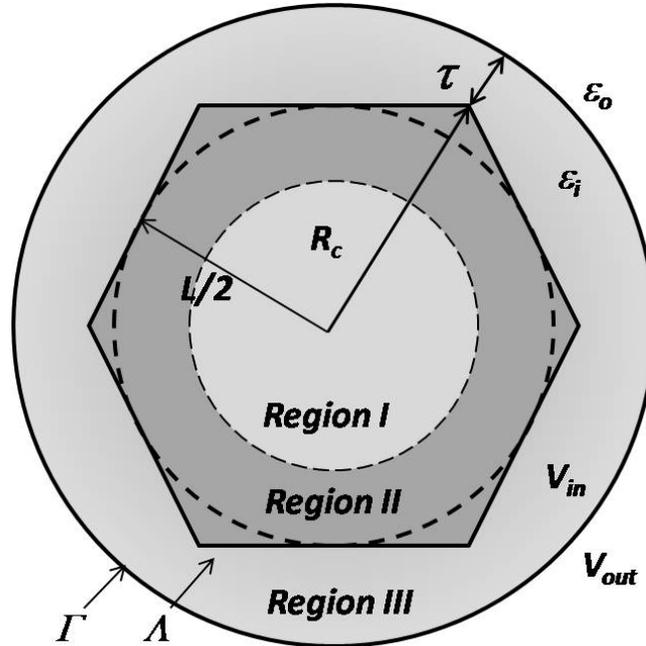


Figure 2.3: Schematic representation showing how the simulation box for the ICSM is divided into three parts inside the spherical cavity of volume, Λ , and surface area, Γ . Regions I and II fill the space within the truncated octahedron box (TO-box) that is placed at the center of the spherical cavity. Region III fills the remaining space as a buffer layer within the spherical cavity and outside the TO-box characterized by length, τ . The water molecules in Region II are imaged into region III using periodic boundary conditions applied to the TO-box. Therefore, water molecules in region III are not associated with dynamic equations of motion. Region I defines the productive volume for simulations, processing translational invariant properties of bulk water consistent with simulations in large periodic systems. In practice, the majority of region II also represents bulk water well.

The buffer layer (Region III) is filled with water molecules that are periodically imaged from Region II in the box. The purpose the buffer layer is to reduce the artifacts near the spherical boundary due to the discontinuity of dielectric constant. The size of τ is a priori unknown, and it is presumably model dependent, but it should have some minimal value that reflects the linear size dimension of water molecules themselves. It should be thick enough to protect the TO-box from being affected by the unphysical artifacts, but no thicker. The central part in the TO box is not imaged and is labeled as Region I. The TO-box combines with buffer layer, forming the RF cavity Γ . All charges

in Γ contribute to the RF, computed using the image charge method. Region I with correct bulk water properties can be used for biomolecular simulations and the size is $d = L(\sqrt{3} - \sqrt{5}/2) - 2\tau$. According to the formula, for $L=4.5$ nm and $\tau=0.5$ nm, for example, the box allows simulations of solute molecules with diameter < 1.7 nm.

To minimize computational cost, it is desirable to place the cavity boundary that separates the implicit solvent from the simulation space of explicit molecules as close as possible to the TO-box boundary. Because it is common to use the molecular surface of a solute molecule of interest as the boundary between two dielectric media, it is natural to view the space between the TO-box and the boundary of the spherical cavity as wasted. Leaving this space as vacuum makes the properties of water deviate far from known results. Instead, periodic boundary conditions on the TO-box are used to fill this space with imaged water. Based on previous systematic MD simulations, the optimal range of τ is between 4 to 6 Å, which gives some flexibility on speed/accuracy tradeoff. This range is largely independent of the size of the TO-box. Moreover, the rate at which accuracy is lost below 4 Å is large. Conversely, little gain is seen in accuracy beyond 6 Å. As the buffer layer is increased, additional computational cost is incurred because the atomic charges from the imaged water in the buffer region produce image charges outside the spherical cavity. However, because the water in this buffer region is imaged, they do not require equations of motion. The buffer layer characterized by thickness τ allows the diverging values of the RF (near the interface due to the discontinuous dielectric constants) to be ignored entirely. Therefore, the buffer layer is critical to shield the artifacts induced by the explicit/implicit interface.

2.3 Boundary conditions for water molecules:

The boundary condition applied to water molecules at the surface of the spherical cavity might be important, as they might change the nature of the artifacts from the explicit/implicit interface. Three common boundary conditions are considered: Atom-based, group-based and Orientational Disorder Limit (ODL) based rules [47, 48], as summarized in Fig. 2.4. In our simulations we compared both situations of atom based and ODL limit for water molecules. We would like to check how sensitive the results are to the different boundary conditions. In the limit that $L \rightarrow$ infinity, any of these boundary condition rules should not matter. However, for finite size systems, one boundary condition may require the least amount of finite size corrections, which will show up as differences in the electrostatic or structural properties as a function of simulation cell size.

The atom-based criteria take into account the direct Coulomb interactions for any atom within the spherical cavity independent of the locations of the other atoms within the same water molecules. The group-based criteria, however, depends on the distances of groups of charges, such as a water molecule. The distance can be calculated between two oxygen atoms or between centers of mass of two water molecules. Orientational disorder limit (ODL) method is based on the positions of oxygen atoms. Besides, the orientations of complete disordered situations of the water molecules are also considered [48].

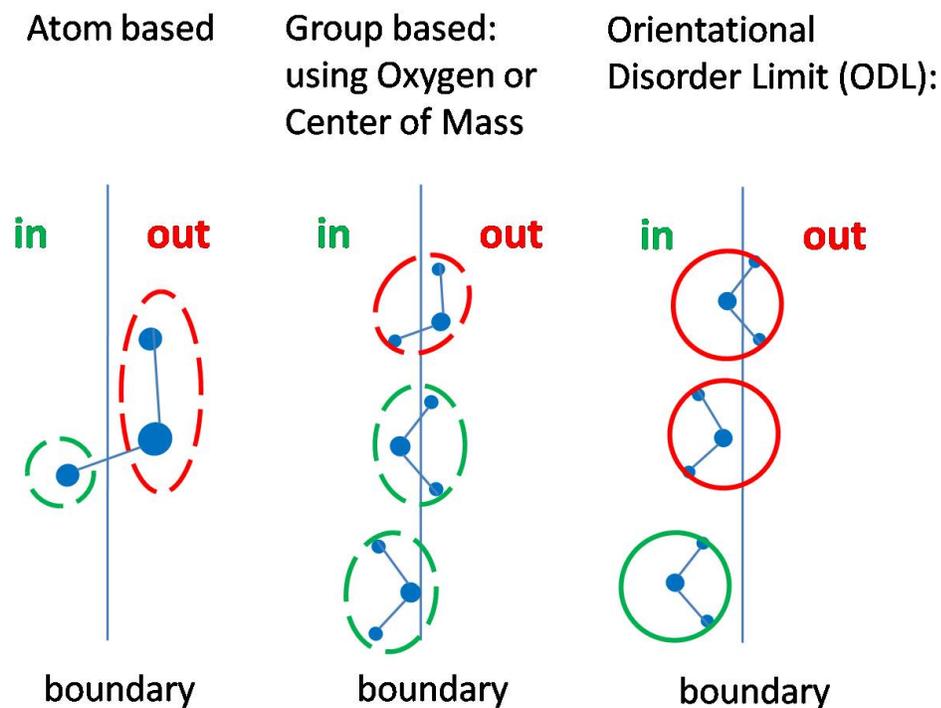


Figure 2.4: Three different truncation criteria are used for water molecules. Red circle means part or entire of the water molecule is treated outside the boundary. Green circle means the molecules are inside the boundary.

Fig. 2.4 shows whether a water molecule will be treated as inside a boundary according to their positions. For the atom-based case, because the atoms are treated individually, so the left hydrogen atom is inside the boundary while the oxygen and other hydrogen atoms are outside. Atom based rules allow the total net charge inside and outside to fluctuate. For the group-based case, if judging by oxygen atoms, the oxygen of the first water molecule is outside, which makes the entire molecule be treated as outside the boundary. Water molecules are treated as inside the boundary if their oxygen atoms are inside, even though one or two hydrogen atoms might be outside. The group-based truncation may cause artifacts in RF method [49]. For ODL-based case, although the oxygen atoms of the first two water molecules are inside boundary, the molecules are still treated outside. That is because they will overlap with the boundary if the molecules

have different orientations using oxygen atom as an origin. The water molecule can be considered as a sphere of radius of H-O bond length and with oxygen atom as the origin, as shown in right of Fig. 2.4.

CHAPTER 3: CHARACTERIZATION OF REACTION FIELD

3.1 Introduction

Although the image charge method [15, 37] has been used in hybrid explicit/implicit model for electrostatic calculation [50], the combination of multiple image charges [40-42] and PBC in the ICSM is a new construction for a better solvent model [9, 39]. The ICSM faithfully reproduces the properties of bulk TIP3P waters [39] leading us to investigate its essential characteristics. The former work of ICSM shows that the dielectric constant falls to very low value without a RF, while all other important bulk properties of water are still correct [39]. So the questions that need to be answered are: (i) Why is the RF important, and how does it affect the behavior of ICSM? (ii) How important is the buffer layer in reducing artifacts due to the discontinuous dielectric constant at the cavity boundary? (iii) What is the advantage of using more image charges to systematically increase the accuracy of the RF within the ICSM? (iv) How does the system respond to using different levels of accuracy for RF and different boundary conditions for molecule? All the questions above will be mainly answered by analyzing the histogram distributions of electrostatic forces and torques through the entire simulation cavity.

3.2 Methods

Histogram distributions for electrostatic forces and torques on water molecules are constructed based on MD trajectories derived from the simulation of bulk water in thermodynamic equilibrium at room temperature ($T = 300$ K) using ICSM. The Tip3p water model is used, in which oxygen atom has a charge of -0.834 and hydrogen 0.417. The length of HO bond is fixed at 0.9572 Å. All details for performing the MD simulations in this work are as published previously [39]. After equilibration, typical MD trajectories are 4 to 6 ns long, and frames were sampled at 0.2 ps apart.

The net force on a water molecule is the sum of pairwise forces on all its atoms.

The pairwise force is given by: $\vec{F} = \frac{1}{4\pi\epsilon_i} \left(\frac{Q_1 Q_2}{|r|^3} \right) \vec{r}$ whether it is a direct interaction between

atoms, or indirect interaction between atoms and image charges that reside outside of the spherical cavity. The water in the buffer layer is considered as contributing to the direct part of the electrostatic force. The RF derives from a linear superposition of all forces from the image charges that lie outside the spherical cavity. The torque on a water molecule is given as:

$$\vec{\Gamma} = (\vec{r}_{H_1} - \vec{r}_{COM}) \times \vec{F}_{H_1} + (\vec{r}_{H_2} - \vec{r}_{COM}) \times \vec{F}_{H_2} + (\vec{r}_O - \vec{r}_{COM}) \times \vec{F}_O \quad (\text{Eq. 3.1})$$

where \vec{r}_{COM} is the center of mass (COM) of the water molecule, and \vec{F}_O , \vec{F}_{H_1} , \vec{F}_{H_2} are the electrostatic forces on the oxygen and two hydrogen atoms.

Different types of distributions for forces and torques on water molecules are calculated. The first type of distribution is for the magnitude of the net forces and torques from electrostatic interactions only (no van der Waals forces are included). The second type of distribution is when these forces and torques are broken down into the direct and

indirect parts. The third type of distribution is for the radial component of these forces and torques, given by: $F_r = \vec{F} \cdot \hat{r}$ and $\Gamma_r = \vec{\Gamma} \cdot \hat{r}$. All three types of distributions are generated for TO-box sizes of $L=30 \text{ \AA}$ and $L=45 \text{ \AA}$, atom-based, group-based and ODL-based molecular boundary rules, and for 0, 1, 2 and 3 image charges per source charge.

The distributions are generated for water molecules as a function of distance away from the origin in the form of shells to check how homogeneous the water properties are throughout the TO-box. That is, the spherical cavity is divided into 100 concentric shells of equal volume so that each shell contains approximately the same number of water molecules, typically 13 to 15 for 30 \AA box, and 44 to 51 for the 45 \AA box. For a particular shell, the electrostatic force and torque values of all water molecules in this shell are calculated and augmented over all the frames from the MD trajectory. The advantage of using shell numbers is that the relative position of a shell in systems of different sizes is the same for the same shell number, which makes comparisons easier, as shown in Fig. 3.1. All of the forces or torques per molecule are accumulated and used to construct the histogram distribution for a given shell. The same procedure is applied to all shells. If the oxygen atom is inside a particular shell, then the entire water molecule is treated as if it is in that shell. Thus each water molecule belongs to one and only one shell for a given frame, although it moves between shells over the trajectory. Note that although the histograms are normalized, they do not represent probability density functions because we simply binned the data. Normalization means that summing the probability over all the bins yields unity.

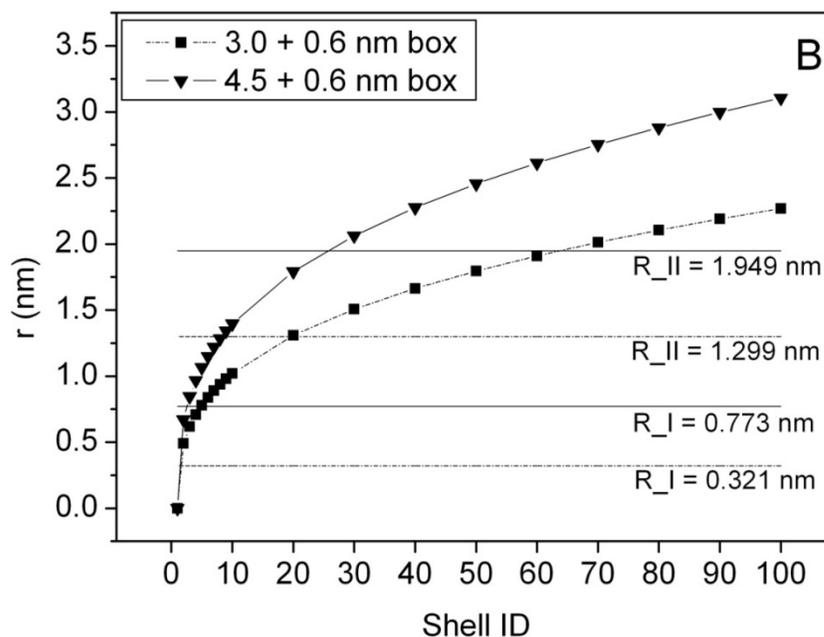


Figure 3.1: The location of each shell compared with the sizes of Region I and II for different simulation box sizes, 30 Å (dash lines) and 45 Å (solid lines).

3.3 Results and Discussions

Force Dependence on Number of Images and System Size

As shown in Fig. 3.2, the statistical distribution of total force on a water molecule for a $L=30\text{\AA}$ TO-box is largely independent of both shell location and the number of image charges per source charge used to represent the RF. The most obvious outlier is for shell 90, which is close to the spherical cavity wall that also defines the outer boundary of shell 100. The force distribution for shells near the spherical cavity boundary within the buffer layer deviate far from the shells located within region I, as illustrated by the result of shell 90. For the 0 image case, a small deviation starts to appear at shell 70. Note that for the $L=30\text{\AA}$ TO-box, shell 41 is the smallest shell that contains only imaged water within region III. Consequently, the histograms for the magnitude of the net electrostatic force on a water molecule within any shell that lies fully or partly within the TO-box are essentially the same, independent of the number of image charges used per source charge.

This result suggests that the number of image charges per source charge (including 0) is not important. If this is the case, then only the buffer layer is playing a significant role in maintaining proper forces within the TO-box.

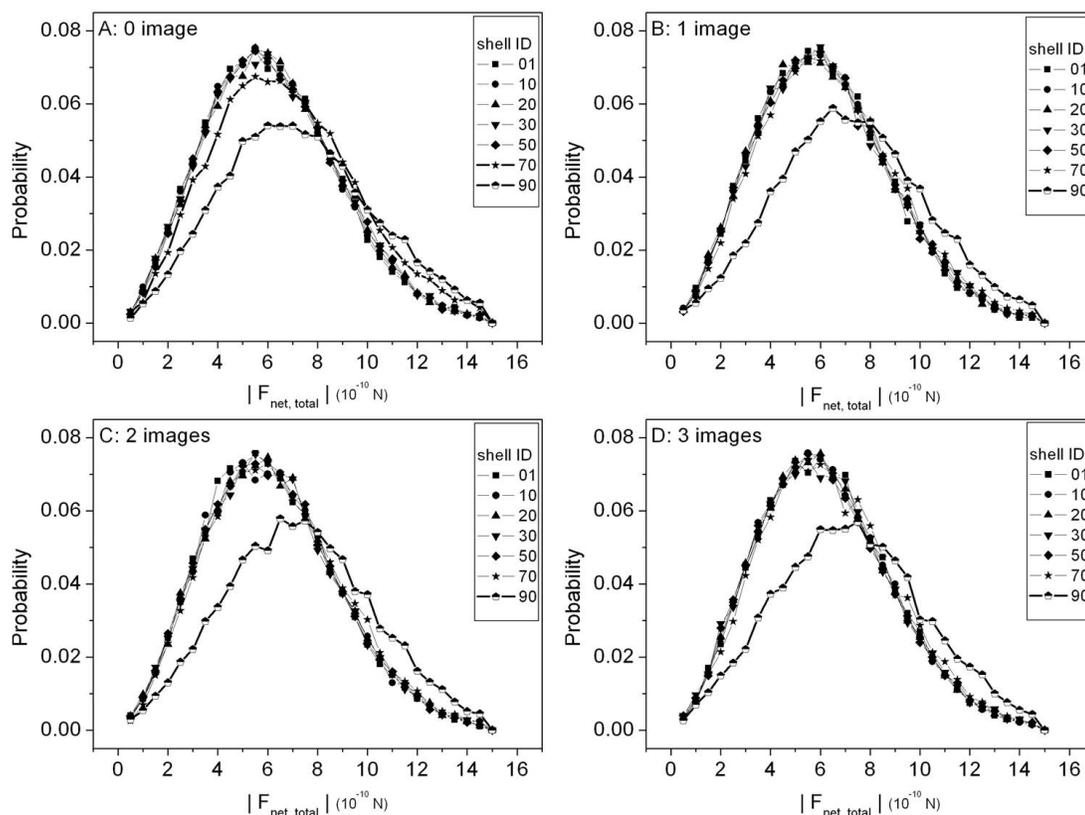


Figure 3.2: Selected distributions within various shells for the magnitude of the net electrostatic force on a water molecule. The total force consists of direct water interactions and the RF component. A box size of $L=30\text{\AA}$ and 6\AA buffer layer is used, and the results are shown for different number of images. Note that the distributions are normalized histograms generated using 30 bins of equal size over the range of 0 to $1.5 \times 10^{-10} \text{ N}$.

Histogram distributions are shown in Fig. 3.3 for the net force on a water molecule that is separated into direct and indirect parts. It is seen that there is strong shell dependence for the RF part, but only a weak dependence on shell position for the direct part, which is associated with all explicit water molecules within the spherical cavity. Simultaneous strong shell dependence on the RF and weak shell dependence on the total

force is possible because of separation of scales. The RF forces in comparison to direct forces typically provide less than a 2% effect, which is shell dependent. Within the productive region, the relative importance is about 1% while at the end of region II the percentage goes as high as 2%. The initial objective for the ICSM was to reproduce correct physical properties of water in the productive region (region I as defined in Fig. 2.3). Pleasingly, the simulations for bulk water properties [39] and for calculating the ion solvation free energy [9] produced results that preserved homogenous characteristics of the system under study markedly well within the TO-box, which includes region II. Because the relative strength of the direct to indirect RF forces is typically greater than 50, it appears that the main reason why the ICSM does well is because of the buffer layer. This result suggests that the buffer layer should be employed, and perhaps the image charges that control the RF forces are not necessary!

As the shells move further out into the buffer range, the relative contribution is 36% by shell 70 for 30 Å box, compared with 16% for 45 Å box, as shown in Table 3.1. The useful shells in regions I and II are extended outwards using a larger TO-box and the outliers again only occur well within the buffer region. This result shows that the buffer layer made up of imaged water of the TO-box is absorbing large RF forces, but dynamics is not affected by these forces. We tested to see if the buffer layer by itself is sufficient to generate an accurate RF, meaning all image charges in the domain outside of the spherical cavity are simply dropped. After dropping image charges outside the cavity, the buffer layer alone yields very large erroneous results in the dielectric response [39], while other structural properties of water did not deviate much. Moreover, we find that to insure

high accuracy in the RF throughout the TO-box, the imaged water in the buffer region must also generate image charges in the domain outside of the cavity (data not shown).

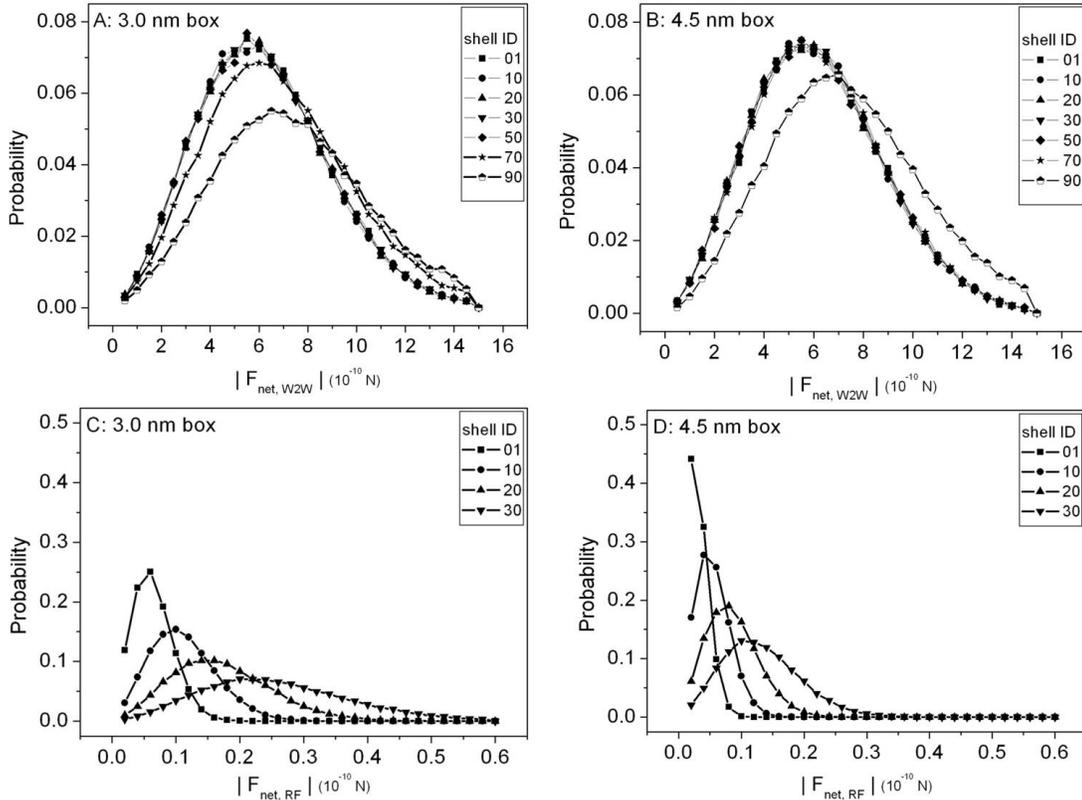


Figure 3.3: The net electrostatic force is split into two parts originating from direct interactions and from the RF. The bin size for the direct component is the same as used in Fig. 3.2, and 30 bins of equal size over the range of 0 to 0.6×10^{-10} N was used for the RF component. The top two panels show the direct interactions, while the bottom panels show the RF component of the net force. The left column shows results for the $L=30\text{\AA}$ TO-box, and the right column shows results for the $L=45\text{\AA}$ TO-box. Both box sizes use a 6\AA buffer layer.

The role of the buffer layer is therefore essential to preserve local structure of water consistent with homogeneous bulk water properties, which includes the long-range electrostatic force that is in part induced by the waters in the buffer layer. This leads to the question, why is the RF essential (using at least one image charge) despite having a relative force contribution that is typically less than 2%? Since the dielectric response is related to the polarization of water, a natural quantity to characterize is the torques on

water molecules, which governs the orientation of the permanent electric dipoles. Another question is whether it is possible that the forces and/or torques generated by the RF are sensitive to the particular molecular boundary condition?

Table 3.1: For different shells, the ratio of the RF part to the total electrostatic forces (RF part + direct part) per water molecule. The unit of force is 10^{-10} N.

Shell ID	30+6 Å Total force	30+6 Å RF force	Ratio RF/Total	45+6 Å Total force	45+6 Å RF force	Ratio RF/Total
01	7.6640	0.0787	0.010	6.5267	0.0400	0.006
10	7.4940	0.1195	0.016	6.4476	0.0649	0.010
20	7.1199	0.1808	0.025	6.4917	0.0967	0.015
30	6.7418	0.2692	0.040	6.4319	0.1435	0.022
50	6.3996	0.6353	0.100	6.4946	0.3351	0.052
70	6.6274	2.3622	0.356	6.5268	1.0676	0.164

Force and Torque Dependence on Molecular Boundary Conditions

No systematic shell dependence on the distributions for forces and torques of any type (i.e. total, direct or RF) were found when different molecular boundary conditions were employed. As defined in Fig. 2.4, the two extreme cases are the atom-based and ODL-based rules. In Fig. 3.4, we show two representative examples of distributions for the radial components of the forces and torques using atom-based and ODL-based rules. Some differences are found in the shells within the buffer layer close to the spherical cavity walls as illustrated by shell 70. However, even qualitative features remain the same (comparing panels C and D for example). Therefore, as far as force and torque distributions are concerned, the molecular boundary rule that is applied along the spherical cavity wall is not important provided the buffer layer is large enough, which

was previously determined to be 4 to 6Å [39]. Within the buffer layer, the differences in forces and torques on the imaged water molecules that depend on the choice of molecular boundaries are not reflected in the dynamics of the molecules. Although data is not shown for group-based boundary rules (the original implementation [39]), in summary, no obvious difference between any of these three boundary conditions is found in the histogram distributions of forces and torques for bulk water.

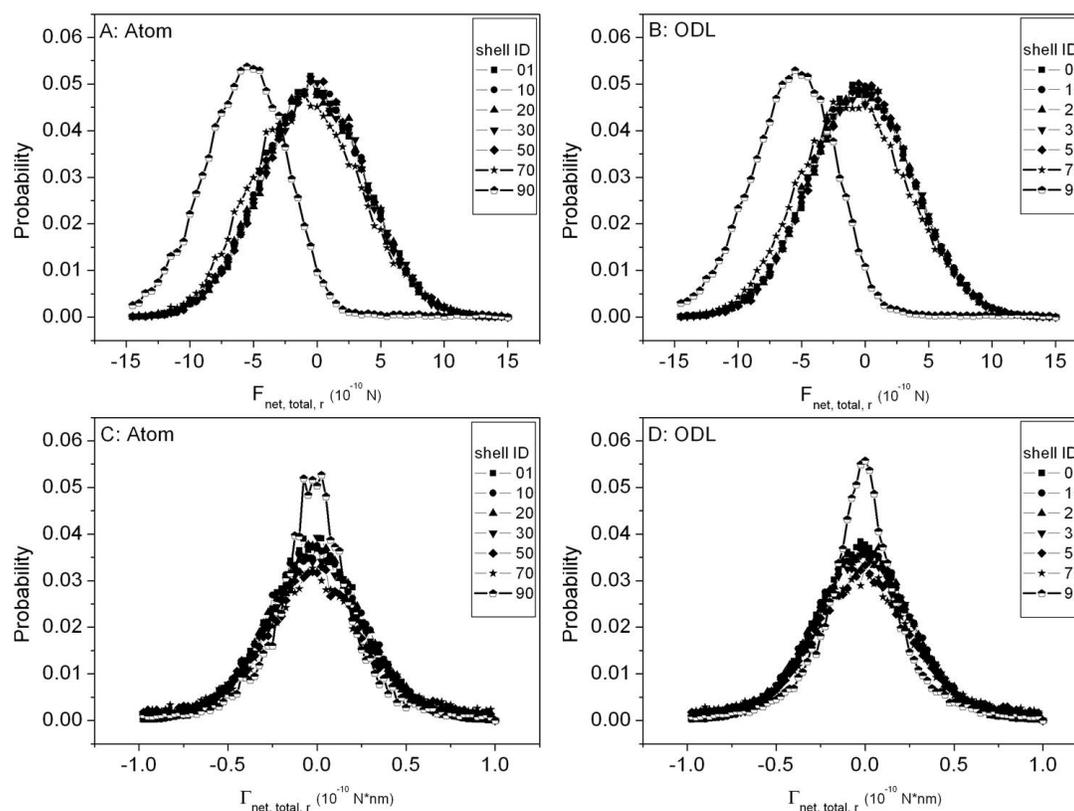


Figure 3.4: Comparing distributions for the radial component of the total force (shown in top row) and the radial component for the total torque (shown in bottom row) using the $L=30\text{\AA}$ TO-box and one image charge. The (left, right) columns show the histograms for (atom-, ODL-) based rules.

Torque Dependence on Number of Images and System Size

In Fig. 3.5, it is shown that for 1 image charge, the histogram distributions for the magnitude of torques that originate from only direct water-to-water interactions carry

slight shell dependence. This result is independent of system size. The shell dependence is dramatically increased once the shells enter the buffer region, as expected. Similar histogram distributions are shown for the magnitude of net torques that originate from only the RF. In this case, the shell dependence is very strong throughout the entire spherical cavity, both inside and outside the TO-box --- similar to what was found for forces. However, the relative scales are much closer. Whereas the forces from the RF compared to direct interactions is typically a 1% contribution within region I, the torques due to the RF compared to the direct interactions is about a 20% contribution, and increase from there as the shell radius increases. This result suggests that the importance of the image charges is not related directly to forces, but more specifically to torques, which is responsible for the orientation of the water molecules, and thus the local polarization properties.

This result is physically intuitive, and it motivated a further check on the significance of the number of image charges on affecting the torque distributions (recalling forces were not sensitive at all). In Fig. 3.6, the effect of using a different number of image charges per source point (0 to 3) on the histogram distributions for net torque that include both direct and RF components are compared for the ODL-based molecular boundary case. It is somewhat surprising to see that the affect of the RF is hardly visible. Using 1 to 3 image charges per source point show virtually no differences within statistical noise. When using no images, a more noticeable change does occur for the outer shells, but no deviation is discernable for any shell up to shell 20, which ends region II. This result again suggests that the buffer layer of imaged water is sufficient to model bulk water, dropping the image charges altogether. As mentioned above, this

experiment was indeed performed in earlier work [39]. Most of the physical quantities that was checked did not depend strongly on whether the image charge was present or not, except for the dielectric constant, which dramatically changed between the 0 and 1 image case.

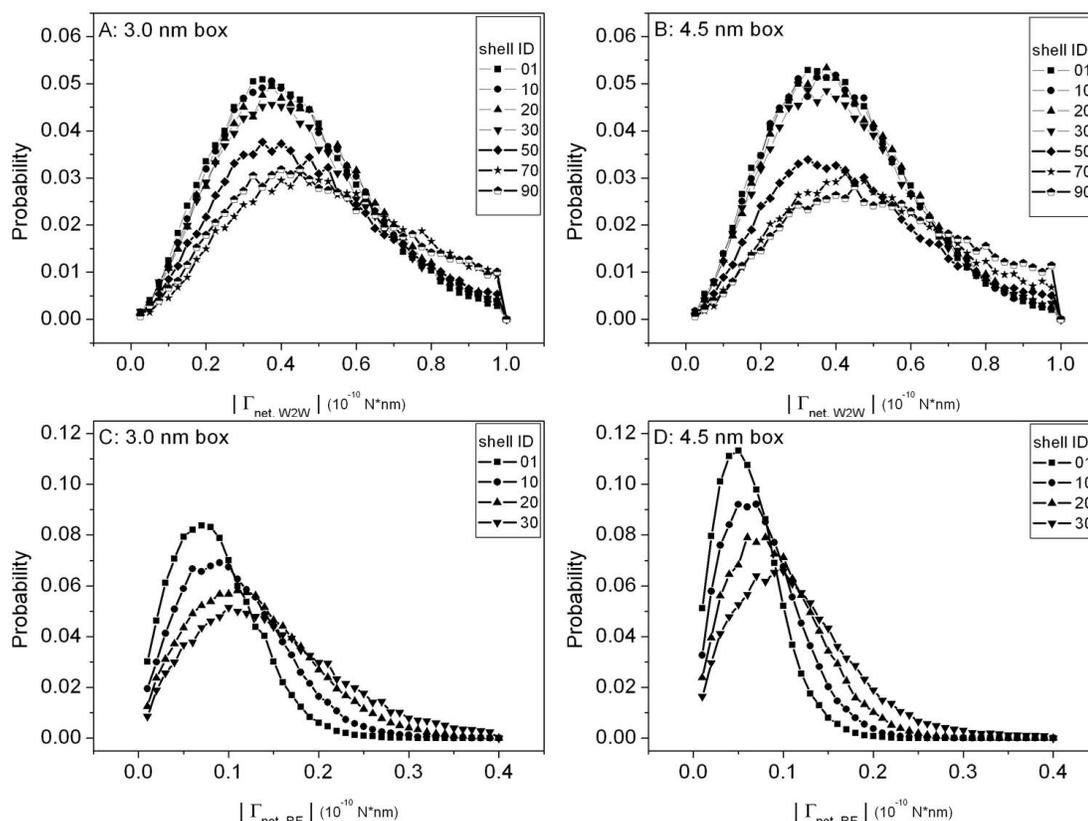


Figure 3.5: The net torque on water molecules due to electrostatic forces is split into two parts that originate from direct interactions and the RF. To produce the histograms, 40 bins are used over the range 0 to 1×10^{-10} Nnm for the direct component, and over the range 0 to 0.4×10^{-10} N nm for the RF component. The top two panels show the results for the direct interactions, while the bottom panels show the results for the RF component of the net torque. The left column shows results for the $L=30\text{\AA}$ TO-box, and the right column shows results for the $L=45\text{\AA}$ TO-box. Both box sizes use a 6\AA buffer layer.

In Fig. 3.7, similar results are shown only for the radial part of the net torques. It was our expectation that a difference in torque distributions would be detected when comparing the 0 and 1 image charge case. In our previous studies, group-based molecular boundary rules were employed [39], while the results shown here are for ODL-based

rules that were incorporated in the ICSM in subsequent work [9]. As we show below, the molecular boundary condition is not the reason for not seeing a difference between the 0 and 1 image charge case.

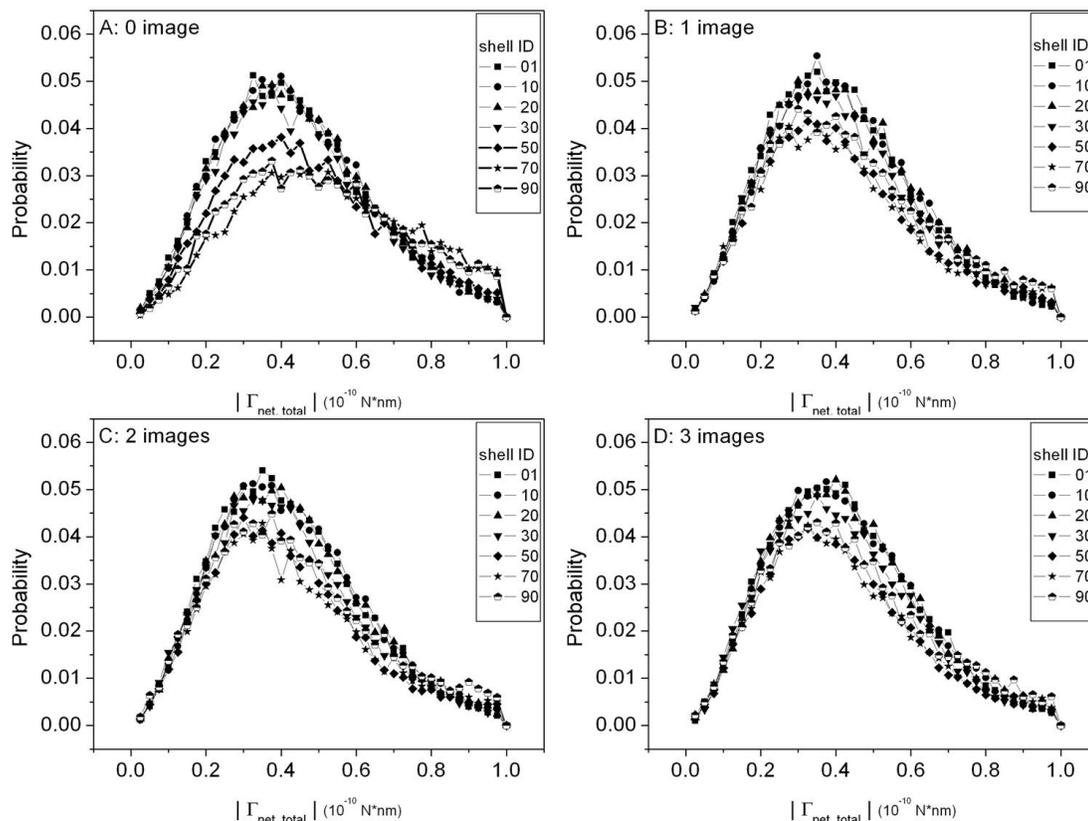


Figure 3.6: Histogram distributions for the magnitude of the net torque due to the electrostatic forces originating from both the direct and RF parts is shown for four cases invoking 0 to 3 images in the MD simulation using the $L=30 \text{ \AA}$ TO-box with a 6 \AA buffer layer. To produce the histograms, 40 bins are used over the range 0 to $1 \times 10^{-10} \text{ Nm}$.

A possible reason why the torque distributions do not reflect the differences we found earlier between the 0 and 1 image charge cases is because the differences that are present are spatially averaged out by using concentric shells. Even the radial component of the torque distribution is subject to this possibility. However, because the radial pair distribution function (between oxygen atoms) is not very sensitive to using the 0- and 1-image charge cases (data not shown), this implies that during the MD simulation the

collective behavior of the water molecules restore the net total forces and torques to statistically support the same histogram distribution. However, this still leaves a perplexing dilemma, because at least 1 image charge is necessary to maintain the correct dielectric constant. These results suggest that our analysis is missing correlations in orientations between pairs of water molecules.

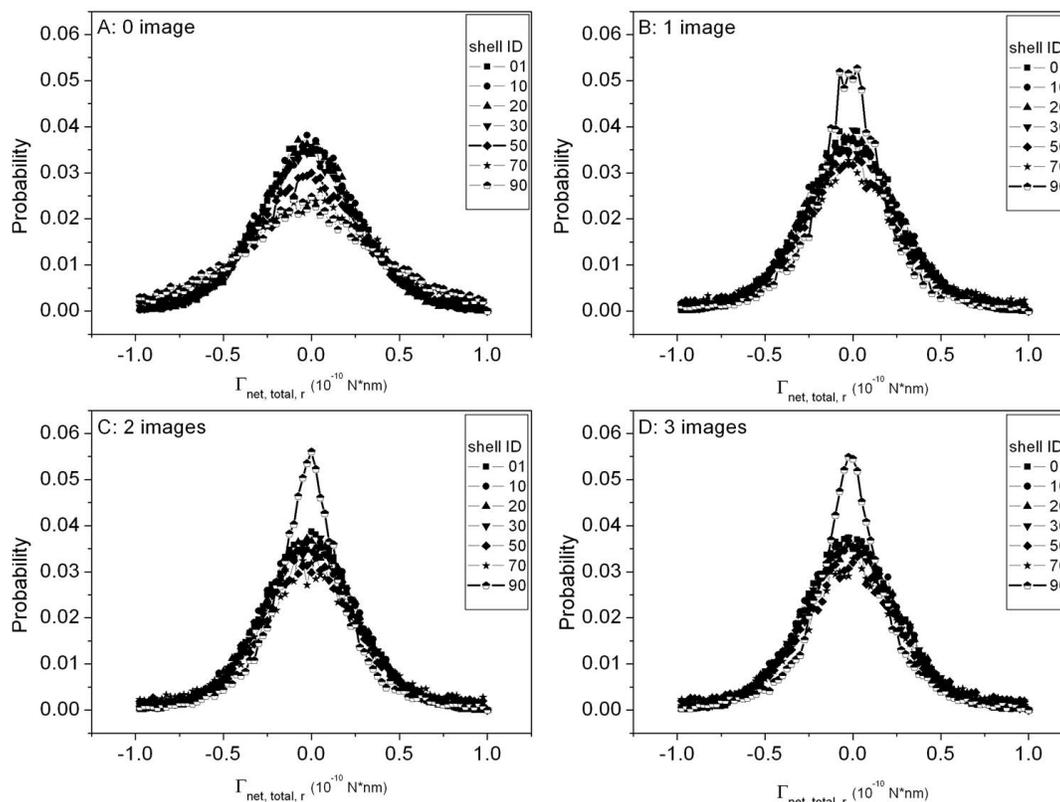


Figure 3.7: Histogram distributions for the radial component of the net torque due to the electrostatic forces originating from both the direct and RF parts is shown for four cases invoking 0 to 3 images in the MD simulation using the $L=30 \text{ \AA}$ TO-box with a 6 \AA buffer layer. To produce the histograms, 40 bins are used over the range 0 to $1 \times 10^{-10} \text{ Nm}$.

To demonstrate that there is indeed an important difference between using 0 and 1 image charge per source charge, the net dipole moment for a spherical ball centered at the origin as a function of radius is shown in Fig. 3.8 for both the atom-based and ODL-based rules applied at the spherical cavity walls. Error bars are calculated to show that $N_i > 0$ cases are all equivalent to one another within statistical errors, but these results are very

different from the no image charge case. The formula for the total dipole moment as a function of radius r with respect to the origin of the TO-box and spherical cavity is given in Eq. 3.2, where a spherical ball contains all water molecules in frame, j , that have its oxygen atom within the radius, r , considered.

Make sure you send PDF and make sure equations are correct. I hate word. This equation is all screwed up again. Check all equations in your PDF file before you send it off to the committee!!!! In fact, check all equations, and figures, and make sure there are no formatting problems. You can have someone else do this, as it is checking for formatting errors, not content.

$$\langle P^2 \rangle_{total} = \sum_{j=1}^J \langle \vec{P}_j \cdot \vec{P}_j \rangle \quad (\text{Eq. 3.2})$$

where $\vec{P}_j = \sum_{k=1}^K (q_{O,k} \vec{r}_{O,k,j} + q_{H_1,k} \vec{r}_{H_1,k,j} + q_{H_2,k} \vec{r}_{H_2,k,j})$

The dipole moment is averaged over all j -frames from the MD trajectory, where the atom position vectors for the k -th molecule in the j -th frame weight the oxygen and hydrogen atoms. Because the first non-zero multipole is the dipole moment, the result is independent of choice of origin. Thus, the expression given in Eq. 3.2 is equivalent to defining the origin to be at the COM or at the oxygen atom for each water molecule.

The total dipole moment is needed to calculate the dielectric constant, and it represents a quantity that reflects fluctuations in the orientations of water molecules that captures correlations that are presumably missed in the torque distributions shown above. This dipole moment calculation was done for both the atom-based and ODL-based boundary rules. These results confirm the previous study using group-based boundary rules. Therefore, any of the three common molecular boundary conditions considered

here have little to no influence on the molecular force and torque distributions regarding properties of bulk water. This is not to say that there are not advantages to employ one boundary condition over the other for certain physical quantities or deficiencies may exist in certain cases as pointed out previously [39, 49].

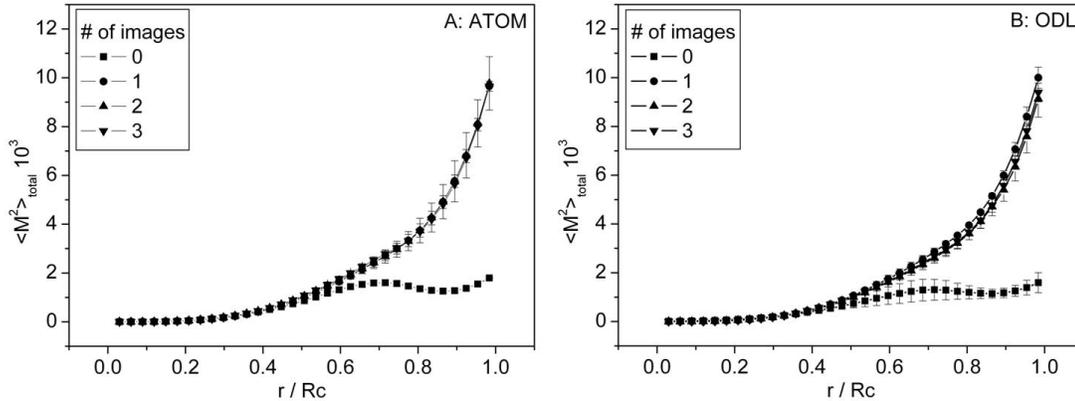


Figure 3.8: The total dipole moment for spherical balls of radius r is plotted for four cases invoking 0 to 3 images in the MD simulation using the $L=30$ Å TO-box with a 6 Å buffer layer. R_c is the radius of entire spherical cavity. The overlapping error bars show that $N_i=1, 2, 3$ cases are producing the same results, based on the 6.0 ns trajectory for atom-based and 4.8 ns for ODL-based cases. The error bars on the 0 image curve in the atom-based case are smaller than the size of the symbols. A: Atom-based and B: ODL-based boundary rules are used.

Table 3.2: For different shells, the ratio of the RF part to the total electrostatic torques (RF part + direct part) per water molecule. The unit of torque is 10^{-10} N nm.

Shell ID	30+6 Å Total torque	30+6 Å RF torque	Ratio RF/Total	45+6 Å Total torque	45+6 Å RF torque	Ratio RF/Total
01	0.4373	0.0910	0.208	0.4479	0.0677	0.151
10	0.4552	0.1127	0.248	0.4368	0.0823	0.188
20	0.4737	0.1346	0.284	0.4374	0.0986	0.225
30	0.5533	0.1610	0.291	0.5140	0.1227	0.239
50	0.7224	0.2773	0.384	0.8412	0.2230	0.265
70	0.7267	0.5059	0.696	0.8469	0.4017	0.474

In Table 3.2, both total electrostatic torque and torque from RF are growing when the size of shell increases. In shell 1, the ratio is 20.8% and 15.1% for 30 Å and 45 Å boxes, respectively. These ratios are much larger than those of electrostatic force shown in Table 3.1, which are no larger than 2.0% before shell 10. The comparison between Table 3.1 and 3.2 elucidates that, in region I, electrostatic torques from the RF play a more important role than electrostatic force to maintain the correct dielectric properties. Therefore, the importance of the image charges is not related directly to forces, but more specifically to torques, which is responsible for the orientation of the water molecules, and thus the local polarization properties. This result is physically intuitive, and it motivated a further check on the significance of the number of image charges on affecting the torque distributions (recalling forces were not sensitive at all).

The interesting feature identified is that within region I, there is essentially no difference between using 0 or 1 image charge per source charge. The buffer layer on its own already makes region I resemble bulk properties of water. However, region II retains the properties of bulk water as well. As Fig. 3.8 shows, a non-negligible deviation starts to occur just at the end of region II (a radius of 12.9Å ends region II for a L=30Å TO-box.). For bulk water simulations [39] and for ion solvation [9], we previously reported that in practice region II also maintained the properties of bulk water well when 1 image charge per source charge is used in the ICSM. For practical purposes, dropping the single image will surely increase unwanted finite size effects because physical correlations involving water orientations will be lost, and therefore is not recommended. Moreover, 1 image charge per source charge appears to be sufficient for high accuracy.

Importance of the Reaction Field and the Buffer Region

Fig. 3.9 shows how the dielectric constant change depends on different numbers of image charges [39]. When reaction field is gone $N_i=0$ case, the curve that corresponds to this state is very different from those computed from $N_i > 0$ situations. It falls back to small value when approaching to the edge of the sphere. The plot reveals that the reaction field is essential for maintaining a uniform dielectric response throughout the simulation box. Our aim is to find out how RF affects the dielectric properties with different numbers of image charges.

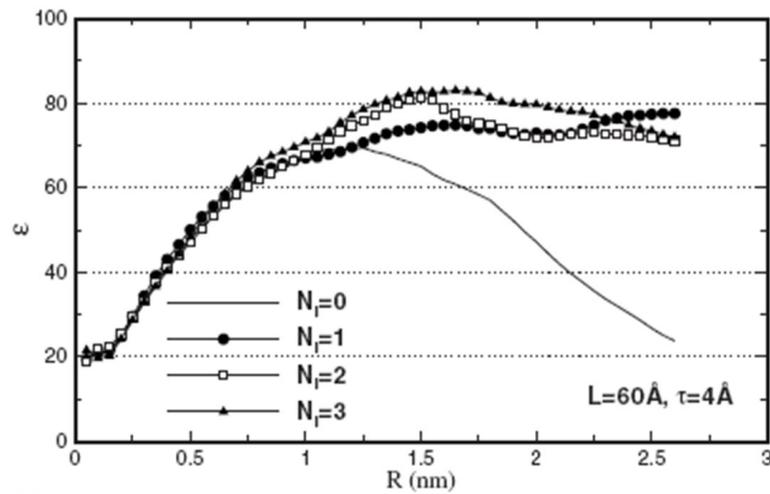


Figure 3.9: Dielectric constant $\epsilon(R)$ for increasing number of image charges N_i . Convergence occurs at $N_i=1$. The reaction field ($N_i \geq 1$) is necessary to produce uniform dielectric properties across the simulation box.

The result in Fig. 3.9 shows that, when the number of image charges N_i is equal to or bigger than 1, the entire system will not have significant changes in dielectric constants. And no reaction field (zero image charge) will give the wrong dielectric constant. It also provides more proof to the conclusion that, for certain accuracy, only one image charge (Kelvin image charge) is necessary for the RF [39].

Building on prior works [9, 39], combined with the above detailed analysis of force and torque distributions, we have shown that the importance of the RF manifests

directly through the net torques on the water molecules. The RF component represents a long-range effect that works in a subtle way to create correlations in the orientation of water molecules, and thus affects the polarizability of the solvent. The observation that torque plays a central role in describing proper bulk properties of water has been exploited in the local molecular field (LMF) [51] theory approach. In general, boundary conditions that arise at the interfaces between different types of medium are difficult to deal with due to the long-range nature of the electrostatic force. As such, simple cut-off procedures create large errors. These errors have been mitigated within the LMF approach [51] by accounting for the net long-range electrostatic forces in terms of reorienting torques on water molecules, which is consistent with our more accurate calculations. As such, this suggests that the success of the LMF approach is precisely because the dominant effect has been accounted for well.

The results presented above also suggest that if 1 image charge per source charge is kept the buffer layer can be reduced, since all shells maintained the same force and torque distributions within region I and most (if not all) of region II. Although the distributions were not shown, it was already clearly demonstrated [39] that the buffer layer thickness is critical to maintaining accurate structural and dynamic properties of bulk water. As soon as the buffer layer is reduced (say to 4Å) some properties of water deviate from the desired target properties of bulk water. This prior observation is understandable because the relative importance of both torques and forces from the RF compared to direct interactions increases when a water molecule comes closer to the spherical cavity wall. From a practical point of view, a buffer layer of 4Å may introduce errors that are tolerable for many applications, but 6Å will ensure high accuracy

whenever needed. This finding leads us to point out a serious concern involving artifacts found in any explicit/implicit model.

The method of images [40] is a rigorous way of solving the Poisson equation for a system with spherical geometry that we have employed in the ICSM. It is worth noting that this method can be extended to accurately solve the linearized Poisson-Boltzmann equation [52] with arbitrary ionic strength. In the case of bulk water, we show here that the Poisson equation can be solved more accurately using multiple images (2 or 3 per source point for example) than is required because the physical quantities that can be calculated do not gain in any detectable accuracy. This is because the discontinuous model for the dielectric to change from within the cavity to the continuous dielectric medium outside of the spherical cavity walls is a mathematical idealization. As such, the forces and torques near this boundary are unphysical. To counter this problem, the buffer layer is introduced to absorb idealized-model induced errors. The buffer layer is therefore an essential part of the ICSM.

It is important to realize that multiple image charges (more than one) is required for accurate calculation of the electrostatic forces for charges in the simulation box near the spherical dielectric interface where the Kirkwood expansion converges extremely slowly. However, this region is where the buffer layer is located. The ICSM uses imaged water molecules in the buffer layer, which do not have associated dynamical equations of motions. Increasing the number of image charges will dramatically help increase the accuracy of the water molecules within the buffer layer, but our results show that the accuracy is already sufficient for the water molecules outside of the buffer layer. In other words, improving accuracy for the forces and torques on water molecules within the

buffer layer is not relevant. Due to the scaling of how big the buffer layer is relative to the spherical shell radius, it may be that more image charges will be needed for very large box sizes. However, even up to an 80Å box with $\epsilon_o = 80$, one image charge proved sufficient [39]. Also, we explored whether more than one image charge is ever needed for bulk water simulations using different ϵ_o values that go as low as $\epsilon_o = 10$. In all cases checked for bulk water, we find one image charge is good enough to recover accurate properties of bulk water. However, also note that this single image charge is slightly modified from the classical Kelvin charge, and partly accounts for the imaged line charge in addition to the classical Kelvin charge.

The significance of the presence of a buffer layer of a minimum thickness to ensure accurate calculations of electrostatic interactions is more far-reaching than a simulation protocol for modeling bulk water. Rather, the analysis presented here demonstrates that any model that employs a discontinuous change in the dielectric constant within the physical domain of interest will create large spurious forces and torques on explicitly modeled atoms. Therefore, when developing a hybrid model that interfaces continuum electrostatics with explicit atomic systems, a buffer layer should be included to better model structural details at the interface [53], and to allow a way for artifacts caused by the unphysical boundary conditions to die out far enough away from the parts of the system that are under study.

3.4 Conclusions

Molecular forces due to the RF are typically less than 2% of the total force, while in comparison, molecular torques due to the RF are typically 20% of the total torque. Since molecular torques affect water orientation, and thus polarization, the role of the RF

is essential in describing the correct molecular response to electrostatic forces. In particular, the RF is responsible for generating correlations in the orientations of water molecules, which affects the local dipole moment. In addition, it is shown that using a discontinuous dielectric model produces large artifacts at the interface, which is why a buffer layer is needed. The role of the buffer layer used in the ICSM is critical to maintaining high accuracy since the spurious forces and torques that are generated near the interface walls do not affect the dynamics of the molecules near the walls because they are imaged. More generally, the analysis suggests that it is important to create a buffer layer in multi-scale hybrid models, and/or to completely avoid discontinuous change in the dielectric within the physical domain of interest. Also the results show that very little influence has been observed with multiple image charges and different boundary conditions for water molecule. Although the break-even point for the ICSM to outperform the PME is about 30,000 particles [54], it is slower than PME for smaller systems. Considering the inefficiency in computational cost due to the buffer layer, we would like to improve the discontinuity of dielectric at the boundary to possibly shrink the buffer layer. This is the subject of the next chapter.

CHAPTER 4: OPTIMIZATION OF IMAGE CHARGES

4.1 Introduction

The implementation of the buffer layer has received attention for reducing the artifacts of the hybrid explicit/implicit model [32, 33, 55, 56]. A carefully constructed buffer layer will provide a good approximation for the dynamic properties for a region that are far from the boundary. An early treatment of the buffer layer was suggested in terms of Langevin stochastic algorithm for the dynamics of the molecules, described by Brooks and Karplus [57]. The ICSM, however, constructs the buffer layer in a simpler way by using imaged water in the simulation box through PBC where the water molecules in the buffer have no equation of motion. Nevertheless, the buffer layer molecules do have electrostatic interactions on the explicit solvent and induce the image charges for RF.

The analytical solution for representing implicit solvent effects assumes that the dielectric constant near the boundary has unphysical discrete change from 1 to 80, and we call this version of ICSM Discontinuous Dielectric Model (DDM). The goals of this project are to optimize the DDM to improve accuracy while reducing finite size effects. If successful, this would also make the productive region larger and the buffer layer thinner to improve the simulation efficiency.

The approach is to use a different set of image charges to reflect a continuously changing dielectric profile near the boundary to minimize artifacts caused by unphysical

discontinuous dielectric constants that were applied for a convenient mathematical idealization.

4.2 Methods:

Solving inverse problem:

The optimization can be solved as an inverse problem. Firstly, the electrostatic properties (such as potential, forces and force tensors, we use potential here for example) is calculated from a MD simulation of a large periodic system using GROMACS [58]. These calculations are taken as exact because the system size of the periodic system is greater than the correlation length governing water-water interactions. Secondly, by using least square error method, the image charges from DDM model will be modified so that the dielectric potential have the best match to that large periodic system, quantified by the smallest least squares error. The assumption is that the interface potential of the spherical cavity is a transition region that matches an explicit model of solvent to an implicit model of the same solvent (in this case, pure water), and that this property will not be modified by the solute particles that are contained within region I of the simulation box.

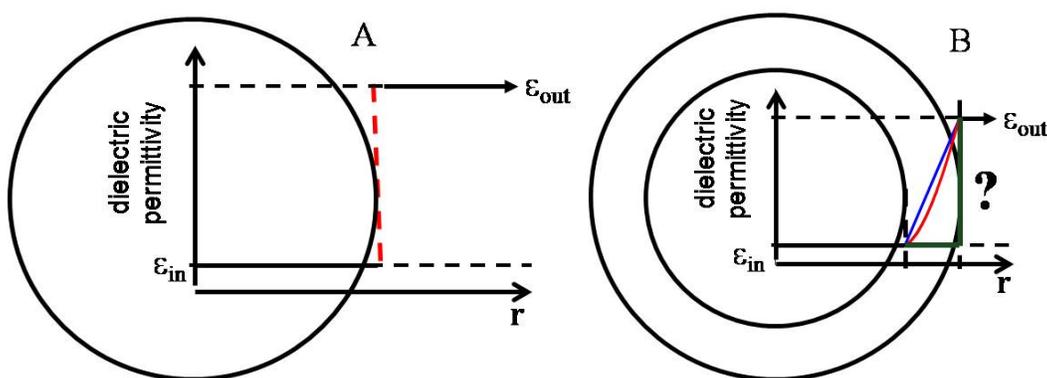


Figure 4.1: A) Discontinuity of dielectric permittivity for DDM model in which a sudden jump of dielectric constant occurs near the boundary of the cavity. B) Continuous transition of dielectric permittivity through the buffer layer for optimized model.

The DDM has an analytical solution using Gauss-Radau quadrature points [40-42], as shown in Fig. 4.1 A. The formula for the Kelvin image charge is given by,

$$q_{kelvin} = \gamma q_s a / r_s, \quad r_{kelvin} = a^2 / r_s, \quad \text{where } \gamma = \left(\frac{\epsilon_i - \epsilon_o}{\epsilon_i + \epsilon_o} \right) \quad (\text{Eq. 4.1})$$

However, this solution is calculated by boundary conditions with discontinuous dielectric constants. In Fig. 4.1 A, for actual calculation, ϵ_{in} is 1 and ϵ_{out} is 80 for water solvation. The dielectric permittivity changes suddenly from 1 to 80 close to the edge of sphere, which makes the system unphysical near the boundary. To avoid the unphysical forces, a buffer layer is added outside the simulation box so that the center of the sphere will be protected from the boundary area and can be treated as bulk water, shown in Fig. 4.1 B. In the simulation, only the center part of the sphere defined by region I in Fig. 2.3 can be used as bulk water, because the solvation in outer shells is still affected by the buffer layer. For a truncated octahedron box with diameter of 3.0 nm and 0.6 nm buffer layer, the usable region is only 0.63 nm in diameter.

The idea in Fig. 4.1 B is that, without knowing the dielectric profile between a spherical cavity and the outside continuum dielectric medium, we can match the electrostatic potential in the explicit region directly to the value defined by the target solvation model by optimizing the image charges. A Least Squares Error (LSE) method is employed to carry out this optimization. After finding the optimal image charges, MD simulation is performed using ICSM with the optimized image charges. Based on the MD simulations, the model-dependent dielectric profile can be calculated, but the important point is that it does not need to be known a priori. Thus the inverse problem avoids having to define the precise dielectric profile at the interface of the spherical cavity, which would be dependent on many model details.

Least squares error method:

In this work, we fix the locations of the image charges from the analytical solution, changing their magnitudes to minimize the LSE with respect to the exact solution of RF potential V_{exact} .

$$Error = \sum_i^N (V_{exact,i} - V_{estimate,i})^2 \quad (\text{Eq. 4.2})$$

Here the variable “ i ” represents field positions within the region of space of interest, and this will be described in more detail below. This method employed can be viewed as a variation problem. We assume the form of the RF is the same as Eq. 2.4, and that the quadrature points are exactly the same, and the only difference is in what the weights are in representing the line charge density. As such, this translates to simply finding unknown image charges as the variables we solve for when obtaining the LSE. In other words, this procedure is equivalent to finding the best fit for the RF to V_{exact} as a reference. Moreover, because the LSE method samples a large number of field points in the sphere, it tends to spread the error uniformly throughout the cavity.

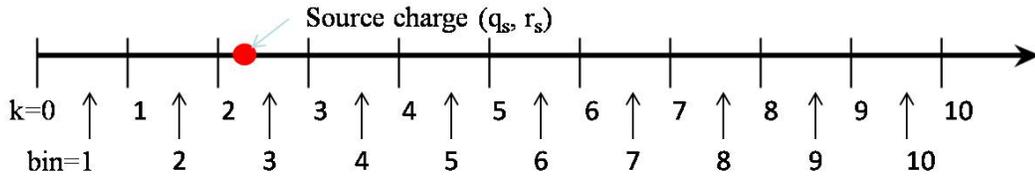
Interpolation method:

Because we want to reproduce the electrostatic potential of the target water model, in the context of LSE we minimize the differences of the potential between exact physical value Φ_{RF} and the value of image charge solvation model with optimized image charges Q_m . However, this would require knowing what the image charge needs to be for any source charge location, which would require an infinite number of fitting variables since the source charge can be anywhere within the spherical cavity. To get around this problem, we parameterize the equations using a variety of interpolation methods, which

only requires a small number of variables. A straightforward approach of defining a set of bins, and doing an interpolation between the bins was implemented first.

$$Err(Q) = \sum_{n=1}^N \left(\sum_k \sum_{m=0}^M \frac{Q_m(r_s | k, k+1)}{4\pi\epsilon_i |r_n - x_m(r_s)|} - \Phi_{RF}(r_n) \right)^2 \quad (\text{Eq. 4.3})$$

where N are N field points for electrostatic potential calculation, K is the number of bins for interpolation, M means how many image charges will be used for one source charge. $M=0$ means use Kelvin image charge only. The procedure to find the minimum in the LSE is essentially looking for a set of optimized image charges: $Q_m = (Q_0, \mathbb{L}, Q_M)$, so that $Err(Q_m) = \min(Err(Q))$.



The schematic above shows how we did the interpolation and optimization. We divided the radius of the sphere in the image charge salvation model into several bins. In practice, 10 bins are good enough for optimization. The number of bin edges ranges from 0 to 10 and the 11 optimized charges are located at these edges. The spherical cavity is full of source charges (water molecules). Suppose one source charge (r_s, q_s) falls in bin 3, and the optimizing charges that are located at the edges of bin 3 are $q_{k=2}$ and $q_{k=3}$. If Q_s is the image charge of q_s , $Q_{k=2}$ and $Q_{k=3}$ are the image charges formed by $q_{k=2}$ and $q_{k=3}$, according to the definition of linear interpolation, the source charges Q_s could be calculated by $Q_{k=2}$ and $Q_{k=3}$:

$$Q_s = Q_{k=2}(1 - \lambda) + Q_{k=3}\lambda, \text{ where } \lambda = \frac{r_s - r_{k=2}}{r_{k=3} - r_{k=2}} \quad (\text{Eq. 4.4})$$

In the RF method, when the source charge is close to the center of sphere (r_s is very small), the distance of its image charge could be infinite. To reduce the influence of this effect, we optimized $u_s = Q_m \times r_s$ instead of Q_s directly. So the new linear interpolation formula for u_s is:

$$u_s = u_{k=2}(1 - \lambda) + u_{k=3}\lambda, \text{ where } \lambda = \frac{r_s - r_{k=2}}{r_{k=3} - r_{k=2}} \quad (\text{Eq. 4.5})$$

Besides linear interpolation, we also tried non-linear ways to express the image charges. Here, u_s is expressed as a certain order of polynomial terms, so that the coefficients of each term will be optimized. Let R be the radius of the system sphere,

$$u_s = Q_m \times r_s = \sum_{j=0}^{\text{order}} c_j x^j, \quad (\text{Eq. 4.6})$$

where $x = r_s / R$, c_j is the coefficient of the polynomial term x^j

$$\delta Q = \delta u / r_s = \left(\sum_{j=0}^{\text{order}} \delta c_j x^j \right) / r_s \quad (\text{Eq. 4.7})$$

Suppose we express u_s in a 6-order polynomial terms, then

$$u_s = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + c_5 x^5 + c_6 x^6, \text{ where } x = r_s / R \quad (\text{Eq. 4.8})$$

After seeing the results from the linear, polynomial and other ways we interpolated, we noticed the solution looks similar to a $\tanh()$ function. Therefore, we directly parameterized the image charges as a $\tanh()$ function to fit the curve. In this procedure we define:

$$u_s = a + b \times \tanh(cR),$$

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$

$$\delta u_k = u_k - u_{k-1} = \delta a_k + \text{coeff}1(R, c_{k-1}) \delta b_k + \text{coeff}2(R, c_{k-1}) b_{k-1} \delta c_k$$

$$\text{coeff}1(R, c_{k-1}) = \tanh(c_{k-1}R)$$

$$\text{coeff}2(R, c_{k-1}) = \frac{R}{\cosh^2(c_{k-1}R)}, \quad \cosh x = \frac{1}{2}(e^x + e^{-x})$$

Since the $\tanh()$ function is a nonlinear function, we assume an initial set of parameters $\{a, b, c\}$, and linearize the equation, which is minimized in terms of small changes in $\{a, b, c\}$. Then, this procedure is repeated until the differences in the variables stop changing. Instead of using 11 parameters (11 bins) in linear interpolation, we are able to reduce the problem to optimize 3 parameters in $\tanh()$ interpolation, a and b decide the start and end position of the curve, c decides the slope of the curve. Based on the optimizing curve from linear interpolation results, we think that in the application of running MD simulations, it will be more convenient and easier to calculate image charges from optimized u_s if a continuous formula is used rather than performing interpolations.

Using LAPACK to solve inverse problem

The LSE optimization problem reduces to a standard linear algebra problem that can be solved numerically using the Linear Algebra PACKage (LAPACK) mathematical library, which is a software package provided by Univ. of Tennessee; Univ. of California, Berkeley; Univ. of Colorado Denver; and NAG Ltd.. The equations are

$$u_s = Q_m \times r_s$$

$$Ax = B \Rightarrow A \delta Q_m = \delta \Phi_{RF} \Rightarrow \frac{A}{r_s} \delta u_s = \delta \Phi_{RF} \quad (\text{Eq. 4.9})$$

where u_s is the parameters needed to be optimized. A/r_s is a matrix, the row number is field points \times xyz directions, for column it is the number of parameters. In optimization iterations, we are actually optimizing u_s because we use $u_s = r_s \times Q_m$ to make the calculations numerically stable when the source charge is very close to the center of sphere. Take linear interpolation for example,

$$\begin{pmatrix} A_{1,1} & A_{1,2} & A_{1,3} \\ \dots & \dots & \dots \\ A_{N,1} & A_{N,2} & A_{N,3} \end{pmatrix} \times \begin{pmatrix} \delta u_{k=0,m=0} \\ \dots \\ \delta u_{k=10,m=0} \\ \delta u_{k=0,m=1} \\ \dots \\ \delta u_{k=10,m=1} \\ \delta u_{k=0,m=2} \\ \dots \\ \delta u_{k=10,m=2} \end{pmatrix} = \begin{pmatrix} \delta \Phi_1 \\ \delta \Phi_2 \\ \dots \\ \delta \Phi_N \end{pmatrix}, \quad (\text{Eq. 4.10})$$

The element A_{ij} is defined as:

$$\frac{A_{i,j}}{r_s} = -\frac{1}{4\pi\epsilon_0} \frac{1}{r_{s,p,q}} \left[\frac{(1-\lambda_{k=0,1,p,q})\beta_{k=0,1}}{|r_i - x_{k=1,j,p,q}|} \left\{ \frac{\lambda_{k=0,1,p,q}\beta_{k=0,1}}{|r_i - x_{k=1,j,p,q}|} + \frac{(1-\lambda_{k=1,2,p,q})\beta_{k=1,2}}{|r_i - x_{k=2,j,p,q}|} \right\}, \right. \\ \left. \frac{\lambda_{k=1,2,p,q}\beta_{k=1,2}}{|r_i - x_{k=2,j,p,q}|} + \frac{(1-\lambda_{k=2,3,p,q})\beta_{k=2,3}}{|r_i - x_{k=3,j,p,q}|} \right\}, \dots, \frac{(1-\lambda_{k=9,10,p,q})\beta_{k=9,10}}{|r_i - x_{k=10,j,p,q}|} \right] \quad (\text{Eq. 4.11})$$

A_{ij} means j -th image charge (Kelvin image charge) interacting with the i -th field point.

$$r_k \leq r_{s,p,q} \leq r_{k+1}, \lambda_{k,k+1,p,q} = \frac{r_{s,p,q} - r_k}{r_{k+1} - r_k}$$

$$\beta_{k,k+1} = \begin{cases} 1, & \text{if } r_k \leq r_{s,p,q} \leq r_{k+1} \\ 0, & \text{other situation} \end{cases}$$

$x_{k,j,p,q}$ is the position of the j -th image charge for one source charge, p is the frame id and q is the atom id in that frame p . For one particular source charge, it will fall into one

of the 11th bins along the radial direction, for example, k -th bin with r_k and r_{k+1} as edges. Then $\beta_{k,k+1}$ will equal to 1. As a result, all the other elements in A_{ij} matrix will be zero because the b will be zero, except for the bin in which the source charge falls. The A matrix will be calculated for each of the q atoms in p frames, totally $p \times q$, and summed together.

Optimize the boundary conditions of the DDM model:

We have two parts of the optimization procedure in our work. One is the optimization for the magnitude of the image charges while keeping their locations unchanged. The other one is to optimize the boundary conditions according to the DDM model. In DDM model, all the water molecules in the simulation box and the buffer zone contribute to the image charges located outside buffer layer for RF. In our optimization, we decide to divide the entire sphere into different shells so that we treat each shell differently.

As shown in Fig. 4.2, in DDM model, outside the buffer layer we have only one sphere $R_3 = D/2 + \tau$, where D is diameter of the TO-box and τ is the thickness of buffer layer. To optimize the boundary, we introduce two more shells with radius R_1 and R_2 . R_1 , R_2 and R_3 are flexible and can be changed to find the optimized combination.

Based on the ODL truncation, to protect the field points inside the TO-box, we use the water molecules outside R_0 for electrostatic forces calculation. $R_0 = D/2 + \delta$, where δ is the H-O bond length for water molecule and set as 0.11 nm.

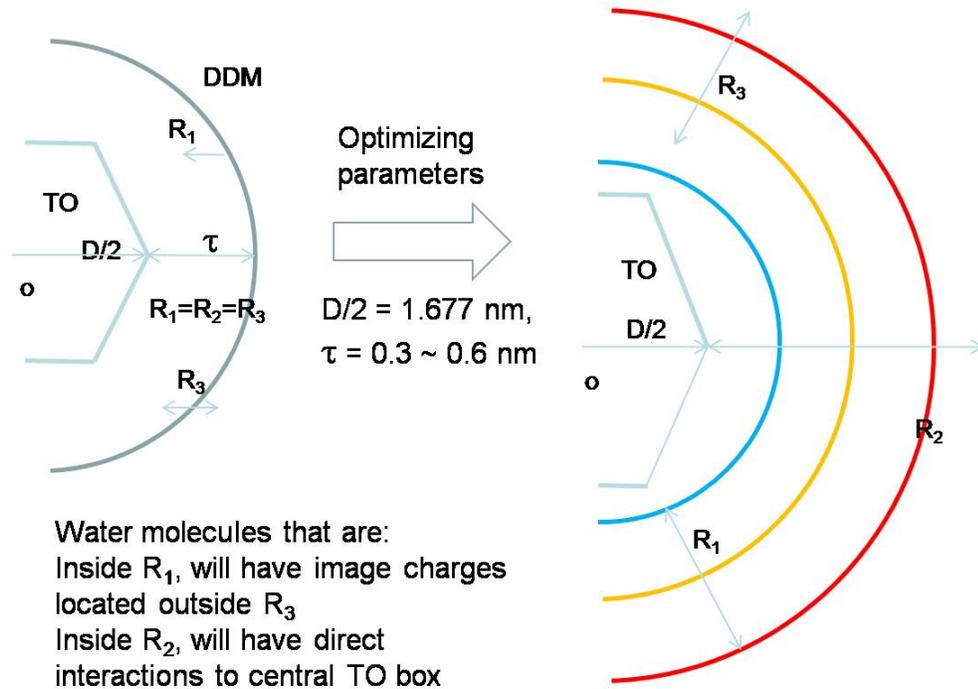


Figure 4.2: The boundary of DDM model is split into three different boundaries to look for the best combination of the R_1 , R_2 and R_3 . R_1 should be larger than TO-box, so $R_1 \geq D/2$. R_2 can change between TO-box and the edge of the spherical cavity, thus $D/2 \leq R_2 \leq D/2 + \tau$. R_3 is for image charge calculation, and should be $R_3 \geq R_1$.

Fig. 4.3 shows how water molecules are treated differently depending on their locations. For a field points inside truncated octahedron box, it can feel both the RF interactions from image charges locating outside R_3 and direct interactions from other water molecule surrounding it inside R_2 . For RF part, water molecules (q_s, r_s) , (q_{s1}, r_{s1}) and (q_p, r_p) are inside R_1 sphere according to ODL truncation criteria and they will generate image charges outside R_3 . For direct interaction, water molecules (q_s, r_s) , (q_{s1}, r_{s1}) , (q_p, r_p) , (q_b, r_b) and (q_{t1}, r_{t1}) will generate direct electrostatic forces on field points.

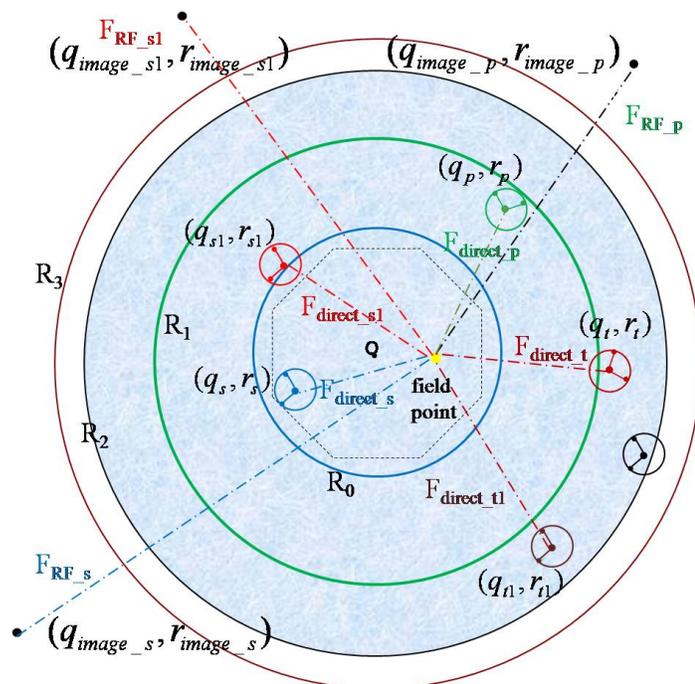


Figure 4.3: How water molecules in different shells contribute to the electrostatic force felt by field point, from direct interaction and RF, respectively, in ODL truncation. Water molecules s , $s1$ and p that are inside R_1 will generate image charges outside R_3 in continuum dielectric medium and contributing to RF. Water molecules s , $s1$, p , t and $t1$ that are inside R_2 will have direct interaction to the field point.

In Fig. 4.4, R_0 is fixed and it only depends on the size of TO-box. For water molecules in the shell between R_0 and R_1 , they have image charges located outside R_3 for RF. So R_1 should not be larger than R_3 . For water molecules in shell between R_0 and R_2 , they have direct electrostatic interaction to the field points inside TO-box. R_2 should not be larger than the size of the entire sphere, $D/2 + \tau$. However, for R_3 , as long as it is larger than R_1 , it can be either smaller or larger than R_2 . Trajectories of all the combinations of the three boundaries R_1 , R_2 and R_3 with different t values are generated.

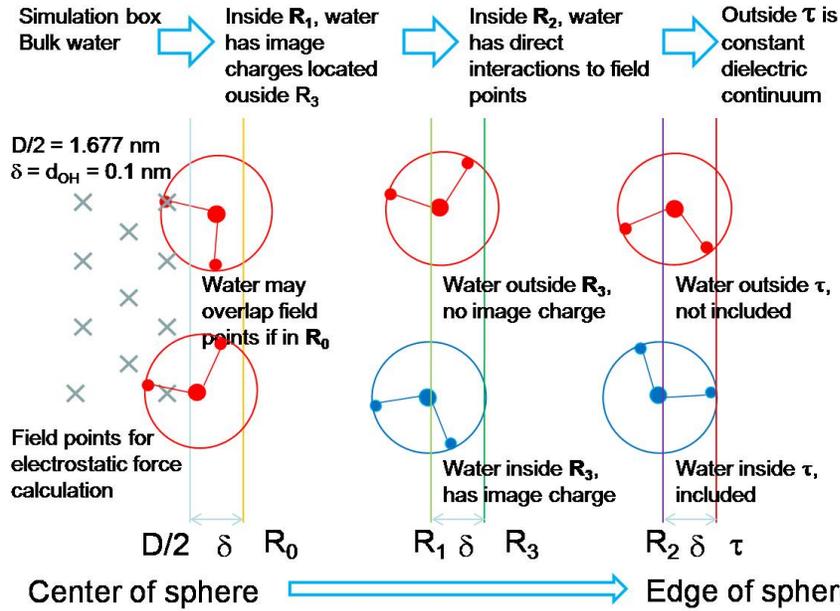


Figure 4.4: How water molecules behavior near the three boundaries based on ODL truncation criteria. The region inside R_0 is bulk water area, in which the field points are located for electrostatic calculation. The area between R_0 and $D/2+\tau$ will be divided by R_1 and R_2 . Atoms inside R_1 will have image charges for RF in bulk water. Atoms inside R_2 will have direct coulomb interactions to bulk water.

$$F_{RF}^{TARGET} = F_{RF}^s + F_{RF}^{s1} + F_{RF}^p,$$

$$F_{direct}^{TARGET} = F_{direct}^p + F_{direct}^t + F_{direct}^{t1} + F_{direct}^s + F_{direct}^{s1} \quad (\text{Eq. 4.12})$$

Iteration process:

In the n th iteration, the optimized image charges calculated from $(n-1)$ -th step will be used to recalculate $V_{estimate}^{(n)}$, until the results converge. The very first step is based on the existing image charge solvation model. The iteration steps are useful in two respects. First, for nonlinear interpolations, such as with polynomial terms, the optimized result cannot be gotten in one step. More steps are necessary to make least square error smaller. Second, our initial dielectric constants for first step are $\epsilon_{in} = 1$ and $\epsilon_{out} = 80$. During the iteration steps we can recalculate ϵ_{out} and optimize again based on the new dielectric

constant. When the system becomes converged, we can get accurate dielectric constant consistently.

4.3 Results and Discussions:

Using large box model to calculate V_{exact} :

The purpose of this study is to reproduce the dielectric potential of the reference solvation model by optimizing the DDM we developed before [39]. To use the LSE method, we should first have the actual potential from the reference model so that we can minimize the difference between the exact potential and our calculated ones.

We developed a simple way to calculate the actual values, which is called the large box method, as shown in the Fig. 4.5.

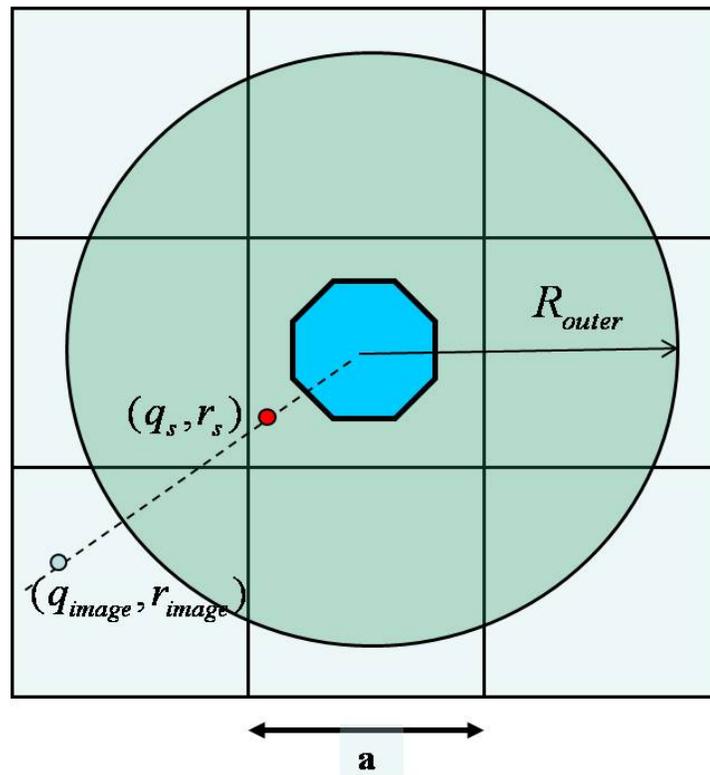


Figure 4.5: The large box model with infinite number of periodic boxes of size a

The truncated octahedron (TO) box shown in blue is the simulation box in DDM model. As mentioned previously, the central region that is far from the boundary of the cavity has bulk water properties. To get the accurate electrostatic potential of bulk water in TO-box, we use DDM model with a very large buffer layer. The entire sphere has a radius of R_{outer} . If R_{outer} becomes infinite, the physical potential of the central box will be very accurate. The space is filled with infinite number of periodic boxes. Inside R_{outer} , the solvent is treated explicitly, while outside is represented by continuum dielectric medium, filled with image charges from explicit water in the cavity.

$$F_{direct}^{R_{outer}} + F_{RF}^{R_{outer}} \Rightarrow F_{exact}^{total} \Rightarrow F_{direct}^{TARGET} + F_{RF}^{TARGET} \quad (\text{Eq. 4.13})$$

The center part of the equation is the actual physical value of the electrostatic potential inside the truncated octahedron box, which is independent on model types. Right part of the equation represents our optimizing model, consisting of electrostatic forces from both RF and water molecules. The optimization needs to use actual physical value as references, however, we cannot get the exactly value directly. So we build a model and try to make the electrostatic forces calculated by this model will be as close as possible to the actual physical value.

Take the PME summation for example as a reference model. The cubic box is simulated with the PME method, which uses periodic boundary conditions. The real simulation area in our optimized DDM model will be the truncated box shown in the center of the sphere. To make this area usable for biomolecule simulation, we need to make sure the bulk water properties are correct. Firstly, according to the definition of periodic boundary conditions, we fill the space with cubic box and its image boxes in three dimensions. In this way, we assume the space is full of an infinite number of water

molecules. Secondly, a spherical cavity is created which contains the truncated box in the middle. Based on the ICSM, the effect of the explicit water molecules outside the cavity could be replaced by serials of image charges formed by all source charges inside the cavity. Let the radius of this spherical cavity be R_{outer} . According to the result of DDM [39], the larger R_{outer} we use, the more accurate dielectric potential we can get in the truncated box in the center of sphere. The dielectric interactions V_{inner} inside truncated area contain two parts: the RF from image charges outside sphere V_{RF} and the direct interaction from the explicit water between truncated area and sphere V_{direct} .

$$V_{inner} = V_{RF} + V_{direct} \quad (\text{Eq. 4.14})$$

Note that when the size of the sphere R_{outer} becomes infinite, V_{inner} is most accurate but this will also increase the calculation time dramatically. In practice, we need to find a suitable size of the sphere for both accuracy and efficiency. In this paper we try to optimize our model to reproduce the dielectric properties of PME method, so we compare the dielectric force in truncated area in different size of sphere with PME results.

Fig. 4.6 shows the variations of electrostatic force with different sizes of spheres and different methods. The unit of force is $10^{-10} N$. 2.5a means the radius of sphere R_{outer} is 2.5 times bigger than the length a in the original cubic box. For each atom inside truncated area, the electrostatic force on this atom will be calculated under different conditions: 2.5a, 1.5a, 1.0a and PME cases. Using the 2.5a case as a reference model we calculate the residuals between the reference and 1.5a and 1.0a cases. From the scatter plot we can see that the residuals of the 1.0a case are widely spread while the 1.5a case is more concentrate near the zero point, showing that the force of the 1.0a case results in

bigger differences than the 1.5a case compared with the 2.5a case. This result suggests that the value of electrostatic force in the truncated area depends on the size of the outer sphere R_{outer} . To check if 2.5a size is big enough to match the PME model, we plot the residuals between 2.5a and PME cases. The right part of Fig. 4.6 shows a good match between the two values of the forces.

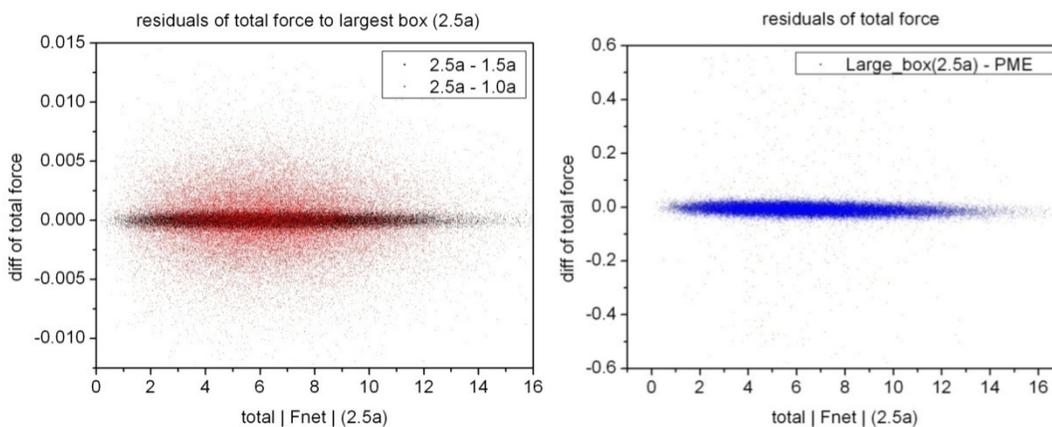


Figure 4.6: Plots of residuals between different sizes of large box models and residuals between large box and PME model. On the left, the accuracy of large box method depends on the size of R_{outer} . On the right, $R_{outer}=2.5a$ case provides a good approximation of electrostatic force to the PME model.

To make sure large box model can reproduce the electrostatic interactions of PME method statistically, we also compared the histogram distributions of electrostatic forces of all the atoms, shown in Fig. 4.7. The agreement of two distributions suggests that the large box model is an effective way to calculate V_{exact} accurately in LSE formula. Furthermore, the large box model can be applied to any other solvation models besides PME.

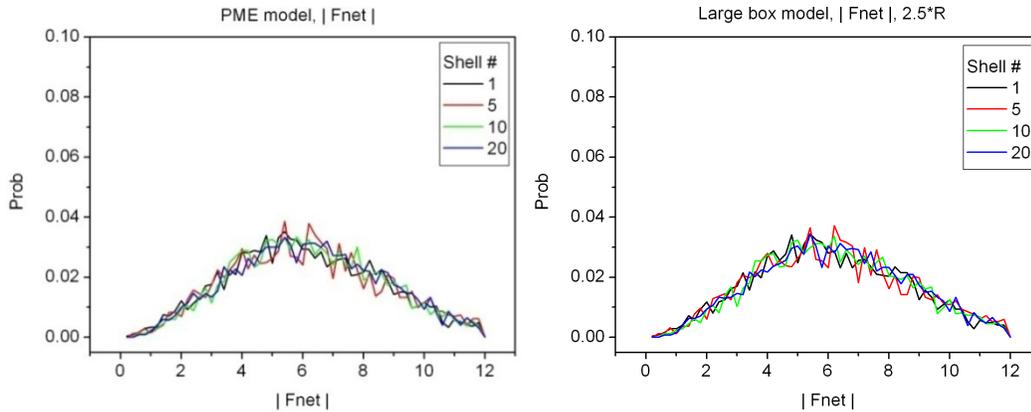


Figure 4.7: Comparison of the histogram distributions of electrostatic forces of atoms in PME and Large box models along the radial direction. The shell 1 is the central sphere, while the shell 20 is close to the edge of sphere. Each shell has the same volume, which means the similar number of water molecules for statistics.

So in our work, when $R_{outer} = 2.5a$, we treat V_{inner} of large box model as the accurate potential V_{exact} in bulk area of the PME model:

$$V_{inner} = V_{exact} \quad (\text{Eq. 4.15})$$

To make the problem simpler, we use a sphere area of radius R_{inner} to replace the truncated box in the center of the system. The R_{inner} sphere is now inside the real simulation box of size a .

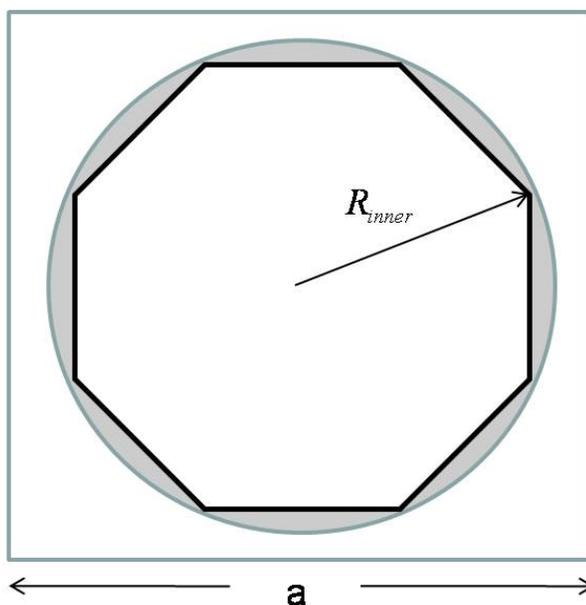


Figure 4.8: The simplified model will be optimized to match PME results. After optimization, the non-physical artifacts of the boundary can be reduced. The thickness of the buffer layer will be minimized, which lowers the computational cost or enlarge the bulk water region in TO-box.

In Fig. 4.8, the RF part of electrostatic potential inside the R_{inner} will be V_{inner} , which we calculate in large box model. According to ICSM, all the source charges inside R_{inner} can generate image charges that are located outside R_{inner} and create RF $V_{estimate}$ in the small sphere. We are trying to match $V_{estimate}$ to V_{exact} . To realize this, we need to optimize the image charge values for the RF so that the final potential will be close to V_{exact} .

Optimizing the model using different interpolation methods:

To evaluate the electrostatic properties inside R_{inner} sphere, we statistically calculate the electrostatic interactions on field points that are distributed uniformly in three directions within the sphere. The size of every grid is 1 angstrom. For each of the field points, we calculate $V_{estimate}$ and V_{exact} . The square value of the summation over the

differences between the exact and estimated potentials for all of the field points will be averaged by the total number, which will be the LSE per field point. Besides the total LSE averaged throughout the sphere, we also check the root mean square (RMS) value of LSE for its radius dependence. The sphere is divided into 5 shells with approximately same volume, which means each of them have similar numbers of statistical field points. The first shell is in the center and the fifth shell is near the edge of the sphere. Because our optimization is based on an existing DDM, we compare our total LSE and RMS values between the DDM and optimized model.

Table 4.1: Shell RMS (LSE) and total LSE of old DDM and new optimized models. Linear interpolation has the minimum RMS value compared to DDM and Polynomials.

Shell #	RMS LSE (DDM)	RMS LSE (linear)	RMS LSE (Polynomials)
1	0.24385	0.14920	0.17047
2	0.26896	0.18143	0.19836
3	0.29776	0.21632	0.22999
4	0.33378	0.25793	0.26889
5	0.36949	0.29796	0.30706
Total LSE	0.08970	0.04799	0.054172

As shown in the Table 4.1, the RMS of the LSE values of the optimized model using linear and polynomial interpolation methods are compared against those obtained with the DDM and. For each of the five shells, optimized models have smaller RMS-LSE and also smaller total LSE averaged throughout all the field points than DDM, which suggests that the new models have smaller averaged deviation from V_{exact} both along the radial direction and totally.

In Fig. 4.9, the results of two different interpolation of u_s are shown. There are a total of 960 frames used in the statistical calculation. We divide the entire frames into two parts, each with 480 frames. We do the optimizing process for different parts of the frames to check the convergence of the system. R_1 is the radius of the spherical cavity and r_s / R_1 gives the relative position of the source charge along the radial direction. The curves from three interpolation ways have similar shapes, suggesting a convergence among the different methods. For each the plot, the three curves have some deviations from each other especially when r_s / R_1 is close to zero, which is near the center of sphere. This could be caused by lack of enough statistical data because there are not enough water molecules close to the central point. When we use 3840 frames, the deviation becomes much smaller. For linear interpolation, there is a big gap between the first ($r_s / R_1 = 0.0$) and second optimizing point ($r_s / R_1 \approx 0.45$). Because there is very few water molecules locating very close to the origin of sphere, we cannot get enough statistics even if we sample more time in this gap area. For polynomial interpolation, because the optimized result is a continuing formula, we can find interpolated u_s values for every position along the radius, as shown in the plot. However, according to the results shown in Table 4.1, the linear interpolation has a smaller total LSE and RMS of the LSE for each shell than the polynomial method.

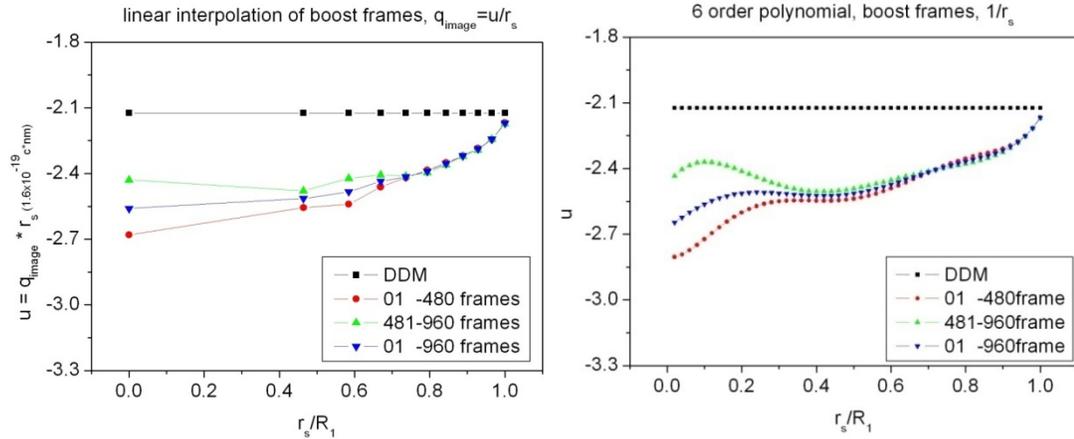


Figure 4.9: The optimized u values from two interpolation methods, using different number of frames. The straight line in black represents the DDM that has not been optimized. The colorful curves show the optimization with different statistics. The beginning parts of the optimized curves, which are near the edge of the spherical cavity, are in good overlap. However, near the center of the sphere, because of the lack of enough water molecules for statistics, the error bars are increasing as approaching to the center.

Check improvement after optimization:

After optimization using 3840 frames, we get a set of optimized u_s values, which can be used to recalculate image charges of every source charge based on DDM model. To test if the sets of u_s can improve the DDM, we apply them to 240 new frames to calculate the RF, which have no correlation with the initial 3840 frames. Again, we divide the sphere into 5 shells. For each shell in 240 frames, the RMS (LSE) value of field points in that shell will be calculated and these 240 RMS LSE values are analyzed by comparing histogram distributions.

For linear, polynomial and sine/cosine interpolation methods we compare their histogram distributions with the one for DDM.

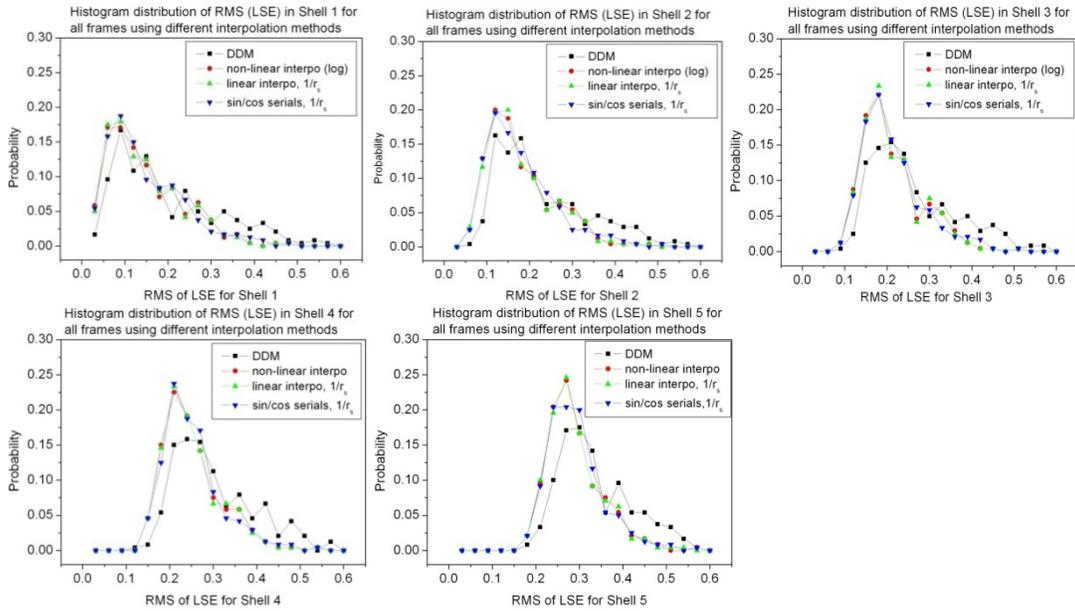


Figure 4.10: The histogram distributions of shell RMS (LSE) show the differences among four models. Near the center of sphere all models are similar. The differences between DDM and the other three optimized models become larger when the shells approach to the edge of the cavity.

The comparison of histogram distribution per shell is shown in Fig. 4.10. For shell 1 and 2, when the field points are close to the center of sphere, the curves of all four models are similar and no obvious improvements are found. For shell 3, 4 and 5, when the field points are close to the edge of sphere (note that because each shell has same volume, the radius of each shell is not located uniformly), there are differences between the three new models and the DDM model. The new models have smaller mean values and variations of the curves. This suggests that our optimizing methods mainly improve the RF near the edge of sphere, which is where the problems are located.

Optimize boundary conditions for DDM model using ODL truncation

Because there are three boundaries to optimized, R_1 , R_2 and R_3 , we fix one of them and change the rest two to find the best combination.

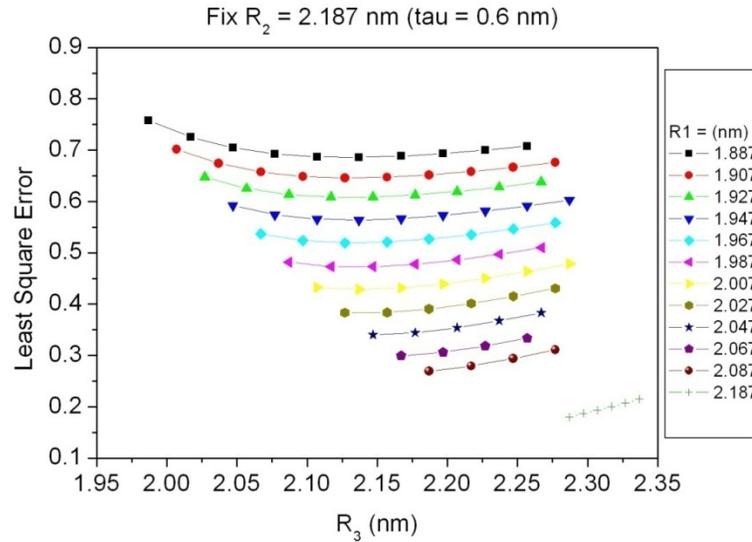


Figure 4.11: The LSE between optimized and target model. When the size of the cavity is fixed (R_2 is fixed), with different sizes of R_1 , there will be a minimum value of LSE for each curve if R_3 has a proper size.

Take $\tau=0.6$ nm for example, make R_2 fixed, and $R_2 = D/2 + \tau - \delta = 2.187$ nm. As shown in Fig.4.11, curves with different colors represent situations with different R_1 . R_3 in x-axis is changing from R_1+d to outside R_2 . The total averaged LSE is shown in y-axis. From the plot we can see the smallest LSE happens on the bottom curve when $R_1=2.187$ nm and $R_3 = 2.287$ nm. This means when R_2 is fixed, R_1 and R_3 should get close to R_2 if minimum LSE can be got.

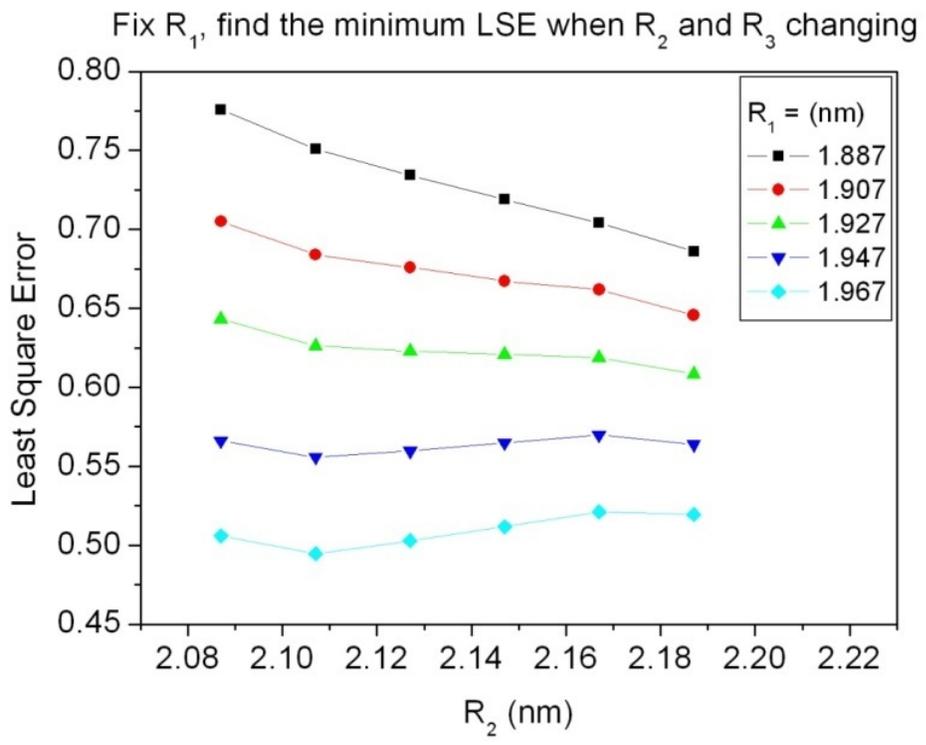


Figure 4.12: R_2 and R_3 will change simultaneously and $R_3=R_2+\delta$. For different R_1 , an optimized value of R_2 and R_3 can be found to get the minimum LSE.

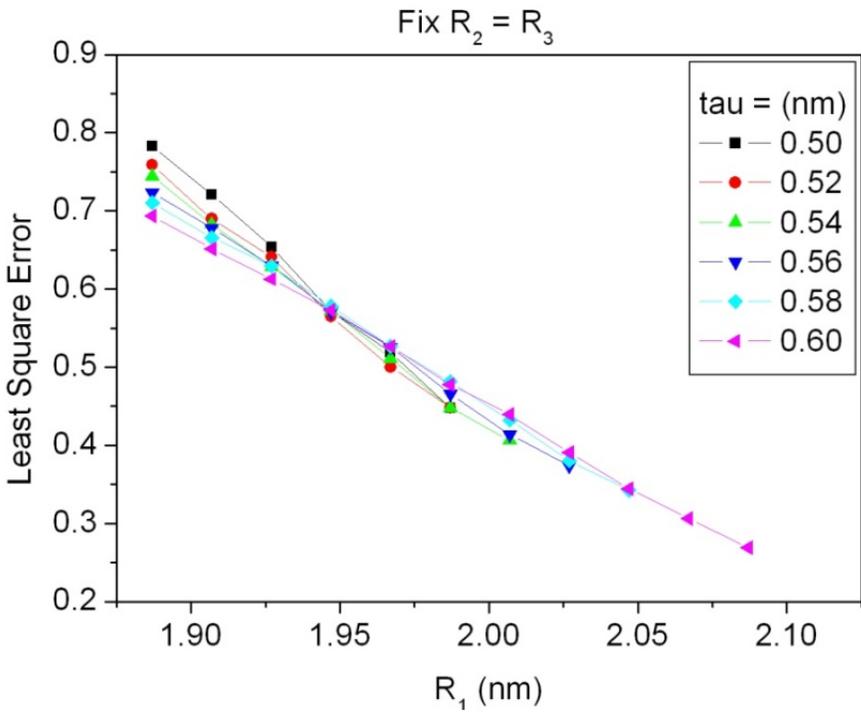


Figure 4.13: When fixing $R_2=R_3$, to get minimum LSE, R_1 should increase to get very close to R_2 .

When R_1 is fixed, R_2 changes from R_1+d to $D/2 + \tau - \delta$, and R_3 always equals to $R_2 + \delta$, the LSE plot is shown in 4.12. The curve with smallest LSE is the curve with $R_1 = 1.967$ nm. Notice that when R_2 is getting bigger, the τ value is also increasing.

In Fig. 4.13 of the third case, if $R_2=R_3$ are fixed to certain value, LSE is getting smaller as R_1 approaches to R_2 . From the three plots shown above, we can conclude that, for ODL truncation, to get the minimum LSE, $R_2 = R_3 - \delta_{OH}$, $R_1 = R_2$.

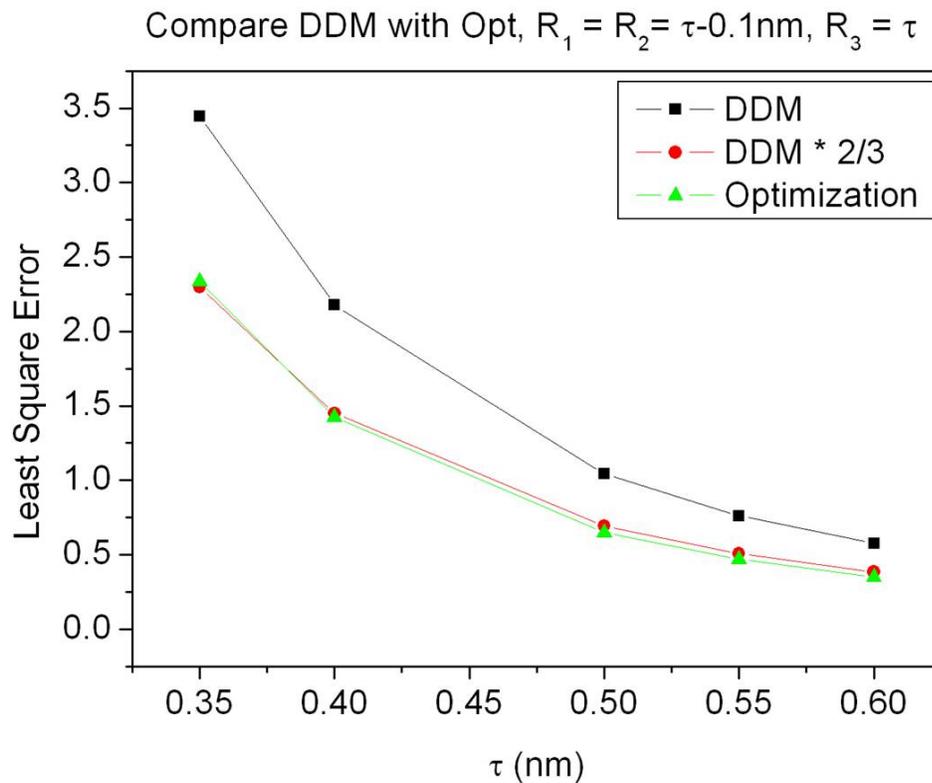


Figure 4.14: How is LSE reduced by optimized model compared with DDM. A 33% decreasing of the LSE has been achieved

Based on the conclusion from boundary optimization, we set $R_2 = R_3 - \delta_{OH}$, $R_1 = R_2$, and compare the LSE between DDM and Optimized model. As shown in Fig. 4.14, with different sizes of the buffer layer (τ value), the LSE of the Optimized model (Green) are almost 2/3 of the original DDM (Black), which is very close to the reference curve (Red)

shown 2/3 of DDM. This means the optimized model has a 33% improvement of LSE compared to DDM.

Interpolation using $TANH()$ function

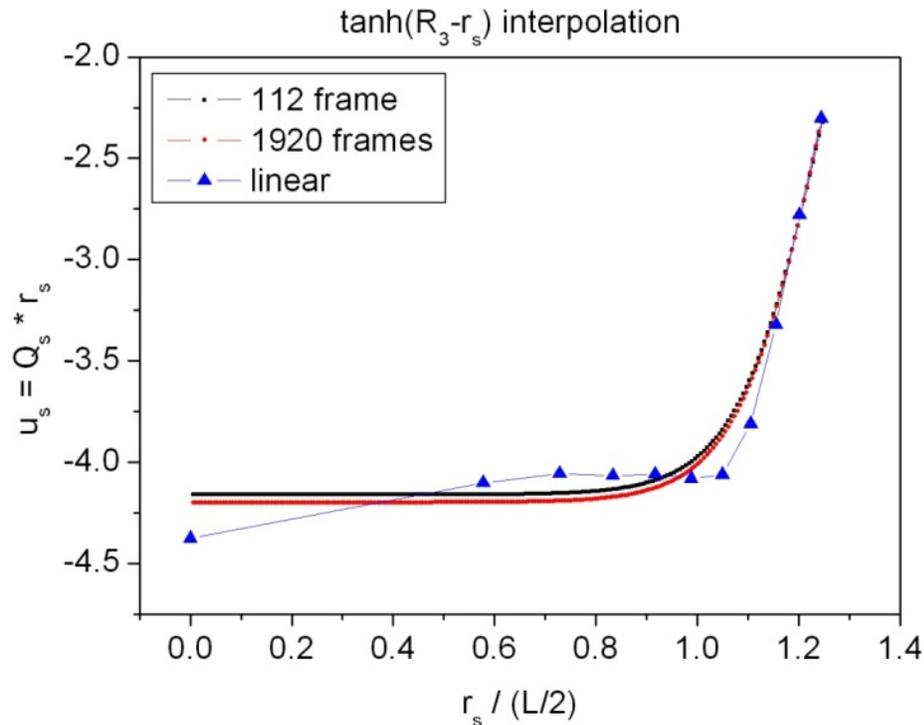


Figure 4.15: Continuous curve using $Tanh()$ interpolation compared with linear interpolation. Although the main shapes are similar, the $Tanh()$ and linear curves do not overlap each other, which leads to the slight differences of LSE.

From Fig. 4.15 we can see that there are small differences between linear and $tanh()$ curve, which could result in slight difference in LSE. Two curves are plot based on different sizes of statistics samples. The results are based on ODL truncation and $tanh()$ function interpolation. The formula is $a_0 + b_0 \times \tanh(c_0 \times (R_3 - r_s))$, where a_0 , b_0 and c_0 are the optimized parameters, R_3 is the radius for image charge location and r_s is the positions of source charges. The large box approximation we developed to calculate the exact electrostatic potential is proved to be effective and accurate. To verify our results, we checked the residuals of electrostatic forces from RF between PME and Large box

approximation. The comparison shows that when the size of the outer sphere is large enough, the central truncated box area could be treated as bulk water and its electrostatic properties are close to the exact value.

Table 4.2: Comparing the LSE of DDM and two interpolation results for different size of the simulation system

TO + t (Å)	a_0	b_0	c_0	LSE(Tanh)	LSE(Linear)	LSE(DDM)
30+2.5	2.143	3.619	3.057	11.781	9.664	12.442
30+3.0	2.185	3.012	3.156	4.969	4.312	6.018
30+4.0	2.242	2.304	3.409	1.574	1.422	2.177
30+4.5	2.268	2.080	3.496	1.018	0.934	1.472
30+5.0	2.304	1.895	3.576	0.702	0.649	1.042
30+5.5	2.340	1.738	3.672	0.501	0.469	0.762
30+6.0	2.375	1.700	3.642	0.370	0.351	0.575

It is shown in the Table 4.2 that after optimization, the LSE of $\tanh()$ and linear interpolations are much smaller than original DDM model. Linear interpolation has a little bit smaller LSE than $\tanh()$. Our image charge optimization is based on an analytical RF model with discrete dielectric constants between explicit and continuous media. In this model, the positions and values of the image charges for RF are determined analytically [39]. The LSE method with different interpolation schemes will give the optimized numerical values of image charges with the positions being fixed. By comparing the total LSE and RMS of the LSE of each shell in TO-box of optimized

results with those of analytical model, we found that the new optimized models have smaller total LSE and variations of error. This suggests the improvement for the new model to reproduce the exact electrostatic properties of bulk water.

When implemented different interpolation methods for image charges, we tried both linear and polynomial ways and compare the two results. The interpolating curves look similar and agree with each other in big shape, although detailed LSE and RMS of the LSE per shell calculations show that linear interpolation gives a closer approximation to the exact value.

Right now we are only optimizing a special case in our model, in which the boundaries of explicit water (buffer layer) and RF are the same size. This is the simplest way for optimization. However, it may not be best. Some problems occurred in our final step of optimization when we implement the optimized image charges back to MD simulation using ICSM. The dielectric constant of the bulk water region is much larger than expected, although all other characteristics including radial distribution function, water density and diffusion coefficient are in proper values with tolerable error bars. The dielectric response is directly related to the average fluctuation in the dipole moment averaged over a small but finite volume. The average dipole moment should be zero within this volume based on symmetry. Surprisingly, a nonzero total dipole moment is found within the spherical cavity, well beyond expected fluctuations, shown in Fig. 4.16. By symmetry of the problem, the average must be zero. This means there is spontaneous symmetry breaking occurring in the model, indicating there is an interesting phase transition like what occurs in liquid crystals, but this is unphysical for water at room temperature.

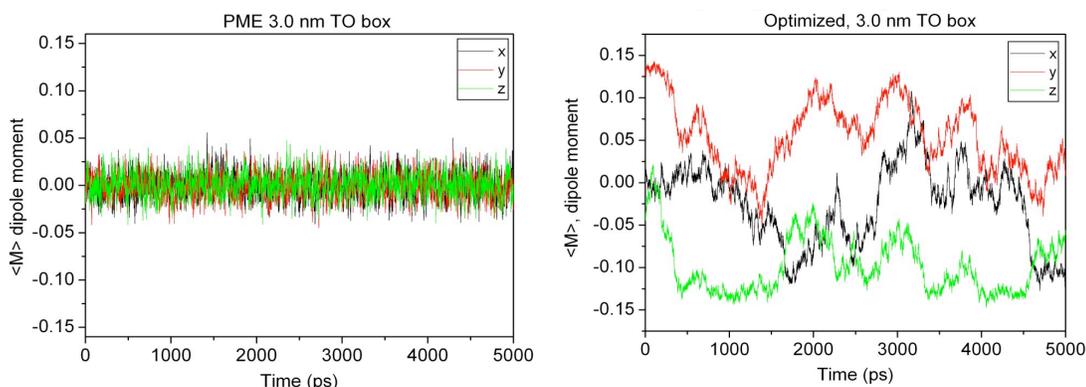


Figure 4.16: Comparison of averaged dipole moment for PME and optimized model

Besides the possibility of a potential bug in the code, which I checked many times, I looked at the problem more carefully to better characterize why this is occurring. What is the cause of the abnormal fluctuations in dipole moment? Why are problems not showing up in any of the other quantities we have calculated?

The most obvious difference between the optimized image charges compared to those from the DDM is that the former are almost twice as big for most of the source charges inside the spherical cavity except those very near the spherical boundary. As an interesting benchmark for a comparison, I arbitrarily increased the image charges in the DDM to see what would happen. Interesting, the same kind of symmetry breaking phenomena occurs. This suggests that a possible cause is that the total charge outside the cavity due to image charges is important, and there might need to place a constraint on their magnitude.

After intensive investigation of the abnormal fluctuations, we found out that it is the boundary conditions of the water molecule that affect the dielectric properties in the optimization model and show the results in Fig. 4.17.

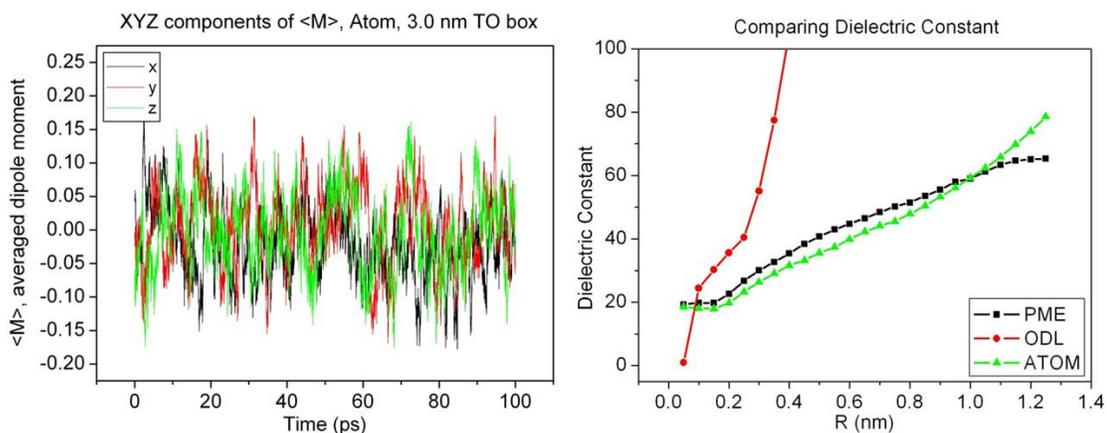


Figure 4.17: A) improved averaged dipole moment using atom based rule and B) Dielectric constant matches PME for atom based situation.

Fig. 4.18 shows that the optimized image charges using atom-based conditions are less negative than those in ODL based case. This matches the prediction we made above, that the non-random orientation of water molecules is caused by too strong interactions from image charges with too bigger opposite charges.

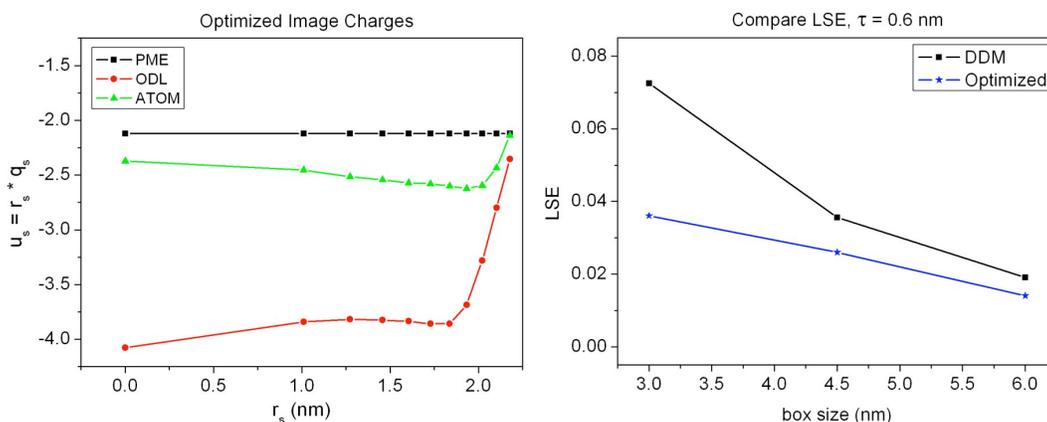


Figure 4.18: A) Modified image charges from ODL and atom based rules. B) More than 25% improvement in reducing the LSE with atom based rule.

Because we are matching our model to PME calculation, we plot the dielectric constant of PME results for a comparison to our optimized results. In Fig. 4.19, the left figure shows how the dielectric constant increases as the volume of sample sphere for calculation using tip3p water [59]. When the size of TO box is small, for example, 3.0 nm,

the dielectric constant keeps on increasing and not converged. This is also the problem we had when analyzing our optimized result based on 3.0 nm PME cubic box. As the box size increasing from 3.0 nm to 6.0 nm, the curve tends to converge near the edge of the box. In comparison, the dielectric constant for tip4p water[59] is smaller and converges faster than tip3p. To prove that our optimizing algorithm works, one possible way is to use larger PME box as reference trajectory for potential calculation. Another way is using tip4p water model rather than tip3p. If either way shows the convergence of dielectric constant using optimized DDM model, that means our optimization works, and it can reproduce the electrostatic properties of the reference potential we chose, including the dielectric constant.

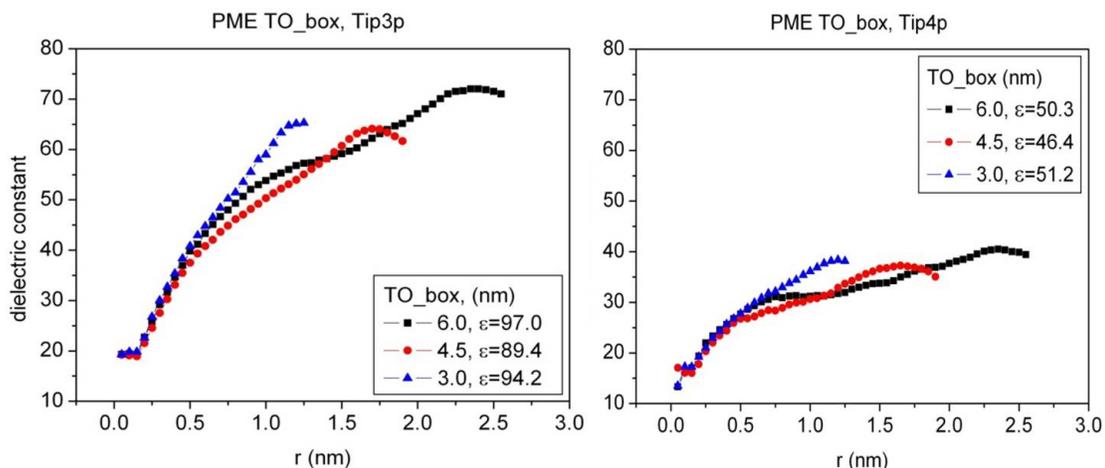


Figure 4.19: How dielectric constants of tip3p and tip4p water model change in a sample sphere with radius r , using PME calculation for different sizes of TO box

On the other hand, we are also improving our optimized model to get a better result. In the beginning, when PME trajectory is used to generate accurate electrostatic potential, we use cubic box, which has a larger size than the simulation box in DDM. As shown in Fig. 4.20, according to the original DDM model, the buffer layer should be filled with imaged water from the Region II, to take care of the short range interaction. However, we

did not do that at the beginning of our projects, instead, we use bigger cubic simulation boxes. Now after repeating all the optimization steps using imaged water in the buffer layer, by using exactly the same size of truncated octahedron boxes, we can compare the two different situations.

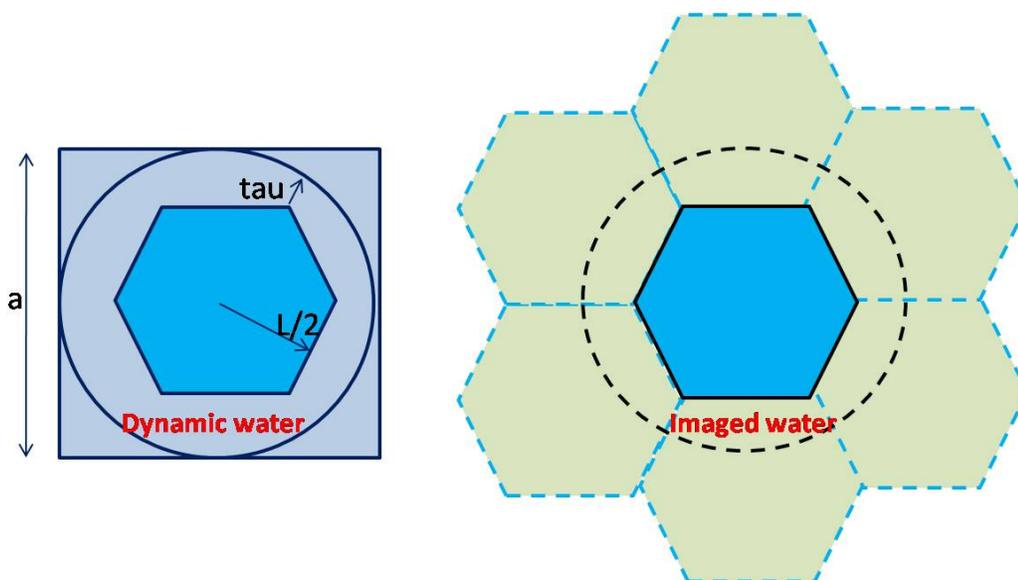


Figure 4.20: DDM model with different kinds of water molecules in buffer layer. Left: the buffer layer is part of the cubic PME simulation box. Right: the buffer layer is part of the imaged TO box generated by periodic boundary conditions.

The result shown in Fig. 4.21 suggests that the imaged water case has a very different optimization curve from the dynamic water. They have large deviation near the center of the sphere, although they share similar slope at the edge of sphere. It turns out this shared region might be very important because the image charges from these water molecules locate very close to the system and have stronger effects to potential inside simulation box. For the image charges of those water near the center of box, their image charges are far away outside the spherical edge, which have less effect on the bulk region. The dielectric constants of optimized results are close to the PME value although near the

edge of the simulation box they are not converged. DDM underestimates the value compared to PME, as shown in Fig. 4.22.

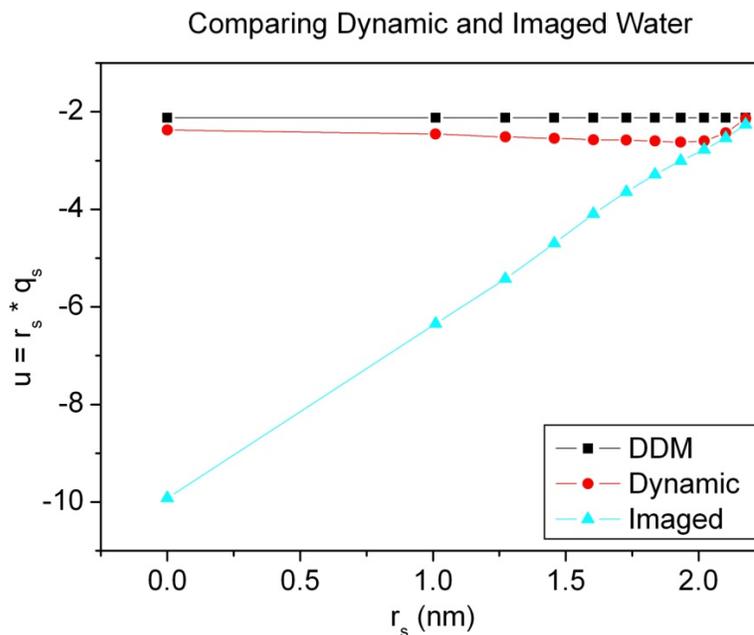


Figure 4.21: The two different optimized results from dynamic (red) and imaged (blue) water in buffer layer, compared with original DDM (black)

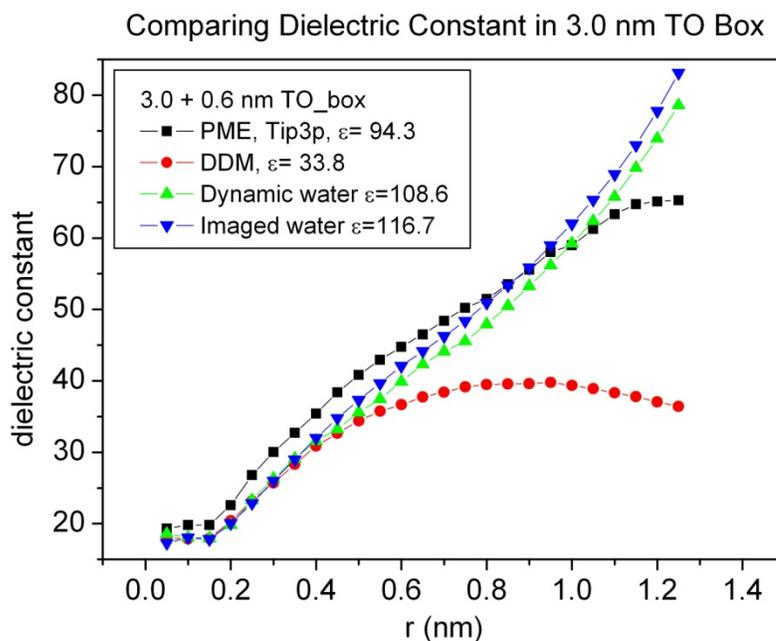


Figure 4.22: Comparison of dielectric constant among PME, DDM and optimized model using dynamic and imaged water.

4.4 Conclusions

Due to the discontinuous dielectric profile in the DDM, surface artifacts is induced and a buffer layer between the explicit solvent and continuum medium is needed to maintain the bulk properties in central productive region. We try to optimize the image charge formula to achieve more accurate RF so that the artifacts are reduced with the new continuous dielectric profile and the buffer layer will shrink for higher efficiency. According to the characterization results in the first project, various physical properties, including electrostatic potentials, forces and torques (force tensors), are used as the accurate reference values for error minimization. With LSE and multiple interpolation methods, a new set of image charges are generated with modified magnitudes and unchanged positions, to reproduce the accurate electrostatic properties from PME. Compared to DDM, the optimize model shows more than 25% reduction of the error matching to PME value when using atom based rule for water. The MD simulation coupled with modified image charges gives the expected continuous dielectric profile, which is close to the profile in PME simulation. However, we realize that the artifacts come from long range and short range interactions together, and the surface effects can drop dramatically only when both of them are optimized. So the buffer layer might be reduced only after short range forces also being optimized. Luckily, a computational procedure to reproduce electrostatic properties with optimized image charge method has been constructed, for selected water model tip3p and OPLS-AA force field. This makes the future work easier on optimization of short range forces and different water models. This project focuses on the most accurate RF ever and tries to reduce the buffer for better efficiency, although further studies are needed to realize that.

CHAPTER 5: OPTIMIZATION OF BOUNDARY CONDITIONS IN ICSM

5.1 Introduction

The important role played by the RF and buffer layer was demonstrated in chapter 3. In chapter 4, the critical dependence on the molecular boundary rules was demonstrated, and a general method to optimize the image charge model tuned to specific water models was developed. Unfortunately, in practice the optimized image charges do not result in increased computational efficiency. This means, the overall efficiency is still less than we would like. However, if accuracy for arbitrary solutions could be demonstrated, then the ICSM as described in chapter 2 would be a superior model compared to PME, at least for large systems consisting of 30,000 atoms or more [54]. When the ICSM is used to simulate high ionic concentration solutions, non-uniform density of ions occur due to correlations between the ions and their periodic images. One way to avoid this problem is to increase the cavity radius by a factor of 5 or more. However, this will make PME the most accurate and most efficient method available for any practical problem of interest, and the ICSM will never be the model of choice.

Because of the PBC in the ICSM, unphysical correlations between ions that appear at finite ionic strength severely shrink the productive volume. The ions have to be constrained inside productive region I, defined in Fig 2.3, to avoid penetration into region II and producing extra images of ions in region III. This makes the model inefficient because region II is still simulated but can be used for production.

The objective of this project is to replace the PBC in ICSM with NPBC for higher efficiency and accuracy in ionic simulation. This new focus creates a dilemma because all previous hybrid models used non-periodic boundary conditions (NPBC), and several papers starting from 1983 continuing on beyond 2008 claim that the construction of a good hybrid model has been solved. However, it should be noted that none of these prior models were demonstrated to work in a scaled up version for a general purpose MD simulation. Armed with a good understanding of the RF and the role of the buffer layer, it was in our view warranted to make an attempt to create such a hybrid model with NPBC, albeit this has been proven to be a challenging problem based on our literature review, and the lack of demonstrated-results on large size systems from all prior models published to date.

Several models have been devised and implemented using non-periodic spherical cavities, where care is given to reproduce the bulk properties of water within the sphere, with minimum amount of surface effects. For example, Brooks and Karplus used the mean field force approximation (MFFA) to generate a soft boundary force and a stochastic buffer region [57]. King and Warshel developed the surface constrained all-atom solvent (SCAAS) model to deal with the surface polarization effect [32]. Juffer and Berendsen treated the atoms near the surface boundary dynamically with a position restraint to represent the effects of atoms from the outside surface [60]. Abagyan and Totrov developed a modified image charge approximation (MIMEL) with a position-independent correction [15]. Petraglio *et al.* developed a simpler model for short range forces involving a van der Waals (or purely repulsive) force from the wall, and simulated

bulk water and acetonitrile [50]. Brancato *et al.* introduced explicit solvent shells to ensure the bulk behavior of solvent near the boundary [61].

The implementation of a new ICSM having non-periodic boundaries shares similar aspects with these previous models. Unlike previous work, we demonstrate that the results from our new ICSM with NPBC are valid for both small and large systems. Previous works were focused on solvation properties of small molecules and did not consider extending their method to large systems. We found early on that just about any reasonable implementation will work very well at small system sizes, but as the system size increases, the model most likely will fail miserably if even one critical physical property isn't right. In other words, sensitivity to the exact model details matters more as the system size increases, which surprised us. What we found is that the properties of the buffer layer must resemble that of bulk water with a high accuracy otherwise these errors propagate into the productive volume.

Several novel ideas were implemented as described below in detail, of which some were similar to past works and others very different. At the end, we developed a working model—demonstrated for the first time in the field—that reproduces the bulk properties of water in the productive volume that would be obtained using PME for various system sizes. Moreover, the simulation can be performed in the NVT and NPT modes, where the latter constant pressure simulation case (NPT) has never been done before. In particular, the model requires a buffer layer and constraint forces are applied to the water molecules in this buffer layer. For constant pressure simulations, the spherical wall is movable. The performance of the new ICSM in simulations of TIP3P water has been carefully studied,

and all structural, static, and dynamic properties with different systematic parameters match the PME results markedly well.

5.2. Methods

5.2.1 Computational Details

The Tip3p [59] all-atom model is used to characterize the bulk water properties in this work. The multiple image-charge method[40-42] and the fast multipole method (FMM) [20-22] are used to calculate the electrostatic potential efficiently. The positions and velocities of the particles are generated according to the velocity Verlet algorithm coupled with Nosé–Hoover thermostat [62]. The dielectric constant of the external field is set to 80 and the cutoff distance of van der Waals interaction is 1.0nm. The simulations are all performed over 4 ns, at 250, 300 and 350 K, using three different sphere radii of 3.0, 4.0 and 5.0 nm.

The diffusion constant for bulk water, D , is calculated by expressing the Einstein relation [63] as a random walk, and then track autocorrelations in displacement vectors. We perform the averages over all water molecules that have their center of mass within a sphere, for example, a radius of 2.5 nm for a 5.0 nm sphere. These molecules are followed for all time afterward, which form the set A .

$$\lim_{t \rightarrow \infty} \left\langle \left\| r_i(t) - r_i(0) \right\|^2 \right\rangle_{i \in A} = \lim_{t \rightarrow \infty} \left\langle R_i^2(t) \right\rangle_{i \in A} = 6Dt \quad (\text{Eq. 5.1})$$

Let $t = N\Delta t$, then $\vec{R}(N) = \sum_{k=1}^N \vec{d}_k$ where \vec{d}_k is the k -th displacement of the center of mass of a given water molecule in time Δt . The k -index defines a random walk step, such that $k=1$ is the first step, $k=2$ is the second step, and so forth until $k=N$, which ends the

random walk of the water molecule. The mean squared displacement $\langle R^2(t) \rangle$ in Eq. 5.1 is rewritten as:

$$\langle R^2(N) \rangle = \left\langle \sum_{k=1}^N \vec{d}_k \cdot \sum_{j=1}^N \vec{d}_j \right\rangle = \sum_{k=1}^N \sum_{j=1}^N \langle \vec{d}_k \cdot \vec{d}_j \rangle = \sum_{k=1}^N \langle d_k^2 \rangle + 2 \sum_{k=1}^N \sum_{j>k}^N \langle \vec{d}_k \cdot \vec{d}_j \rangle \quad (\text{Eq. 5.2})$$

then the diffusion constant is given by:

$$D(N) = \frac{\langle d_1^2 \rangle}{6\Delta t} + \frac{1}{3\Delta t} \sum_{k=2}^N \left(1 - \frac{k-1}{N}\right) \langle \vec{d}_1 \cdot \vec{d}_k \rangle \quad (\text{Eq. 5.3})$$

The moments are obtained by averaging over all water molecules in set A and over time. Once the autocorrelation is calculated, it is clear when Eq. 5.3 converges, and at what time this takes place. As such, this method is accurate and gives assurance when there is enough sampling data. We found that 4 ps are more than sufficient time to obtain converged results. Moreover, the results are not dependent on the arbitrary choice of Δt , for which we used 0.2, 0.4, 0.8, 1.0 and 2.0 ps for different Δt as a check. Therefore, despite our simulations have finite boundaries, the size of all simulation cells we used were large enough to handle the correlation time of the diffusing water.

5.2.2 Model Details

ICSM with Finite Boundaries

In recent papers [9, 39], we introduced the Image Charge Solvation Model (ICSM), which is a hybrid solvation model combining both explicit and implicit solvent representations. The ICSM consists of a spherical cavity of explicit solvent embedded in a continuum dielectric medium and can solve the electrostatic interactions accurately and efficiently for large systems. The continuum dielectric medium generates a reaction field (RF) on the explicit solvent, which is computed using the multiple image-charge method

[40-42], and fast multipole method [20, 21]. The ICSM model has been tested in simulation of pure water and ion solvation. The bulk properties for pure water are found to be virtually identical to those obtained using PBC with PME for large systems [39]. For ion solvation, the model gives superior accuracy for simulation of mixed media where the charging free energy of a single sodium and potential of mean force for ion pairs were calculated [9]. A key advantage of the ICSM is that the water molecules in this buffer region require no dynamical equations to solve.

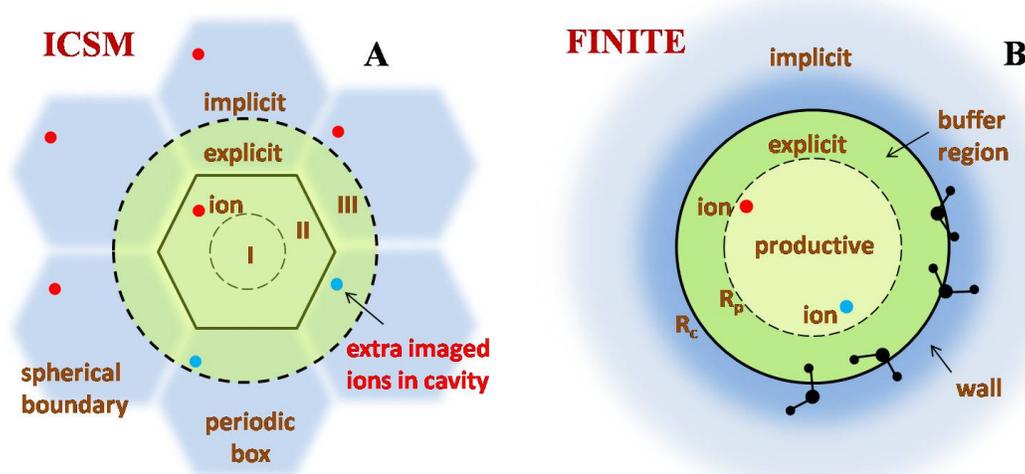


Figure 5.1: A) ICSM with periodic boundary condition. In the center it is the truncated octahedron (TO) simulation box, and the surrounding are the periodic images of the central box. The spherical cavity contains not only the whole TO box (Region I and II), but also parts of the periodic neighbors (Region III, as a buffer layer with periodic water), which are generated from Region II of the original TO box. If an ion moves from Region I to Region II, extra periodic images of the ion will appear in Region III immediately, causing non-physical ion correlations and inconsistent total charges in the system. B) Finite boundary condition model. A buffer layer with dynamic water is introduced to isolate the central productive region from the implicit solvent outside. The water molecules follow the group based boundary condition near the surface. The wall potential will be calculated using mean field approach. Ions can now move inside cavity without worrying about correlation problems caused by periodic boundary conditions.

In spite of the success of the ICSM described above, it has limitations for more general cases. As shown in Fig. 5.1 A the buffer layer (region III) is between the

truncated octahedron (TO) simulation box and spherical boundary, which is generated from explicit solvent in region II by the PBC. For simulations involving diffusing ions, phantom ions appear/disappear in the buffer layer when ions enter/leave region II from the productive region I. Phantom ions introduce unphysical correlations that are spatially persistent and they create fluctuations in the total charge of the system. A possible resolution to the problem is to reduce the size of the productive region, but already the productive region compared to regions II and III is small. For a TO box of 4.5nm in length, the diameter of region I that allows simulations of solute molecules is less than 1.7 nm, which means most of the computational time is used to calculate non-usable regions.

The advantage of replacing the PBC with a finite boundary eliminates the problem of phantom ions, and the efficiency is greatly improved. Unfortunately, removal of the PBC requires treating water molecules in the buffer region with dynamical equations of motion (Fig. 5.1 B). As such, both the RF and van der Waal forces from solvent outside the spherical cavity need to be modeled explicitly, which presents difficulty due to the discontinuous jump in the dielectric constant, and the switch from an explicit all-atom model to a continuum model [43]. In the following subsections, several steps are described that were implemented to solve the problems that arise from finite boundaries. These steps are (i) developing appropriate molecular boundary rules; (ii) calculation of an effective spatially dependent density of solvent outside the spherical cavity; (iii) using this effective density, determine the mean van der Waal force on atoms inside the cavity; (iv) define a diffusive force to maintain constant atomic density within the buffer layer; (v) define a torque on electric dipoles to maintain proper orientation of water molecules

in the buffer layer; (vi) define a spatially distributed thermostat to maintain constant temperature in the buffer region; (vii) incorporate a self-adjusting movable spherical cavity wall to maintain constant pressure, and (viii) location of image charge cavity and buffer layer thickness to best reproduce bulk water properties. Note that in this work, it is the PME results pertaining to tip3p water model that is considered to define “true” bulk properties.

(i) Molecular Boundary Rules

The center of mass of a water molecule is used as a point of reference when applying constraint forces and torques. When the number density is calculated within a sample volume, such as a shell, the center of mass is required to fall inside the sample volume to be counted. This means that water molecules near a boundary will typically have one hydrogen atom inside the sample volume and the other hydrogen atom outside. Furthermore, when taking the radial component of the dipole moment of a water molecule, the ray extending from the origin of the spherical cavity to the molecule’s center of mass defines the radial direction. Note that because an effective vdW force is employed (described below) to control confinement of water molecules, the center of mass of a water molecule will frequently exceed the spherical cavity radius, R_c . This allowance is critical to maintain uniform bulk density of oxygen and hydrogen atoms separately within the physical domain, defined by the spherical cavity up to R_c .

(ii) Effective Density outside Spherical Cavity

The effective density of water molecules in the continuum medium is calculated based on having uniform density of bulk water inside the spherical cavity, which represents the physical domain. Demanding uniform density outside the cavity is not

physically relevant, and would lead to a mean van der Waals force from the implicit solvent that creates a depletion layer within the physical domain. Since outside the cavity is not the physical domain, the only requirement is that its properties create atomic forces on the water molecules inside the cavity to ensure bulk properties are simulated. Therefore, the depletion region in density should be on the outside of the spherical cavity. This spatially dependent density will be reflected in the Radial Distribution Function (RDF) for bulk water, which will be consistent with having uniform density of bulk water inside the spherical cavity. The effective density is calculated using the RDF within a mean field treatment. The free energy G_{xy} between two atoms is given as:

$$G_{xy} = -k_B T \ln(g_{xy}(r | T)) \quad (\text{Eq. 5.4})$$

where k_B is Boltzmann constant, T is the temperature of the system, g_{xy} is the RDF between two types of atoms x and y . Generally, adding free energy of all possible pairs of atoms in the system overestimates the total entropy. However, the additivity assumption is equivalent to neglecting higher level of correlations within a Mean Field Approximation (MFA). Employing the MFA, the total free energy G_x for atom x at a certain position is given by:

$$G_x(\vec{r}) = -k_B T \sum_y \ln(g_{xy}[|\vec{r} - \vec{r}_y|, T]). \quad (\text{Eq. 5.5})$$

For an oxygen atom located outside R_c , the total free energy G_O contributed from all oxygen and hydrogen atoms that are uniformly distributed within the spherical volume is given by:

$$G_O(r) = -k_B T \left(n_O \iiint_{V_{in}} \ln[g_{OO}(|r\hat{r} - \vec{r}'|)] d^3r' + n_H \iiint_{V_{in}} \ln[g_{OH}(|r\hat{r} - \vec{r}'|)] d^3r' \right) (\text{Eq. 5.6})$$

where n_O and n_H are the number densities of oxygen and hydrogen atoms of bulk water. Since correlations between O-O and O-H atoms disappear after a distance of 1.5 nm, the integrals were performed numerically as shown in Fig.5.2 A. Once the excess free energy relative to a bulk value on the outside of the cavity is calculated, the effective number density of oxygen at distance r from the center of the sphere is given as:

$$n_{eff}(r) = n_w \exp\left(-\frac{G_O(r)}{k_B T}\right) \quad (\text{Eq. 5.7})$$

where n_w is bulk water density.

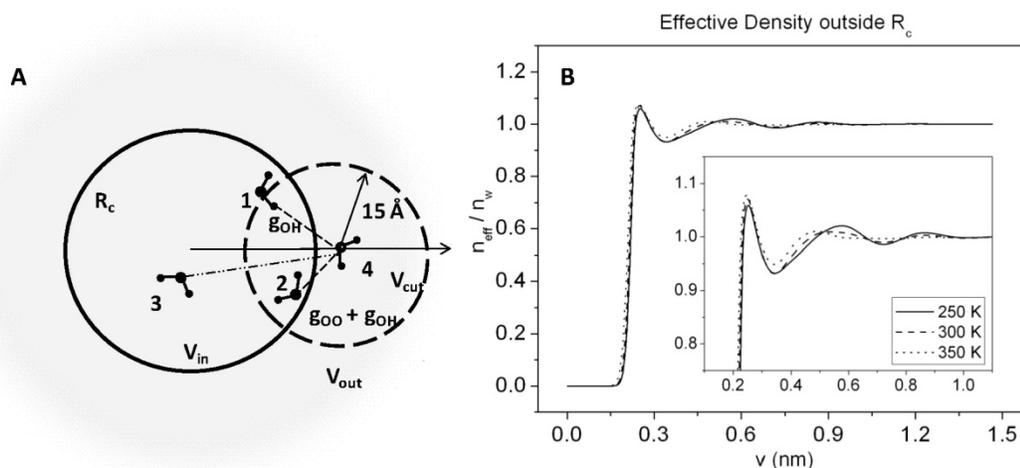


Figure 5.2: A) The way to calculate the effective density of water outside R_c . Only the atoms inside the crossing region between physical domain V_{in} and cutoff sphere V_{cut} (eg. one hydrogen atom in water 1 and all atoms in water 2) have effective contributions to water density in V_{out} . Neglecting the interaction from water 3 which is beyond the cutoff distance can improve the efficiency of calculation and will not change the final result of effective density. B) The effective density curve outside R_c at different temperatures. v is the distance from R_c to any point outside cavity along the radial direction.

In Fig. 5.2 B, starting from R_c , the relative density of water is zero until about 0.15 nm before increasing to around 1.0, because water molecules on both sides of the boundary R_c cannot get too close to each other. At different temperatures, the shapes of the effective density curves are similar, with only minor variations about 1. According to

Fig. 5.2 B, a step function might be a sufficiently accurate approximation for the effective density profile outside the spherical cavity. This simplification was evaluated (described below), and we conclude that an appropriate model only requires a uniform bulk density of oxygen everywhere inside and outside of the spherical cavity, except for a narrow shell just outside the spherical cavity that contains no oxygen.

(iii) Effective van der Waals Force inside Spherical Cavity

The net vdW force acting on an oxygen atom (using tip3p water model) at position \vec{r} inside the spherical cavity is given by:

$$\begin{aligned}\vec{F}_{vdW}(\vec{r}) &= \iiint_{V_{out}} \vec{F}_{LJ}(|\vec{r} - \vec{r}'|) n_w d^3 r' \\ &= \iiint_{V_{out}} \vec{F}_{LJ}(|\vec{r} - \vec{r}'|) n_w d^3 r' + \iiint_{V_{in}} \vec{F}_{LJ}(|\vec{r} - \vec{r}'|) (n_w - n_w) d^3 r'\end{aligned}\quad (\text{Eq. 5.8})$$

where \vec{F}_{LJ} is the 12-6 Lennard-Jones (LJ) potential for tip3p water model. The volume integral over entire space is zero, and Eq. 5.8 can be simplified and calculated numerically:

$$\vec{F}_{vdW}(\vec{r}) = -\iiint_{V_{in}} \vec{F}_{LJ}(|\vec{r} - \vec{r}'|) n_w d^3 r' \quad (\text{Eq. 5.9})$$

Because the force is conservative and the system is spherical symmetric, x, y or z axis that goes through the center of sphere is the same as each other. The effective force must be radial direction when \vec{r} is along \hat{z} :

$$F_{net\ x} = F_{net\ y} = 0 \quad \text{and} \quad F_{net\ z} = n_w \iiint_{V_{in}} \left(\frac{\partial U_{vdW}}{\partial z} \right) d^3 r' \quad (\text{Eq. 5.10})$$

Define two types of integrals:

$$A_n(r) \equiv \iiint_{V_{in}} \frac{1}{|\vec{r} - \vec{r}'|^n} d^3 r' \quad \text{and} \quad B_n(r) \equiv \iiint_{V_{in}} \frac{\vec{r}' \cos \theta}{|\vec{r} - \vec{r}'|^n} d^3 r' \quad (\text{Eq. 5.11})$$

Solve explicitly for both terms:

$$C_n(r) \equiv rA_n(r) \quad (\text{Eq. 5.12})$$

The solution will be:

$$F_r(r | R, n_w) = 6n_w \varepsilon \sigma^6 \left(2\sigma^6 (B_{14}(r) - C_{14}(r)) + (C_8(r) - B_8(r)) \right) \quad (\text{Eq. 5.13})$$

where $F_r(r | R, n_w)$ is radial the net force and

$$B_{14}(r) = \frac{\pi}{5280r^2} \left(\frac{11R^3 - 121R^2r + 33Rr^2 - 3R^3}{(R-r)^{11}} - \frac{11R^3 + 121R^2r + 33Rr^2 + 3R^3}{(R+r)^{11}} \right)$$

$$B_8(r) = \frac{\pi}{120r^2} \left(\frac{5R^3 - 251R^2r + 15Rr^2 - 3R^3}{(R-r)^5} - \frac{5R^3 + 25R^2r + 15Rr^2 + 3R^3}{(R+r)^5} \right)$$

$$C_{14}(r) = \frac{\pi}{660} \left(\frac{11R+r}{(R+r)^{11}} - \frac{11R-r}{(R-r)^{11}} \right) \text{ and } C_8(r) = \frac{\pi}{60} \left(\frac{5R+r}{(R+r)^5} - \frac{5R-r}{(R-r)^5} \right)$$

Fig. 5.3A shows the effective vdW forces calculated from implicit solvent outside the spherical cavity. When oxygen atoms are close to the center of the sphere, the vdW force due to the solvent outside the spherical cavity is negligible. The effect of the net mean vdW force from the solvent is critical for water molecules near the spherical cavity boundary. As a water molecule approaches the boundary, there is first an attractive force reaching a maximum value of about $8.0 \text{ kJmol}^{-1}\text{nm}^{-1}$ at a distance of 0.11 nm from the boundary, and the net vdW force rapidly switches from attractive to repulsive. The repulsive force near the boundary forms a “wall”, which confines the explicit water molecules to the physical domain. Notice that there is very little difference in the effective vdW force when calculated by the actual effective density compared to the step function density in Fig. 5.3 B. Moreover, this difference is small for any temperature between 250K and 400K, suggesting that the essential characteristic required to

incorporate into the model is a gap from the spherical cavity, which has been employed in most previously models [33, 56, 61, 64].

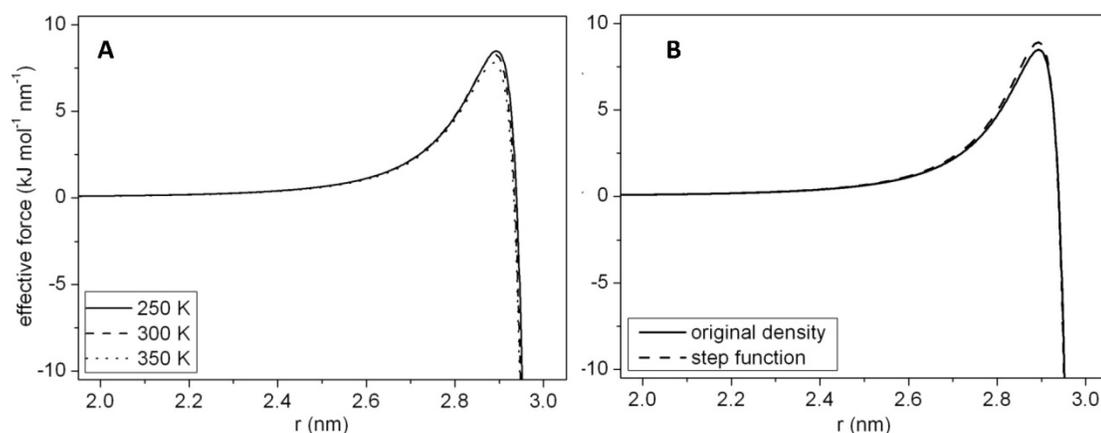


Figure 5.3: The effective vdW forces calculated from A) effective water density at three different temperatures and B) effective density and step function curve at 300 K

(iv) Diffusion Force for Density Control

While we found that the mean vdW force calculated above is essential to maintain uniform bulk density of water, it was not sufficient. A depletion layer of water molecules forms within the physical domain near the spherical cavity boundary, and excess water increases the bulk density by ~10% from the correct density in the productive volume. Relaxing the location of the spherical shell does not in any way improve the density profiles. To reduce the surface effects from the boundary, we introduced a self-adapting diffusive force in the buffer layer. Other models have implemented manipulating forces of various kinds to maintain uniform density within the spherical cavity [50, 56, 64-66]. However, in this work, the manipulating forces are restricted to the water molecules in the buffer region, preserving the productive region. The buffer region is divided into multiple shells, with varying thickness so that each shell has the exact same volume, holding approximately 100 water molecules. For each time step, the density of water in each shell is calculated and compared with the bulk water density within the productive

volume. If a density gradient is found between consecutive shells, a diffusion force will be applied to drive the center of mass of the water from high density to low density shells. A collection of Lagrange multipliers are introduced for all shells within the buffer layer for the diffusion force as:

$$S_{\rho}(i) = S_{\rho}(i) + \alpha_{\rho} \frac{n(i-1) - n(i+1)}{n_w} \quad (\text{Eq. 5.14})$$

$$F_{\rho}(i) = S_{\rho}(i)F_0$$

where $S_{\rho}(i)$ is the Lagrange multiplier for the i -th shell, α_{ρ} is an empirically adjusted relaxation constant, n_w is the bulk water density of the productive volume, and $n(i \pm 1)$ is the number densities for the $(i \pm 1)$ -th shells. F_0 is a convenient reference force to set a reasonable scale, which is set to be $8.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. The Lagrange multipliers are accumulative and self-corrected with a short enough relaxation time for the system to adapt to density fluctuations. The diffusion force $F_{\rho}(i)$, which equals to the unit force times the Lagrange multiplier, is the same for each water molecule in the same shell. The diffusive constraint force attempts to maintain uniform bulk density within the buffer layer, which is within the physical domain.

(v) Torques for Molecular Orientation

For finite boundary conditions, a common behavior of the water molecules near the boundary is that the water dipole prefers to point outwards [55], rather than randomly orientated as bulk water. This is because part of the electrostatic interactions is missing due to the absence of water outside cavity. The discrete change in dielectric constant that forms the basis of the image charges and the calculation of the RF is a mathematical idealization that also creates unphysical effects [43]. As such, finding a way to restore the

orientation of the water molecules to be that of bulk water within the buffer layer is required, and has been noted to be problematic [32, 33, 67]. The orientation of a water molecule is characterized only by its electric dipole moment. To maintain the random orientation of water molecules, constraint torques are introduced in the buffer region to maintain the radial dipole moment to be that of bulk water. By symmetry, there is no reason why the dipole moments directed perpendicular to the radial direction would be anything but random, and this was confirmed in simulations. As such, the constraint torques only couple to the radial direction of the dipole moments of the water molecules.

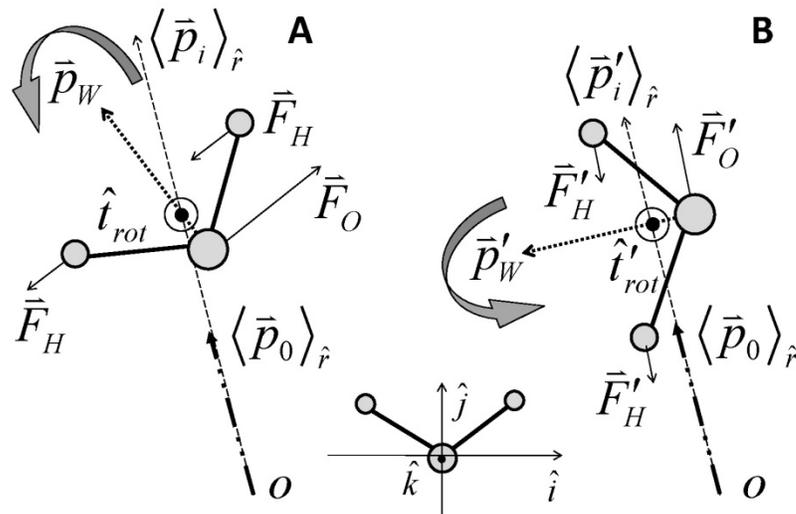


Figure 5.4: 2D schematic plot that shows how external torque force rotates a single water molecule. A) before and B) after rotation.

The setup of how torques are applied is given in Fig. 5.4. $\langle \bar{p}_i \rangle_{\hat{r}}$ and $\langle \bar{p}_0 \rangle_{\hat{r}}$ are the radial components of the actual and expected dipole moment in i -th shell based on COM, \hat{p}_W is the dipole moment of a water molecule in i -th shell. \hat{t}_{rot} is the direction of torque that is applied to the COM, \vec{F}_O and \vec{F}_H are the torque forces on Oxygen and Hydrogen atoms. O is the central point of the sphere. As shown in Fig. 5.4 A, when $\langle \bar{p}_i \rangle_{\hat{r}}$ is larger

than $\langle \bar{p}_0 \rangle_{\hat{r}}$, the water molecules inside the i -th shell need to be rotated to reduce the averaged radial dipole moment. The axis of rotation \hat{t}_{rot} is defined as:

$$\hat{t}_{rot} = \begin{cases} \hat{i} \cos(\varphi) + \hat{k} \sin(\varphi) & \forall |\sin(\theta_{rp})| < 0.01 \\ \frac{\hat{r} \times \hat{p}_W}{|\hat{r} \times \hat{p}_W|} & \text{otherwise} \end{cases} \quad (\text{Eq. 5.15 a})$$

For normal cases, the axis equals to the cross product of dipole moment of a water molecule and the radial direction that is defined using the center of mass of the molecule. For special cases that the radial direction \hat{r} and water dipole \hat{p}_W are almost parallel or anti-parallel to each other ($|\sin(\theta_{rp})| < 0.01$), a rotation matrix is used to avoid singularities, where \hat{i} , \hat{j} and \hat{k} are defined as orthogonal internal coordinates of the water molecule shown in Fig. 5.4. Rotating with a torque about this axis is the most efficient way to move the dipole vector to change its radial component. The direction of the torque forces must be calculated. This direction must be perpendicular to the rotation axis and the dipole moment direction. Thus, this direction is given by: $\hat{d}_F = \hat{t}_{rot} \times \hat{p}_W$. The magnitude of the force is controlled by the parameter λ :

$$\lambda = \left(\left(\chi (\hat{p}_W \cdot \langle \bar{p}_i \rangle_{\hat{r}}) + 1 \right) / 2 \right)^2, \quad \vec{F}'_0 = \langle \lambda F_0 \rangle_{\hat{d}} \quad (\text{Eq. 5.15 b})$$

where $\chi = \text{sign}(\langle p_0 \rangle_{\hat{r}} - \langle p_i \rangle_{\hat{r}})$. $\chi = -1$ when the argument of sign is negative, and $\chi = 1$ otherwise, including zero. Fig. 5.4 B shows that after the rotation, the averaged radial dipole moment should be smaller than before and the orientation of the water molecule is changed.

The torque is broken down into a force component that is perpendicular to a moment arm that is always of constant length. As such, a torque is expressed as torque-

force, which allows us to directly compare different constraint forces applied on the water molecules within the buffer.

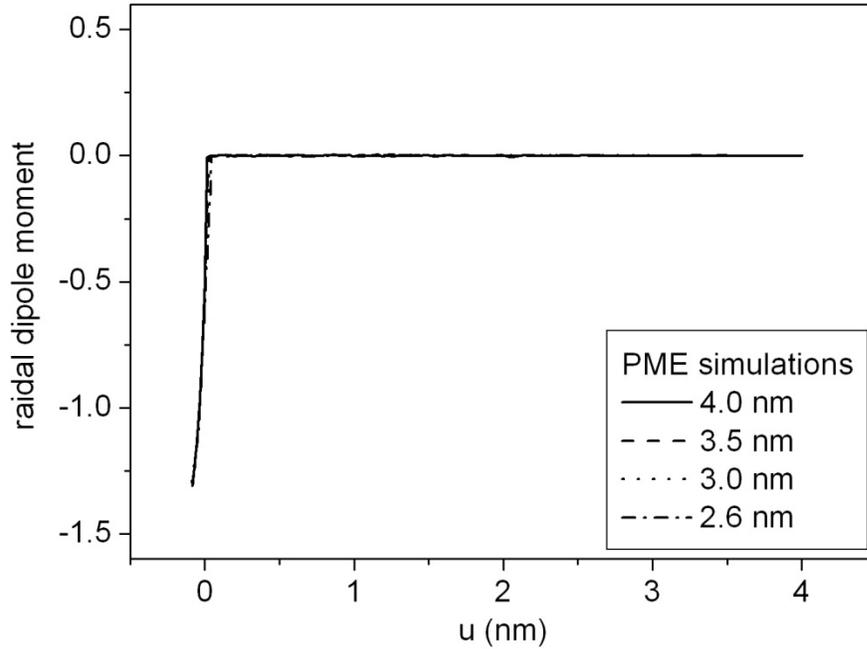


Figure 5.5: Target values for the radial dipole based calculated from PME trajectories of different sizes of boxes. They all overlap each other near the boundary without size dependence.

Being that the torque is applied on the center of mass of a water molecule, only its orientation is affected without translating its center of mass. The Lagrange multiplier for dipole control is:

$$S_M(i) = S_M(i) + \alpha_M (\langle p_i \rangle_{\hat{r}} - \langle p_0 \rangle_{\hat{r}}) \quad (\text{Eq. 5.15 c})$$

$$F_M(i) = F_O(i) = S_M(i) F'_0$$

where $S_M(i)$ is the Lagrange multiplier for the i -th shell, α_M is a relaxation rate, $F_O(i)$ is the force on oxygen atom, $\langle p_i \rangle_{\hat{r}}$ is the radial component of the average dipole moment of all the water molecules in the i -th shell, and $\langle p_0 \rangle_{\hat{r}}$ is the expected averaged radial dipole moment of bulk water. In Fig. 5.5, we show the target values for the radial dipole based

on PME results. Notice that the average dipole moment is zero everywhere inside R_c , but water molecules can move past R_c . The radial dipole moment is not zero there, because the radial dipole moment is only calculated based on the water molecules that have at least one atom within the spherical cavity. As such, this gives a net radial dipole moment that is negative. Bulk properties of explicit water should reproduce the PME results, and this target is only calculated once. Actually, it is very difficult for the constraint torque-forces to succeed in matching the target dipole moment for $r > R_c$. Nevertheless, the attempt of driving the orientation of the water molecules to align with what bulk properties would have (for the same tagged particles) helps bring non-bulk properties to bulk properties most rapidly, meaning the smallest possible buffer layer. The torque-force $F_M(i)$ is the same for each water molecule in the same shell.

(vi) Thermostats for Temperature Control

Because the diffusion and torque forces applied on water molecules introduce external work to the water molecules within the buffer layer, the kinetic energy of the particles in the buffer will increase dramatically. Having a single thermostat control keeps the average temperature correct. However, a temperature gradient forms, and the temperature of the water within the productive region is much lower to compensate for high temperature in the buffer region. In order to solve this problem an independent thermostat assigned to each shell in buffer layer maintains constant temperature. A single thermostat is assigned to the productive region.

(vii) Moveable Boundary for Pressure Control

To perform constant pressure simulations within the productive region, the boundary R_c is allowed to vary. The calculation of pressure proved to be simple and

sufficiently accurate by simply applying the macroscopic definition of pressure on the bulk water contained in the productive region, which is a spherical volume of radius, R_p . Note that $R_c = R_p + \tau$ where τ is the thickness of the buffer region, which remains constant. The productive volume will fluctuate as the spherical cavity defined by R_c varies. While imperfections can be tolerated in the buffer layer, the productive region must simulate true bulk properties. Therefore, pressure is calculated based on the productive volume, and not the total volume of the simulation. Making use of the spherical geometry, hydrostatic pressure is given by F_r / A where the radial component of the total force, F_r , on the system (everything inside the sphere) from the universe (everything outside the sphere), where A is the surface area of the sphere. Since all forces between atoms within a molecule cancel out, only the net forces on the center of mass of the water molecules need to be calculated. The total force by the system on the sphere (the reaction force) is given by:

$$F_r(R_p) = - \sum_i^{N(R_p)} \vec{F}_{cm,i} \cdot \hat{r}_i \quad (\text{Eq. 5.16})$$

where R_p is the radius of the spherical productive region. Technically, all internal forces between molecules inside the productive region cancel out as well, but $\vec{F}_{cm,i}$ is a force already known at no additional computational cost. The change of R_c is calculated as:

$$\Delta R_c = \alpha_p \frac{P_0 - P}{|P|} \frac{R_p^2}{2R_c} \quad (\text{Eq. 5.17})$$

where α_p is a relaxation rate and P_0 is the target reference pressure, which is 1.0 bar here.

(viii) Location of Image Cavity and Buffer Layer Thickness

In the analytical solution of the method of multiple image charges, the locations of the image charges start from the edge of the spherical boundary to infinity [41, 42]. The position of the Kelvin image charge (same as for a conductor) r_k is defined as $r_k = R_{RF}^2/r_s = R_c^2/r_s$, where r_s is the position of source charge in the cavity and R_{RF} is the radius of dielectric sphere for image charge calculation. However, due to the group-based rule, the water molecules inside the cavity can move across the boundary R_c into the region where image charges are generated. To avoid singularities between explicit and image charges, for those source charges that are very close to the boundary, the R_{RF} have to be shifted outwards for a distance b . Similar procedures have been seen to prevent the too strong electrostatic force to break the simulation when the particles approach the boundary, such as surrounding the droplet model with vacuum buffer [56, 64], applying cutoff for forces [34] and using a buffer with explicit solvent [39]. The position of the image charge is defined as:

$$\begin{cases} r_k = R_c^2/r_s & , \text{if } r_s \leq R_p \\ r_k = (R_c + b)^2/r_s & , \text{if } r_s > R_p \end{cases} \quad (\text{Eq. 5.18})$$

For atoms that are inside and beyond the productive region, the R_{RF} for their image charge calculation is R_c , and $R_c + b$, respectively. For the results reported in this paper, the shifting distance b for the image charges is 0.15 nm.

According to our study in this work, we tried other formulas similar to Eq. 5.18 except they allowed b to change continuously (a linear interpolation starting from R_c) or the value of b and its derivative changed continuously (a quadratic interpolation starting from R_c). We also tried a $\tanh()$ function to transition from 0 to b . It did not appear that

any of these models produced better or worse results after an optimized b was found for the model in question, but the discontinuous case presents the easiest model to reflect why the buffer region thickness must be a function of R_c . The image charge solution is scale invariant. This means the electrostatic force accuracies can be expressed as percentages of where the source charge is in relation to R_c . As R_c increases, the atoms near the boundary of the productive region and the buffer layer will be influenced differently by the electrostatic forces, unless the ratio R_p/R_c is fixed. As such, we let the accuracies decrease within the buffer region, but remain a constant within the productive region. After running multiple simulations and comparing their results, an optimal proportion of 1:5 for the thickness of buffer layer over the radius of sphere has been used. For 3.0, 4.0 and 5.0 nm spheres, the thicknesses of buffer layer are 0.6, 0.8 and 1.0 nm, respectively. In Table 5.1, the numbers of solvent atoms that can be simulated inside productive region of the ICSM and finite boundary model are compared, based on the bulk water density at 300 K, which shows a much bigger usable region can be achieved in the new model.

Table 5.1: Compare the number of atoms simulated in the productive regions of ICSM and finite boundary model, within different sizes of spheres. The thickness of the buffer layer is 0.6 nm for ICSM/PBC and $R \times 20\%$ for ICSM/NPBC

Sphere Size R (nm)	ICSM/PBC (# of atoms)	ICSM/NPBC (# of atoms)	Speeding Ratio
3.0	154	5930	29
4.0	848	13440	12
5.0	2460	26251	8

5.3. Results and Discussions

To validate the performance of the ICSM with finite boundaries, key properties of the solvent are characterized such as; relative number density; mean radial dipole moment; the RDF; fluctuation of dipole moment and diffusion constant. These bulk properties are investigated with different systematic parameters, radii of the spherical cavity and temperatures. In addition, the Lagrange multipliers used for the constraint forces are monitored and the temperature profile of the entire simulation box is monitored. Beyond time averaged quantities, instantaneous quantities such as all three dipole moment components, volume fluctuations and pressure variations are monitored and compared to corresponding quantities using PME method. The PME method [16] combined with PBC is used as a reference to define the “exact” bulk properties to evaluate the capability of the new model to obtain bulk properties of water within the productive region. Each PME trajectory comes from 10 ns simulation in NPT ensemble after equilibration, performed by the GROMACS program package [58, 68].

The requirement of the proposed model is to provide bulk properties as accurate as PME within the productive region, where solute will be solvated and simulated. Taking advantage of spherical symmetry, properties of water molecules are calculated as a function of distance away from the origin in the form of shells, to check how homogeneous these properties are throughout the sphere. That is, the spherical cavity is divided into concentric shells, each of them with the same volume so that each shell contains approximately the same number of water molecules, typically between 100 and 150.

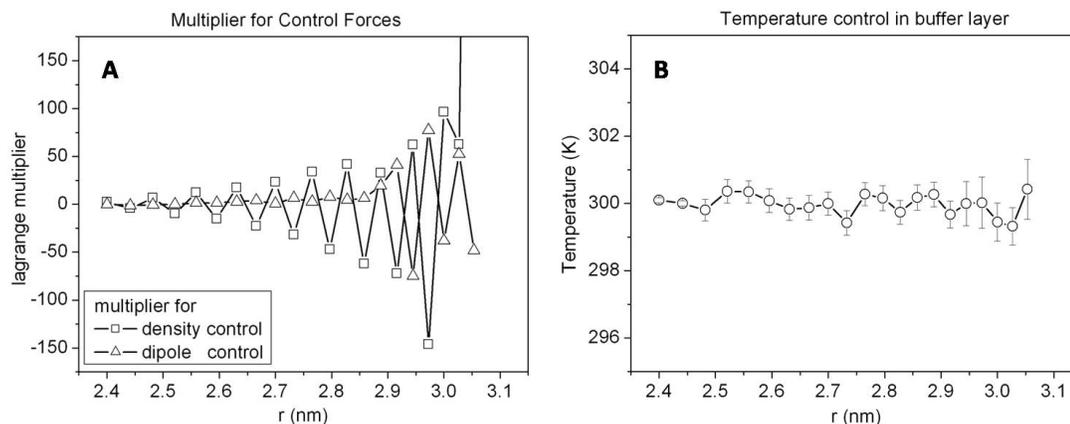


Figure 5.6: A) The Lagrange multipliers for diffusion and torque force in the buffer region of a 3.0 nm sphere. The diffusion force changes in opposite directions between two neighboring shells and the magnitude increases gradually when approaching to the boundary. The torque force is much smaller compared to diffusion force and remains small value for most part of the buffer region. Its sudden increase suggests that there is a noticeable change in the orientation of water dipoles near the surface. B) The temperatures of different shells in the buffer layer after multiple thermostats at 300 K.

For a particular shell, the properties, eg, density of water in this shell, are calculated and averaged over all the frames from the MD trajectory, and the same procedure is applied to all shells. Based on the center of mass location, each water molecule belongs to one and only one shell for a given frame, although it transverses shells along a trajectory. To evaluate the multiple constraints inside buffer region, the calculations of the behaviors of Lagrange multiple for constraint forces, fluctuation of the multiple temperatures in the buffer and the time dependent pressure in productive region have been performed. Fig. 5.6 A shows how Lagrange multipliers change from the edge of productive region R_p to the boundary of the system R_c for both density and dipole constraint forces in a sphere of 3.0 nm radius. When the water molecules are close to the productive region ($r = 2.4$ nm), the multipliers are relatively small and close to zero. As the molecules are getting closer to the boundary, the magnitudes of the multipliers are increasing gradually, suggesting that bigger constraint forces are necessary to maintain

the correct density and orientation of the water near the surface. In general, from R_p to R_c in the buffer layer, stronger surface effects require stronger control forces to manipulate water molecules. As shown in Fig. 5.6 B, in the 3.0 nm sphere, the very first data point ($r = 2.4$ nm) represents the average temperature in the productive region, and the rest ones represent temperatures in different shells in the buffer layer. The total average temperature of the system is around 300 K after multiple thermostats. An example of the pressure calculation using the formula above is shown in Fig. 5.7, which gives an average value of 11 bar in 4.0 nm sphere, compared to 21 bar got from PME.

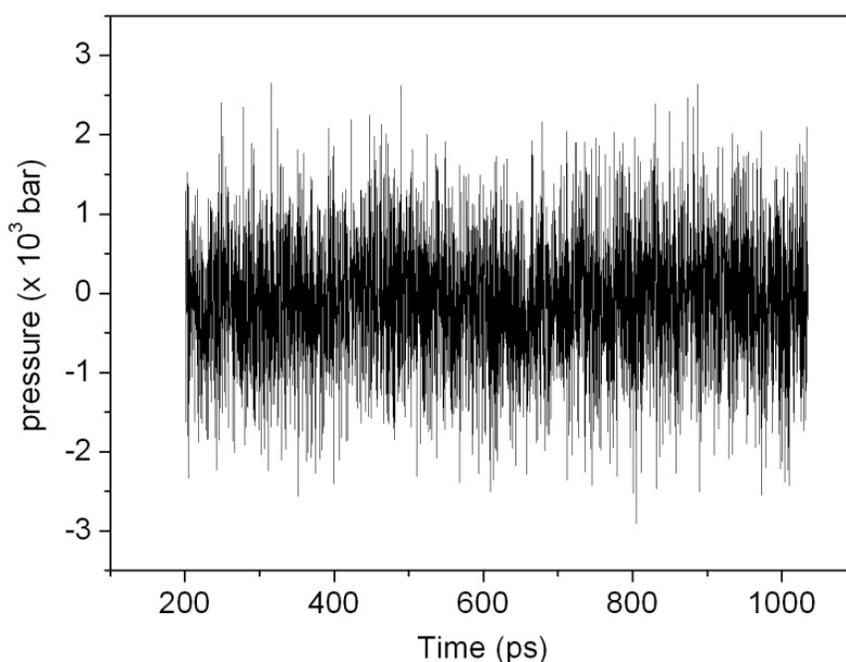


Figure 5.7: Pressure calculation in the productive region of 4.0 nm sphere at 300 K, with first 200 ps trajectory thrown away.

As shown in Fig. 5.8, in different sizes of systems, Lagrange multipliers for density and dipole constraint forces have similar behaviors. From the boundary of cavity ($u=0$) to the edge of productive region towards the center of sphere, the magnitude and shape of the curves are similar for 3 sizes of systems. For density controlling in Fig. 5.8 A, there

might be a minimum depth in the buffer layer required by the diffusion forces to effectively adjust the density of water. When approaching to the productive region of sphere, eg, $u=1.0$ nm, the magnitudes of multipliers drop to zero. However, no strong position dependence is observed for the multipliers or torque forces. Only near the surface of the boundary R_c , larger dipole control forces are needed to maintain the orientation of the water dipoles, as shown in Fig. 5.8 B.

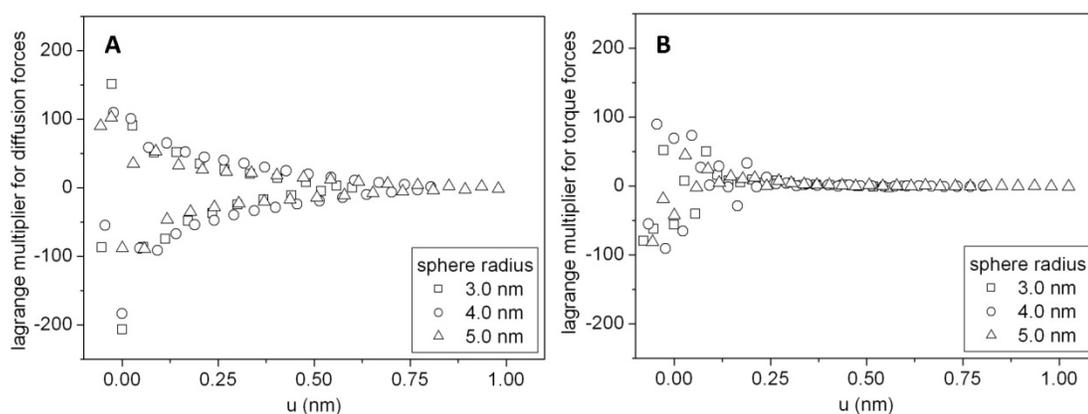


Figure 5.8: Size independence of the Lagrange multiplier for A) diffusion forces and B) torque forces. u is defined as the distance from R_c to the center of sphere and make it clearer to compare the properties near the boundary in different cases. For example, for a sphere with 3.0 nm radius, $u = 0$ nm is the position right at R_c , $u = 0.6$ nm is the boundary between buffer and productive region and $u = 3.0$ nm is the origin of the sphere.

a. Demonstrated Uniform Density in Productive Region

Fig. 5.9 A compares the results before and after applying density control constraint forces in the buffer region. In the sphere with a radius of 5.0 nm, without constraint the relative water density near the center of sphere is 8 % higher than bulk density and the fluctuations are much bigger near the boundary. With density control forces, the density in each shell within the productive region matches the PME density well, and the buffer layer has much smaller fluctuations. In Fig. 5.9 B, to check the size independence of the model, densities in three different sphere sizes are calculated and agree well with the

PME result. The non-accurate parts in the density curves start to become appreciable in the buffer layer, but the errors increase near the cavity radius, R_c .

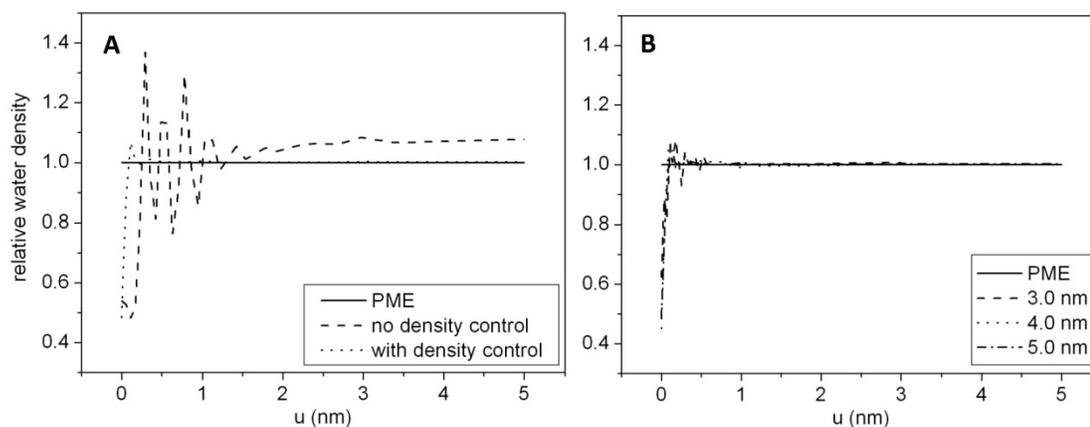


Figure 5.9: A) Relative density of water in simulations with and without diffusion forces in 5.0 nm sphere. B) With diffusion forces, the density matches the PME curve and does not depend on the sizes of the spheres.

b. Demonstrated Randomly Orientated Dipoles

As explicit water molecules from physical domain approach to the finite boundary of the cavity, their dipole moments tend to point outwards. Fig. 5.10 A shows that without dipole control forces, the water molecules in buffer region which are close to the surface have strong orientation preference and point to opposite directions in two consecutive shells. Due to the characteristics of long-range interactions, inhomogeneous arrangement of the water dipoles in buffer layer definitely affect the water dipoles in productive region, which results in the non-random orientation of water in the center of sphere. However, after applying dipole control forces, not only the radial dipole moment near the center of sphere is close to zero, but also the fluctuations of radial dipole in the buffer region are much smaller. In Fig. 5.10 B the radial dipole moments in three different sphere sizes are calculated and they all agree with the PME value in their productive regions. Again, noticeable deviation from expected value only happens in the

buffer layer. For the smallest simulation box, the water dipole moment in the productive region is most affected by the strong surface effects.

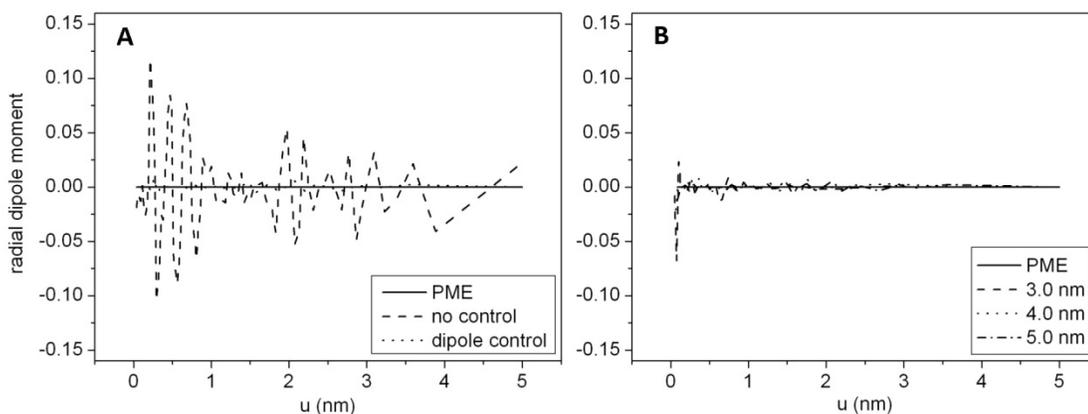


Figure 5.10: A) Radial dipole moment of water in simulations with and without torque forces in 5.0 nm sphere. B) With torque forces, the radial dipole moment matches the PME curve and does not depend on the sizes of the spheres.

c. Demonstrated Bulk Water Structure

The bulk water structure is examined by calculating the RDF for three different size systems. Generally, all RDF curves follow the shape of PME results well, with small deviations, suggesting that the structure of the water solution is not very sensitive to the system size. However, although the curves are all similar to the PME results, when zoomed in the slight differences become apparent. In Fig. 5.11, the RDF curve for the 3.0 nm spherical cavity was observed to have the biggest deviation from the PME curve. The larger size spherical cavities of 4.0 and 5.0 nm have RDFs that match better to the expected values. As the size of the system becomes larger, on one hand, the thickness of the buffer layer is also bigger in proportion, which allows the constraint forces to be activated in a larger volume and makes it easier to match the bulk water properties. On the other hand, bigger system size allows the central productive region be farther away from the surface artifacts.

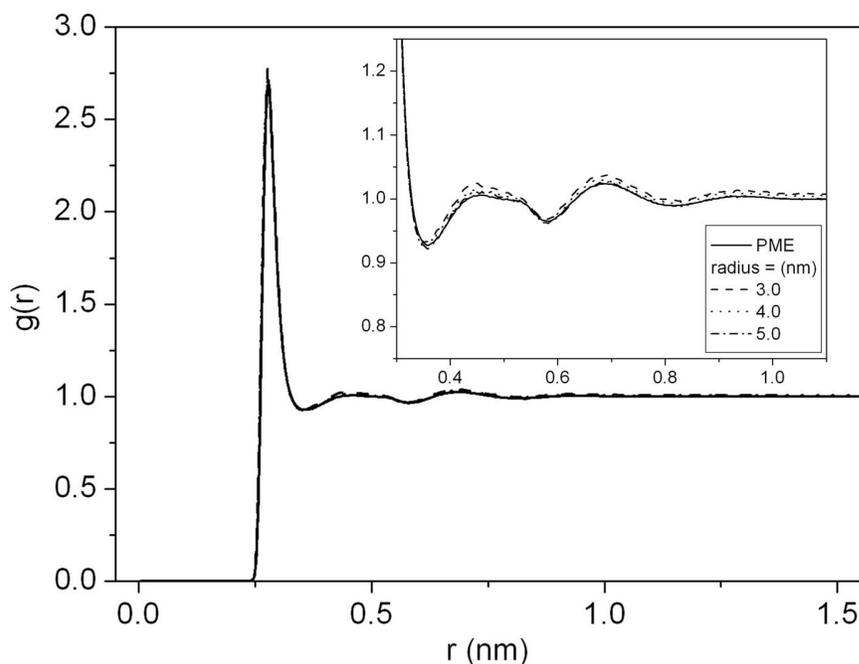


Figure 5.11 Oxygen-oxygen RDF in different sizes of spheres at 300K. As the size of system increases, the curve matches PME results better.

d. Diffusion Coefficient Retains Weak System Size Dependence

In addition to the structural properties, we access how well the ICSM with a finite boundary describes the dynamical properties of the simulated water. The self-diffusion coefficient D is calculated, which is an important dynamical characteristic of a liquid [63]. As shown in Table 5.2, the diffusion coefficient of the finite boundary model approaches to the desired value of PME as the size of sphere becomes larger, revealing size dependence in the diffusion property. When the system is small, according to our settings, only a small part of the water molecules located near the center of the sphere will be selected for the calculation of D . The time range t for motion in the calculation is short ($t = 4.0$ ps for 3.0 nm sphere) to make sure most of the selected water molecules will not travel into the buffer layer during this time, otherwise the error of the calculation may be increased. We calculate the numbers of water molecules in the productive region

of our model in different sizes of sphere, and simulate the same numbers of molecules using PBC and PME. The results show that D will increase a little bit as the size of box increases. However, in our model, with larger systems and larger number of sampled molecules near the center, the self-diffusion coefficient adequately approaches the PME value.

Table 5.2: Diffusion coefficient with error estimate (unit of D is $10^{-9}\text{m}^2\text{s}^{-1}$).

Size of Sphere	PME ^[69]	3.0 nm	4.0 nm	5.0 nm
Diffusion Constant	5.88(+/-0.05)	5.39(+/-0.02)	5.54(+/-0.01)	5.83(+/-0.01)

e. Demonstrated Bulk Properties of Dipole Moment Fluctuations

Dielectric properties are evaluated by computing the fluctuation of the net dipole moment $\langle M(R)^2 \rangle$ for a spherical ball centered at the origin as a function of radius R . The total dipole moment represents a quantity that reflects fluctuations in the orientations of water molecules. The formula for the total dipole moment with respect to the origin of spherical cavity is given in Eq. 5.19, where a spherical ball contains all water molecules in frame, j , that have its center of mass within the radius R .

$$\langle M^2 \rangle_{total} = \sum_{j=1}^J \langle \bar{M}_j \cdot \bar{M}_j \rangle \quad \text{Eq. (5.19)}$$

$$\text{where } \bar{M}_j = \sum_{k=1}^K (q_{O,k} \vec{r}_{O,k,j} + q_{H_1,k} \vec{r}_{H_1,k,j} + q_{H_2,k} \vec{r}_{H_2,k,j})$$

The dipole moment is averaged over all j -frames from the MD trajectory, where the atom position vectors for the k -th molecule in the j -th frame weight the oxygen and hydrogen atoms.

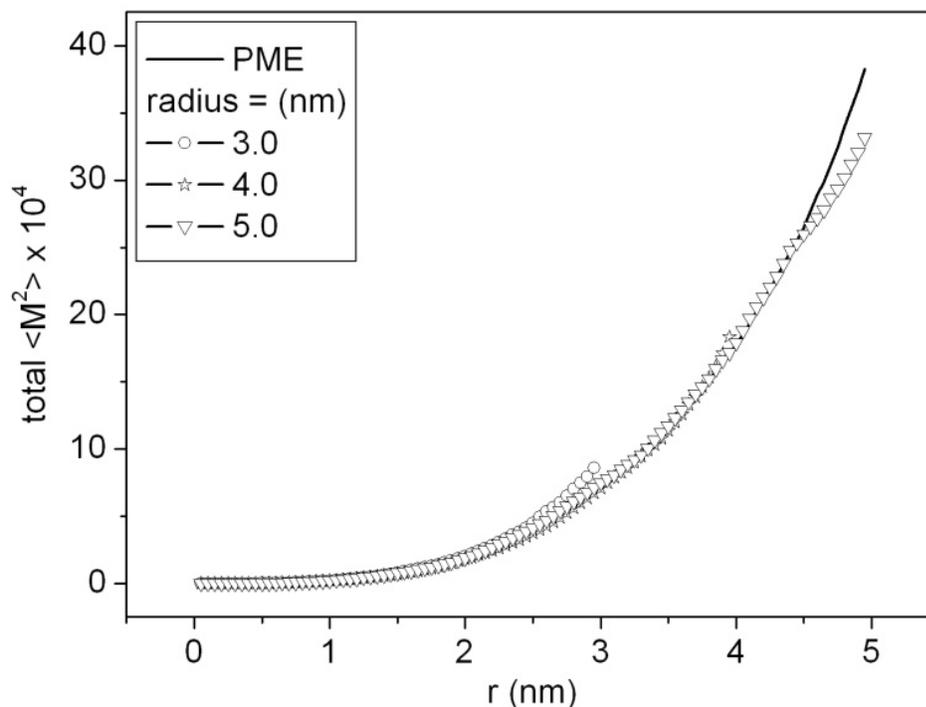


Figure 5.12: The fluctuation of total dipole moment for spherical balls of radius r in different sizes of spheres compared to PME results.

As shown in Fig. 5.12, for the productive region of different sizes of system, their total fluctuations of dipole moments match the PME result very well. For all three cases, when approaching the edge of the buffer layer, deviation from expected value is observed but still acceptable. The matching of $\langle M^2 \rangle$ represents the correct orientation of water dipole in all of the productive volume and most part of the buffer region in the finite boundary model.

f. Temperature Independence for Matching Bulk Properties of PME

As shown, bulk properties of Tip3p water are reproduced well in different size systems at 300 K. However, the efficiency of the effective vdW and constraint forces at multiple temperatures also has to be evaluated for the robustness of the model. To characterize the temperature independence, bulk properties are calculated at 250 K and 350 K, and compared with PME results. As shown in Fig. 5.13 A and B, the relative

water density and radial dipole moment almost overlap each other at 250 K, 300 K and 350 K, suggesting the correct concentration and orientation of water molecules have been achieved at different temperatures. For RDF and fluctuation of total dipole moment, although the curves from PME data vary among themselves according to the temperature, the finite boundary model matches to the PME results accurately (Fig. 5.13 C-F). Although the curves of the dipole moments deviate from the PME values within the buffer region, we can still guarantee a virtually perfect matching in the productive region.

5.4 Conclusions

An ICSM that employs NPBC has been introduced to simulate aqueous solutions. This is a hybrid model that combines the explicit solvent and solute molecules in a spherical cavity with a continuum dielectric medium for solvent around the cavity. Systematic studies of different parameters, including the size of simulation sphere and the temperature of simulation, have been performed. With adjustable thickness for the buffer layer, there is no strong size dependence for the essential properties of water. Spherical cavities with nominal radii of 3.0, 4.0 and 5.0 nm have been studied, and no intrinsic systematic errors are found related to the size of simulation box. With the self-consistent constraint forces that are only applied in the buffer layer, properties including water density, RDF and dipole moment in the central productive region are in excellent agreement with PME results. After introducing the constraint forces, multiple thermostats are necessary to maintain the correct temperature in the productive region and buffer layer. The diffusion constant is smaller than bulk value for the small system due to the nature of its definition, but it approaches to the PME result quickly as the system size increases. The new model is equipped with a movable cavity wall to perform both NVT

and NPT simulations. Water properties are calculated at 250, 300 and 350 K to verify that the model will give similar results as PME at different temperatures and too ensure the boundary conditions are robust. The temperature dependent properties of water are accurately reproduced in the productive region, revealing that the model is indeed stable and reliable in multiple conditions.

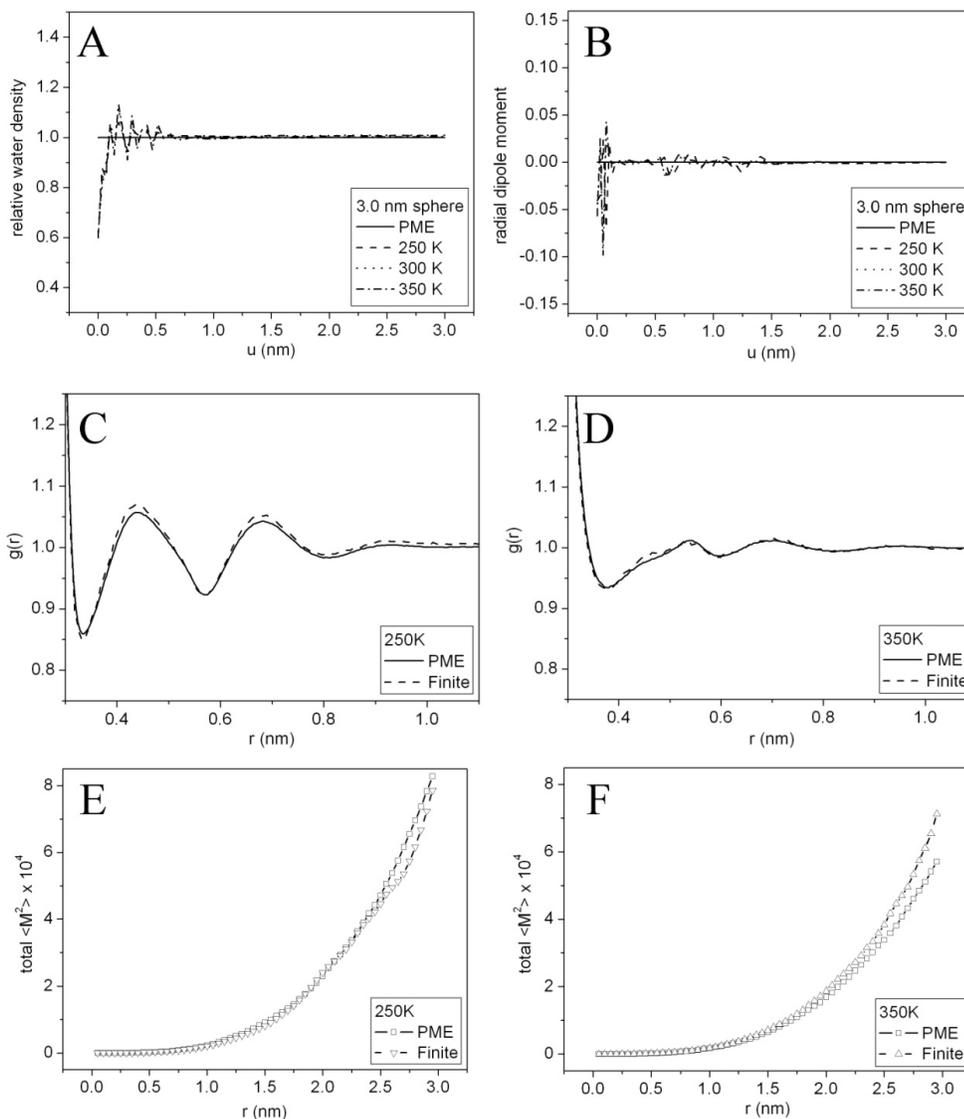


Figure 5.13: Check the temperature independence of the model in 3.0 nm sphere. A) and B) relative water density and radial dipole moment at three temperatures, C) and D) radial distribution functions compared to PME results at 250K and 350K, E) and F) fluctuation of total dipole moments compared to PME results at 250K and 350K.

When boundary conditions are to be distinguished, we refer to the (new, original) model as (ICSM/NPBC, ICSM/PBC). The advantages of ICSM/NPBC are that it does not allow unphysical correlations between phantom ions or solute molecules caused by periodic images, and a much larger productive volume is achieved for the same size cavity. We validated ICSM/NPBC by checking its performance on liquid water. Only pure water is considered in this work because many critical aspects were explored and resolved that were not explored/resolved in prior works on bulk water. In fact, the problem of obtaining bulk properties of water when the substance is confined to a finite space because of mathematical abstraction only has been very challenging, as testimony by the large number of previous works on this very same topic, including ours. After an extensive investigation, excellent agreement has been found between the ICSM/NPBC and PME simulation results. The larger the simulation box, the better agreement is achieved. Characterizing ionic solutions using the ICSM/NPBC will be subsequently published. The biggest achievements of this project are the improvement of both accuracy (correction of ion problem) and efficiency (much larger usable region) in modeling aqueous solution compared to ICSM/PBC.

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

Electrostatic interaction is important in biomolecular processes, and it is the bottleneck in molecular dynamics simulation due to its long-range nature, which means expensive calculations. How to improve the accuracy and efficiency for its calculation has been a long-standing goal in the field of computational biology, and finding better models is still a very active area of research within the field. Intensive studies have been performed with multiple choices of solvent models, the boundary conditions and the techniques to deal with the electrostatic interactions. Among all these works, the ICSM/PBC, which is a hybrid model developed in our group, has been proved to be successful to handle the pure water and dilute ionic solutions. This dissertation gives a broad overview, detailed characterization and comprehensive optimization of the existing ICSM/PBC, and also introduces a new improved ICSM/NPBC for accurate and efficient way to simulate solvations. Novel research was performed to 1) intensively investigate the physical nature of the RF and effects of buffer layer, which reveals the essential contribution of RF on the orientation of water molecules and the indispensable role of buffer layer; 2) optimize the image charge method in a self-consistent procedure that side steps knowing a priori dielectric profile, by which the RF can reproduce the accurate electrostatic field and reduce the artifacts; and 3) improve the unphysical ion correlations and small productive region in the ICSM/PBC due to its periodic boundary conditions, by non-trivial conversion to the non-periodic boundaries (ICSM/NPBC) employing multiple

constraint forces, for higher efficiency and more flexibility in multi-component solvents.

In Chapter 3, we focus on why the RF is critical to maintain the correct dielectric constant and how exactly the RF influences the dielectric properties. To answer these questions, we paid attention to the two physical variables, force and torque, which controls the translation and rotation of a water molecule, respectively. Histogram distributions of both variables over the concentric shells from the origin to the edge have been plotted associated with different box sizes, number of image charges and boundary conditions for water molecule. It turns out that the electrostatic forces due to the RF are only 2% of the total net electrostatic forces, compared to the fact that the electrostatic torques due to the RF is more than 20% of the total net torques. This result, which is first demonstrated in our work, reveals the importance of the RF for its strong effects on orientation of water dipoles. In addition to the evaluation of the RF, the force and torque distributions in buffer layer are also studied. Strong surface artifacts have been exerted on the water molecules in buffer by the boundary. However, the dynamic properties of the water close to wall have not been disturbed because they are imaged water. Numbers of image charges for accuracy of RF and three boundary conditions for water molecules have been tested by ICSM/PBC, without showing significant influence on the performance of the model.

In Chapter 4, the artifacts due to the discontinuous dielectric profile near the boundary in DDM, which also have been the most common problems for all the implicit and hybrid models, is optimized by matching the RF to accurate properties from PME directly. The multiple image charge method used in DDM has analytical solution with the assumption that the dielectric constant is discontinuous at the surface, which equals to 1

for inside explicit solvent but 80 for outside implicit solvent. However, this assumption is mathematical and the actual dielectric profile is unknown. The magnitudes of the image charges have been modified by least square error method to minimize the difference between the RF and the target values. 25% of LSE decreasing has been achieved for the optimized RF properties suggesting the reduction of artifacts from surface. These optimized image charges are then implemented back to MD simulation for the continuous dielectric profile calculation, which looks close to the PME results. In this inverse problem the numerical optimization of RF to represent a continuous dielectric profile at the boundary, which tries to solve the artifacts problem from the beginning, has not been done before.

In Chapter 5, a new ICSM/NPBC has been developed to overcome the problems of the extra ion correlations and too small productive region due to the PBC. ICSM/PBC shows its success in dealing with pure water and dilute ionic solutions. However, with high ionic strength, PBC will introduce non-physical ion correlations between the ions and their periodic images in buffer layer and the usable volume is limited in the center of sphere. Rather than stick to the PBC, we implemented NPBC instead. A buffer layer of explicit dynamic water molecules has been introduced to reduce the surface effects between two solvents near the boundary. According to our study, the incorrect behaviors of the water molecules near the wall will lead to wrong properties of water farther away from the wall because of the long range nature of the interactions. Consequently, we had to make the water in the buffer layer, which is subjected to intense erroneous forces due to boundary conditions, have bulk properties of water. We achieved this goal using multiple constraint forces, including effective vdW forces, diffusion forces for density,

torque forces for dipole, and multiple thermostats for temperatures, have to be applied to the water molecules inside buffer region to maintain their properties in a reasonable range. The simulation box is either fixed or flexible, allowing a NVT or NPT simulation based on pressure calculation. The new model is tested within 3.0, 4.0 and 5.0 nm radius of sphere at 250, 300 and 350 K. All important characteristics of water, including water density, RDF, diffusion constant, radial component and the fluctuation of the dipole moment are calculated and compared with bulk properties calculated from PME trajectories. In the case of ions, we only know at this point that we completely removed unphysical correlations of ions due to their images. Although the thickness of the buffer layer increases with box size, higher accuracy of properties and much bigger productive volume can be obtained compare to typical ICSM.

The future work can be continued in two aspects. 1) For the optimization project of DDM, not only the electrostatic forces, but also the short-range forces can be optimized, using the similar procedure developed in Chapter 4. As long as all the interactions from the wall are optimized, the artifacts will be minimized and the thickness of buffer layer may be decreased as originally intended, with more accurate RF and increased usable region. 2) For the ICSM/NPBC project, the solutions of single ion and high ionic strength should be tested in the new model. Extra constraints might be introduced to keep the ions staying in productive region mostly. The pair distribution function, solvation free energy and other thermodynamic properties of the ions and solutions need to be compared to PME results.

REFERENCES

- [1] B. J. Bennion and V. Daggett, "The molecular basis for the chemical denaturation of proteins by urea," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 5142-5147, 2003.
- [2] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glittli, P. H. Hunenberger, M. A. Kastenholtz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, and H. B. Yu, "Biomolecular Modeling: Goals, Problems, Perspectives," *Angew. Chem. Int. Ed.*, vol. 45, pp. 4064-4092, 2006.
- [3] H. Schreiber and O. Steinhauser, "Cutoff size does strongly influence molecular dynamics results on solvated polypeptides," *Biochemistry*, vol. 31, pp. 5856-5860, 1992.
- [4] D. M. York, W. Yang, T. Darden, and L. G. Pedersen, "Toward the Accurate Modeling of DNA: The Importance of Long-Range Electrostatics," *J. Am. Chem. Soc.*, vol. 117, pp. 5001-5002, 1995.
- [5] R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik, "Coupling atomistic and continuum hydrodynamics through a mesoscopic model: Application to liquid water," *J. Chem. Phys.*, vol. 131, p. 244107, 2009.
- [6] W. Weber, P. H. Hünenberger, and J. A. McCammon, "Molecular Dynamics Simulations of a Polyalanine Octapeptide under Ewald Boundary Conditions: Influence of Artificial Periodicity on Peptide Conformation," *J. Phys. Chem. B*, vol. 104, pp. 3668-3675, 2000.
- [7] M. A. Villarreal and G. G. Monticha, "On the Ewald Artifacts in Computer Simulations. The Test-Case of the Octaalanine Peptide With Charged Termini," *J. Biomol. Struct. Dyn.*, vol. 23, pp. 135-142, 2005.
- [8] M. A. Kastenholtz and P. H. Hunenberger, "Influence of Artificial Periodicity and Ionic Strength in Molecular Dynamics Simulations of Charged Biomolecules Employing Lattice-Sum Methods," *J. Phys. Chem. B.*, vol. 108, pp. 774-788, 2004.
- [9] Y. Lin, A. Baumketner, W. Song, S. Deng, D. Jacobs, and W. Cai, "Ionic solvation studied by image-charge reaction field method," *J. Chem. Phys.*, vol. 134, p. 3530094, 2011.
- [10] B. Roux, "Implicit solvent models," *Computational Biochemistry and Biophysics*, vol. Marcel Dekker, New York, 2001.

- [11] B. Honig and A. Nicholls, "Classical electrostatics in biology and chemistry," *Science*, vol. 268, pp. 1144-1149, 1995.
- [12] N. A. Baker, "Improving implicit solvent simulations: a Poisson-centric view," *Curr. Opin. Struct. Biol.*, vol. 15, pp. 137-143, 2005.
- [13] J. Wang, C. Tan, Y. Tan, Q. Lu, and R. Luo, "Poisson-Boltzmann Solvents in Molecular Dynamics Simulations," *Comm. Comput. Phys.*, vol. 3, pp. 1010-1031, 2008.
- [14] P. P. Ewald, "Die Berechnung optischer und elektrostatischer Gitterpotentiale," *Ann. Phys*, vol. 369, pp. 253-287, 1921.
- [15] R. Abagyan and M. Totrov, "Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins," *J. Mol. Biol.*, vol. 235, pp. 983-1002, 1994.
- [16] U. Essmann, L. Perera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen, "A smooth particle mesh Ewald method," *J. Chem. Phys.*, vol. 103, p. 8577, 1995.
- [17] T. A. Darden, D. M. York, and L. G. Pedersen, "Particle Mesh Ewald: an Nlog(N) method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, pp. 10089-10092, 1993.
- [18] J. Shimada, H. Kaneko, and T. Takada, "Efficient calculations of coulombic interactions in biomolecular simulations with periodic boundary conditions," *J. Comput. Chem.*, vol. 14, pp. 867-878, 1993.
- [19] B. Luty, M. Davis, I. Tironi, and W. Gunsteren, "A comparison of particle-particle, particle-mesh and Ewald methods for calculating electrostatic interactions in periodic molecular systems," *Mol. Simul.*, vol. 14, pp. 11-20, 1994.
- [20] L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations," *J. Comput. Phys.*, vol. 73, pp. 325-348, 1987.
- [21] L. Greengard., "The rapid evaluation of potential fields in particle systems," *MIT Press, Cambridge*, 1988.
- [22] L. Ying, G. Biros, and D. Zorin., "A Kernel-independent Adaptive Fast Multipole Method in Two and Three Dimensions," *J. Comput. Phys.*, vol. 196, p. 591, 2004.
- [23] C. Sagui and T. A. Darden, "Molecular dynamics simulations of biomolecules: long-range electrostatic effects," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 28, pp. 155-179, 1999.

- [24] K. E. Schmidt and M. A. Lee, "Implementing the fast multipole method in three dimensions," *J. Stat. Phys.*, vol. 63, pp. 1223-1235, 1991.
- [25] A. Okur and C. Simmerling, "Hybrid explicit/implicit solvation methods," *Annual Reports in Computational Chemistry*, vol. 2, p. Chapter 6, 2006.
- [26] J. G. Kirkwood, "Theory of solutions of molecules containing widely separated charges with special application to Zwitterions," *J. Chem. Phys.*, vol. 2, pp. 351-361, 1934.
- [27] W. Im, S. Berneche, and B. Roux, "Generalized solvent boundary potential for computer simulations," *J. Chem. Phys.*, vol. 114, p. 2924, 2001.
- [28] M. Lee, F. Salsbury, and M. Olson, "An efficient hybrid explicit/implicit solvent method for biomolecular simulations," *J. Comput. Chem.*, vol. 25, pp. 1967-1978, 2004.
- [29] D. Bashford and D. Case, "Generalized Born models of macromolecular solvation effects," *Ann. Rev. Phys. Chem.*, vol. 51, pp. 129-152, 2000.
- [30] A. Okur and C. Simmerling, "Hybrid Explicit/Implicit Solvation Methods," *Ann. Rep. Comp. Chem.*, vol. 2, pp. 97-109, 2006.
- [31] P. Yang, S. Liaw, and C. Lim, "Representing an Infinite Solvent System with a Rectangular Finite System Using Image Charges," *J. Phys. Chem. B*, vol. 106, pp. 2973-2982, 2002.
- [32] G. King and A. Warshel, "A surface contained all atom solvent model for effective simulations of polar solutions," *J. Chem. Phys.*, vol. 91, pp. 3647-3661, 1989.
- [33] D. Beglov and B. Roux, "Finite representation of an infinite bulk system: solvent boundary potential for computer simulations," *J. Chem. Phys.*, vol. 100, p. 9050, 1994.
- [34] H. Alper and R. Levy, "Dielectric and thermodynamic response of a generalized reaction field model for liquid state simulations," *J. Chem. Phys.*, vol. 99, p. 9847, 1993.
- [35] J. Kirkwood, "Statistical Mechanics of Liquid Solutions," *Chem. Rev.*, vol. 19, pp. 275-307, 1936.
- [36] J. C. Maxwell, "Treatise on Electricity and Magnetism," *Dover, New York*, 1954.
- [37] H. Friedman, "Image approximation to the reaction field," *Mol. Phys.*, vol. 29, pp. 1533-1543, 1975.

- [38] P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," *McGraw-Hill, New York*, 1953.
- [39] Y. Lin, A. Baumketner, S. Deng, Z. Xu, D. J. Jacobs, and W. Cai, "An image-based reaction field method for electrostatic interactions in molecular dynamics simulations of aqueous solutions," *J. Chem. Phys.*, vol. 131, p. 154103, 2009.
- [40] W. Cai, S. Deng, and D. Jacobs., "Extending the fast multipole method to charges inside or outside a dielectric sphere," *J. Comput. Phys.*, vol. 223, p. 846, 2007.
- [41] S. Deng and W. Cai, "Extending the fast multipole method for charges inside a dielectric sphere in an ionic solvent: High-order image approximations for reaction fields " *J. Comput. Phys.*, vol. 227, pp. 1246-1266, 2007.
- [42] S. Deng and W. Cai, "Discrete Image Approximations of Ionic Solvent Induced Reaction Field to Charges," *Commun. Comput. Phys.*, vol. 2, p. 1007, 2007.
- [43] W. Song, Y. Lin, A. Baumketner, S. Deng, W. Cai, and D. J. Jacobs, "Effect of the reaction field on molecular forces and torques revealed by an image-charge solvation model," *Commun. Comput. Phys.*, vol. 13, pp. 129-149, 2013.
- [44] S. Deng, W. Cai, and D. J. Jacobs, "A comparable study of image approximations to the reaction field," *Comput. Phys. Commun.*, vol. 177, pp. 689-699, 2007.
- [45] C. Neumann, "Hydrodynamische untersuchen nebst einem anhang uber die probleme der elektrostatik und der magnetischen induktion," *Teubner, Leipzig*, vol. 279, 1883.
- [46] A. Okur and C. Simmerling, "Hybrid explicit/implicit solvation methods," *Annual Reports in Computational Chemistry*, vol. 2, pp. 97-109, 2006.
- [47] M. A. Kastholz and P. H. Hunenberger, "Computation of methodology-independent ionic solvation free energies from molecular simulations. II. The hydration free energy of the sodium cation " *J. Chem. Phys.*, vol. 124, pp. 224501-224520, 2006.
- [48] M. A. Kastholz and Philippe H. Hunenberger, "Computation of methodology-independent ionic solvation free energies from molecular simulations. I. The electrostatic potential in molecular liquids.," *J. Chem. Phys.*, vol. 124, pp. 124106-124127, 2006.
- [49] B. Ni and A. Baumketner, "Effect of atom- and group-based truncations on biomolecules simulated with reaction-field electrostatics," *J. Mol. Model.*, vol. 17, pp. 2883-2893, 2011.

- [50] G. Petraglio, M. Ceccarelli, and M. Parrinello, "Nonperiodic Boundary Conditions for Solvated systems," *J. Chem. Phys.*, vol. 123, p. 044103, 2005.
- [51] J. M. Rodgersab and J. D. Weeks, "Interplay of Local Hydrogen-Bonding and Long-Ranged Dipolar Forces in Simulations of Confined Water," *PNAS*, vol. 105, pp. 19136-19141, 2008.
- [52] Z. Xu, S. Deng, and W. Cai, "Image charge approximations of reaction fields in solvents with arbitrary ionic strength," *J. Comput. Phys.*, vol. 228, pp. 2092-2099, 2009.
- [53] D. S. Ceruttia, N. A. Baker, and J. A. McCammon, "Solvent reaction field potential inside an uncharged globular protein: A bridge between implicit and explicit solvent models?," *J. Chem. Phys.*, vol. 127, pp. 155101-155112, 2007.
- [54] K. Baker, A. Baumketner, Y. Lin, S. Deng, D. Jacobs, and W. Cai, "ICSM: An order N method for calculating electrostatic interactions added to TINKER," *Comput. Phys. Commun.*, vol. 184, pp. 19-26, 2013.
- [55] A. C. Belch and M. Berkowitz., "Molecular Dynamics Simulations of TIPS2 Water Restricted by a Spherical Hydrophobic Boundary," *Chem. Phys. Lett.*, vol. 113, 1985.
- [56] A. Wallqvist, "On the Implementation of Friedman Boundary Conditions in Liquid Water Simulations," *Mol. Simul.*, vol. 10, pp. 13-17, 1993.
- [57] C. L. Brooks and M. Karplus, "Deformable stochastic boundaries in molecular dynamics," *J. Chem. Phys.*, vol. 79, p. 6312, 1983.
- [58] H.J.C. Berendsen, D. van der Spoel, and R. v. Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," *Comp. Phys. Comm.*, vol. 91, pp. 43-56, 1995.
- [59] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, pp. 926-935, 1983.
- [60] A. H. Jufferab and H. J. C. Berendsen, "Dynamic surface boundary conditions," *Mol. Phys.*, vol. 79, pp. 623-644, 1993.
- [61] G. Brancato, N. Rega, and V. Barone, "A hybrid explicit/implicit solvation method for first-principle molecular dynamics simulations," *J. Chem. Phys.*, vol. 128, p. 144501, 2008.
- [62] S. Jang and G. A. Voth, "Simple reversible molecular dynamics algorithms for Nose-Hoover chain dynamics," *J. Chem. Phys.*, vol. 107, p. 9514, 1997.

- [63] M. P. Allen and D. J. Tildesley, "Computer Simulation of Liquids," *Oxford University Press, Oxford*, 1987.
- [64] L. Wang and J. Hermans, "Reaction field molecular-dynamics simulation with Friedman's image charge method," *J. Chem. Phys.*, vol. 99, pp. 12001-12007, 1995.
- [65] R. Blaak and J. P. Hansen, "Dielectric response of a polar fluid trapped in a spherical nanocavity," *J. Chem. Phys.*, vol. 124, p. 144714, 2006.
- [66] J. A. C. Rullmann and P. T. V. Duijnen, "Analysis of discrete and continuum dielectric models; application to the calculation of protonation energies in solution," *Mol. Phys.*, vol. 61, pp. 293 - 311, 1987.
- [67] G. Brancato, A. Di Nola, V. Barone, and A. Amadei, "A mean field approach for molecular simulations of fluid systems," *J. Chem. Phys.*, vol. 122, p. 154109, 2005.
- [68] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl., "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation," *J. Chem. Theory Comput.*, vol. 4, p. 435, 2008.
- [69] P. Mark and L. Nilsson, "Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K," *J. Phys. Chem. A*, vol. 105, pp. 9954-9960, 2001.