Molecular-dynamics simulations of self-assembled monolayers (SAM) on parallel computers

Satyavani Vemparala
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Physical Sciences and Mathematics Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/2258

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
MOLECULAR-DYNAMICS SIMULATIONS OF SELF-ASSEMBLED MONOLAYERS (SAM) ON PARALLEL COMPUTERS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Physics and Astronomy

by

Satyavani Vemparala
M.Sc., University of Hyderabad, Hyderabad, India, 1996
M.Tech., Indian Institute of Technology, Kharagpur, India, 1998
December 2003
ACKNOWLEDGMENTS

I am thankful to my advisors Drs. Priya Vashishta, Aiichiro Nakano and Rajiv Kalia for the opportunity to work in an exciting field of molecular simulations and for all the support and encouragement I received as a member of their group. I am especially grateful to Dr. Nakano, for the all his invaluable guidance and untiring patience without which most of this work would not have resulted.

I am very fortunate to have found friends like Hideaki Kikuchi, Sanjay Kodiyalam and Jabari Lee at LSU, who not only helped me gain perspective towards my research, but also in life. I am indebted to Dr. Kikuchi for all help rendered to me during my years as a student in the group.

Thanks to Dr. Bijaya Karki, for his valuable input, without whose participation crucial parts of this work would have taken much longer time. A heartfelt thanks to members of CCLMS and CACS: Cindy Rountree, Ashish Sharma, Maxim Makeev with whom I have often shared many interesting discussions. I’ve had the opportunity to interact with past and present members of CCLMS and CACS: Phillip Walsh, Tim Campbell, Laurent van Brutzel, Brent Neal, Xiaotao Su, Elefterios Lidorikis, Paulo Branicio, Gurcan Aral, Xinlian Liu, Cheng Zhang, Weiqiang Wang, Zhen Lu, Naoto Umezawa, Izabela Szlufarska, Kenichi Nomura. I am grateful to all the members of Computational Chemistry List (http://www.ccl.net/chemistry/) who have answered most of my ‘stupid’ questions during the initial phase of my research, when other sources of help were scarce.

I am grateful to Drs. Richard Kurtz, Joel Tohline, and Grover Waldrop for serving on my thesis committee.

Many thanks to the systems managers Monika Lee and Hortensia Valdes for their assistance regarding the computer systems at LSU. I am grateful to all the past and present secretaries and staff at LSU and USC: Arnell Jackson, Beverly Rodriguez, Jade Ethridge,
Sabrina Feeley, Shemeka Law, Karen Richard, Cathy Mixon, Ophelia Dudley, Conner Campion, for the numerous ways they have provided assistance.

I would like to acknowledge the financial support I have received from the Teaching Assistantship from the Department of Physics and Astronomy. I am also grateful for being awarded a Graduate Enhancement award and to Dr. Juhan Frank for bringing this opportunity to my attention. DOE, NSF, AFOSR, USC-LSU Multidisciplinary University Research Initiative, Biological Computation and Visualization Center (BCVC) funded the research projects on which I based my thesis. Simulations were performed on the 160-processor PC cluster in the Concurrent Computing Laboratory for Materials Simulations, 256-processor dual Intel Xeon cluster at Biological Computation and Visualization Center and 1024-Linux cluster SuperMike at Louisiana State University, IBM SP3 and SP4 at Naval Office (NAVO).

I am deeply indebted to my parents for their love and continuous support that has made my education possible. Ushi, thanks for just being my sister and sharing all the pains and joys of growing up and your friendship means more to me than I let you know. Thanks to Kalyan for introducing me to the wonderful world of Harry Potter and all the countless hours of excited discussions about the same. It is my good fortune to have a best friend like Ravi who is a safe harness to fall back in troubled times, and has been my companion in this long journey. Thanks to Ravi’s parents for their understanding and support. And finally thanks to Bill Waterson for Calvin & Hobbes whose wisdom and wit have brightened many a dark and gloomy days.
# TABLE OF CONTENTS

ACKNOWLEDGMENTS .................................................................................. ii

LIST OF TABLES .......................................................................................... vi

LIST OF FIGURES ...................................................................................... vii

ABSTRACT .................................................................................................. xii

CHAPTER 1. INTRODUCTION .................................................................... 1

CHAPTER 2. MOLECULAR DYNAMICS SIMULATIONS ........................... 6
  2.1 Molecular Dynamics (MD) Methodology ....................................... 6
  2.2 Interatomic Potentials for Organic Systems .................................... 7
     2.2.1 Bonded Potential Terms ..................................................... 8
     2.2.2 Non-bonded Potential Terms ............................................. 11
  2.3 Periodic Boundary Conditions ...................................................... 13
  2.4 Thermal Quantities and Statistical Mechanics .............................. 15
  2.5 Structural Correlations ................................................................. 16

CHAPTER 3. SPACE-TIME MULTIREsolution ALGORITHMS ........... 21
  3.1 Integration Algorithms .................................................................. 21
  3.2 Short-Range Interactions ............................................................... 24
     3.2.1 Bonded Interactions ............................................................ 24
     3.2.2 Non-Bonded Interactions .................................................... 25
  3.3 Long-Range Interactions ............................................................... 27
     3.3.1 Ewald Summation Method ................................................ 28
     3.3.2 Fast Multipole Method ....................................................... 30
  3.4 Multiple Time-Scale Method ......................................................... 35

CHAPTER 4. PARALLEL MD AND Scalability TESTS ....................... 38
  4.1 Parallelization Paradigms ............................................................... 38
  4.2 Space Decomposition ................................................................... 41
  4.3 Parallel FMM: Scalability Tests and Results .................................. 42
  4.4 Parallel MMD: Scalability Tests and Results .................................. 44
  4.5 Summary ......................................................................................... 48

CHAPTER 5. SIMULATIONS OF ALkanethiol SELF-ASEMbled MONOLayer SYSTEMS ......................................................... 50
  5.1 Background .................................................................................... 50
  5.2 Interatomic Potential Parameters for Alkanethiol SAMs ............... 55
  5.3 Initial Setup .................................................................................... 57
  5.4 Results and Discussion ................................................................. 59
     5.4.1 Effect of Temperature .......................................................... 59
# LIST OF TABLES

4.1 Results of accuracy tests of FMM on Linux PC cluster................................. 43

4.2 Comparison of the Linux cluster *SuperMike*, IBM SP4 *Marcellus*, and
Compaq SC45 *Emerald*........................................................................................................ 45

5.1 Force field parameters for system of methyl terminated alkanethiol SAMs...... 56

6.1 Force field parameters for system of PEG terminated alkanethiol SAMs......... 74
LIST OF FIGURES

2.1  Comparision of harmonic and Morse potentials for bond stretching............... 9
2.2  Harmonic bend angle potential................................................................. 10
2.3  Torsion angle potential (a) definition of torsion angle (b) form of torsion angle potential.......................................................... 11
2.4  Comparision of Lennard-Jones and Buckingham potential forms. ............... 12
2.5  Two-dimensional periodic boundary conditions in MD simulations............... 14
3.1  Short-range nature of bond, bend and torsional potential terms.................... 25
3.2  Shifted and unshifted Lennard-Jones potential............................................ 26
3.3  Linked-cells, neighbor cells and Verlet-list radius in 2D MD box.................... 27
3.4  Charge distribution in the Ewald sum with original point charges and Gaussian screening distribution........................................ 29
3.5  Far-field multipole approximation in FMM. ............................................. 31
3.6  Octree division in two-dimensions each color represents a different level of octree......................................................................... 33
4.1  Difference between atom decomposition and spatial decomposition methods. The atoms that a processor is responsible for are also shown. The colors signify the processor they belong to: red-processor A, blue-processor B, green-processor C and purple-processor D  (a) system at t=0 (b) system at t=T and movement of atoms(c)system under atom decomposition method (d) system under spatial decomposition method........ 39
4.2  Schematic of spatial decomposition in MMD algorithm. The figure shows the primary skin $S_p(p)$ of thickness $r_p$ and the secondary skin $S_c(p)$ of thickness $r_c$ for processor 1. Atom $t$ has migrated from processor 3 to processor 4 due to time-stepping procedure.................................................. 41
4.3  Timings as function of number of processors. (a) Total execution (solid symbols) and communication (open symbols) times are plotted for the number of particles per processor, $N/P = 10^5$ (squares), $5\times10^5$ (triangles), and $10^6$ (circles) on IBM SP3.  (b) The same as (a) on IBM SP4. ................. 43
5.4 Schematic representation of a tilted alkanethiol molecule: $\square$ is the tilt angle of the molecule away from the surface normal, $\square$ is the precession angle about the surface normal that represents the tilt direction; and $\square$ is the twist (rotation about the molecular axis) angle that defines the orientation of the chain backbone plane relative to the tilt direction. ............................. 54

5.5 The initial configuration of the system of alkanethiol SAMs. Atoms in red, black and blue are sulfur, carbon and hydrogen species respectively. (a) the system in $x$-$y$ direction (top view) and (b) the system in $x$-$z$ direction (side view). ......................................................... 58

5.6 Time evolution of the collective tilt angle during thermalization for $n = 13$. .... 59

5.7 (a) Temperature variation of the collective tilt angle: The simulation results from heating run and cooling run are shown by circles and diamonds, respectively, for the systems of a hundred thousand (open symbols) and a million atoms (solid symbols) for $n = 13$. (b) Temperature
variation of tilt direction: The simulation results from heating run and cooling run are shown by circles and diamonds, respectively. 60

5.8 Calculated twist angle distribution plotted as a function of twist angle at $T = 100, 300$ and $450K$ for $n = 13$. 61

5.9 Calculated radial distribution function, $g(r)$, at $T = 100, 300$ and $450K$ for $n = 13$. The curves for $T = 100$ and $300K$ are shifted vertically for clarity. 62

5.10 Calculated structure factors for (a) only odd carbons in a chain are considered (b) all the atoms in the chain, including sulfur, are considered. 63

5.11 Calculated structure factors for (a) $T = 0K$ (b) $T = 200K$. (c) $T = 400K$. 64

5.12 Calculated density profiles along the surface normal at $T = 100, 300$ and $450K$ for $n = 13$. The curves for $T = 100$ and $300K$ are shifted vertically for clarity. 65

5.13 The collective tilt angle as a function of lattice spacing. The simulation results are shown by diamonds at $0K$, squares at $200K$ and open circles at $300K$. The room temperature experimental data for Langmuir (air/water) monolayer are shown by solid circles. The solid line shows $\bar{D}_{vdw}$ (see Eq. (5.9)). Shown in the inset is the vdw energy as a function of inter-chain separation where $a_{min} = 4.7A$. 66

5.14 Tilt angle as a function of chain length for $T = 50, 200$, and $300K$. 68

5.15 The van der Waals energy/atom as a function of chain length for $T = 50, 200$, and $300K$. 68

5.16 (a) Torsion angle distribution for $n = 7$ at $T = 200, 300$ and $400K$. (b) Torsion angle distribution for $n = 23$ at $T = 200, 300$ and $400K$. The gauche defects are marked by $g$ and $g^*$ and trans configuration by $t$. 69

5.17 (a) Twist angle distribution for $n = 7$ at $T = 0, 200$ and $300K$. (b) Twist angle distribution for $n = 23$ at $T = 0, 200$ and $300K$. 70

6.1 Temperature variation of the collective tilt angle: The simulation results from heating run and cooling run are shown by circles and squares, respectively. 78

6.2 Temperature variation of tilt direction: circles and squares show the simulation results from heating run and cooling run respectively. 78

6.3 Calculated partial radial distribution function, $g_{rr}(r)$ at $T = 50, 200, 300$, and $400K$ in heating cycle. The curve for different temperatures are shifted vertically for clarity. 79
6.4 Calculated partial radial distribution function, \( g_{oo}(r) \) at \( T = 204\text{K} \) in heating cycle and \( T = 197\text{K} \) in cooling cycle. The curve for heating cycle is shifted vertically for clarity................................................................. 80

6.5 Calculated density profiles along the surface normal for \( T = 100, 200 \) and 300K for \( n = 13 \). The curves for \( T = 100 \) and 300K are shifted vertically for clarity................................................................. 81

6.6 The effect of lattice spacing on the radial distribution function of oxygen-oxygen atoms in PEG part of SAMs. The peak about 3 Å is due to the gauche conformation about CC bond in PEG part and becomes prominent with lattice spacing. The lack of any specific peaks after 4 Å for \( a = 5.8 \) Å, indicates the amorphous nature of PEG part as lattice spacing is increased..... 82

6.7 The torsion angle distribution in PEG part for different lattice constant at \( T = 200\text{K} \)........................................................................................................ 82

6.8 Effect of PEG chain length on the gauche conformation in \( g_{oo}(r) \) .................. 83

7.1 Effect of electric field on the gauche conformation about CC bond in PEG.
(a) for positive field strengths(b) for negative field strengths......................... 87

7.2 Calculated density profiles along the surface normal for (a) positive field
(b) negative field are shown. ................................................................. 88

7.3 Effect of electric field on tilt structure for (a) positive field (b) negative
field are shown......................................................................................... 89

7.4 2D structure factors for PEG-terminated SAMs. (a) field = 0V/Å (b) field
= 2V/Å (c) field = -2V/Å........................................................................ 90

7.5 2D surface plots of the tilt direction of all the chains in (a) \( E_z = 0 \) (b) \( E_z = +2\)V/Å. (c) 2D surface plot of polarity of chains in Y-direction............... 91

7.6 Surface layer 1Å below the highest \( z \) for (a) field=0.5 V/Å(b) field=1V/Å
(c) field= 1.5V/Å(d) field=2V/Å............................................................. 92

7.7 The atomic configuration color coded by charges: oxygen (red), hydrogen
(white), methyl terminal carbon(yellow) and carbon in PEG (green) in top
6Å layer of the system (a) \( E_z = 0 \) (b) \( E_z = +2\)V/Å (c) \( E_z = -2 \) V/Å. .............. 93

7.8 The top 6Å layer of the system (a) before (b) after application of positive
electric field. (c) shows the same layer after the removal of electric field...... 94

7.9 Plot of the tilt angle as a function of time for (a) \( E_z = +2\)V/Åand (b) \( E_z = -2\)V/Å................................................................. 95
7.10 Reversible transition of trans-gauche conformations with and removal of electric field. .............................................................. 96
ABSTRACT

The purpose of this dissertation is to investigate the properties of self-assembled monolayers, particularly alkanethiols and Poly (ethylene glycol) terminated alkanethiols. These simulations are based on realistic interatomic potentials and require scalable and portable multiresolution algorithms implemented on parallel computers.

Large-scale molecular dynamics simulations of self-assembled alkanethiol monolayer systems have been carried out using an all-atom model involving a million atoms to investigate their structural properties as a function of temperature, lattice spacing and molecular chain-length. Results show that the alkanethiol chains tilt from the surface normal by a collective angle of 25° along next-nearest neighbor direction at 300K. At 350K the system transforms to a disordered phase characterized by small tilt angle, flexible tilt direction, and random distribution of backbone planes. With increasing lattice spacing, \( a \), the tilt angle increases rapidly from a nearly zero value at \( a = 4.7\text{Å} \) to as high as 34° at \( a = 5.3\text{Å} \) at 300K.

We also studied the effect of end groups on the tilt structure of SAM films. We characterized the system with respect to temperature, the alkane chain length, lattice spacing, and the length of the end group. We found that the gauche defects were predominant only in the tails, and the gauche defects increased with the temperature and number of EG units. Effect of electric field on the structure of poly (ethylene glycol) (PEG) terminated alkanethiol self assembled monolayer (SAM) on gold has been studied using parallel molecular dynamics method. An applied electric field triggers a conformational transition from all-trans to a mostly gauche conformation. The polarity of the electric field has a significant effect on the surface structure of PEG leading to a profound effect on the hydrophilicity of the surface. The electric field applied anti-parallel to the surface normal causes a reversible transition to an ordered state in which the
oxygen atoms are exposed. On the other hand, an electric field applied in a direction parallel to the surface normal introduces considerable disorder in the system and the oxygen atoms are buried inside.
CHAPTER 1
INTRODUCTION

Self-assembly is a process in which molecules or aggregates of molecules ordered structures spontaneously. The process of self-assembly was long present in the fields of materials science, chemistry and biology before it gained importance as a novel technological tool for synthesis of new kind of materials. Examples include crystals [1], lipid bilayers [2], protein folding [3] and self-assembled monolayers [4]. Understanding self-assembly is important both from the scientific and technological perspective for various reasons. First it gives a glimpse into the workings of the way the complex structures such as cells, proteins etc., self-assemble to perform the various functions. Technologically, self-assembly offers a new strategy to generate nanostructures.

For the self-assembly to occur and be viable, the following components are essential. First we need an aggregate of molecules in a less ordered state, which when interact with each other, go to a more ordered state. Secondly, the nature of the interactions, which determines the final self-assembly, is usually non-covalent in nature and can include van der Waals, Coulomb, hydrophobic, hydrogen bond type of interactions. Most of the self-assembly usually takes place in a solution, and hence the environment can play a significant role both in the time scale of the formation and also on the final structure of the self-assembly.

Most of the self-assembled structures that are studied today are in the nanoscale to macro scale range because of the ease of preparation [5] and their potential applications in materials science and engineering. From the present technology point of view, self-assembly is possibly the only way to generate nanoscale 3D structures [6], as lithographic techniques, which dominate the fabrication industry, are only planar in nature. The other issues concerning conventional lithography are that at nanometer scale, the current technology is prohibitively expensive and
time consuming. So there is an increasing need to develop alternative technologies when dealing with nanostructures. Recent experiments towards this direction [7] have been successful in making nanoscale templates made of genetically engineered ring structured proteins. Many issues regarding self-assembly are still unclear. They include the range of formation of these structures, the thermal stability, nature of the defects, etc.

Monolayers of alkanethiols on gold are the most studied SAMs, which were first prepared by Nuzzo and Allara in 1983 [8] by adsorbing di-\textit{n}-alkyl disulfides from dilute solutions. SAMs with their ease of preparation and manipulation of the surface properties by just changing the end groups, provide design flexibility and opportunity to study interactions at interfaces. SAMs have applications in diverse fields such as wetting [9], adhesion, lubrication [10] etc.; one of the more recent applications of SAMs is in the field of molecular recognition [11]. Macromolecules such as proteins and DNA can be immobilized on monolayers. The resistance of SAMS with certain end groups (oligo ethylene glycol) to protein adsorption [12] has opened up new avenues where SAMs can be used to tailor which macromolecules bind on the surface.

Self-assembly of alkanethiol films are attractive in diverse fields due to the possibility of tuning their properties by selectively modifying specific functional groups (i.e., end groups) without modifying the entire chain. For instance, by changing just the end group of alkanethiol-based monolayer [e.g., \(S(CH_{2})_{n}CH_{3}\)] from \(CH_{3}\) to \(OH\), the property of the resulting chemical surface can be varied from hydrophobic to hydrophilic. With the emerging technologies in biomaterials, the protection of the surfaces from contamination by proteins, cells, blood, etc has become essential [13,14]. Many biomaterials lack the ability to protect the surfaces and this contamination could compromise their bulk mechanical properties, which is undesirable [15].
The reduction of adsorption of proteins on surfaces is important because, it has been shown that the absorption is the first step in immune response of the body on any foreign material. Many strategies have been tried and used to control the non-specific adsorption of proteins on surfaces, and poly (ethylene glycol)(PEG) has met with considerable success [106,108]. Its low toxicity and low immunogenicity render it to be an important coating on many biomaterials. PEG is a water solvable polymer [16] and is one of the most biocompatible polymers [12,107]. Its high water solvability is attributed to the helical nature of its backbone, which maximizes the hydrogen bonding, while at the same time hiding the hydrophobic parts [115]. PEG-terminated alkanethiol SAMs exhibit interesting protein adsorption properties [110] and as such are useful for interfacing biological materials with inorganic surfaces. For example, it is possible to immobilize a single biomolecule or an array of biomolecules on the surface of SAMs for biosensor applications.

SAMs exhibit rather complex structural and dynamical behaviors, which are not well understood. Although the molecular chains show collective tilt away from the surface normal, the atomistic details of the tilt structure are not known. Several fundamental issues such as mentioned below require investigation, the understanding of which will help us in not only understanding the mechanisms of self-assembly but also aid in designing new applications. Some of the questions that are not yet well understood are (1) the nature of phase transitions and the order to disorder transitions (2) what impact does the end group have on the structure and properties of SAMs (3) what kind of structures are formed and what are the parameters that characterize the system and structure (4) what are the effects of internal parameters like chain length or choice of substrate and the effects of external parameters such as temperature and electric field. Among these parameters, an externally applied electric field is technologically
most important due to its control over surface properties e.g., reversible switching of hydrophilicity has been observed in SAMs [113]. To optimize surface properties, it is important to atomistically understand the dependence of surface structures on an applied electric field. Unfortunately the effect of electric fields on the structure of SAMs at the atomic level is not well known. Atomistic simulations based on molecular dynamics (MD) technique are expected to provide some insight into the behavior of the system.

Computer simulations play an increasing role, along with experiments, in understanding many of the issues in fields of science and engineering. Simulations are widely accepted as an important scientific investigative tool, which complements the traditional experimental methods [17]. Many of the nanoscale systems are still beyond the reach of experiments but are accessible to simulations, which can play a crucial role in bridging the gap between theory and experiments. Different methodologies are developed to study relevant phenomena at microscopic level, some of which are density functional theory [18], Monte Carlo methods [19], molecular dynamics methods [20], \textit{ab initio} molecular dynamics method [21] and tight-binding molecular-dynamics method [22].

For many purposes, classical treatment of the atoms (separation of nuclei and electronic motion) using empirical interatomic potentials suffices. Molecular dynamics (MD) approach is one such approach, where the problem is reduced to solving a system of $3N$ coupled Newton’s equations of motion. The key ingredient of MD approach, which provides phase-space trajectories, is the interatomic potential. Most of the potentials are developed by a combination of \textit{ab initio} structure calculations and experimental data. Realistic potentials, in general, have very complicated functional forms and can be computationally intensive.
Computing power is the next key issue that needs to be considered, when simulations of the order of nano-scale are to be studied. A. Rahman performed the first molecular dynamics simulations with realistic potential on 864 argon atoms in 1964 on a CDC 3600 [23]. With emerging technologies and advances in computer hardware, it is now possible to simulate system sizes of the order of $\sim 10^8$ atoms [24]. Most of the large-scale system simulations are possible because of computers with massive parallel architectures. Recent construction of Beowulf clusters with off-shelf PCs provides a cost-effective viable alternative to the high-end supercomputers such as IBM SP machines [25, 26]. For example, Louisiana State University (LSU) has recently acquired a Linux cluster [27] consisting of 512 dual Intel Xeon 1.8 GHz nodes (i.e., 1,024 processors) connected by Myricom’s Myrinet interconnect.

The focus of this thesis is the large-scale parallel MD simulations of self-assembled monolayers (SAM) of alkanethiol systems. We have investigated the effects of temperature, substrate, chain lengths of SAMs on the tilt structure of the system. We have also studied the effect of end groups and the effect of electric field on the overall structure of SAMs.

The outline of this thesis is as follows: Chapter 2 deals with description of MD methods, interaction potential; Chapter 3 contains description of multiresolution algorithms including Fast Multiple algorithm; Chapter 4 contains details of parallel implementation of our MD algorithm for organic systems and our results regarding the scalability tests performed on high-end parallel machines; MD simulations of alkanethiols are reported in Chapter 5 and MD simulations of PEG-terminated SAMs are described in Chapter 6; finally Chapter 7 contains results on the effect of electric field on the structure of SAMs.
CHAPTER 2
MOLECULAR DYNAMICS SIMULATIONS

In molecular dynamics (MD) simulation, the time evolution of interacting atoms is determined by integrating the Newton’s equations of motion. The trajectory of \( N \)-atoms is calculated in a \( 6N \) phase space (\( 3N \) positions and \( 3N \) momenta). Molecular dynamics is a statistical mechanical simulation method, in which macroscopic physical properties of the system can be predicted and the time evolution at microscopic level can be determined. In Section 2.1 the basics of MD are discussed, followed by a discussion of inter-atomic potentials in Section 2.2. Sections 2.3 and 2.4 discuss the boundary conditions and evaluation of physical properties, respectively.

2.1 Molecular Dynamics (MD) Methodology

In MD simulations we consider a system of \( N \) atoms with coordinates \( \{ \mathbf{r}_i \mid i = 1, \ldots, N \} \) and velocities \( \{ \mathbf{v}_i \mid i = 1, \ldots, N \} \) interacting via an inter-atomic potential \( V(\{ \mathbf{r}_i \}) \). The Hamiltonian of the system of atoms treated as classical point-like objects can be written as

\[
H = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + V(\{ \mathbf{r}_i \}).
\]  

(2.1)

Substituting (2.1) in the Hamiltonian equations of motion,

\[
\mathbf{r}_i = \frac{\partial H}{\partial \mathbf{p}_i} \quad \text{and} \quad \mathbf{p}_i = \frac{\partial H}{\partial \mathbf{r}_i}
\]

(2.2)

yields the Newton’s equations of motion

\[
m_i \ddot{\mathbf{r}}_i = \frac{\partial}{\partial \mathbf{r}_i} V(\{ \mathbf{r}_i \}) = \mathbf{F}_i \quad \{i = 1, \ldots, N\},
\]

(2.3)

where \( \mathbf{F}_i \) is the force on atom \( i \).

The use of the classical equations of motion at atomistic level instead of quantum mechanics (at least for most purposes) can be justified through two considerations: Born-
Oppenheimer approximation and de Broglie wavelength. In 1923, Born and Oppenheimer [28] noted that nuclei are much heavier than electrons, resulting in a considerable difference in the velocities. It is therefore reasonable to regard the nuclei as fixed when compared to the electrons and express the Hamiltonian of the system only in terms of nuclei variables, with the fast motion of electrons being averaged out. The de Broglie wavelength [29] is defined as

\[ \lambda = \sqrt{\frac{2\hbar^2}{Mk_B T}} \]  

(2.4)

where \( M \) is the atomic mass (or nuclei mass) and \( T \) is the temperature. The classical approximation is valid if \( \lambda/a \ll 1 \), where \( a \) is the inter atomic separation. Some typical values of \( \lambda/a \) are 0.97 for hydrogen and 0.054 for sodium (at triple-point temperatures) [2]. So we see that the quantum effects will be small in the case of sodium, but are significant in case of hydrogen.

2.2 Interatomic Potentials for Organic Systems

For realistic simulation of materials, the most important ingredient in a MD simulation is the reliable inter-atomic potential function, \( V \) in Eq. (2.1). The selection of the potential function determines the accuracy of the results, and the feasible time scale of the simulation is determined by the complexity of the potential function selected. The potential function can be generally written as a sum of one-body, two-body and higher order potential functions:

\[ V(r_1, \ldots, r_N) = \sum_i V^{(1)}(r_i) + \sum_{j>i} V^{(2)}(r_i, r_j) + \sum_{k>j>i} V^{(3)}(r_i, r_j, r_k) + \ldots \]  

(2.5)

The forces acting on the individual atoms then are derived from the potential as:

\[ \mathbf{F}_j = \frac{\partial}{\partial r_j} V(r_1, \ldots, r_N) \]  

(2.6)

which are then used in classical equations of motion to advance the positions and momenta of atoms.
In Eq. (2.5), the first term represents the effect of an external field, if any. The second term, a pair-wise potential, is the most important part of the potential function, which depends only on the relative separation between the two atoms and hence can be replaced by \( V^{(2)}(r_{ij}) \), where \( r_{ij} = \| \mathbf{r}_i - \mathbf{r}_j \| \). The third and fourth terms in Eq. (2.5) are computationally more expensive operations, and hence are not used in most of the simulations of liquids and crystalline materials. These terms, however, are essential when one is dealing with molecular systems such as polymers and proteins. One can still follow an atomistic approach to the molecular systems; only the chemical bonds will be replaced by potential terms. The chemical bonds can be treated rigid, in which case constraint dynamics needs to be applied, or a harmonic potential well can be applied to allow a small change in the chemical bond length. A detailed description is given in the following Sections. The interaction potentials for the polymeric systems can be divided into two kinds: bonded interactions and non-bonded interactions.

2.2.1 Bonded Potential Terms

The bonded potential terms are bond stretching, bond bending and torsion terms.

\[
V_{Bonded} = V_{Stretched} + V_{Bent} + V_{Torsion}
\]  

(2.7)

The bond stretch and bond angles terms can be evaluated as though the atoms are connected through springs. Once this assumption is made, Hooke’s law gives the description of the potential terms.

Consider two bonded atoms \( i \) and \( j \) with parameters, the natural bond length \( r_{ij}^0 \) and the spring force constant \( K_{ij} \). The displacements from the natural bond length can be formulated as

\[
V_{Stretched} = \frac{1}{2} K_{ij} (r_{ij} - r_{ij}^0)^2
\]  

(2.8)
The parameters $r_{ij}^0$, $K_{ij}$ are specific to a particular force field used. The force constant $K_{ij}$ is important to consider because it determines how ‘deep’ the well is or how rigid the bond can be held. Another form of potential common in literature for the bond stretching potential term is the Morse potential [30] which has the form

$$V_{Morse} = D_{ij}[e^{\frac{\alpha}{2}(r_{ij} - r_{ij}^0)} - 1]^2$$

(2.9)

where $D_{ij}$ is the bond energy for the bond between atoms $i$ and $j$, $\alpha$ is related to $K_{ij}$ and $D_{ij}$ as

$$\alpha = \frac{K_{ij}^2}{D_{ij}}$$

(2.10)

The two potentials are shown in Figure 2.1. The Morse potential has an anharmonic well and is more suited for problems that involve the dissociation or breaking of bonds. Morse potential is typically used for systems in which breaking of bonds never occurs, but it might result in stretching the bonds to unrealistic lengths. In all our simulations, a harmonic potential form is used for bond stretching.

![Figure 2.1: Comparision of harmonic and Morse potentials for bond stretching.](image-url)
The bond angle potential has also the form of a harmonic potential as stretch potential (see Figure 2.2), now the angles replacing the bond lengths:

\[ V_{\text{Bend}} = \frac{1}{2} K_{ijk} q_{ijk} q_{ijk} - q_{ijk}^0 \]  

(2.11)

where \( K_{ijk} \) is the bending force constant and \( q_{ijk}^0 \) is the natural bend angle between the bonds \( i-j \) and \( j-k \), and \( j \) is the central atom. \( q_{ijk} \) can be defined in terms of \( r_{ij} \) and \( r_{jk} \) as

\[ q_{ijk} = \cos^{-1} \left( \frac{r_{ij} \cdot r_{jk}}{r_{ij} r_{jk}} \right) \]  

(2.12)

![Figure 2.2: Harmonic bend angle potential](image)

The torsion angle potential is a four body potential, which cannot be neglected in organic systems as it gives rise to conformational transformations. The torsion angle, which is also called dihedral angle of four atoms in sequence, \( i-j-k-l \) is defined as the angle between the plane of atoms \( i-j-k \) and plane of atoms \( j-k-l \) as shown in Figure 2.3(a) and in Eq. (2.13):

\[ \square = \cos^{-1} \left( \frac{r_{ij} \cdot r_{jk}}{r_{ij} r_{jk}} \right) \left( \frac{r_{jk} \cdot r_{kl}}{r_{jk} r_{kl}} \right) \]  

(2.13)
The form of the torsion potential is usually periodic (harmonic) in nature (see Figure 2.3(b)). With respect to the angles made by the two planes, we can define two possible minima in the potential viz., trans and gauche conformations [31,32]. The torsion angle is said to be in trans conformation, when the torsion angle $\phi$ is $0^\circ$ and in gauche conformation when $\phi$ is $\pm 120^\circ (g^-)$ or $\pm 120^\circ (g^+)$. The torsional potential usually takes either of the forms in Eq. (2.13) or (2.14).

\[
V(\phi) = a_0 + \sum_n a_n \cos^n(\phi) \tag{2.14}
\]

\[
V(\phi) = a_0 + \sum_n a_n \cos^n(\phi - \phi_0) \tag{2.15}
\]

Figure 2.3: Torsion angle potential (a) definition of torsion angle (b) form of torsion angle potential.

### 2.2.2 Non-bonded Potential Terms

The non-bonded potential is a pair potential, which depends only on the distance between the atoms. The non-bonded potential usually has a ‘van der Waals’ term, core repulsion term and coulomb term. The van der Waals forces are relatively weak electrostatic forces between uncharged atoms and van der Waals solids have lower melting points than those held by covalent
or ionic bonds. The atoms, though are electrically neutral, may be slightly polar in nature introducing dipoles in the system. The weak attractive forces between opposite dipoles accounts for the van der Waals potential, which has the form of $\frac{1}{r_{ij}}$. The core repulsion term due to Pauli’s exclusion principle can be represented by both $\frac{1}{r_{ij}^{12}}$ form or the exponential form. Shown in Figure 2.4 are two such representations of van der Waals and core repulsion terms: Lennard-Jones potential [32] or the Buckingham potential [33] form.

The Lennard-Jones potential can be divided into two parts: the core repulsive part at short distance due to Pauli’s exclusion principle. The second part of the Lennard-jones potential is the attractive tail at long separations, which can be attributed to the correlation between the electron clouds surrounding the atoms.

$$V(r_{ij}) = 4\varepsilon \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6}$$

The Buckingham potential has an exponential term to represent the core repulsion terms. At short distances, the potential behaves like a hard sphere potential, but softens as distance increases.
\[ V(r_{ij}) = A_{ij} \exp(B_{ij}r_{ij}) \left[ C_{ij} \frac{1}{r_{ij}^6} \right] \]  

(2.17)

In a system of charges, Coulomb potential, in addition to the pair potentials described above, is needed to describe electrostatic interactions of the system. The Coulomb interaction between two charged particles is given by

\[ V_{\text{Coul}}(r_{ij}) = \frac{q_i q_j}{\epsilon r_{ij}} \]  

(2.18)

where \( q_i \) is the charge on atom \( i \) and \( \epsilon \) is the dielectric constant of the medium. Different methods to calculate Coulomb interaction will be described in next chapter. The description of specific interaction forms and the potential parameters for the system of self-assembled monolayers are given in chapters 5 and 6.

2.3 Periodic Boundary Conditions

Modern MD simulations investigate the properties of a system of a few thousand to million atoms, but still are far from the thermodynamic limits \( (N = 10^{23}) \). For these system sizes, one cannot neglect the effect of periodic boundary conditions. For a three dimensional system of \( N \) atoms, the fraction of atoms at the surface is proportional to \( N^{3/3} [34] \) which is not a small number. For example, in a simple cubic lattice of 1000 atoms, \(~49\%\) atoms are on the surface. To minimize the problem of surface effects, periodic boundary conditions (PBC) are employed [35]. When PBC is used, the atoms are enclosed in a box that is repeated in all the three directions to infinity. Any atom \( i \) interacts with not only all the atoms in the simulation box, but also with all the images (which are infinite!). During the course of simulation, if an atom \( i \) leaves the box boundaries, its image will enter the simulation box from the opposite face (see Figure 2.5). Thus the number of atoms is conserved. The PBC virtually eliminates the surface effects and also the
simulation box itself has no more significance, as a translation of the box boundaries to the next lattice positions will leave the dynamics unchanged.

The PBC is accompanied with some problems, such as the increase in the number of atoms (because of their images). Furthermore the finite sums of the potential energies should now be converted to infinite sums because of infinite images. In calculating the forces and the energy on an atom \( i \) in the simulation box, we have to sum over all its interactions with \( N-1 \) atoms in the box. But with PBC, in principle we need to include the interactions of atom \( i \) with all the infinite images of the \( N-1 \) atoms also, which in practice is impossible to do. If the potential functions used in the simulation are short-ranged in nature, we can make an approximation called minimum image convention [36]. Here an atom \( i \) in the simulation box will interact only with the periodic images of the other \( N-1 \) atoms that lie within the simulation box. In short range potentials, for any atom \( i \) the interactions with all atoms \( j \) which are within a distance \( r_c \) dominate. In a simulation with minimum image convention, \( r_c \) is less than half of the box dimension. If \( r_c \) violates this condition, any atom \( i \) might interact with more than one periodic image of another atom which is incorrect.
The introduction of the cutoff distance \( r_c \) also helps in further reduction of the number of interaction terms. An atom \( i \) now interacts with only atoms that fall within the sphere of radius \( r_c \), a reduction of \( 4\pi r_c^3 / 3L^3 \) number of neighbors, which can be significant. However, the truncation of the potential at \( r_c \) may introduce errors in the calculation of potential energy. The error due to the introduction of the cutoff radius can be reduced by choosing the \( r_c \) as large as possible. The error can increase exponentially unless the potential energy \( V(r) \) decays much more rapidly than \( r^{-6} \) in three dimensions. This is true for the van der Waals potential function, which decays as \( r^{-6} \), and hence truncated form can be used. But in Macromolecular systems, where the Coulomb and dipolar terms are very important, such a truncation method cannot be used (they decay as \( r^{-1} \)). Hence PBC is a real problem when it comes to electrostatic interactions. Methods to deal with this problem are described in the next chapter.

### 2.4. Thermal Quantities and Statistical Mechanics

Some of the thermal quantities that define a macro state are temperature \( T \), pressure \( P \), Number of atoms \( N \), and volume \( V \). A set of above thermodynamic parameters defines a particular ensemble, which is used to calculate the average thermal properties that define the system. For example, a microcanonical ensemble is defined by \( N \), \( V \), and \( E \). Statistical mechanics on the other hand deals with the ensemble averaged values of the physical quantities. As described earlier, a system of \( N \) atoms spans a phase space, \( [q^{3N}(t),r^{3N}(t)] \) of 6N dimensions, and any physical quantity \( A \) will be calculated in ensemble average as:

\[
A = \langle A \rangle_{\text{ensemble}} = \sum_\square A(\square) \rho_{\text{ensemble}}(\square),
\]

(2.19)

where \( \rho_{\text{ensemble}}(\square) \) is the phase space density associated with the ensemble under consideration. This calculation of ensemble averages, however, is a very exhaustive process, as all the different
states in the phase space have to be scanned to calculate the average of thermodynamic quantities.

Under certain circumstances, time average becomes equivalent to an ensemble average. The probability density of an ensemble should not change with time, and any reasonable starting configuration should tend to a stationary state as time proceeds. In MD simulations, with proper choice of initial configuration, and proper selection of time step, both conditions can be satisfied. If the simulation time is long enough, even ergodicity can be nearly ensured. The time average of physical quantity $A$ can be calculated as

$$A = \langle A \rangle_{\text{ensemble}} = \frac{1}{t_{\text{simulation}}} \sum_{t=1}^{t_{\text{simulation}}} A(t)$$  \hspace{1cm} (2.20)

### 2.5 Structural Correlations

One of the correlation functions which is often calculated in MD simulations is the particle radial distribution function (RDF), which measures the deviation from complete randomness or complete crystallinity. If the system is isotropic, the particle distribution function reduces to a pair radial distribution function (RDF), which only depends on the relative separation between the atoms. If the separation is much longer than the inter-particle potential range, RDF approaches the ideal gas limit. The RDF can be directly measured by the x-ray or neutron-scattering experiments, which is one of the reasons it is an important quantity calculated in MD simulations. In addition to the experimental relevance, many thermodynamic quantities can also be calculated. In the case of $N$ identical atoms in a volume $V$, the 2-particle RDF can be defined as [29]:

$$g_N^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = \frac{\langle \delta^{(2)}(\mathbf{r}_1, \mathbf{r}_2) \rangle}{\langle \delta^2 \rangle} = \frac{1}{V} \left\langle \prod_{i<j} \delta(\mathbf{r}_i - \mathbf{r}_j) \delta(\mathbf{r}_2 - \mathbf{r}_j) \right\rangle.$$  \hspace{1cm} (2.21)
where \( \rho_N^2(r_1, r_2) \) is the particle density function and \( \rho = \rho_N = N/V \) is the particle density of a homogeneous system. If the system is translational invariant, the RDF depends only on the relative separation \( r = r_1 - r_2 \):

\[
g_N^{(2)}(r) = \frac{\rho_N^2(r)}{\rho^2} = \frac{\rho_N^2(r) V}{N} \tag{2.22}
\]

We now derive the expression for \( \rho_N^2(r) \) using the properties of delta functions [37].

\[
\rho_N^2(r)V = \rho_N^2(r + r_i - r_j)\,dr_i\,dr_j = \left\langle \delta\left(\mathbf{r} - \mathbf{r}_i\right)\delta\left(\mathbf{r} - \mathbf{r}_j\right)\delta\left(\mathbf{r} - \mathbf{r}_i\right)\delta\left(\mathbf{r} - \mathbf{r}_j\right)\,d\mathbf{r}_i\,d\mathbf{r}_j \right\rangle = \left\langle \delta\left(\mathbf{r} + \mathbf{r}_j - \mathbf{r}_i\right)\right\rangle \tag{2.23}
\]

Here we use Eq. (2.21) and the identity, \( \int f(x_i)\delta(x_1 - x_2)\,dx_i = f(x_2) \). Now \( g_N(r) \) can be written as

\[
g_N^{(2)}(r) = \frac{1}{N\rho^2} \left\langle \delta\left(\mathbf{r} + \mathbf{r}_j - \mathbf{r}_i\right)\right\rangle = \frac{V}{N^2} \left\langle \delta\left(\mathbf{r} + \mathbf{r}_j - \mathbf{r}_i\right)\right\rangle \tag{2.24}
\]

where \( \mathbf{r}_j = \mathbf{r}_i - \mathbf{r}_j \). In case of isotropic system, one can average over all directions of \( r \) to yield

\[
g_N^{(2)}(r) = \frac{V}{4\rho^2N^2} \left\langle \delta\left(\mathbf{r} + \mathbf{r}_j - \mathbf{r}_i\right)\right\rangle \tag{2.21}
\]

In simulations, the range of potential function (or for better accuracy, half the box length) is divided into bins and the delta functions are replaced by a histogram compiled of all pair separations falling within each such bin. In multi species system, the partial distribution functions give more useful information. Consider a system of two species \( \mathcal{A} \) and \( \mathcal{B} \) and the corresponding number of atoms \( N_{\mathcal{A}} \) and \( N_{\mathcal{B}} \) such that the total number of atoms is \( N = N_{\mathcal{A}} + N_{\mathcal{B}} \) in a volume of \( V \). The partial RDF \( g_{\mathcal{AB}}(r) \) for a homogeneous and isotropic system is defined as

\[
g_{\mathcal{AB}}(r) = \frac{V}{4\rho^2N_{\mathcal{A}}N_{\mathcal{B}}} \left\langle \delta\left(\mathbf{r} + \mathbf{r}_j - \mathbf{r}_i\right)\right\rangle \tag{2.22}
\]
The ensemble average in Eq. (2.22) can be replaced by a time average as explained in Section 2.4. Typically in simulations, the RDF is calculated every $k$ steps and if the number of times the RDF calculated is $n_{rdf}$, a time average is taken as follows:

$$\langle g(r) \rangle_{time} = \frac{1}{n_{rdf}} \sum_{i=t_0}^{t_0+k} g(r)$$

(2.23)

where $t_0$ is updated every $k$ steps.

The static structure factor of $N$ atom system is defined as the autocorrelation function

$$S(k) = \frac{1}{N} \langle \overline{\rho}(k) \overline{\rho}(k) \rangle$$

(2.24)

where $\overline{\rho}(k)$ is the Fourier transform of local density of $\overline{\rho}(r)$ as

$$\overline{\rho}(k) = \sum_{i=1}^{N} \exp(i \mathbf{k} \cdot \mathbf{r}_i)$$

(2.25)

Substituting Eq. (2.25) in Eq. (2.24), we get

$$S(k) = \frac{1}{N} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \exp(i \mathbf{k} \cdot \mathbf{r}_i) \exp(i \mathbf{k} \cdot \mathbf{r}_j) \right)$$

$$= \frac{1}{N} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \exp(i \mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)) \right).$$

(2.26)

Equation (2.26) can be split into two parts:

$$S(k) = \frac{1}{N} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \exp(i \mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)) \right) + \frac{1}{N} \left( \sum_{i=1}^{N} \sum_{j=1}^{N} \exp(i \mathbf{k} \cdot (\mathbf{r}_i \mathbf{r}_j)) \right)$$

(2.27)

and

$$\left( \sum_{i=1}^{N} \sum_{j=1}^{N} \exp(i \mathbf{k} \cdot (\mathbf{r}_i \mathbf{r}_j)) \right) = N$$

(2.28)

so, substituting Eq. (2.28) in Eq. (2.27), we get
\[ S(k) = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \prod_{j=1}^{N} \exp[i \mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)] \right\} \] (2.29)

Using the definitions of delta functions,

\[ S(k) = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \prod_{j=1}^{N} \delta(\mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)) \right\} \] (2.30)

Using Eq. (2.21) in Eq. (2.30), we get

\[ S(k) = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i \mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)] \] (2.31)

\[ = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i \mathbf{k} \cdot (\mathbf{r}_i - \mathbf{r}_j)] \] (2.32)

\[ = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i \mathbf{k} \cdot \mathbf{r}_i] \] (2.33)

We can rewrite Eq. (2.33) as

\[ S(k) = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i \mathbf{k} \cdot \mathbf{r}_i] \] (2.34)

Using the definition of delta function of \( \delta(\mathbf{k}) = \frac{1}{(2\pi)^3} \int \exp[i \mathbf{k} \cdot \mathbf{r}] d\mathbf{r} \), Eq. (2.34) can be written as

\[ S(k) = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i \mathbf{k} \cdot \mathbf{r}_i] \] (2.35)

The delta function in Eq. (2.35) corresponds to the forward scattering of radiation, and can be ignored, hence

\[ S(k) = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i \mathbf{k} \cdot \mathbf{r}_i] \] (2.36)

If the system is isotropic, structure factor depends only on \( k = |\mathbf{k}| \), thus

\[ S(k) = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i k \cos \theta] \] (2.34)

\[ = 1 + \frac{1}{N} \left\{ \prod_{i=1}^{N} \delta(\mathbf{k} \cdot \mathbf{r}_i) \right\} \exp[i k \cos \theta] \] (2.34)
and using \( \int \exp(ikr \cos \theta) \cos \theta \, dr = 2 \frac{\sin kr}{kr} \), Eq. (2.34) can be written as

\[
S(k) = 1 + 4 \int \int r^2 \left[ g(r) \right] \frac{\sin kr}{kr} \, dr
\]

Partial static structure factors are calculated from the Fourier transforms of corresponding partial pair-distribution functions [38]

\[
S_\alpha\beta(k) = \int \int \left[ g_\alpha\beta(r) \right] \frac{\sin kr}{kr} \, dr
\]

where \( c_\alpha\beta = \frac{N_\alpha N_\beta}{N} \).

The x-ray structure factor can be obtained by weighting the partial structure factor with form factors \( b_\alpha \) as

\[
S_n(k) = \frac{\int \int \left[ b_\alpha b_\beta \left( c_\alpha c_\beta \right)^{1/2} \right] S_\alpha\beta(k) \, dr}{\int \int \left[ b_\alpha b_\beta \right] \left( c_\alpha c_\beta \right)^{1/2} \, dr}
\]
Molecular dynamics of realistic materials involve system sizes of $\sim 10^5 \parallel 10^6$ atoms and complex potential functions (see Section 2.2). The potential functions describe phenomena at different time scales, and an incorporation of algorithms, which span time scales, is essential. The integration algorithms themselves have to satisfy stability and conservation requirements, and the error bounds on particularly chosen integration algorithm is particularly relevant. The choice of time step determines the simulation time and is difficult to choose given the variation of time scales in potential field. In this chapter we describe the various space–time resolution based algorithms developed to deal with above-mentioned issues. In Section 3.1, we describe the integration algorithms used in MD simulations. In Section 3.2, bonded-list management algorithms are described, and in Section 3.3, lists for non-bonded short-range interactions are described. Long-range Coulomb interactions (Fast Multipole Method) are discussed in Section 3.4, and finally in Section 3.5 multiple-time-scale methods applied in this study are described.

3.1 Integration Algorithms

The time trajectory of atoms in the phase space is described through the integration algorithms, which are based on finite difference methods. The time in finite difference methods is discretized on a grid, with a separation between grid points being $\Delta t$, the time step of the algorithm. The choice of $\Delta t$ determines the length of the simulation run and also depends on the integration algorithm. Choosing longer time step enables one to simulate for longer times but the accuracy is an important issue to consider. Energy conservation is also another important consideration. There are two types of energy conservation, short time and long time. Most of the time integration algorithms are good at short time energy conservation, but the long time energy
conservation is more difficult to achieve. The extreme dependence of the final states on the small changes in the initial conditions is called the Lyapunov instability [7] and the time scale should be selected such that the total simulation time is much longer than the time for the instability to develop. The reversibility of the integration algorithms is also an important issue as the Newton’s equations of motion are reversible. The two popular algorithms used in MD simulations are the Verlet algorithm [39,40] and predictor-corrector [41] algorithms.

Verlet algorithm and the velocity version of the same are Taylor’s expansion of atom position and velocities. To derive the Verlet algorithm, we start from the Taylor’s expansion of position:

\[ \mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \frac{\partial \mathbf{r}(t)}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 \mathbf{r}(t)}{\partial t^2} (\Delta t)^2 + \frac{1}{3!} \frac{\partial^3 \mathbf{r}(t)}{\partial t^3} (\Delta t)^3 + O((\Delta t)^4) \]  

(3.1)

and

\[ \mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \frac{\partial \mathbf{r}(t)}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 \mathbf{r}(t)}{\partial t^2} (\Delta t)^2 - \frac{1}{3!} \frac{\partial^3 \mathbf{r}(t)}{\partial t^3} (\Delta t)^3 + O((\Delta t)^4). \]  

(3.2)

Adding Eqs. (3.1) and (3.2) we get:

\[ \mathbf{r}(t + \Delta t) + \mathbf{r}(t - \Delta t) = 2 \mathbf{r}(t) + \frac{\partial^2 \mathbf{r}(t)}{\partial t^2} (\Delta t)^2 + O((\Delta t)^4). \]  

(3.3)

Rewriting \( \frac{\partial^2 \mathbf{r}(t)}{\partial t^2} = \mathbf{a}(t) = \frac{\mathbf{F}(t)}{m} \), where \( \mathbf{a}(t) \) is the acceleration, \( \mathbf{F}(t) \) is the force and \( m \) is the mass:

\[ \mathbf{r}(t + \Delta t) = 2 \mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{F}(t)}{m} (\Delta t)^2 + O((\Delta t)^4). \]  

(3.4)

The velocities can be derived by subtracting the two equations:

\[ \mathbf{v}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{2\Delta t} + O((\Delta t)^2) \]  

(3.5)
The positions are accurate to the order of $\|t^1$ but the velocities are accurate to only $\|t^2$. However the velocities are used only to calculate the kinetic energy and temperature and loss of accuracy in velocity calculation will not affect the dynamics. Verlet algorithm is reversible and has very little long time energy drift. The usage of memory was minimal as only positions and velocities are stored. One drawback of the Verlet algorithm is that the velocities and positions are not calculated at the same time. We can cast the algorithm in a different way so that the velocities are calculated at the same time as the positions, which is called velocity-Verlet algorithm [42]. The positions and velocities are expanded in Taylor’s form:

$$v(t + \Delta t) = v(t) + \frac{\partial v(t)}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 v(t)}{\partial t^2} \Delta t^2 + O(\Delta t^3). \quad (3.6)$$

Rewriting the velocity equation in terms of acceleration:

$$v(t + \Delta t) = v(t) + a(t) \Delta t + \frac{1}{2} \frac{\partial a(t)}{\partial t} \Delta t^2 + O(\Delta t^3). \quad (3.7)$$

We can use the Taylor’s expansion of acceleration as:

$$a(t + \Delta t) = a(t) + \frac{\partial a(t)}{\partial t} \Delta t + O(\Delta t^2) \quad (3.8)$$

Multiplying the above equation with $\Delta t$ and rewriting the equation we get:

$$\frac{\partial a(t)}{\partial t} \Delta t^2 [a(t + \Delta t) - a(t)] \quad (3.9)$$

Substituting the above equation in velocity expression we obtain:

$$v(t + \Delta t) = v(t) + \frac{1}{2} (a(t + \Delta t) + a(t)) \Delta t^2 + O(\Delta t^3) \quad (3.10)$$

The velocity form of the Verlet algorithm can thus be written as:

$$r(t + \Delta t) = r(t) + v(t) \Delta t + \frac{1}{2} \frac{F(t)}{m} \Delta t^2 + O(\Delta t^3) \quad (3.11)$$
\[ \mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \frac{1}{2} \left( \mathbf{F}(t + \Delta t) + \mathbf{F}(t) \right) \Delta t^2 + O(\Delta t^3) \] (3.12)

In our simulations in this thesis, we have used the velocity-version of the Verlet algorithm.

### 3.2. Short-Range Interactions

#### 3.2.1 Bonded Interactions

Macromolecules usually have several structural units connected by covalent bonds. Covalent bonds are formed by sharing electrons between two or more atoms and are very short-range in nature (1-2 Å) and are characterized by strengths of the order of 100 Kcal/mol. Depending on the number of valence electrons or position in the periodic table, an atom can form a certain number of bonds with other atoms. The valences of some of the most commonly occurring atoms in macromolecular chemistry are one for hydrogen, two for oxygen, three for nitrogen and four for carbon. These covalent bonds also have the property of directionality, which dictates the orientation of these bonds with respect to each other. The bond lengths and the bond angles determine how these multi-valent atoms form bonds and form crystalline solids. Usually chains of atoms, which form the backbone of the polymer, are connected further with side atoms or chains giving the macromolecule its unique properties. There exists yet another structural feature in macromolecules, which determines their properties, the amount of rotation possible around any of the covalent bonds, which is called torsion motion.

Two atoms are involved in a bond formation, three atoms in angle formation and four atoms in torsion angle formation. The distance range of three bonded interactions are 1-2 Å for bond length, 1-3 Å for bend angle and 2-4 Å for torsion interactions, which are all short-range, see Figure 3.1. In our simulations, the bonded atom information is used to compute the other two (bend and torsion) interaction lists.
3.2.2 Non-Bonded Interactions

For van der Waals contribution to the non-bonded potential energy, a cutoff distance \( r_c \) is introduced beyond which the potential and force are set to zero. Though the introduction of \( r_c \) reduces the computational complexity, it can cause other difficulties such as non-conservation of energy whenever a pair of molecules crosses \( r_c \). In order to circumvent the truncation errors we shift the potential and forces so that they are continuous at \( r_c \).\[43\]

\[
V(r_{ij}) = V(r_{ij}) - V(r_{c}) \bigg|_{r_{ij} = r_{c}} - \int_{r_{ij} - r_{c}} dV(r_{ij}) = 0, \quad r_{ij} > r_{c} \tag{3.13}
\]

Figure 3.2 shows the two curves for the shifted and original Lennard-Jones potential. Once the cutoff is introduced, we can take advantage of the fewer number of atoms that are present within the cutoff distance. Verlet [39] introduced a book keeping technique called neighbor lists to do exactly this. A list is constructed periodically by going over all the atoms and calculating which atoms are within the cutoff radius, which can be used for a certain number of MD steps before it is reconstructed again. The number of times the list is reconstructed depends on the time step and the average velocity of the atoms. This method reduces the time spent in searching all the atoms (which is \( O(N^2) \)) at every time step. The interval between the updates can be fixed at the beginning of the simulation. A more refined method of updating the lists, by
monitoring the displacements of atoms is also used in simulation [44], which is more time consuming.

In case of macromolecules, because of presence of bonded interactions, certain exclusion rules need to be applied when calculating which pair of atoms is non-bonded. The parameters for the bond, bend-angle and torsion potential interactions are derived such that all the requisite information (van der Waals and Coulomb) is embedded in them. Hence in macromolecules, two atoms are considered non-bonded pair only if they are separated by three bonds or belong to two different molecules. This rule is called 1-4 exclusion and should be checked before inclusion of any atom in the Verlet neighbor list of an atom.

Construction of Verlet neighbor lists is not very efficient method, as the search for the atoms within the cutoff radius still requires going over all the atoms. As the number of atoms in the system to be simulated increases, the search process is very exhaustive and to alleviate the process of searching all the atoms, the concept of linked-cells was introduced [45,46]. In the linked-cell-list method, the physical system is divided into $M_x \times M_y \times M_z$ cells of cell edge $> r_c$.
There are on average $N_c = N/(M_x M_y M_z)$ atoms in each cell. Because the cell edge is at least $r_c$, an atom in a particular cell $m$ has to search only the 26 neighboring cells in addition to cell $m$ to find all the atoms within the distance $r_c$. This computation scales as $27NN_c$, compared with the $O(N^2)$ brute force search.

![Figure 3.3: Linked-cells, neighbor cells and Verlet-list radius in 2D MD box](image)

### 3.3 Long-Range Interactions

Calculation of long-range coulomb interactions is very important in MD simulations of organic and biologically oriented systems due to the presence of solvent (typically water) atoms. Cutoff schemes used for dealing with the van der Waaals interactions cannot be used for Coulomb interactions, and it has been shown that cutoffs introduce significant errors in MD simulations. The introduction of periodic boundary conditions minimizes the unwanted surface effects but at the same time increases the number of ‘particles’ as the central system is replicated in three dimensions infinitely. The total coulomb energy for a system of $N$ particles in a cubic box of size $L$ can be written as:

$$E_{Coul} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{|r_{ij} + n|}$$

(3.13)
where $q_i$ and $q_j$ are the charges of any two atoms. The sum over $n$ represents the sum over all the replicated cells and $n = (n_x, n_y, n_z)L$, where $n_x, n_y, n_z$ are all integers. The prime on the summation over $n$ indicates that when $n = 0, i = j$ term is excluded. The problem with Eq. (3.13) is that the series is conditionally convergent, i.e., the result depends on the order in which the sums in Eq. (3.13) are added. From computational point of view, the calculation of Eq. (3.13) involves terms of the order $O(N^2)$, which is not scalable with increase of the number of atoms. There are different methods proposed to deal with both the issues, the most popular method being the Ewald summation method [5]. It has been shown that this method scales as $O(N^{3/2})$ [47], an improvement over the direct method. Other methods include particle mesh Ewald method [48] which scales as $O(N \log N)$ and more recently fast multipole methods have been developed [49] which scale as $O(N)$. We have implemented a scalable version of FMM which will be discussed in Section 3.3.2.

### 3.3.1 Ewald Summation Method

To improve the convergence of the summation in Eq. (3.13) the expression is rewritten in terms of the charge density. Each point charge is surrounded by a charge distribution of opposite charge such that the total charge of this distribution cancels the point charge. This introduction of a charge distribution screens the point charge and the long-range point charge interactions are now reduced to short range interactions between the charge distributions. But since we want the result due to the point charge interactions, the artificial introduction of charge distributions is canceled by introduction of second charge distribution of equal and opposite magnitude to the first one. The distribution is taken to be a Gaussian. This is shown schematically in Figure 3.4.
Figure 3.4: Charge distribution in the Ewald sum with original point charges and Gaussian screening distribution

The sum over the second Gaussian distribution is performed in reciprocal space to solve the Poisson equation. We also need to correct for the self-interaction between a point charge and charge distribution at the point charge. This is a self-interaction term and is a constant. The Eq. (3.13) can now be written as a sum of two rapidly converging series in real and reciprocal space along with a constant self interaction term:

\[ E_{\text{Coul}} = E_{\text{Ewald}} = E_{\text{real}} + E_{\text{reciprocal}} + E_{\text{self}} \]  \hspace{1cm} (3.14)

where

\[ E_{\text{real}} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} q_i q_j \frac{\text{erfc}\left(\frac{r_{ij} + n}{a}\right)}{r_{ij} + n}, \]  \hspace{1cm} (3.15)

\[ E_{\text{reciprocal}} = \frac{2}{V} \prod_{k} \frac{1}{k^2} e^{(k^2/4a^2)} \sum_{i=1}^{N} q_i e^{i r_i \cdot k}, \]  \hspace{1cm} (3.16)

\[ E_{\text{self}} = \prod_{j=1}^{N} q_j^2. \]  \hspace{1cm} (3.18)
The volume of the MD box is $V$ and $k = \left[ \frac{2\pi n_x}{L_x}, \frac{2\pi n_y}{L_y}, \frac{2\pi n_z}{L_z} \right]$ is the reciprocal vector. $\text{erfc}(x)$ is the complimentary error function and is defined as $\text{erfc}(x) = 1 - \text{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ which falls to zero as $x$ increases (long distances), i.e., $\text{erfc}(x) \xrightarrow{x \to \infty} 0$.

The Gaussian parameter $\square$ appears in both the real space and reciprocal parts of the Ewald summation. For the real space part, a large value of $\square$ makes the real space part converge faster, but increases the number of $k$ vectors. By choosing the value of $\square$ properly, one can make the real space part extends no further than the radius of Verlet neighbor lists, thus calculating the more time consuming part of the Ewald calculations more efficiently. Some of the parameters that should be taken into consideration include the system size, the proper choice of cutoff in real space and the desired accuracy. The above method of Ewald summation leads to $O(N^{3/2})$ order algorithm. More efficient methods have been developed recently where the reciprocal part is performed using fast fourier transforms thus reducing the complexity to $O(N \log N)$.

3.3.2 Fast Multipole Method

The fast multipole method (FMM) developed by Greengard and Rokhlin [49] calculates electrostatic energies and forces for a collection of $N$ charged particles with $O(N)$ operations and with predictable error bounds, whereas the optimal Ewald method is an $O(N^{3/2})$ algorithm (see Section 3.3.1). The FMM enables atomistic simulations of realistic materials involving millions to billions of charged particles under various settings of boundary conditions [50].

The idea behind the FMM is to divide the potential (force) exerted on an atom into near field and far-field component. The near-field contribution is calculated directly whereas multipole expansion approximates the far-field contribution.
We consider a collection of $N$ particles with charges \( \{q_i \mid i = 1, \ldots, N\} \) enclosed in a sphere of radius $a$ (see Figure 3.5). The potential due to this collection of charges at a distance $r$ such that $r > a$ is given by

$$V(r) = \sum_{i=1}^{N} \frac{q_i}{r} = \sum_{i=0}^{l} \frac{M_i^m}{r^{i+1}} Y_i^m(\hat{r}, \hat{r})$$

(3.19)

where $M_i^m = \sum_{i=1}^{N} q_i r_i^m Y_i^m(\hat{r}, \hat{r})$

(3.20)

where $^*$ denotes the complex conjugate. The functions $Y_i^m$ are the spherical harmonics and $M_i^m$ are the moments of a multipole expansion about the origin of the potential at $r$ due to charges located in a sphere of radius $a$. The $Y_i^m$ used in this discussion are defined from the associated Legendre polynomials:

$$Y_i^m(\hat{r}, \hat{r}) = \begin{cases} \prod_{m=1}^{l} (\hat{r})^m, & m \geq 0 \\ \sqrt{(l+m)!} P_i^m(\cos \hat{r}) e^{im\hat{r}}, & 1, m < 0 \end{cases}$$

(3.21)

$$P_i^m(x) = \frac{1}{2^l l! (l+m)!} \left( \prod_{m=2}^{l} (l^2 x^2)^l \right)^{l+m/2} \frac{d^{l+m}}{dx^{l+m}} \left( x^2 \right)^l$$

(3.22)

The following recursion formulas are useful to calculate $P_{lm}(x)$

[Figure 3.5: Far-field multipole approximation in FMM]
\[ P_{nm}(x) = (1)^m [(2m \bigotimes 1)(2m \bigotimes 3) \cdots l \bigotimes x^2]^{m/2} \quad \text{for } m \geq 1 \]
\[ P_{00}(x) = 1 \]
\[ P_{m+1,m}(x) = x(2m + 1)P_{mn}(x) \quad \text{for } m \geq 0 \]
\[ (l \bigotimes m)P_{lm}(x) = x(2l \bigotimes 1)P_{l1,m}(x) \bigotimes (l + m \bigotimes 1)P_{l2,m}(x) \quad \text{for } m \geq 0 \]
\[ P_{l0,m}(x) = (1)^m P_{lm}(x) \quad \text{for } m \geq 1 \]

To evaluate the Coulomb interactions under FMM paradigm, we need to another quantity \( L^m_l \), which are the coefficients of a local Taylor-type expansion of the potential about the origin due to distant distribution of charges at \( r \):
\[ L^m_l(a) = q \frac{P^m_l(\cos \theta)}{a^{l+1}} \exp(i m \theta) \quad (3.24) \]

For efficient calculation of FMM, three types of transformations are defined on \( M^m_l, L^m_l \):
\[ M_{lm}(\vec{A} \bigotimes \vec{b}) = \bigotimes_{j=0}^{l} \bigotimes_{k=0}^{j} T^\text{MM}_{l, j, m, k} (\vec{b}) M_{jk} (\vec{A}) \quad \text{[multipole-to-multipole]} \quad (3.25) \]
\[ L_{lm}(\vec{b} \bigotimes \vec{A}) \sim \bigotimes_{j=0}^{p} \bigotimes_{k=0}^{j} T^\text{ML}_{j+1, k+1, m} (\vec{b}) M_{jk} (\vec{A}) \quad \text{(p = maximum of l) [multipole-to-local]} \quad (3.26) \]
\[ L_{lm}(\vec{c} \bigotimes \vec{A}) = \bigotimes_{j=0}^{p} \bigotimes_{k=0}^{j} T^\text{LL}_{l, j, k, m} (\vec{b} \bigotimes \vec{c}) L_{jk} (\vec{b} \bigotimes \vec{A}) \quad \text{(p = maximum of l) [local-to-local]} \quad (3.27) \]

with the operators,
\[ T^\text{MM}_{l, j, m, k} (\vec{b}) = M_{l, j, m, k} (\vec{b}) \quad (3.28) \]
\[ T^\text{ML}_{j+1, k+1, m} (\vec{b}) = L_{j+1, k+1, m} (\vec{b}) \quad (3.29) \]
\[ T^\text{LL}_{l, j, k, m} (\vec{b}) = M_{l, j, k, m} (\vec{b}) \quad (3.30) \]

In FMM, the Coulomb potential is calculated in a hierarchical fashion. The simulation box (level \( l = 0 \)) is recursively divided into eight child cells (four in 2D, see Figure 3.6), until the size of the smallest cell (level \( l = L \)) is same as that of the linked cell (see Section 2.4). Since this
division of simulation box resembles a tree data structures, the cells at the finest level are called leaf cells.

![Image of Octree division in two-dimensions](image)

Figure 3.6: Octree division in two-dimensions; each color represents a different level of octree.

The FMM calculates Coulomb potentials of charged particles contained in a simulation box in five steps [49,47,51]:

(i) The simulation box with dimensions \((h_x, h_y, h_z)\) is successively subdivided. Let \(l_{bot}^{x}, l_{bot}^{y},\) and \(l_{bot}^{z}\) be the prescribed numbers of recursive subdivisions along the \(x, y,\) and \(z\)-axes. At the finest subdivision level \(l_{bot} = \max(l_{bot}^{x}, l_{bot}^{y}, l_{bot}^{z})\), the simulation box is decomposed into cells with dimensions \(\left(\frac{h_x}{2^{l_{bot}^{x}}}, \frac{h_y}{2^{l_{bot}^{y}}}, \frac{h_z}{2^{l_{bot}^{z}}}\right)\). The largest division is the simulation box itself at level 0. At level \(l\), the simulation box is composed of cells with dimension \(\left(\frac{h_x}{\max(2^{l_{bot}^{x}}l_{bot}^{x}+1), \frac{h_y}{\max(2^{l_{bot}^{y}}l_{bot}^{y}+1), \frac{h_z}{\max(2^{l_{bot}^{z}}l_{bot}^{z}+1)}\right)}\).

(ii) Compute multipole moments of all the cells at the finest level of subdivision. Sweep up from the smallest cells to largest cell to obtain multipole moments of cells at all subdivision levels using the multipole-to-multipole transformation formula, Eq. (3.25).

(iii) Sweep down from the largest cell to cells at the next level of subdivision to obtain local expansion coefficients in the smallest cells: First, transform local expansion coefficients of larger cell to cells at the next level of subdivision using the local-to-local
transformation formula for shifting the origin of a local expansion, Eq. (3.27). Second, add to these local expansion coefficients contribution from cells at the next level of subdivision, which have not been included and are well-separated from the cell being considered, using the multipole-to-local transformation formula, Eq. (3.26).

(iv) Once the preceding step has reached the finest subdivision level, evaluate the potential for each particle, Eq. (3.19), using the local expansion coefficients, $L_{lm}(\bar{A})$, of the smallest cell containing the particle.

(v) Add contributions from other charges in the same cell and the near neighbor cells by direct computations.

The FMMP performs steps (i)-(iv). The direct computations in step (v) should be performed separately in a subroutine supplied by the user. Such a separation in the FMM steps is appropriate since usual MD simulation codes have a linked list of neighboring particles for fast computations of short-range forces and the linked list can be exploited for the direct calculations in step (v).

In FMMP, the most time-consuming part is the multipole-to-local transformation operation in step (iii). If we take terms only up to $|j + l| < p$ in Eq. (3.26), the computation time becomes approximately half with some loss of accuracy [51].

Forces acting on particles are calculated by differentiating the potential, using the following formulas:

\[
\frac{\partial M_{lm}}{\partial x} = M_{lm}\left[\frac{lx}{x^2 + y^2} + \frac{im}{x^2 + y^2}\right] M_{l+1,m} \frac{zx}{x^2 + y^2},
\]

\[
\frac{\partial M_{lm}}{\partial y} = M_{lm}\left[\frac{ly}{x^2 + y^2} + \frac{im}{x^2 + y^2}\right] M_{l+1,m} \frac{yz}{x^2 + y^2},
\]

(3.31) (3.32)
\[ \frac{\partial M_{lm}}{\partial z} = M_{[l \parallel m]} \cdot \]  

(3.33)

The results of the scalability tests performed on the FMM code are described in next chapter.

3.4 Multiple Time-Scale Method

The techniques such as linked-cell lists and Verlet neighbor lists reduce the computational complexity for the non-bonded interactions, which are the most time consuming part in MD simulations. The next problem is to choose the time discretization unit, \( \Delta t \), for numerical integration. The bonded interactions vary much more rapidly than the non-bonded interactions and \( \Delta t \) needs to be chosen smaller than the smallest time scale. The multiple time-scale (MTS) method reduces the number of force computations significantly by separating the fast (small time-scale) modes from the slow (large time-scale) modes. We use an MTS algorithm called rRESPA [52,53,54], which is reversible since it is based on the symmetric Trotter expansion of the time evolution operator. The rRESPA algorithm is also symplectic, i.e., the phase space volume occupied by the atoms is a loop invariant, which results in long time stability of solutions. We now derive the Trotter’s expansion of the Liouville operator.

Consider any function \( f(p^N(t), r^N(t)) \), which is a function of the phase space. A time derivative of such function is defined as

\[ \frac{\partial f}{\partial t} = \frac{\partial f}{\partial r} \frac{\partial r}{\partial t} + \frac{\partial f}{\partial p} \frac{\partial p}{\partial t} = \frac{\partial f}{\partial r} p \frac{\partial r}{\partial p} + \frac{\partial f}{\partial p} p \frac{\partial r}{\partial p} \]  

(3.34)

\[ = iL f , \text{ where } iL = \frac{\partial f}{\partial r} p \frac{\partial r}{\partial p} + \frac{\partial f}{\partial p} p \frac{\partial r}{\partial p} \]  

(3.35)

where \( iL \) is the Liouville operator.

After integrating Eq. (3.35), we get

\[ f(p^N(t), r^N(t)) = \exp(iLt) f(p^N(0), r^N(0)) \]  

(3.36)
We can now define a classical propagator $U(t) = \exp(iLt)$ and using the definition of force $F = \dot{p}$, the classical propagator can be written as

$$U(t) = \exp\left(\frac{i}{\hbar} \frac{\partial}{\partial r} + F \frac{\partial}{\partial p}\right)$$  \hspace{1cm} (3.37)$$

If the Liouville operator is divided into two parts, $iL = i(L_1 + L_2)$, the Trotter’s expansion for a small $\Box t$ is given as

$$U(\Box t) = \exp(i(L_1 + L_2)\Box t)$$  \hspace{1cm} (3.38)$$

$$= \exp(iL_1 \frac{\Box t}{2}) \exp(iL_2 \frac{\Box t}{2}) \exp(iL_1 \frac{\Box t}{2}) + O(\Box t^3).$$  \hspace{1cm} (3.39)$$

In systems with different time scales, we categorize the bond and bond angle forces as the short time-scale forces, torsion forces as medium time-scale forces and non-bonded forces as long time-scale forces. The Liouville operator accordingly can be written as

$$iL = i(L_1 + L_m + L_s)$$  \hspace{1cm} (3.40)$$

$$= \Box r \frac{\partial}{\partial r} + (F_l + F_m + F_s) \frac{\partial}{\partial p}$$  \hspace{1cm} (3.41)$$

where

$$iL_l = F_l \frac{\partial}{\partial p}$$  \hspace{1cm} (3.42)$$

$$iL_m = F_m \frac{\partial}{\partial p}$$  \hspace{1cm} (3.43)$$

$$iL_s = \dot{r} \frac{\partial}{\partial r} + F_s \frac{\partial}{\partial p}$$  \hspace{1cm} (3.44)$$

Using the Trotter’s expansion in Eq. (3.39),

$$U_{\text{short}}(t) = \exp(iL_l \frac{\Box t}{2}) \exp(iL_m \frac{\Box t}{2}) \exp(iL_s \Box t) \exp(iL_m \frac{\Box t}{2}) \exp(iL_1 \frac{\Box t}{2})$$  \hspace{1cm} (3.45)$$

$$= \exp(F_l \frac{\partial}{\partial p} \frac{\Box t}{2}) \exp(F_m \frac{\partial}{\partial p} \frac{\Box t}{2}) \exp(iL_s \Box t) \exp(F_m \frac{\partial}{\partial p} \frac{\Box t}{2}) \exp(F_l \frac{\partial}{\partial p} \frac{\Box t}{2})$$  \hspace{1cm} (3.46)$$
The center Liouville operator can be expanded as

\[
\exp(iL_s t) = \exp(F, \frac{\partial}{\partial p}) \exp(r \frac{\partial}{\partial r}) \exp(F, \frac{\partial}{\partial p}) \exp(r \frac{\partial}{\partial r}) \exp(F, \frac{\partial}{\partial p})
\]

(3.47)

The MTS scheme gives considerable improvement in the performance of the code.
CHAPTER 4
PARALLEL MD AND SCALABILITY TESTS

In molecular dynamics, the calculation of forces is the most time consuming computational task. For systems of large sizes, this force computation can almost take up as much as 90% of computation time. Therefore there is a continual need to develop new algorithms to speed up the force calculation. A potential solution for simulation of systems of large sizes is the utilization of latest technology of parallel computers [55,56,57]. The basic principle is to assign a specific computational task to each of the set of $P$ processors, which in principle provides a cost-effective solution to the difficulty posed by the force computations in large systems. In the simulation of organic systems, which is the essential part of this thesis, significant progress has been made in developing parallel MD simulations [58,59,60,61]. This chapter describes some of the algorithms used in parallel MD field and also some of the results of the scalability tests that were performed on our algorithm Macro Molecular Dynamics (MMD). Section 4.1 describes some common parallelization paradigms and models of parallel computing. In Section 4.2, we describe the parallel FMM and the results of scalability tests while Section 4.3 describes the parallel MMD algorithm and results of scalability tests.

4.1 Parallelization Paradigms

The forces and positions of all atoms can be updated simultaneously, which renders a natural parallelism in molecular dynamics. The most popular methods of parallelization in literature are (1) atom decomposition [62,63,64] and (2) spatial decomposition [65,66]. In both the methods, each processor is responsible for a fixed number of force and position updates. The difference in the two methods lies in the different sets of atoms that each processor is responsible for, see Figure 4.1.
Figure 4.1: Difference between atom decomposition and spatial decomposition methods. The atoms that a processor is responsible for are also shown. The colors signify the processor they belong to: red-processor A, blue-processor B, green-processor C and purple-processor D. (a) system at $t=0$ (b) system at $t=T$ and movement of atoms (c) system under atom decomposition method (d) system under spatial decomposition method.

In the atom decomposition method, a pre-determined set of atoms is assigned to each processor and this assignment remains invariant throughout the simulation irrespective of the movement of the atoms spatially. For the calculation of all the forces acting on any atom, identical copies of all atoms must exist on all the processors; hence sometimes this method is also called replicated-data method. Atom decomposition method is a popular method for molecular simulations involving organic systems, as it provides a straightforward method of
calculating 3- and 4- body force calculations. Parallel implementations of commercial packages like CHARMM [67], AMBER [68] and GROMOS [69] use this method. The drawback of this method is the $O(N)$ memory requirement on each processor, which inhibits the ability to simulate large systems due to memory limitations. The communication costs also scale as $O(N)$ as global communication is required.

In the spatial decomposition method, each processor is responsible for updating forces and positions of atoms that are within its domain. Atoms are reassigned to processors, as they move spatially in the system. To compute all the forces, each processor not only needs all the information in its domain, but also of neighboring nodes. Hence the communication is local in nature in contrast to the atom-decomposition methods. The memory requirement is also reduced to $O(N/P)$ enabling large system size simulations. More will be discussed in next Section.

There are many computational models that exist for parallel computation [70], of which message passing model [71] is considered in this thesis. The message-passing model is defined by a set of processors having only local memory, and the processors communicate by sending and receiving messages using pre-defined library subroutines through interconnection network. Several kinds of parallel computers are developed along different methodologies, SIMD (Single Instruction, Multiple Data) [72,73] and MIMD (Multiple Instructions, Multiple Data) [74,62] are popular. In SIMD, a single instruction is executed simultaneously on each processor, whereas in MIMD, each processor computes separately from its neighbor, and in most modern applications, MIMD machines are most popular. In this thesis, all the computations are performed on SIMD machines.
4.2 Space decomposition

The MMD algorithm has been parallelized using spatial decomposition [5,25], in which the physical system is divided into $P$ subsystems, $P_x \times P_y \times P_z$ in the $x$, $y$, and $z$ directions, and each subsystem is assigned to a processor. Each processor $p$ thus stores arrays containing the positions $\vec{r}_i (i = 1, \ldots, N_p)$, velocities $\vec{v}_i (i = 1, \ldots, N_p)$, and species $s_i (i = 1, \ldots, N_p)$ of atoms, where $N_p$ is the number of atoms residing in the $p$th subsystem. When time-stepping procedure updates the atomic positions, some atom $i$ in subsystem $p$ may have moved out to a neighboring subsystem $p'$. This atom is ‘migrated’ to processor $p'$, i.e., the information, $\vec{r}_i$, $\vec{v}_i$ and $s_i$, is removed from processor $p$ and sent to processor $p'$, where it is appended to position, velocity and species arrays.

![Figure 4.2: Schematic of spatial decomposition in MMD algorithm. The figure shows the primary skin $S_p (p)$ of thickness $r_p$ and the secondary skin $S_t (p)$ of thickness $r_c$ for processor 1. Atom $i$ has migrated from processor 3 to processor 4 due to time-stepping procedure.](image)

To calculate bonded and van der Waals interactions as well as the near-field contribution to Coulombic interaction in processor $p$, all atoms, which are in the neighboring processors but are close to $p$, are ‘cached’. Namely the positions and species of these atoms must be received
from the neighboring processors, and then appended to local arrays in $p$. More precisely a skin of thickness $r_c$ (the cutoff radius of the van der Waals potential function, $V_{vdw}(r)$) is defined around each processor $p$, and cache the information about the atoms within this skin from neighboring processors to $p$. The set of atoms in this ‘secondary’ skin, $S_s(p)$, is denoted as $B_s = \{i \in S_s(p)\}$. To minimize the communication time for the bonded-force calculations, another skin, ‘primary’ skin $S_p(p)$, of thickness $r_p (< r_c)$ is defined, where $r_p$ is at least twice the longest bond length. The set of atoms in this skin is denoted as $B_p = \{i \in S_p(p)\}$. Figure 4.2 schematically shows these two skins. With the spatial decomposition, the computation scales as $N/P$ while communication scales as $(N/P)^{2/3}$.

### 4.3 Parallel FMM: Scalability Tests and Results

The principles of FMM were described in Section 2.4.2. In this Section, some of the tests on the parallel implementation of the FMM are described. To determine the accuracy of the FMM calculations, a test was performed and the results are compared with those of the Ewald summation method and direct calculation.

In the accuracy test, a system consisting of 1000 atoms was considered, assined charges of $\pm 1$ randomly, and the atoms were scattered in a cubic box, such that the overall charge neutrality is maintained. The simulation box is decomposed into $P = 8$ nodes and the level of the FMM tree is chosen to be 3. The test was performed on a Linux PC cluster (8 nodes, 100BaseTX-connected PentiumIII/600MHz) and a PGI Fortran77 compiler is used. The results of the accuracy test are shown in Table 4.1. The test compares the FMM results of the potential fields felt by the particles (that include nearest neighbor contribution) with the direct calculation results and the averaged values of the relative errors are also shown in the Table 4.1. In the present case, the averaged relative errors of the potential fields are on the order of 0.01%.
Table 4.1: Results of accuracy tests of FMM on Linux PC cluster.

<table>
<thead>
<tr>
<th>IPBC</th>
<th>rel_error</th>
<th>t_setup</th>
<th>t_comm.</th>
<th>t_up</th>
<th>t_down</th>
<th>t_wrap</th>
<th>T_pfs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.1077e-3</td>
<td>1.45e-3</td>
<td>5.38e-2</td>
<td>5.67e-2</td>
<td>5.68e-1</td>
<td>3.00e-6</td>
<td>3.17e-3</td>
</tr>
<tr>
<td>1</td>
<td>0.4799e-3</td>
<td>1.42e-3</td>
<td>5.49e-2</td>
<td>5.77e-2</td>
<td>1.15e+0</td>
<td>1.47e-1</td>
<td>3.16e-3</td>
</tr>
</tbody>
</table>

The second set of the tests examines the scalability of the FMMP on massively parallel computers. The calculations were performed on an IBM SP3, called “HABU”, at the U.S. Naval Oceanographic Office (NAVO) Major Shared Resource Center. The HABU is configured with 375 MHz Power3 CPUs and has 334 nodes with 4 CPUs and 4 GB of memory per node, a total of 1,336 processors. It runs AIX 4.3 operating system and IBM XL Fortran 7.1 compiler. The same calculations were repeated on another parallel computer, IBM SP4, called “Marcellus” at NAVO. The Marcellus is configured with 1.3 GHz Power4 CPUs and has 148 nodes with 8 CPUs and 8 GB of memory per node, total of 1,184 processors.

Figure 4.3: Timings as function of number of processors. (a) Total execution (solid symbols) and communication (open symbols) times are plotted for the number of particles per processor, \( N/P = 10^5 \) (squares), \( 5 \times 10^5 \) (triangles), and \( 10^6 \) (circles) on IBM SP3. (b) The same as (a) on IBM SP4.
The compute nodes for this test were chosen as \((nx, ny, nz) = (1, 1, 1), (2, 2, 2), (4, 4, 4),\) and \((8, 8, 8)\) with the number of particles treated in each node, \(N/P\), fixed. The values of \(N/P = 100,000, 500,000,\) and \(1,000,000\) were chosen. Accordingly, the values of \(lbotx = lboty = lbotz\) were also varied from 4 to 7.

Figure 4.3(a) shows the results on the IBM SP3. The total execution time of the FMMP code with a fixed value of \(N/P\) is nearly constant, when the total number of particles, \(N\), is increased by varying the number of compute nodes, \(P\). The parallel efficiency for the case of \(P = 512\) and \(N/P = 1,000,000\) is as high as 0.98. The communication time is only a small fraction of the total execution time. Figure 4.3(b) shows the results on the IBM SP4, which is almost 2.5 times faster than the IBM SP3 in terms of computation, and a slight improvement in terms of communication time. It can also be observed in Figure 4.3, that the total execution time is not proportional to \(N/P\) for a fixed value of \(P\), but is of the form \(a + b(N/P)\), where \(a >> b\).

4.4 Parallel MMD: Scalability Tests and Results

In this Section the scalability tests performed on MMD code are described. To study the effect of granularity \(N/P = 5280, 10560,\) and \(21120\) values were used. The number of processors \(P\) varies from 1 to 1024 and the maximum number of atoms is 21.6 million (the system is self assembled monolayers of alkane-thiolates on gold surface, simulations of which will be described in detail in Chapters 5, 6, and 7) The tests are performed on two multi Tera-flop machines, IBM SP4 Marcellus at the Naval Oceanographic Office (NAVO) and Intel Xeon-based Linux cluster SuperMike at Louisiana State University (LSU). A comparison of the two systems is given in Table 4.2. In SuperMike 512 dual-processor nodes are connected by Myrinet network, which consists of 24 switch units and 64 nodes are connected within the same switch.
Marcellus on the other hand uses a proprietary network and IBM’s Colony II switch for communication.

Table 4.2: Comparison of the Linux cluster *SuperMike*, IBM SP4 *Marcellus*, and Compaq SC45 *Emerald*

<table>
<thead>
<tr>
<th></th>
<th>Linux cluster</th>
<th>IBM SP4</th>
<th>Compaq SC45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>1.8GHz Intel Xeon</td>
<td>1.3GHz Power 4</td>
<td>1.0GHz Alpha 21264</td>
</tr>
<tr>
<td># of processors</td>
<td>1024</td>
<td>1184</td>
<td>512</td>
</tr>
<tr>
<td>Peak speed</td>
<td>3.7 Tflops</td>
<td>6.1 Tflops</td>
<td>1.0 Tflops</td>
</tr>
<tr>
<td>L2 Cache</td>
<td>512KB/processor</td>
<td>1.4MB/2-processor-chip</td>
<td>8MB/processor</td>
</tr>
<tr>
<td>Memory</td>
<td>2GB/2-processor-node</td>
<td>8GB/8-processor-board</td>
<td>4GB/4-processor-node</td>
</tr>
<tr>
<td>Latency</td>
<td>9 μsec</td>
<td>21 μsec</td>
<td>5 μsec</td>
</tr>
<tr>
<td>Band width</td>
<td>250MB/proc</td>
<td>90MB/proc</td>
<td>125MB/proc</td>
</tr>
<tr>
<td></td>
<td>(bidirectional)</td>
<td>(bidirectional)</td>
<td>(bidirectional)</td>
</tr>
</tbody>
</table>

Figure 4.4 shows the scalability of the MMD algorithm as a function of the number of processors on IBM SP4. The number of atoms per processor \((N/P)\) was also varied to study the effect of granularity. The parallel efficiency for \(P\) processors, which is defined as the ratio of time on a single processor to time on \(P\) processors, increases with \(N/P\). Up to 512 processors, the parallel efficiency is 80%, 83% and 87% for \(N/P\) values of 5280, 10560 and 21120, respectively. The observed degradation in parallel efficiency for smaller \(N/P\) is due to larger communication to computation ratio. The computation time scales as \(N/P\) and the communication time scales as \((N/P)^{2/3}\), and hence the communication to computation ratio, \((N/P)^{1/3}\), is a decreasing function of \(N/P\). It can also be seen from Figure 4.4 that the communication time is much smaller compared to the total execution time in all the three cases.
Figure 4.4: Total execution time (solid symbols) and communication time (open symbols) of MMD as a function of \( P \) for different work loads: \( N/P = 5280 \) (squares), 10560 (circles) and 21120 (triangles).

On IBM SP4, two processors share L2 cache on a chip, four chips on a board share memory, and four boards constitute a physical unit called LPAR (Logical PARtition). Thus each LPAR has 32 processors. To study the performance degradation due to the sharing of the L2 cache bindUtils, a set of tools was used that enables the use of only one processor per chip, thereby reducing the congestion caused by the sharing of the cache.

Figure 4.5: Total execution time (squares) and communication time (circles) of MMD as a function of \( P \) with (open symbols) and without (solid symbols) bindUtils for \( N/P = 10560 \).
In Figure 4.5, solid squares show the execution time when two processors share L2 cache and open squares show the execution time when only one processor per chip is used. Sharing of L2 cache increases the execution time as much as 10%. The small jump in the execution time from 32 processors to 64 processors occurs due to communication bottleneck across two LPARs.

![Figure 4.6: Total execution time (solid symbols) and communication time (open symbols) of the HIO-MD algorithm as a function of P with N/P = 10560 on the Linux cluster at LSU-BCVC, using one processor per node (squares and blue) and two processors per node (circles and red).](image)

This result suggests a detrimental effect of sharing main memory by the two processors within a node on the Linux cluster. To quantify this effect, Figure 4.6 compares the results of two sets of benchmark tests on the 256-processor Linux cluster at LSU-BCVC. In the first set, only one processor per dual-processor node is used, whereas in the second set, both processors are used. The performance degradation due to sharing main memory in the execution time is nearly constant (~ 90%) from 2 to 128 processors. Figure 4.6 also shows performance degradation in the communication time, because the shared memory is used for communication between two processors in the same node. The resulting congestion in memory and/or internal bus degrades the performance.

On the Linux cluster, the effect of choice of compiler on performance of MMD algorithm was studied. For this purpose we used PGI 4.0 and Intel 6.0 compilers and the result is shown in
Figure 4.7. The Intel compiler makes better use of architecture of the cluster, including the Streaming SIMD Extensions 2 (SSE2) that augments the floating-point functional unit to deliver two results per cycle in the ideal case. Accordingly the total execution time with Intel compiler is smaller than that with PGI compiler, but with no significant difference in communication time. In both the cases, one processor per node was used up to 512 processors, and 2 processors per node for 1024 processor case.

![Graph showing total execution time and communication time](image)

**Figure 4.7:** Total execution time (squares) and communication time (circles) of MMD as a function of $P$ for different compilers, Intel (open symbols) and PGI (solid symbols) for $N/P = 10560$.

### 4.5 Summary

Results of scalability tests on two algorithms Fast Multipole Method (FMM) and macro-molecular dynamics (MMD) algorithms were described in this chapter. The MMD algorithm was based on a space-time multiresolution approach combining the FMM and multiple time-scale method as well as dynamic management of distributed linked cells, neighbor lists, and atomic $n$-tuples. The results of scalability tests on FMM were presented first, in which the parallel efficiency of FMM on a system size of 500 million was shown to be as high as 0.98. The tests on FMM were performed on two tera-flop machines IBM SP3 and SP4. The scalability and portability of MMD algorithm were described next. The tests on MMD were performed on 1024-
processor Intel Xeon-based Linux cluster, 1184-processor IBM SP4 and 512 Compaq Alpha Server. The parallel efficiency of the MMD code was shown to be 0.87 on IBM SP4. The effects of memory- and cache-sharing were also demonstrated and their effect was shown to be significant. Such a scalable parallel macro-molecular simulation algorithm is expected to play an important role in the design of hybrid quantum-device/biological-cell systems and ‘virtual-cell’ technologies.
CHAPTER 5
SIMULATIONS OF ALKANETHIOL SELF-ASSEMBLED MONOLAYER SYSTEMS

5.1 Background

Self-assembled monolayers (SAMs) of functionalized long-chain molecules on the surface of solid substrates have been the subject of extensive theoretical and experimental research [5] due to their broad range of applications in mechanical (e.g. lubrication [75]), chemical, electronic, and biotechnological fields [10]. Self-assembly of organic films are attractive in diverse fields due to the possibility of tuning their properties by selectively modifying specific functional groups (i.e., end groups) without modifying the entire chain. For instance, by changing just the end group of alkanethiol-based monolayer [e.g., S(CH₂)nCH₃] from CH₃ to OH, the property of the resulting chemical surface can be varied from hydrophobic to hydrophilic. Oligo-ethylene-glycol-terminated alkanethiol SAMs exhibit interesting protein adsorption properties [110] and as such are useful for interfacing biological materials with inorganic surfaces. For example, it is possible to immobilize a single biomolecule or an array of biomolecules on the surface of SAMs for biosensor applications.

SAMs can be formed by spontaneous adsorption of different types of molecules from solution onto a variety of solid substrates such as gold, silver and hydroxylated silicon. A self-assembled monolayer can be divided into three parts (see Figure 5.1): head group, alkyl chains and end group. The head-group – substrate interactions are usually very strong and exothermic in nature: the chemisorption bond that forms between gold and sulfur is of the order of ~40-45 Kcal/mol [76] and the bond is stable up to temperatures of ~475K [77]. This chemisorption process is very fast and occurs in a few seconds. Once the head group atoms are pinned in correct place on the substrate, van der Waals and electrostatic forces (depending on the presence
of polar groups) help form a closed packed and ordered assembly. The terminal functionality (end groups) can be changed to achieve different surface properties and are usually associated with conformational defects (gauche) at room temperature.

![Image of SAM components](image)

Figure 5.1: Schematic and different components that make up a self-assembled monolayer

Alkanethiols on Au(111) surface are most widely studied mainly because gold is easy to clean, chemically inert. The most popular method for growing monolayers of alkanethiols, in terms of the ease of preparation, is in solution. Another method of preparation is vapor deposition technique, where the gold surface is exposed to alkanethiol vapors in ultra high vacuum. The schematic of the preparation of thiols on gold surface in solution is shown in Figure 5.2. A fresh, clean gold substrate is immersed in a dilute solution (most commonly used solvent is ethanol), with concentrations of the order ~1-2 mM. The growth of SAMs progresses in stages. In the preliminary stages, a low-density striped phase is formed, where the alkane chains lie parallel to the surface (see Figure 5.2(b)). This is followed by an intermediate stage (see Figure 5.2(c)) and after several hours of immersion times, a closed-packed oriented monolayer is formed on the Au(111) surface (see Figure 5.2(d)).
Earlier experimental studies showed that this system forms a well-ordered \((\sqrt{3} \times \sqrt{3})R30^\circ\) triangular lattice with lattice constant \(a = 4.97 \pm 0.05 \text{Å}\) with one chain per unit cell [78,79]. The \((\sqrt{3} \times \sqrt{3})R30^\circ\) can be explained in terms of Wood’s notation as follows. When describing the reconstructed surfaces, the surface adsorbate overlayers are defined as

\[
\begin{array}{ccc}
\hat{b}_1 & b_2 & \hat{b} \\
\hat{a}_1 & a_2 & \hat{a}
\end{array}
\] (5.1)

where \(a_1\) and \(a_2\) are real space surface lattice parameters of the original surface (gold) and \(b_1\) and \(b_2\) are the lattice parameters for the reconstructed surface (sulfur) and \(\hat{\theta}\) is the rotation angle of the reconstructed layer with respect to the original layer. In case of alkanethiols on gold, \(a_1 = a_2 = a_0\), and \(b_1 = b_2 = \sqrt{3}a_0\) and \(\hat{\theta} = 30^\circ\) (see Figure 5.3), using Eq. (5.1), the reconstructed sulfur lattice is \((\sqrt{3} \times \sqrt{3})R30^\circ\) relative to the underlying gold surface.
The lattice spacing of bulk gold is 2.88 Å, the \(\sqrt{3} \times \sqrt{3} R 30^\circ\) structure corresponds to a sulfur-sulfur separation of \(\sim 5\) Å \((\sqrt{3} \times 2.88\) Å\), and the area per molecule corresponds to \(\sim 21.6\) Å. The closed packed minimum energy configuration (alkane chains perpendicular to the substrate) of alkane chains corresponds to an area 18.4 Å. This difference in area per molecule suggests that the chains need to tilt by an angle of \(\cos^{-1}\frac{18.4}{21.6} \approx 32^\circ\) to maintain the minimum energy for the alkane chains. Infrared spectroscopy studies have established that the alkanethiol chains tilt on average by 30±10° [80,81]. Subsequent experiments observed super-lattice structures, superimposed on the triangular lattice, such as two-chain [82] and four-chain unit cell \(c(4 \parallel 2)\) reconstructions [83,84,85,86]. Ulman et al. [87] used grazing angle Fourier transform infrared spectroscopy to compare monolayers on silver surfaces, and the spectra from silver surface (chains have a lattice of \(a = 4.65 \pm 0.15\) Å) suggests that the tilt angle of \(\sim 7^\circ\), much smaller than that on gold, though the lattice constant of bulk silver is 2.89 Å, and is very similar to that of gold.
The chains are more densely packed on silver than on gold, and the area per molecule is \(\sim 18.5 \, \text{Å}^2\) per molecule on silver leading to much lesser value of the tilt angle. This suggests that the tilt structure in SAMs is a balance between chain-chain interactions and binding of head-group to the substrate. In gold, the head-group to substrate binding is optimized at the cost of inter-chain potential minimum, and hence the chains tilt, whereas, on silver, the chain-chain interactions are optimized at the cost of binding, hence smaller tilt.

The tilted alkanethiol chains are characterized by three angles: tilt angle (\(q\)), angle made by the chain with respect to the surface normal, twist angle (\(f\)), rotation of the chain about its axis and tilt direction (\(c\)), projection of the chain in the substrate plane, see Figure 5.4.

![Figure 5.4: Schematic representation of a tilted alkanethiol molecule:](image)

SAMs exhibit rather complex structural and dynamical behaviors, which are not well understood. Although the molecular chains show collective tilt away from the surface normal,
the atomistic details of the tilt structure are not known. Computer simulations can provide a theoretical basis for answering some of the questions as well as interpreting experimental results. Molecular dynamics (MD) and Monte Carlo (MC) methods were previously used to simulate SAMs of alkanethiol molecules chemisorbed on gold substrate [88,89,90,91]. Although these studies have given significant insight into various properties of SAMs, the simulated system sizes are relatively small (< 5,000 atoms or 100 chains) and many of these simulations are based on united-atom models, where CH₂ and CH₃ units are treated as single sites. In this chapter, large-scale parallel MD study of alkanethiol SAM with system size up to a million atoms based on an all-atom model (hydrogen atoms are treated explicitly) is described. The structural properties of the SAMs as a function of temperature, lattice spacing and molecular chain length are studied in this chapter.

5.2 Interatomic Potential Parameters for Alkanethiol SAMs

The interatomic potential (V) model used in SAM simulations can be split into bonded, non-bonded and surface interaction energies:

\[ V = V_{\text{bonded}} + V_{\text{nonbonded}} + V_{\text{surface}}. \] (5.2)

Here \( V_{\text{bonded}} \) represents the bonded interactions that arise from bond-stretching, bending, and torsions:

\[ V_{\text{bonded}} = \sum_{ij} \frac{1}{2} k_{ij}^b (r_{ij} - r_{ij}^0)^2 + \sum_{ijk} \frac{1}{2} k_{ijk}^b (\theta_{ijk} - \theta_{ijk}^0)^2 + \sum_{ijkl} \frac{1}{2} k_{ijkl}^b \{1 + \cos 3(\phi_{ijkl} - \phi_{ijkl}^0)\}, \] (5.3)

where two atoms \( i \) and \( j \) define a covalent bond of length \( r_{ij} \); three atoms \( i, j \) and \( k \) define bond angle \( \theta_{ijk} \), and four atoms \( i, j, k \) and \( l \) define a dihedral angle \( \phi_{ijkl} \). For determining the non-bonded pairs a 1-4 exclusion rule, where a pair of atoms \( i \) and \( j \) are non-bonded if they belong to different molecules or are separated by more than three bonds in the same molecule, is used in
SAM simulations. The non-bonded interactions (i.e., long-range repulsions and van der Waals interactions) are of the form

\[ V_{\text{nonbonded}} = \sum_{i<j} \left( A_{ij} \exp(B_{ij} r_{ij}) - C_{ij} r_{ij}^6 \right), \quad (5.4) \]

which are truncated at a cut-off radius of \( r_c = 9.0 \text{Å} \):

\[ V_{\text{nonbonded}}(r) = V_{\text{nonbonded}}(r_c) - V_{\text{nonbonded}}(r_c) \frac{dV_{\text{nonbonded}}}{dr} \bigg|_{r=r_c}. \quad (5.5) \]

In the simulations, the gold atoms are not treated explicitly, but an interaction potential of the form in Eq. (5.6) is used to model the interactions between gold surface atoms and backbone atoms (carbon and sulfur atoms) [88]. The effect of the surface potential is strongest for the sulfur and the first carbon atoms, and decreases rapidly with increase in \( z \) position.

\[ V_{\text{surface}} = \sum_s \frac{1}{2} k_s^{SF} d_s^2 + \sum_{pq} \left[ k_{12}^{SF} (z_p - z_q)^{12} - k_3^{SF} (z_p - z_q)^{3} \right], \quad (5.6) \]

The index \( s \) in (5.6) runs over sulfur atoms, \( p \) counts the number of chains or molecules, \( q \) runs over sulfur and carbon atoms on each molecule, and \( d_s \) is the distance of sulfur atom from its original position on the x-y plane. The first term constrains the motion of sulfur atoms on the x-y surface whereas the second term involves \( z \) positions of sulfur and carbon atoms.

The model parameters [88,92,93] used in our MD simulations for the functional forms in Eqs. (5.3), (5.4) and (5.6) are given in Table 5.1.

<table>
<thead>
<tr>
<th>( ij )</th>
<th>( r_{ij}^{\beta} ) (Å)</th>
<th>( k_{ij}^* ) (Kcal/mol/Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C—C</td>
<td>1.53</td>
<td>617.8</td>
</tr>
<tr>
<td>H—H</td>
<td>1.09</td>
<td>655.2</td>
</tr>
</tbody>
</table>

(Table Cont’d)
5.3 Initial Setup

The system of self-assembled monolayers consists of a triangular lattice of sulfur atoms (see Figure 5.5), bonded to a substrate, which are the head group atoms of $S \square (CH_2)_n \square CH_3$ chains, where $n$ is varied from 7 to 25. (Most of the results in the following Sections are for $n = 13$.) The lattice spacing $a$, of the triangular lattice is also varied from 4.7 to 5.3Å. To study the size effect in our MD simulations, we consider two lateral system sizes, one comprising of 26 $\square$
44 = 2,288 chains and the other comprising of 80 \( \times \) 144 = 23,040 chains. The \( x \)- and \( y \)-dimensions of the MD box corresponding to these systems are 225.17 \( \times \) 220Å and 692.82 \( \times \) 720Å, respectively.

![Figure 5.5: The initial configuration of the system of alkanethiol SAMs. Atoms in red, black and blue are sulfur, carbon and hydrogen species respectively. (a) the system in \( x-y \) direction (top view) and (b) the system in \( x-z \) direction (side view).](image)

Periodic boundary conditions are applied in the surface plane. In order to remove interactions between periodic images in \( z \) direction, a vacuum space (longer than \( r_c \)) is inserted above the free surface of the SAM. An initial angle of \( \theta \sim 2^\circ \) is given to all the molecular chains, and the backbone planes of all the chains are rotated by \( \theta \sim 2^\circ \) with the nearest neighbor direction (parallel to the \( x \)-axis) to break the symmetry of the triangular lattice. The conformation of each chain is chosen to be \textit{all-trans} with its backbone plane in a herringbone structure.

All the atoms are given random initial velocities, according to the Maxwell-Boltzmann distribution, corresponding to temperature of \( T = 300K \). We thermalize the system for 150ps during which the molecular chains increasingly tilt away from the surface normal towards one of the next-nearest neighbor directions and the tilt structure finally reaches a steady state (Figure
5.6. The statistical averages of potential energies and structural parameters are calculated, after the thermalization stages, over the next 50ps.

![Tilt Angle vs Time](image)

Figure 5.6: Time evolution of the collective tilt angle during thermalization for \( n = 13 \).

5.4 Results and Discussion

5.4.1 Effect of Temperature

The effect of temperature on the structure of the alkanethiol system with chain length \( n = 13 \), lattice constant \( a = 5\AA \), and the number of atoms is 100,672 (2,288 chains) was studied. Starting from the initially thermalized system (in which the temperature saturated to a value of 150K), the atom velocities were scaled to vary the system temperature from 150 to 500K in steps of 50K. For each 50K step, the temperature is increased over 30ps, the system is thermalized for 50ps, and statistical averages obtained over another 50ps. To study possible hysteresis, the temperature is decreased from 500K to 200K. The 0K system was obtained by gradually decreasing the temperature from the initially thermalised system.
As shown in Figure 5.7(a), with increasing temperature, the average tilt angle ($q$) of the chains initially decreases slowly and then rapidly until it saturates at a low value for temperatures higher than 350K. At 300K the calculated tilt angle ($25^\circ$) agrees fairly well with experimental tilt angle of $30\pm10^\circ$ [82,83] and $33.7\pm0.8^\circ$ [86], which was measured for $n=12$ using grazing incidence x-ray diffraction. The tilt direction is along the next-nearest neighbor (NNN) direction.
(\(\varphi\sim 0\)) below 300K and at temperatures above 300 K it shifts away from the NNN direction by about 15° (see Figure 5.7(b)). However, the system does not show any clear precession about the surface normal even at very high temperatures. This is in contrast to previous MD studies based on united atom model [89], in which the tilt direction was no longer locked at \(T > 300\) K. Figure 5.7(a) also shows the results of million atom MD simulations, which are nearly identical to hundred thousand atom system, showing negligible effect of system size. Figure 5.7(a) also shows the presence of a hysteresis between 250 and 350K.

![Twist Angle Distribution](image)

Figure 5.8: Calculated twist angle distribution plotted as a function of twist angle at \(T =100\), 300 and 450K for \(n = 13\).

The twist angle (\(\varphi\)) distribution is found to change dramatically with temperature (Figure 5.8). At low temperatures, the twist angle distribution has two major peaks (i.e., the molecular backbone planes have two well-defined orientations with twist angles at ±50° with respect to tilt direction) and two minor peaks at ±130°. This is consistent with the experimental result of 46° at 160K and for \(n = 18\) [94]. With increasing temperature, the minor peaks grow to some extent but the major peaks diminish substantially. The resulting uniform four-site distribution pattern at
300K is similar to that observed in crystalline n-alkane systems [95]. These peaks disappear above 400K, and the resulting flat distribution indicates that the system evolves from a phase in which individual chains are locked in their orientation to a phase where chains are free to rotate about their molecular axes.

Figure 5.9 shows the radial distribution function at $T = 100$, 300 and 450K. The first three peaks correspond to C—H, C—C and S—C bond lengths, respectively, whereas the remaining peaks represent the distances beyond the nearest neighbors. With increasing temperature, the positions of the peaks lying within 3Å remain essentially unchanged but the peaks around 4 and 5Å almost disappear at 450K. This implies the absence of long-range order at high temperatures, which is consistent with the loss of rotational order as seen in Figure 5.8.

![Figure 5.9](image)

**Figure 5.9:** Calculated radial distribution function, $g(r)$, at $T = 100$, 300 and 450K for $n = 13$. The curves for $T = 100$ and 300K are shifted vertically for clarity.

Structural information in X-Y plane can be probed experimentally using x-ray diffraction [96] and surface x-ray scattering studies [97]. The quantity that is measured experimentally in these studies is the structure factor,
\[ S(k) = \frac{1}{N} \langle |\mathcal{F}(\mathbf{k})|^2 \rangle, \]  
where \( \mathcal{F}(\mathbf{k}) \) is the Fourier transform of the local particle density,

\[ \mathcal{F}(\mathbf{k}) = \prod_{i=1}^{N} \exp[i \mathbf{k} \cdot \mathbf{r}_i] = \prod_{i=1}^{N} \exp[i(k_x r_{ix} + k_y r_{iy})] \]  
and \( \mathbf{k} = (k_x, k_y, 0) \) represents a two-dimensional scattering vector.

Six peaks of equal intensity are expected in case of just a hexagonal lattice, which is the underlying lattice in case of alkanethiols. However, the internal structure (i.e., planes formed by the \( CH_2 \) units) modulates the peaks as shown in Figure 5.10(b). In Figure 5.10(a), only odd numbered carbons in all chains were considered, so as to get a nearly linear set of atoms along the surface normal (sulfur atoms are not included). It can be seen that in this case, all the six peaks are of equal intensities, as expected. This is in contrast to Figure 5.10(b), in which all the atoms in a chain are considered and the planes are parallel to x-axis. The shortening of the peaks at \( Ky = 0 \), suggest that the destructive interference due to the plane, reduces the height of peaks.

There is an overall increase in the intensity of the all the peaks (height) because of the increased number of atoms included and also the inclusion of sulfur atoms, which contribute heavily to the delta peaks.

![Figure 5.10: Calculated structure factors for (a) only odd carbons in a chain are considered (b) all the atoms in the chain, including sulfur, are considered.](image_url)
In Figure 5.11, the change in the structure factor with the temperature is shown. Figure 5.11(a) shows $S(k)$ with uniform tilt, at temperature of 0K, whereas Figures 5.11(b) and 5.11(c),
show \( S(k) \) at 200K and 400K, respectively. The peak intensity is a function of both temperature and tilt angle of chains. In case of 200K, the peak intensity decreases from zero-temperature value because of the disorder due to temperature. But in the case of 400K, the heights of all the six peaks are almost the same, because of loss of any directionality of the planes (see Figure 5.9 for the loss of order in twist angle distribution \( \sim 450K \)). The peak heights also increase because of the untilting of the chains.

![Figure 5.12: Calculated density profiles along the surface normal at \( T = 100, 300 \) and 450K for \( n = 13 \). The curves for \( T = 100 \) and 300K are shifted vertically for clarity.](image)

The presence of a doublet pattern in the density plot at 100K is an indication of herringbone structure in the backbone plane (the C-C bond takes an alternative sequence of two orientations, nearly parallel and normal, with respect to the surface normal). As temperature increases, the doublet pattern becomes less pronounced, partly due to increased reorientational motion of the backbone planes about the molecular axes and partly due to decreased tilt angle. Above 300K, the peaks
are uniformly spaced and the profiles extend along the z-direction as a result of the untilting. The peaks become broader, indicating a lower degree of positional order of the backbone atoms at high temperatures.

5.4.2 Effect of Lattice Constant

To study the effect of lattice spacing $a$ on the tilt structure, MD simulations for two additional lattice spacings of 4.7 and 5.3Å were also performed. Figure 5.13 shows the average tilt angle as a function of lattice spacing for $T = 0, 200, \text{ and } 300K$. The chains do not show any significant tilt when the nearest-neighbor spacing is less than 5Å.

![Figure 5.13: The collective tilt angle as a function of lattice spacing. The simulation results are shown by diamonds at 0K, squares at 200K and open circles at 300K. The room temperature experimental data for Langmuir (air/water) monolayer are shown by solid circles. The solid line shows $q_{vdw}$ (see Eq. (5.9)). Shown in the inset is the vdw energy as a function of inter-chain separation where $a_{min} = 4.7Å$.]

With increased lattice spacing, the collective tilt angle rapidly increases, consistent with previous simulations [98] and experiments [95]. Figure 5.13 shows that the effect of temperature is small for $a = 4.7Å$ but is significant for $a = 5Å$. Our results also show that the tilt direction
and twist angle distribution are insensitive to the lattice spacing. The effect of lattice spacing on the tilt angle may be understood as follows. As shown in the inset of Figure 5.13, the van der Waals interactions as a function of inter-chain separation has a minimum at $a_{\text{min}} = 4.7\text{Å}$. For SAMs with lattice spacing $a$, the van der Waals energy can be minimized by tilting the chains by

$$q^{vdw} = \cos^{-1} \left( \frac{3}{4(a/a_{\text{min}})^2} \right)$$

(5.9)

so that the inter-chain separation becomes $a_{\text{min}}$. The $q^{vdw}$ plotted in Figure 5.13 qualitatively explains the calculated lattice space dependence of the tilt angle.

5.4.3 Effect of Molecular Chain Length

MD simulations are performed for alkanethiol SAMs of different chain lengths $n$ between 7 and 25. Figure 5.14 shows the tilt angle as a function of $n$ for $T = 50, 200$ and $300\text{K}$. At low temperatures, the effect of chain length on tilt angle is negligible. However, at higher temperatures, the tilt angle is an increasing function of $n$, and the effect is most pronounced at $300\text{K}$. The tilt angle varies with length more for shorter chains ($n \leq 15$) than it does for longer chains ($n \geq 15$).

The above results may be understood as follows. The tilt angle is determined by a competition between potential-energy and entropy contributions to the free energy. The potential energy contribution favors the tilting as discussed previously, see Eq. (5.9), whereas the entropy contribution favors $q = 0$ due to larger available volume for conformational changes. The potential energy contribution consists of van der Waals and surface energy terms. Since the surface potential in the $z$-axis (second term in Eq. (5.6)) is short ranged ($5\text{Å}$ which is much less than the shortest simulated chain length of $13\text{Å}$ for $n = 7$), its contribution to the energy is almost independent of chain length.
Figure 5.14. Tilt angle as a function of chain length for $T = 50, 200,$ and $300K$

On the other hand, the magnitude of van der Waals potential energy increases, more negative, as a function of chain length, as shown in Figure 5.15. For shorter chains, the surface term dominates the potential energy. However because of the short range nature of the surface term, atoms far from the surface are free to move and hence the entropic contributions are more
important at high temperatures. This explains the large variation of tilt angle with temperature for shorter chains as seen in Figure 5.14.

Figure 5.16: (a) Torsion angle distribution for $n = 7$ at $T = 200, 300$ and 400K. (b) Torsion angle distribution for $n = 23$ at $T = 200, 300$ and 400K. The gauche defects are marked by $g^-$ and $g^+$ and trans configuration by $t$.

Significant effects of chain length on the existence of gauche defects were also observed, i.e., there are more gauche defects in shorter chains than in longer chains. This is seen in Figure 5.16, which shows torsion angle distribution at $T = 200, 300, 400$K in the shortest chain length ($n = 7$, Figure 5.16(a)) and in the longest chain length ($n = 23$, Figure 5.16(b)). Gauche defects ($g^-$
and $g^+$), which can be seen as peaks at $\pm 120^\circ$, start appearing at 300K in $n = 7$ case, while significant gauche peaks are seen only at 400K for $n = 23$.

![Twist angle distribution](image)

**Figure 5.17:** (a) Twist angle distribution for $n = 7$ at $T = 0$, 200 and 300K. (b) Twist angle distribution for $n = 23$ at $T = 0$, 200 and 300K.

We have calculated the chain length dependence of the twist angle distribution. The twist angle distributions for the chain lengths of $n = 7$ and 23 for $T = 0$, 200 and 300K are plotted in Figure 5.17. At 300K, for $n = 7$ (Figure 5.17(a)), the twist angle distribution becomes uniform (chains are free to rotate), whereas some structure remains for $n = 23$ (Figure 5.17(b)). The peak
positions are also different for \( n = 7 \) and 23, even at 0K. We find that for the shorter chain (Figure 5.17(a)), the system at 0K is essentially defect free, with only one peak \( \sim -50^\circ \), but for the case of longer chain (Figure 5.17(b)), some twist defects appear even at 0K (small peak \( \sim 125^\circ \)).

### 5.5 Summary

Large-scale molecular dynamics simulations of self-assembled alkanethiol monolayer systems have been carried out using an all-atom model to investigate their structural properties as a function of temperature, lattice spacing and molecular chain-length. To perform large-scale simulations reported in this paper, we have developed a scalable parallel MD code, which achieved 90% of the perfect speedup on 1,024 IBM SP4 processors. Our MD simulation for a million atoms system showed that the alkanethiol chains tilt from the surface normal by a collective angle of 25\(^\circ\) along next-nearest neighbor direction at 300K. This is consistent with experimental result of 30\(\pm10^\circ\). The tilt structure is found to depend strongly on temperature and exhibits hysteresis. The twist angle distribution is also found to depend on temperature. At 350K the system transforms to a disordered phase characterized by small collective tilt angle, flexible tilt direction, and random distribution of backbone planes, where the chains are free to rotate. The tilt structure also depends on lattice spacing \( a \): With increasing lattice spacing, the tilt angle increases rapidly from a nearly zero value at \( a = 4.7\text{Å} \) to as high as 34\(^\circ\) at \( a = 5.3\text{Å} \) at 300K. Finally, the effects of the molecular chain-length on the tilt structure are significant at high temperatures.
CHAPTER 6
SIMULATIONS OF POLY(ETHYLENE GLYCOL) TERMINATED ALKANETHIOLS

6.1 Background

With the emerging technologies in biomaterials, the protection of the surfaces from contamination by proteins, cells, blood etc has become essential [99,100]. Many biomaterials lack the ability to protect the surfaces and this contamination could compromise their bulk mechanical properties, which is undesirable [101]. The reduction of adsorption of proteins on surfaces is important because, it has been shown that the absorption is the first step in immune response of the body on any foreign material. Many strategies have been tried and used to control the non-specific adsorption of proteins on surfaces, and poly (ethylene glycol)(PEG) has met with considerable success [106,108]. Its low toxicity and low immunogenicity render it to be an important coating on many biomaterials. PEG is a water solvable polymer [102] and is one of the most biocompatible polymers [12,107]. Its high water solvability is attributed to its helical nature of its backbone, which maximizes the hydrogen bonding, while at the same time hiding the hydrophobic parts [115].

Steric repulsion theory was one of the proposed mechanisms through which the PEG can resist cells and proteins [103]. When protein molecules approach PEG coupled surfaces, the available conformational volume space for PEG is reduced and a repulsive force is developed. The steric repulsion theory is based on the amount of coformational freedom available to the PEG chains. In a series of experiments [12,110], Prime and Whitesides, have shown that the PEG terminated alkanethiols also resist the adsorption of proteins. In self assembled monolayer system of PEG terminated alkanethiolates, The chains are densely packed, and the
conformational freedom is highly restricted. The observance of protein adsorption in the case of SAMs brings into question the complete validity of steric repulsion theory. The steric repulsion theory treats the polymer chain as a ball and stick configuration, and ignores the chemical nature of the chains themselves.

Recently Grunze et al. [115] have proposed that in fact the helical nature of PEG on alkanethiolates on gold surface is probably the reason for the resistance to proteins. It was shown that the PEG terminated alkanethiolates with all-trans configuration on silver surface absorbed varying amounts protein. This experiment shows that the nature of PEG chains is important when trying to explain the resistance to absorbing proteins. The study of effect of PEG on the tilt structure of alkanethiol system, and the effects of temperature, lattice constant and PEG chain length are also studied.

6.2 Interatomic Potential Parameters for PEG-Alkanethiol SAMs

We use the interatomic potential model in which the conformational energy (V) of a molecular system is split into bonded, non-bonded and surface interaction energies:

\[ V = V_{\text{bonded}} + V_{\text{nonbonded}} + V_{\text{surface}}. \]  

Here \( V_{\text{bonded}} \) represents the bonded interactions that arise from bond-stretching, bending, and torsions:

\[ V_{\text{bonded}} = \]

\[ \sum_{ij} \frac{1}{2} k^s_{ij} (r_{ij} - r_{ij}^0)^2 + \sum_{ijk} \frac{1}{2} k^b_{ijk} \left( \theta_{ijk} - \theta_{ijk}^0 \right)^2 + \sum_{ijkl} \left[ \begin{array}{c} m \sum_{i} k^0_{ijkl} \prod_{m=1}^{m_{\text{folds}}} k^m_{ijkl} \left( m \left( \Phi_{ijkl} - \Phi_{ijkl}^0 \right) \right) \end{array} \right] \]  

where two atoms \( i \) and \( j \) define a covalent bond of length \( r_{ij} \); three atoms \( i, j \) and \( k \) define bond angle \( \theta_{ijk} \), and four atoms \( i, j, k \) and \( l \) define a dihedral angle \( \Phi_{ijkl} \). A pair of atoms \( i \) and \( j \) are considered to be non-bonded if they belong to different molecules or are separated by more than
three bonds in the same molecule. The resulting non-bonded interactions (i.e., long-range repulsions and van der Waals interactions) are expressed as

\[
V_{\text{nonbonded}} = V_{\text{vib}} + V_{\text{coulomb}} = \sum_{i \neq j} A_{ij} \exp(B_{ij} r_{ij}) C_{ij}^{(p)} + \frac{q_i q_j}{r_{ij}}.
\]  

(6.3)

The Coulomb potential is calculated by the Ewald method. The charges on the PEG atoms are taken from ref.[104]*

Finally, the interactions of sulfur and carbon atoms with gold substrate (i.e., surface interactions) are modeled [88] by

\[
V_{\text{surface}} = \sum_{s} \frac{1}{2} k_{s} S_{s}^2 + \sum_{pq} k_{pq}^{SF} \left[ \frac{\xi_{pq} \xi_{p}^{0}}{\xi_{pq}^{0}} \right]^{12} + \frac{k_{3}^{SF}}{(\xi_{pq} \xi_{p}^{0})^{6}}
\]

(6.4)

where the index \(s\) runs over sulfur atoms, \(p\) counts the number of chains or molecules, \(q\) runs over sulfur and carbon atoms on each molecule, and \(d_s\) is the distance of sulfur atom from its original position on \(x-y\) plane. The first term constrains the motion of sulfur atoms on \(x-y\) surface whereas the second term involves \(z\) positions of sulfur and carbon atoms. The model parameters [88,92,93,104] used in our MD simulations are given in Table 6.1.

<table>
<thead>
<tr>
<th>(ij)</th>
<th>(r_{ij}^0) (Å)</th>
<th>(k_{ij}^{0}) (Kcal/mol/Å(^2))</th>
</tr>
</thead>
<tbody>
<tr>
<td>C—C</td>
<td>1.53</td>
<td>618.0</td>
</tr>
<tr>
<td>H—H</td>
<td>1.09</td>
<td>655.2</td>
</tr>
<tr>
<td>C—O</td>
<td>1.43</td>
<td>739.0</td>
</tr>
<tr>
<td>C—S</td>
<td>1.82</td>
<td>618.0</td>
</tr>
</tbody>
</table>

(Table Cont’d)

* The system considered in the ref. [104] is of the form \(\text{H} \quad \text{CH}_2 \quad (\text{O} \quad \text{CH}_2 \quad \text{CH}_2)_m \quad \text{O} \quad \text{CH}_3\), whereas the PEG terminated alkanethiol system has the form \(\text{S} \quad (\text{CH}_2)_n \quad \text{H}_2\text{C} \cdot \text{CH}_2 \quad (\text{O} \quad \text{CH}_2 \quad \text{CH}_2)_m \quad \text{O} \quad \text{CH}_3\). The atoms in bold are assigned the same charge.
<table>
<thead>
<tr>
<th>$ijk$</th>
<th>$\phi_{ijk}$ (radians)</th>
<th>$k_{ijk}^b$ (Kcal/mol/radian$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C—C—C</td>
<td>1.9373</td>
<td>107.6</td>
</tr>
<tr>
<td>C—C—H</td>
<td>1.9106</td>
<td>85.8</td>
</tr>
<tr>
<td>H—C—H</td>
<td>1.8832</td>
<td>77.0</td>
</tr>
<tr>
<td>C—O—C</td>
<td>1.9471</td>
<td>149.0</td>
</tr>
<tr>
<td>H—C—O</td>
<td>1.9211</td>
<td>112.0</td>
</tr>
<tr>
<td>O—C—C</td>
<td>1.9031</td>
<td>172.0</td>
</tr>
<tr>
<td>S—C—C</td>
<td>1.9967</td>
<td>107.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$ijkl$</th>
<th>$\phi_{ijkl}$ (radians)</th>
<th>nflods</th>
<th>$k_{ijkl}^b$ (Kcal/mol/radian$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C—C—C—H</td>
<td>0</td>
<td>0,3</td>
<td>0.27792</td>
</tr>
<tr>
<td>H—C—C—H</td>
<td>0</td>
<td>0,3</td>
<td>0.27792</td>
</tr>
<tr>
<td>C—C—C—C</td>
<td>0</td>
<td>0,3</td>
<td>0.27792</td>
</tr>
<tr>
<td>C—C—C—O</td>
<td>0</td>
<td>0</td>
<td>2.05</td>
</tr>
<tr>
<td>C—C—O—C</td>
<td>0</td>
<td>1</td>
<td>0.90</td>
</tr>
<tr>
<td>C—C—O—C</td>
<td>□/2</td>
<td>2</td>
<td>0.70</td>
</tr>
<tr>
<td>C—C—O—C</td>
<td>0</td>
<td>3</td>
<td>0.45</td>
</tr>
<tr>
<td>O—C—C—O</td>
<td>0</td>
<td>0</td>
<td>2.72</td>
</tr>
<tr>
<td>O—C—C—O</td>
<td>□</td>
<td>1</td>
<td>-0.08</td>
</tr>
<tr>
<td>O—C—C—O</td>
<td>□/2</td>
<td>2</td>
<td>2.55</td>
</tr>
<tr>
<td>O—C—C—O</td>
<td>0</td>
<td>3</td>
<td>0.25</td>
</tr>
<tr>
<td>C—O—C—H</td>
<td>0</td>
<td>0,3</td>
<td>0.80784</td>
</tr>
<tr>
<td>O—C—C—H</td>
<td>0</td>
<td>0,3</td>
<td>0.27792</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$ij$</th>
<th>$A_{ij}$ (kcal/mol)</th>
<th>$B_{ij}$ (1/ Å)</th>
<th>$C_{ij}$ (Å$^6$ Kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C—C</td>
<td>14976.0</td>
<td>3.09</td>
<td>640.8</td>
</tr>
<tr>
<td>H—H</td>
<td>2649.6</td>
<td>3.74</td>
<td>27.4</td>
</tr>
<tr>
<td>C—H</td>
<td>4320.0</td>
<td>3.42</td>
<td>138.2</td>
</tr>
<tr>
<td>S—S</td>
<td>79937.0</td>
<td>3.18</td>
<td>2002.0</td>
</tr>
</tbody>
</table>

(Table Cont’d)
6.3 Initial Setup

The system of self-assembled monolayers consists of 1500 chains, each with a chemical composition of $S-(\text{CH}_2)_n \square (\text{O-CH}_2 \square \text{CH}_2)_m \square \text{O-CH}_3$. The chains form a triangular lattice with sulfur atoms bonded to a substrate gold (Au111). In our current study $n$ takes values 7, 13, 23 and $m$ takes values 1, 3, and 6 and most of the results that follow are for the case of $n = 13$ and $m = 3$. The lattice spacing $a$ of the triangular lattice is also varied from 4.7 to 5.8Å. Periodic boundary conditions are applied in the surface plane. In order to remove interactions between periodic images in $z$ direction, a vacuum space (longer than $r_c$) is inserted above the free surface of the SAM.

The tilted alkanethiol chain is characterized by tilt angle ($\square$), twist angle ($\square$) and tilt direction ($\square$). An initial angle of $\square \sim 14^\circ$ is given to all the molecular chains and the backbone planes of all the chains are rotated by $\square \sim 10^\circ$ with nearest neighbor direction (parallel to $x$-axis).
to break the symmetry of the triangular lattice. The conformation of each chain was chosen to be *all-trans* in the beginning of the simulation.

All the atoms are given random initial velocities, according to the Maxwell-Boltzmann distribution, corresponding to temperature of $T = 1$K. We thermalize the system for 150ps during which the molecular chains increasingly tilt away from the surface normal towards one of the next-nearest neighbor directions and the tilt structure finally reaches a steady state. The statistical averages of potential energies and structural parameters are calculated, after the thermalization stages, over the next 50ps.

6.4 Results and Discussion

6.4.1 Effect of temperature

We have studied the effect of temperature on the structure of the alkanethiol-EG3-OMe system with parameters of $n = 13$, $m = 3$, lattice constant $a = 5\, \text{Å}$, and the number of atoms is 99,000 (1,500 chains). Starting from the initially thermalized system (in which the temperature saturated to a value of $\sim 50$K), we successively scale the atom velocities to vary the system temperature from 50 to 450K in steps of 25K. For each 25K step, the temperature is increased over 30ps, the system is thermalized for 90ps, and statistical averages obtained over another 50ps. We also obtain the 0K system by gradually decreasing the temperature from the initially thermalized system. To study possible hysteresis, the temperature is decreased from 450K to 0K again in steps of 25K.

As shown in Figure 6.1, with increasing temperature, the average tilt angle ($\theta$) of the chains initially decreases slowly and then rapidly until it saturates at a low value for temperatures higher than 400K. At 300K the calculated tilt angle ($30^\circ$) agrees fairly well with experimental tilt angle of about $30^\circ$ [4,8]. The tilt direction is along the next-nearest neighbor (NNN) direction.
below 300K ($\square \sim 0$) and at temperatures above 300 K, it shifts away from the NNN direction by about 15° (see Figure 6.2). At temperature above 400K, the tilt direction increases to $\sim 65$° and seems to remain locked at that value.

Figure 6.1: Temperature variation of the collective tilt angle: The simulation results from heating run and cooling run are shown by circles and squares, respectively.

Figure 6.2: Temperature variation of tilt direction: circles and squares show the simulation results from heating run and cooling run respectively.

We have also studied the effects of annealing on the system. Figures 6.1 and 6.2 also show the presence of a hysteresis between 100 and 400K. We found that the tilt direction
remains locked at another NNN direction (~60°) at all temperatures in the cooling cycle. This was unlike the case of pure alkanes, where the tilt direction returned to ~0° NNN direction in the cooling cycle [105].

We have also studied the effect of temperature on the number of gauche conformations in the EG part of the chains. It has been established [115] that the unique helical nature (which is due to gauche conformation about CC bond) of PEG is responsible for the hydrophilic nature of the system. The signature of the gauche conformation about CC bond can be seen in the partial radial distribution graph of oxygen-oxygen atoms (see Figure 6.3). For distance below 5Å, the peak ~ 3Å is the indication of gauche conformations and the peak ~ 3.8 Å is the indication of trans conformation about CC bond (OCCO). From Figure 6.3, we can see that at low temperatures (~50K), there is no ‘gauche’ peak indicating that the only conformation about CC bond is trans in nature. As the temperature is increasing, the gauche peak increases, and the intensity of the trans peak decreases showing that the number of gauche conformations in EG increase with the temperature.
Figure 6.3: Calculated partial radial distribution function, \( g_{oo}(r) \) at \( T = 50, 200, 300, \) and \( 400K \) in heating cycle. The curve for different temperatures are shifted vertically for clarity.

We studied the effects of annealing (heating and cooling cycles) on the number of gauche conformation in the EG part of SAMs. We see from Figure 6.4 that, at the same temperatures in heating and cooling cycles, the intensity of gauche and trans peaks are very different. In heating cycle, the trans peak dominates, but in the cooling cycle the gauche peak dominates which can be attributed to the locking of increased gauche conformations at high temperatures in cooling cycles.

![Figure 6.4](image)

Figure 6.4: Calculated partial radial distribution function, \( g_{oo}(r) \) at \( T = 200K \) in heating and cooling cycle. The curve for heating cycle is shifted vertically for clarity.

The line density profile of backbone atoms normal to the surface gives a measure of the positional order and the calculated profiles at \( T = 100, 200, 300K \) are shown in Figure 6.5. The presence of a doublet pattern in the density plot at 100K is an indication of all-trans structure in the backbone plane (the C-C bond takes an alternative sequence of two orientations, nearly
parallel and normal, with respect to the surface normal). As temperature increases, the doublet pattern becomes less pronounced, partly due to increased reorientational motion of the backbone planes about the molecular axes and partly due to decreased tilt angle. At 300 K, the peaks are uniformly spaced and the profiles extend along the $z$-direction as a result of the untilting. The peaks become broader, indicating a lower degree of positional order of the backbone atoms at high temperatures.

![Graph](image)

Figure 6.5: Calculated density profiles along the surface normal for $T = 100$, 200 and 300K for $n = 13$. The curves for $T = 100$ and 300K are shifted vertically for clarity

6.4.2 Effect of Lattice Constant

We studied the effect of spacing between the chains on the tilt structure and the gauche conformations in EG part of SAMs. Figure 6.6 shows the partial radial distribution function of oxygen-oxygen atoms in EG part. As the spacing between the chains is increasing the intensity of gauche peaks increases and the same decreases for the trans peaks. As more volume becomes available for the EG between chains, the gauche conformations increase and the chains are more helical in nature. The same can be seen from the torsion angle distribution graphs shown in Figure 6.7. The gauche peaks increase in intensity with the increase in the chain-chain spacing.
Figure 6.6: The effect of lattice spacing on the radial distribution function of oxygen-oxygen atoms in PEG part of SAMs. The peak about 3 Å is due to the gauche conformation about CC bond in PEG part and becomes prominent with lattice spacing. The lack of any specific peaks after 4 Å for a=5.8 Å, indicates the amorphous nature of PEG part as lattice spacing is increased.

Figure 6.7: The torsion angle distribution in PEG part for different lattice constant at T = 200K

6.4.3 Effect of PEG Chain Length

MD simulations are performed for alkanethiol SAMs of different chain lengths $n$ takes values 7, 13, and 23 and $m$ takes values 1, 3, and 6 (see Figure 6.8). The length of the EG chains
has effect on the oxygen-oxygen partial radial distribution. With increasing length of EG units, the intensity of gauche peaks decreases, suggesting that in longer unit of EG, only the terminal oxygens may be involved in gauche conformations, while the other oxygens may be in trans configuration. The alkane chain length did not have any effect on the structure of EG part of the chain.

![Figure 6.8: Effect of PEG chain length on the gauche conformation in g_{o-o}(r).](image)

6.5 Summary

Large-scale molecular dynamics simulations of Poly (ethylene glycol) (PEG) terminated alkanethiol SAMs have been carried out using an all-atom model to investigate their structural properties as a function of temperature, lattice spacing and molecular chain-length. The presence of polar species such as oxygen, requires the inclusion of Coulomb potential in the force field. The Coulomb interactions were computed by Ewald method. The tilt structure of PEG-SAM system was found to be similar to that of plain alkanethiol system, but the gauche defects in the end groups were significant. While at low temperatures, the chains were predominantly all-trans in nature, with the increase in the temperature, the gauche defects in end groups was found to increase significantly. A clear hysterisis was observed in the case of both tilt angle and tilt
direction in heating-cooling cycles. Unlike the case of alkanethiol system, the tilt direction shifted to another nearest neighbor direction (from 0° with respect to x-axis to 60° with respect to x-axis) at high temperatures and remained locked in that direction in cooling cycle. With increasing lattice spacing, the PEG end group was more helical in nature, as more space was available, which was seen in the increase of gauche peaks in oxygen-oxygen partial radial distribution function. The helical nature of the PEG groups is directly related to the hydrophilicity of the surface layers as the exposure of oxygen atoms at the surface increases with the helical nature. The length of the PEG units also has an effect on the helical nature of PEG, more gauche conformations were found in smaller PEG units than the longer ones.
CHAPTER 7
ELECTRIC FIELD INDUCED SURFACE SWITCHING IN POLY (ETHYLENE GLYCOL) TERMINATED SELF-ASSEMBLED MONOLAYERS

7.1 Background

Protein resistant surfaces play a significant role in biotechnology to stabilize cells [106,107,108] and to prevent the degeneration of the bulk properties of materials upon protein adsorption [109]. Recently poly (ethylene glycol) (PEG) terminated self assembled monolayers (SAM) have attracted much attention because of their successful ability to resist protein adsorption [110]. This surface property is attributed to a large extent to its unique structure, i.e., helical nature [111] and number of exposed oxygen atoms. The surface structure, and consequently the surface properties, is highly sensitive to external parameters such as temperature [112], electric fields [113], and types of solvents [114] and substrate [115].

Among these parameters, externally applied electric field is technologically most important due to its control over surface properties e.g., reversible switching of hydrophilicity has been observed in SAMs [113]. To optimize surface properties, it is important to atomistically understand the dependence of surface structures on applied electric field. Unfortunately the effect of electric fields on the structure of SAMs at the atomic level is not well known. Atomistic simulations based on molecular dynamics (MD) technique is expected to provide some insight into the behavior of the system.

We use the interatomic potential model in which the conformational energy \( V \) of a molecular system is split into bonded, non-bonded, surface interaction and external electric field energies:

\[
V = V_{\text{bonded}} + V_{\text{nonbonded}} + V_{\text{surface}} + V_{\text{electric}}
\]  

(7.1)
The functional forms and potential parameters are described in chapter 6.

The external electric field is modeled by

$$V_{electric} = q_i E_z z_i$$ (7.2)

where $q_i$ is the charge of the $i$th atom, $E_z$ is the electric field strength applied in $z$-direction and $z_i$ is the projection of the position vector on $z$-axis.

7.2 Initial Setup and Simulation Schedule

The PEG terminated SAM system, consists of 1500 chains, each with chemical composition of $S-(CH_2)_{13} \square (O-CH_2 \square CH_2)_{3} \square O-CH_3$, on gold surface. The chains are made in an all trans configuration, tilted initially by about $15^\circ$ from the normal to the surface, and a temperature of 1K is given to all the atoms to break the six-fold symmetry. After a thermalization run of 90ps, the system attained a temperature of ~50K. The system’s temperature is then raised to 200K in steps of 25K, over 200ps. The system is characterized at this temperature and a series of electric fields with varying strengths are applied in $z$-direction. The strength of the electric fields varies form $+2V/\AA$ to $-2 \ V/\AA$, in steps of 0.5 V/Å. For each value of the electric field, the system is thermalized over 120ps. If the final temperatures after thermalization run is higher than 200K, the temperature is decreased to 200K over a 60ps run. To study the reverse transition, the systems with $\pm 2 \ V/\AA$ electric fields, the field is switched off and the system is thermalised over 90ps.

7.3 Results and Discussion

7.3.1 Electric Field Effect on Gauche Conformations

We calculate oxygen-oxygen partial radial distribution ($g_{O-O}$) function in PEG to study the trans to gauche transition in the presence of applied electric field. In the absence of field, $g_{O-O}$ has two significant peaks corresponding to gauche ($\sim 3\AA$) and trans ($\sim 3.8 \ \AA$) conformations about
C–C bond (O=C–C–O) (see dashed curve in Figure 7.1). With no field applied, only the terminal oxygen atom is in gauche conformation, with most of the PEG in nearly all-trans configuration. Accordingly the trans peak ~3.8Å is more significant than the gauche peak ~3Å. With the application of electric field, there is a dramatic change in the ratio of gauche to trans peaks. The intensity of the gauche peaks increases while the intensity of the trans peaks decreases with the magnitude of the field applied. For the field strengths of ± 2V/Å, the trans peaks almost disappear (see Figure 7.1(a) & 7.1(b)). The all-gauche conformation at high electric fields corresponds to a complete helical structure in PEG. Thus the electric field causes a transition from a mixed trans-helical state to a complete helical state.

![Graph showing radial distribution g_0(0) vs. r with different field strengths for (a) negative and (b) positive field strengths.](image)

Figure 7.1: Effect of electric field on the gauche conformation about CC bond in PEG. (a) for negative field strengths (b) for positive field strengths.

We also study the effect of electric field on the entire chain (both alkanethiol and PEG) using the line density profile of backbone atoms normal to the surface and the calculated profiles at negative fields and positive are shown in Figure 7.2(a) and 7.2(b). The presence of a doublet pattern in the density plot at E_z = 0 is an indication of all-trans configuration (the C–C bond takes an alternative sequence of two orientations, nearly parallel and normal, with respect to the surface normal) in the alkanethiol part of the chain. As field strength increases, the doublet
pattern becomes less pronounced indicating appearance of gauche defects even in the alkanethiol part of the chain. The negative fields affect more adversely the alkanethiol structure than the positive fields, which can be seen by complete disappearance of doublet pattern in case negative electric fields.

![Graphs showing density profiles](image)

Figure 7.2: Calculated density profiles along the surface normal for (a) positive field (b) negative field are shown.

### 7.3.2 Electric Field Effect on Tilt Structure

The tilt structure in SAMs is characterized by two angles *viz.*, the tilt angle, which is the angle made by the backbone plane of the chain with respect to the surface normal and the tilt direction, which is the rotation angle about surface normal with respect to x-axis. With the electric field, the tilt angle is affected significantly and the deviation from the initial uniform tilt angle increases with the increase in the magnitude of the field strength (see Figure 7.3(a) and 7.3(b)). The temperature of the system also increases with the application of electric fields because more energy is put into the system. For negative electric fields, the increase in temperature is higher than for positive fields, because of the increased disorder set in the system with negative fields. The negative field affects the tilt structure more adversely than the positive electric field at high field strengths, which can be see from Figure 7.3(a). In both the cases of ±
2V/Å, the tilt angle initially decreases, but in the case of negative electric field, the system does not recover from the initial loss of order in the system (see Figure 7.3(a)), whereas in the case of positive electric field, ~75% of the tilt angle is recovered (see Figure 7.3(b)). This is because the negative field tends to pull the oxygen atom towards the substrate, thus introducing disorder in the alkanethiol part of the chain, which destroys the order. The spacing between the chains (~4.8Å) is not enough for the chain to bend completely as was observed in the Ref.[113], where the spacing was ~5.8Å.

Figure 7.3: Effect of electric field on tilt structure for (a) positive field (b) negative field are shown

Figure 7.4 shows the effect of an electric field on the structure factor of the PEG terminated SAMs. Negative electric fields affect the structure factor (see Figure 7.4(c)) more than the positive fields (see Figure 7.4(b)). The disorder is also more in the case of negative electric fields and the peak heights are increased because of decrease in the tilt angles of the chains (see Figure 7.3(b)).
Figure 7.4: 2D structure factors for PEG-terminated SAMs. (a) field = 0V/Å (b) field = 2V/Å (c) field = -2V/Å.
We also studied the effect of electric field on the tilt direction of chains. Figure 7.5(a) and 7.5(b) show the 2D surface plots of tilt direction of all the chains. For $E_z = 0$, Figure 7.5(a) shows the uniform tilt direction distribution $\sim 10^\circ$. For positive polarity, we observed an alternating ‘stripe’ pattern in the tilt direction of the system (see Figure 7.5(b)). We find a correlation between the observed pattern in tilt direction and in-plane polarization of the chains. Figure 7.5(c) shows the polarization of chains in y-direction, which has a similar pattern as that of tilt direction plot in Figure 7.5(b).

Figure 7.5: 2D surface plots of the tilt direction of all the chains in (a) $E_z = 0$ (b) $E_z = +2\text{V}/\text{Å}$. (c) 2D surface plot of polarity of chains in Y-direction.

### 7.3.3 Electric Field Effect on Surface Oxygen Atoms

With the application of positive electric field more oxygen atoms are exposed near the surface. This can be seen in Figure 7.6, which shows the number of exposed oxygen atoms (white) for different strengths of $E_z$. The thickness of the surface layers shown in Figure 7.6 is 1 Å. We see very clearly from the figures that with the increase of the field strength, more and more oxygen atoms come to the top of the layer. This enhanced exposure of oxygen atoms at the surface will increase the hydrophilicity of the surface.
Figure 7.6: Surface layer 1Å below the highest $z$ for (a) field=0.5 V/Å (b) field=1V/Å (c) field=1.5V/Å (d) field=2V/Å

The exposure of oxygen atoms with positive electric field is also seen in Figure 7.7, which shows atomic configurations of the top 6Å layers of the PEG when $E_z =0$, and ± 2V/Å, where atoms are color coded by the charge. Figure 6 also shows that with the application of negative electric fields, some chains are depressed leaving behind 'holes' in the top layer of PEG system (see Figure7.7(c)).
Figure 7.7: The atomic configuration color coded by charges: oxygen (red), hydrogen (white), methyl terminal carbon (yellow) and carbon in PEG (green) in top 6Å layer of the system (a) $E_z = 0$ (b) $E_z = +2$ V/Å (c) $E_z = -2$ V/Å.

7.3.4 Reversible Structural Transition Under an Electric Field

We also studied whether the electric field induced structural transition is reversible. In Figure 7.8, the atomic configuration of top 6Å layer of the system is shown before the application, after application and after removal of the positive electric field. The reversible effect
can be seen in Figure 7.8(c), where after the removal of electric field, the top layers have more hydrogen than oxygen.

![Figure 7.8: The top 6Å layer of the system (a) before (b) after application of positive electric field. (c) shows the same layer after the removal of electric field.](image)

The dynamics of the tilt structure during the application and removal of electric fields can be seen in Figure 7.9 which plots the tilt angle as a function of time for $E_z = +2V/Å$ (Figure 7.9(a)) and $E_z = -2V/Å$ (Figure 7.9(b)). In both cases the temperature of the system increased both as a result of application and removal of the fields, but the increase is more in case of...
negative fields. We decreased the system temperature after the removal of field, to the original temperature in order to compare the systems. We find that the system to which positive electric field is applied regains the tilt angle (90%), the system with negative electric field could regain only 74% of the tilt angle back.

Figure 7.9: Plot of the tilt angle as a function of time for (a) $E_z = +2\text{V/Å}$ and (b) $E_z = -2\text{V/Å}$

The trans to gauche transition with the electric field is reversible with the removal of the same, which can be seen in Figure 7.10. In Figure 7.10, we compare the partial radial distribution of oxygen-oxygen in PEG before, after application of electric field. The trans peak, which completely disappears with the application of electric field, reappears after the field is removed.
Figure 7.10: Reversible transition of trans-gauche conformations with and removal of electric field

7.4 Summary

Molecular dynamics simulations have been performed on parallel computers to study the effect of electric field on the structure of poly (ethylene glycol) (PEG) terminated alkanethiol self assembled monolayer (SAM) system on gold surface. The electric field triggers a conformational transition from all-trans to a mostly gauche conformation. The polarity of the electric field has a significant effect on the surface structure of PEG. The electric field applied anti-parallel to the surface normal causes a reversible transition to an ordered state where the oxygen atoms are exposed. On the other hand, an electric field applied in a direction parallel to the surface normal introduces considerable disorder in the system and the oxygen atoms are buried. The structural difference caused by the polarity of the electric field has a profound effect on the hydrophilicity of the surface. The parallel field affects the overall tilt structure of SAMs adversely than the anti-parallel field.
CHAPTER 8
CONCLUSIONS

Molecular dynamics simulations of various self-assembled monolayers (SAM) have been described in this thesis. The simulations of properties of SAMs required usage of highly scalable multiresolution algorithms on parallel computers.

Large-scale molecular dynamics simulations of self-assembled alkanethiol monolayer systems have been carried out using an all-atom model to investigate their structural properties as a function of temperature, lattice spacing and molecular chain-length. MD simulations on a million-atom system showed that the alkanethiol chains tilt from the surface normal by a collective angle of 25° along next-nearest neighbor direction at 300K. This is consistent with experimental result of 30±10°. The tilt structure was found to depend strongly on temperature and exhibits hysteresis. The twist angle distribution was also found to depend on temperature and the definitive twist peaks at low temperatures disappear, indicating the rotational freedom of backbone planes, at higher temperature. At 350K the system transforms to a disordered phase characterized by small collective tilt angle, flexible tilt direction, and random distribution of backbone planes, where the chains are free to rotate. The tilt structure also depends on lattice spacing \( a \): With increasing lattice spacing, the tilt angle increases rapidly from a nearly zero value at \( a = 4.7\text{Å} \) to as high as 34° at \( a = 5.3\text{Å} \) at 300K. Finally, the effects of the molecular chain-length on the tilt structure are significant at high temperatures.

Next, large-scale molecular dynamics simulations of Poly (ethylene glycol) (PEG) terminated alkanethiol SAMs have been carried out using an all-atom model to investigate their structural properties as a function of temperature, lattice spacing and molecular chain-length. The presence of polar species such as oxygen, requires the inclusion of Coulomb potential in the force field. The Coulomb interactions were computed by Ewald method. The tilt structure of
PEG-SAM system was found to be similar to that of plain alkanethiol system, but the gauche defects in the end groups were significant. While at low temperatures, the chains were predominantly all-trans in nature, with the increase in the temperature, the gauche defects in end groups was found to increase significantly. A clear hysteresis was observed in the case of both tilt angle and tilt direction in heating-cooling cycles. Unlike the case of alkanethiol system, the tilt direction shifted to another nearest neighbor direction (from 0° with respect to x-axis to 60° with respect to x-axis) at high temperatures and remained locked in that direction in cooling cycle. With increasing lattice spacing, the PEG end group was more helical in nature, as more space was available, which was seen in the increase of gauche peaks in oxygen-oxygen partial radial distribution function. The helical nature of the PEG groups is directly related to the hydrophilicity of the surface layers as the exposure of oxygen atoms at the surface increases with the helical nature. The length of the PEG units also has an effect on the helical nature of PEG, more gauche conformations were found in smaller PEG units than the longer ones.

Electrical field induced surface switching on PEG terminated alkanethiol SAMs was also studied by molecular dynamics simulations have been performed on parallel computers. The electric field triggers a conformational transition from all-trans to a mostly gauche conformation. The polarity of the electric field has a significant effect on the surface structure of PEG. The electric field applied anti-parallel to the surface normal causes a reversible transition to an ordered state where the oxygen atoms are exposed. On the other hand, an electric field applied in a direction parallel to the surface normal introduces considerable disorder in the system and the oxygen atoms are buried. The structural difference caused by the polarity of the electric field has a profound effect on the hydrophilicity of the surface. The parallel field affects the overall tilt structure of SAMs adversely than the anti-parallel field.
To perform the MD simulations discussed above, a parallel, scalable and portable Macro-Molecular Dynamics (MMD) algorithms were also developed as a part of this thesis. Fast Multipole Method (FMM) algorithm was also implemented on parallel machines to be able to co compute Coulomb interactions on system size of the order of millions of atoms. Results of scalability tests on these two algorithms FMM and MMD were described in this thesis. The MMD algorithm was based on a space-time multiresolution approach combining the FMM and multiple time-scale method as well as dynamic management of distributed linked cells, neighbor lists, and atomic n-tuples. The results of scalability tests on FMM were presented first, in which the parallel efficiency of FMM on a system size of 500 million was shown to be as high as 0.98. The tests on FMM were performed on two tera-flop machines IBM SP3 and SP4. The scalability and portability of MMD algorithm were described next. The tests on MMD were performed on 1024-processor Intel Xeon-based Linux cluster, 1184-processor IBM SP4 and 512 Compaq Alpha Server. The parallel efficiency of the MMD code was shown to be 0.87 on IBM SP4. The effects of memory- and cache-sharing were also demonstrated and their effect was shown to be significant. Such a scalable parallel macro-molecular simulation algorithm is expected to play an important role in the design of hybrid quantum-device/biological-cell systems and ‘virtual-cell’ technologies.

Simulations of SAMs on semi-conductor surface of Gallium Arsenide (GaAs) are being investigated. SAMs on GaAs are the basic units, on which complicated structures involving nanoparticles such as ZnS can be added on, which have shown interesting electrical properties [1]. Future work will involve explicit GaAs surface atoms and a hybrid simulation of semiconductor (GaAs), organic system (SAM) and nanoparticles (ZnS) to investigate the structural properties of such hetero structures. The degree of crystallanity in SAM films is important for the mechanical
properties of SAMs and future work involves studying the effects of defects in the SAM films, methods of recovering from the defects, effects of size and shape of defects on the properties of SAM films.
REFERENCES


102


Vemparala S., Karki B., Kalia R.K., Nakano A., Vashsihta P., JCP submitted


VITA

Satyavani (Vani) Vemparala was born on June 17, 1974, in Varangaon, Maharashtra, India, to Subramanya Sarma and Padmavathy Vemparala. In 1996, Vani attained her M.Sc, equivalent to bachelor’s degree in United States, from University of Hyderabad, India. In 1998, she attained her master’s degree in solid state technology from Indian Institute of Technology, Kharagpur, India.

In 1998, Vani entered the dual degree doctoral program in the Department of Physics and Astronomy at Louisiana State University in Baton Rouge, Louisiana. She was awarded a Graduate school Enhancement award in 1999. In 2000, she enrolled in the Systems Science master’s program with the Department of Computer Science. She expects to graduate in both the degrees in 2003 Fall. Her future plans include postdoctoral fellowship at University of Pennsylvania, Philadelphia.