2-1-2013

Molecular dynamics simulation and phase transitions of Tetrafluoromethane

Rasha Ibrahim Abdulshahid

Recommended Citation

APA Citation
Abdulshahid, R. (2013).*Molecular dynamics simulation and phase transitions of Tetrafluoromethane* [Master's thesis, the American University in Cairo]. AUC Knowledge Fountain.
https://fount.aucegypt.edu/etds/1288

MLA Citation
Abdulshahid, Rasha Ibrahim. *Molecular dynamics simulation and phase transitions of Tetrafluoromethane*. 2013. American University in Cairo, Master's thesis. *AUC Knowledge Fountain*.
https://fount.aucegypt.edu/etds/1288

The American University in Cairo

School of Sciences and Engineering

Molecular Dynamics Simulation and Phase Transitions of Tetrafluoromethane

A Thesis Submitted to

Physics Department

In partial fulfillment of the requirements for

the degree of Master of Science

By: Rasha Ibrahim Abdulshahid

Under the supervision of

Prof. Dr. Salah El-Sheikh

Physics department, American University in Cairo, Egypt

Associate Prof. Noha Salem

Engineering, Mathematics and Physics department

Faculty of Engineering, Cairo University, Egypt

January 2013

<u>DEDICATION</u>

This thesis is perhaps one of the most important achievements of my life so far. I finally feel that all the hours of work I spent trying to pull off the results and conclusions I have finally been able to develop were truly worthwhile. However none of this would have been closely possible were it not for the support of a number of persons to whom I heart fully dedicate this thesis and any consequential success. First I dedicate it to my father. I truly thank my dear father for all the support and unconditional love he has always given me. He has always believed in me even when I did not believe in myself and always pushed me towards making this accomplishment. He was always keen to see me at my best and he never ceased to provide me with all the support I needed, both emotional and financial.

I also dedicate it to my Husband. I truly thank him for always being supportive, encouraging and loving. I thank him for giving me all the incentives I needed to finish this work and for always believing that this thesis is only the beginning of a very successful career that he sees for me. If it weren't for him perhaps I would have let my role as a mother completely take over my dreams of a successful career in the scientific community. I would also like to dedicate this thesis to my mother whose love and kindness have always gotten me out of the worst moods, and whose prayers were always there to protect and guide me. Last but not least I would like to dedicate this thesis to my entire family and friends and especially to my daughters who always managed to draw a smile on my face and give me a reason to be successful, just for them to be proud of me.

Thank you all, I would never have made it without your love, support, encouragement and faith.

ACKNOWLEDGEMENTS

This work is not the product of my efforts alone. It was made possible by the efforts of other persons to whom I would like to express my sincerest gratitude and appreciation. First and foremost I would like to thank my advisors who were always there for me every step of the way. I would like to thank Prof. Dr. Salah El Sheikh for always believing in me and giving me one chance after the other. I truly appreciate all his advice, kindness and insistence that I can do anything I put my mind to. Dr. Salah, if it weren't for you I would have quit a long time ago. You are a father, friend, mentor and a true inspiration

I would also like to express my sincerest appreciation to Dr. Noha Salem for all her efforts, all her support and all her advice, whether academic or on the personal level. Dr. Noha, I have always felt that you were more of an older sister to me than an advisor and I always hated to disappoint you. Maybe that was even one further push. Thank you for all your prayers when I needed nothing else.

Besides my respected advisors, I would also like to thank two people for their input in this thesis. First I would like to thank Dr. Nageh Allam for helping me present this thesis in the best possible way. Dr. Nageh, I have only had the privilege of meeting with you a number of times, but the impression you have left will last for very long.

Last but not least I thank Eng. Hany Mustafa, for making it possible for me to finish this work by providing the facilities and help I truly needed. Thank you very much. This thesis is yet another of your many developments.

ABSTRACT

Properties of material under extreme conditions, whether high pressures, high temperatures or low temperatures, are rather interesting. Understanding the properties of any given material at such extreme conditions provides a much clearer conception of the properties of this material in general. Although until recently studying the nature of matter under such extreme conditions was not possible via the traditional experimental or traditional methods, it was made possible by the utilization of computer simulation methods.

Carbon tetrafluoride has been gaining wide interest in recent years due to its important role in the semiconductor processing industries. It is one of the most preferred etchants used today on a number of different silicon films. The importance of CF_4 has resulted in a vast interest in studying the molecule under all possible conditions of pressure and temperature.

The nature of CF_4, it's crystal structure and phase transitions was investigated from 0 K to 450 K and from 0 GPa to 35 GPa, using the Molecular Dynamics simulation technique. Keith Refson's Moldy was the software we decided to use in our research. The results we were able to arrive at were compared with experimental findings and they were quite in agreement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

For Decades, there existed a clear distinction between different sciences. Chemistry, Physics, Biology, Mathematics, Computer science and other classes of science, were being treated and viewed each as a school on its own. That was until scientists started noticing how closely interrelated all the sciences are, and also how vital it is to understand one class in order to fully comprehend the other. It was then that the scientific community started introducing such terminology as physical chemistry, chemical physics, quantum mechanics, biophysics, mathematical physics and computational physics to describe those areas of study where the distinction between two or more sciences seems to diffuse at the interface of these sciences.

This new theory became the basis of many revelations in the scientific world today. As a matter of fact, many of these outbreaks wouldn't have been possible were it not for this concept that I like to refer to as 'the mingling' of the sciences.

As a chemist, I have always been quite fascinated by the nature of matter. However, what fascinated me even more was the fact that Chemistry was often being referred to as the 'central science' for its ability to connect Physics to other branches of science. Perhaps this is one of the major reasons I chose to research this particular topic, 'Phase transitions of a molecule as it is subject to different pressures and temperatures'.

Until a few decades back making such a study was rather difficult, timely and not of very high accuracy. That is mainly due to the fact that studying such molecules was only possible via two methods neither of which could enable us to monitor the molecule during the course of a chemical reaction. If the molecule had been studied experimentally, which is the first of the two methods, not only does the study require very sophisticated equipment, but also there is almost no way to closely monitor the changes that the system undergoes as the reaction proceeds. One can stop the system at any given moment and take the readings they may require, however what

exactly happens as bonds break, form or interchange and how and when the system goes from one phase to another, cannot be monitored.

The second method utilizes complicated mathematical equations to analyze a given system. A mathematical model is devised to find analytical solutions of the system enabling us to predict the behavior of such system given only certain parameters and initial conditions. Such mathematical equations, like the Schrodinger equation, which was used to solve the Hydrogen atom as well as other simple systems, are usually very complicated and very specific. Not all systems can be analyzed by a mathematical equation and not all systems have specifically constructed equations to be used in analytically solving them at any point in time.

Hence, neither the experimental nor the analytical methods proved *very* helpful after all. Although the areas where the two methods fail are vastly different, and although the analytical method may help on many levels, the results derived from both methods always seemed somehow *lacking*. Luckily, for all scientists wishing to explore the world of chemical reactions under different physical conditions, a third alternative has recently emerged.

Computer simulation has, during the past few decades, become the winning horse. As the study of computer science developed and as computer languages became much more comprehensible to non-specialized users, the utilization of computer simulations to study chemical systems has proven very successful. Computer simulations, as will be discussed, were able to blend the previous two methods into a method of higher accuracy, easier manipulation and more tangible results. It requires no sophisticated equipment, usually it requires less time than either of the other two methods, although it requires that a highly trusted computer program be devised.

Having access to such a method, it became possible to study a number of systems which could not have been studied otherwise. With the growing interest in the semiconductor processing industry a number of different molecules became particularly important. A number of organic tetrahedral molecules in are being heavily relied upon in different stages of the

processing of semiconductors. Studying the nature, structure and phase transitions of these molecules makes it possible to utilize them in the most useful way through setting the conditions which optimize their output. Tetrafluoromethane is one of these molecules whose application scope has been widely expanded over the past few decades.

Tetrafluoromethane, usually referred to as Carbon tetrafluoride is an organic tetrahedral halocarbon that is widely used in semiconductor processing as an etchant for a variety of films such as silicon, silicon oxide and silicon nitride. It is also an important precursor of the polymeric films. Furthermore, Carbon Tetrafluoride is utilized in the fabrication of microelectronic and optoelectronic devices. It is used in plasma etching in the gas phase and is also used in dry etching at different temperatures and pressures according to the etching technique being used.

This makes it very important to investigate the phase transitions and record the properties of Carbon Tetrafluoride as well as its structures at all possible phases. CF_4 has already been studied at temperatures between 100K and 300K both experimentally and via simulation methods. It has also been studied at pressures ranging from 0 GPa to 15 GPa. the structure and phase transitions of CF_4 at temperatures and pressures beyond these limits remain vague. Moreover, studying high pressure states probes the interatomic interactions and thus allows broader understanding of chemical systems.

That is why we have decided in this thesis to explore the structure and phase transitions of CF_4 within a large temperature range including areas previously studied as well as areas which have not been tackled before, particularly the temperatures from 0 K to 100 K and from 300 K to 400 K. Furthermore we also expanded our study to pressures higher than 15 GPa.

CHAPTER 1: COMPUTER SIMULATIONS

1.1 THE COMPUTER SIMULATION PROCESS

Computer simulations can be defined as 'a computer program or network of computers that attempts to simulate an abstract model of a particular system'.[1] They have become very functional in mathematical modeling of many natural systems in Chemistry, Physics, Biology as well as many other fields. Computer simulation models are able to solve the most sophisticated of systems taking into accounts all external factors, which were almost completely ignored by the analytical method, as well as all internal interactions, which are completely unresolved by the experimental method. The model can also be adjusted to account for certain errors which may crop up during the course of the simulation thus providing us with *output* results which are highly accurate.

We may be able to further describe the role of the simulation technique if we are able to relate it to other methods of research. Computer simulations can be thought of as an intermediary between theoretical and experimental studies in the sense that one uses the theory to design a valid model and then uses the experimental results to examine the validity of that model. Not only does the simulation approach provide for an alternative method of research, but it may also in some instances be the only usable method. Studies of certain systems under extreme conditions may not be easy if it were not for the simulation method. This method has, for example, made it possible to examine systems under very high/low temperatures and pressures.

The computer simulation approach can be summarized in the following series of steps [2,3]:

1. *Formulating the problem:* During this step, the system to be studied is chosen and relevant information is collected. All

aspects of the problem must be fully understood and accounted for.

2. *Data Collection:* This stage is very important since in order for the model to be accurate one needs to collect all the possible data about the theoretical background. Many aspects of the system to be examined should be made clear. The range of data required for a simulation shall be discussed in more detail at other points in this study.

3. *Tailoring a computer program:* During this stage, a relevant computer program is chosen (if one readily exists) or is tailored to fit the specific needs of the researcher and suit the specific criteria of the system.

4. *Constructing the model:* The models used in computer simulations are referred to as *symbolic models,* where properties of the actual real-life systems are represented by symbols like mathematical equations. At this stage, the model is constructed or built taking into account all internal and external aspects of the system. The system itself is regarded as a collection of entities which are logically related. Each of these entities is of interest to a particular application and must therefore be taken into account when building the model. Features of the system that are important when devising a model are [3]:

   a. *Environment and interdependency:* All external factors which affect the system.

   b. *Sub-systems:* The system itself can be broken down into subsystems the properties of each being of importance.

   c. *Organization:* The system is made up of highly organized elements or components which work closely together in order to ensure functionality of the system.

d. *Change:* The present state or condition of the system may vary with time especially when the system is subjected to certain factors such as the variation in temperature or pressure.

Relevant variables of the system under study are either *'uncontrollable variables*' which are givens and are not to be changed or manipulated, or *'controllable variables'* which are manipulated during the course of study in order to arrive at the solution. The choice of controllable and uncontrollable variables depends upon the system as well as the scope of study.

One must also identify the *'exogenous variables'* and the *'endogenous variables'.* The exogenous variables are those whose values are not affected during the course of the simulation while the endogenous variables are those determined by other variables during the course of the simulation.

5. *Model Validation:* At this stage a series of test runs are executed to ensure that the computer program as well as the stochastic model is fit for the purpose of the study. Usually this is carried out via using results attained from previous studies carried out analytically, experimentally or even by a different computer simulation program. The aim is to determine the degree to which the simulation model represents the real world or the part of the real world we are intending to study.

6. *Implementing the model:* Once validated the model can be implemented to study the characteristics of the system that we wish to explore. If the model proves to be indeed successful it can even be improved and utilized in researching other areas of the system. slight modifications of the model may give a wide range of possible research areas.

7. *Analyzing the solutions:* Output received from the simulation program are closely analyzed, studied and put in its final result format.

This flow chart is a representation of the steps adapted during a computer simulation [4,2].
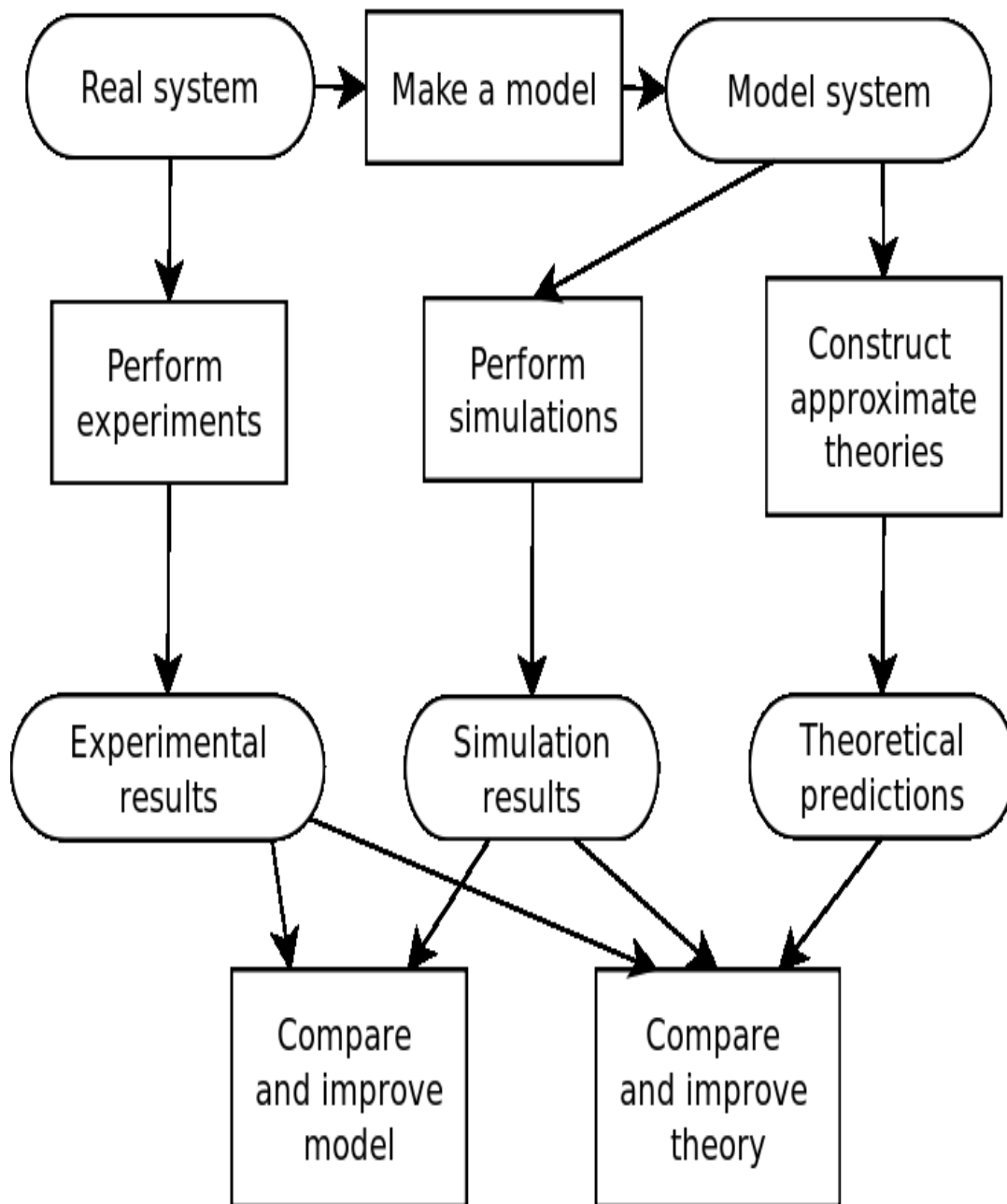
Figure (1.1): The steps adapted in a Molecular Dynamics simulation

1.2 COMPUTER SIMULATION TECHNIQUES

Over the past decades, and since the beginning of utilizing computer simulations in research, a number of developments have been introduced. Not only did a number of different techniques develop but also many developments have taken place within these various techniques. Perhaps the most widely used computer simulation techniques in the field of science are the Monte Carlo simulation techniques and the molecular dynamic simulation techniques.

The Monte Carlo technique in its basic form utilizes probabilities. It relies on repeated random sampling to calculate the results of the computational algorithm studied. It is usually referred to as a 'stochastic strategy' [5]. This particular technique generates a large number of microstates of the equilibrated system via the process of transforming from one microstate to the next within the statistical ensemble. During such process, the positions of the species as well as their orientations and conformations undergo a series of random changes. By averaging the quantities of interest over the microstates being examined one can reach the results being researched.

The Monte Carlo has a number of advantages, which have made it one of the most prominent techniques to date. Such advantages include the ease to extend it to simulate a large number of different ensembles, the flexibility one experiences when choosing the sampling functions and the underlying matrix or trial move, which must satisfy the principle of microscopic reversibility as well as timesaving as only the potential energy is required [6]. The Monte Carlo technique has many applications in computational physics, physical chemistry and other applied fields. The technique is used for statistical field theory calculations as well as for providing solutions to the many-body problem for quantum systems [7].

There is also the Brownian Dynamics technique which is best utilized for simulations involving large polymer molecules or colloidal particles [8,9].

However, the most detailed technique that can be used in molecular modeling is the Molecular Dynamics method. This method was first used by Alder and Wainwright to study the interactions of hard spheres [10,11]. It was introduced during the late 1950's from which point it has had many important breakthroughs concerning the behavior of simple liquids [12]. However, one of the major advances made via the use of this technique was carried out in 1964 by Rahman who was able to carry out the first computer simulation for liquid Argon using a 'realistic' potential [12]. This particular advance was the one which later led to a much more important breakthrough; the first molecular dynamics simulation of a realistic system. This study was performed by Rahman and Stillinger in their simulation of liquid water in 1974 [12]. Since then a large number of simulation techniques have evolved. There are now many specialized techniques for very specific problems, including 'hybrid' quantum mechanical - classical simulations.

Unlike the Monte Carlo technique, this method calculates in principle the real dynamics of the ensemble. Having done so, one can then calculate the properties of the ensemble within a specified time frame. The process by which this is performed can be described as a 'Cartoon Movie' process. A long series of consecutive simulation runs is carried out each giving solutions to the equation of motion at a certain instant in time. When integrated over a time limit it provides positions of the molecules at that time. In other words, Molecular Dynamics can be described a simulation of physical movements of atoms and molecules. The researcher allows the atoms and molecules to interact for a certain period of time, after which in most cases we are able to determine the trajectories of these atoms through numerical solutions of the equation of motion.

Molecular dynamics is perhaps the most detailed molecular simulation method. It proficiently evaluates various properties and dynamics of the system, which cannot otherwise be obtained by applying the Monte Carlo technique. This brings us to the important conclusion that this is a more useful tool, specifically in our field of study, than other available simulation techniques. In physics research MD simulations are used to examine and

explore certain dynamics of phenomenon on atomic level which would otherwise be extremely difficult to observe.

The method allows the prediction of the static and dynamic properties of molecules directly from the underling interactions between the molecules [12]. Molecular dynamics being a deterministic approach actually simulates the time evolution of the molecular system and provides the actual trajectory of the system. This is yet another aspect where molecular dynamics differs from the Monte Carlo technique. Whilst there is no direct relation between successive configurations of Monte Carlo simulations, molecular dynamics provides information about the time dependence of certain properties being examined. This makes it possible to predict these properties at any point of time whether in the future or the past. It also allows us to calculate time correlation functions during a molecular dynamics run. The information generated can be used to fully characterize the thermodynamic state of the system, which gives this technique a depth and versatility that no other technique possesses. For example, molecular dynamics simulations have a kinetic energy contribution to the total energy. The presence of the kinetic energy factor makes it possible to test the system for a total conservation of energy or estimate optimal velocities. In practice, the simulations are interrupted long before there is enough information to derive absolute values of thermodynamic functions; however the differences between thermodynamic functions corresponding to different states of the system are usually computed quite reliably[8].

Having mentioned all the above relative advantages of molecular dynamics, it was chosen as the technique we shall use in our study. Utilizing this method will make it easier for us to view the system at successive points in time and relate its state at one particular point in time to its state at another point in time thus monitoring the changes, the significance of these changes as well as their consequences. MD shall give us *dynamics*; where other methods would have only been able to provide us with an ensemble (smeared picture), MD is able to provide us with movie like results.

CHAPTER 2: MOLECULAR DYNAMICS

2.1 BASICS OF MOLECULAR DYNAMICS (MD)

Molecular Dynamics has provided the methodology for detailed microscopic modeling on a molecular scale. One cannot ignore the fact that the characteristics of matter can only be studied through the understanding of the nature of its constituents and also the dynamics of matter is contained in the solution of the N-body problem. It is only through computational tools that scientists can observe the movement and dynamics of individual molecules and it is best done through molecular dynamics. The theoretical basis of MD has been developed by a number of well known scientists in the analytical mechanics field such as Euler, Hamilton, Lagrange and Newton [13].

Normal equilibrium MD corresponds to microcanonical ensembles of statistical mechanics. However, when we are required to study certain properties at constant temperature and/or pressure the equation of motion is modified to produce the required system.

In its simplest form MD simulation is based on consecutive numerical solutions of the equation of motion (or their Newton-Euler equation of motion). The Newton equation of translational motion is related to force, which is given by [14]:

$$F_i = \sum_j \sum_b \sum_a f_{iajb} \qquad\qquad (2.1)$$

This is the force in a diatomic molecule of atom *a* in molecule *i* on atom *b* in molecule *j*.

For a simple atomic system Newton's equation of motion may be written as [14]:

$$m_i \ddot{\vec{r}}_i = -\frac{\partial U_{\text{pot}}}{\partial \vec{r}_i} = \vec{f}_i \,,$$

(2.2)

where m_i is the mass of the ith atom and f_i is the force acting on the atom, which in this case is assumed to be only due to interactions with other atoms. The force as we see from the above relation is derived from the potential energy of a complete set of 3N atomic coordinate system.

The potential energy is given by [14]:

$$U_{\text{pot}} = \sum_{i=1}^{N-1} \sum_{j>i}^{N} U(\vec{r}_{ij}) \quad, \quad \vec{r}_{ij} = \vec{r}_i - \vec{r}_j \,,$$

(2.3)

Therefore, we can arrive at another important relation for force relating it to the above equation:

$$\vec{f}_i = -\sum_{j(\neq i)} \partial U(r_{ij}) / \partial \vec{r}_i = \sum_{j(\neq i)} \vec{f}_{ij}$$

(2.4)

Thus, the total energy may be written as:

$$E = E_{\text{kin}} + U_{\text{pot}} = \sum_{i=1}^{N} \frac{1}{2} m_i \dot{\vec{r}}_i{}^2 + U_{\text{pot}}$$

(2.5)

Another important factor is torque, the tendency of force to rotate an object about an axis. The laws described here relate the motion of the center of mass of a rigid body to the sum of forces and torques acting on it. The torque may be defined as the turning force of an object. It is related to the center of mass of the molecule (given by $\mathbf{R}_i = \frac{1}{M_i} \sum_a m_{ia} \mathbf{r}_{ia}$) via the following equation:

$$\tau_i = \sum_a (r_{ia} - \mathbf{R}_i) \times \mathbf{f}_{ia}$$

(2.6)

The Euler equation for rotational motion is the one that utilizes the torque as per the following relation: $I_i \cdot \dot{\omega}_i - \omega_i \times I_i \cdot \omega_i = \tau_i$ (2.7)

where ω_i is the angular velocity and i, is the inertia tensor given by:

$$\mathbf{I}_i = \sum m_{ia}(p_{ia}^2 1 - \mathbf{p}_{ia}\mathbf{p}_{ia})$$ (2.8)

where $\mathbf{p}_{ia}$ is the atomic site co-ordinate relative to the center of mass of the molecule and given by $\mathbf{p}_{ia} = \mathbf{r}_{ia} - \mathbf{R}_i$.

2.2 CLASSICAL MOLECULAR DYNAMICS

As mentioned earlier, our interest is in the classical molecular dynamics. In that aspect we are focused on the dynamic behavior of the macro molecular systems. The description is fully microscopic in the sense that the unknowns are positions, $q_i \in \mathbb{R}^3$ momenta, $p_i \in \mathbb{R}^3$, velocities and similar quantities of all the atoms in the system [15]. All atomic trajectories supposedly obey the classical Hamiltonian equations of motion. The use of quantum theory in classical MD is limited to the construction of the interaction potentials in relation to the Born-Oppenheimer approximation. This potential is usually of very special structure. It is made up of a sum of atom-to-atom potentials resulting from various types of interactions between the atoms. Most of the interactions occurring within the molecule (binding forces representing bond structures) are of nonlinear vibrational nature. This means that most typical MD simulations have 'nonlinear highly-oscillatory behavior on multiple time scales in which the fastest vibrations have periods of about a few femtoseconds' [15]. This is one of the problems that may face us while using this technique and we shall fully focus on such aspects in later chapters.

Having mentioned the classical approach we must make it clear that there are three distinct cases to which one can apply the Newton-Euler

equations which we have illustrated above. It is a known fact that these equations describe the combined translational and rotational dynamics of a rigid body.

The first, and perhaps the simplest of cases to which we may apply the Newton-Euler equation, is the MD simulation of hard spheres. This is where we simulate the motion of N colliding particles according to the laws of elastic collision using event-driven simulations. Such simulations are commonly applied to predict properties of physical systems at the atomic level. This includes motion of gaseous molecules, dynamics of chemical reactions, atomic diffusion, sphere packing, phase transitions and front propagation as well as many other domains which involve modeling of particle systems. In this type of system the model is 'idealized'. The particles interact via elastic collisions with each other and with the reflecting boundary. No other forces are exerted, which means that the particles travel in straight lines at constant speed between collisions.

This system is governed by the laws of conservation of energy and linear momentum and the simulation of such system is carried out via one of two techniques. The first technique is referred to as a time-driven simulation. It discretizes time into quanta of size *dt*; the position of the particle is monitored via observations after every *dt* unit of time and the velocities are updated. The second technique is referred to as an event-driven simulation where the focus is on these points in time during which interesting events occur. When this system is simulated we can determine the next pair of spheres to collide, calculate when the collision will occur, calculate the positions of all spheres at the collision time. Furthermore, we can determine the new velocities of the two colliding spheres after collision by applying the principle of conservation of linear momentum.

The second situation where the Newton-Euler equations may be applied is one where the particle is subject to a constant external force between collisions [16]. This is the case where a charged particle moves within a constant uniform electric field. Simulations of such ionic systems

would require inclusion of long-range electrostatic forces. The particle feels a force that is independent of its velocity (perpendicular to its velocity) and hence the path of the particle changes and so does the direction and magnitude of the velocity. The force acting on the particle may be donated by:

$$F_i = q_i \mathbf{E} \qquad (2.9)$$

where q_i is the charge of the particle and $\mathbf{E}$ is the electric field.

The third and definitely most complicated application of the Newton-Euler equations is in situations where there is a dependence on the relative position of the particle. In other words, the force acting on one particle at any point in time depends on the position of that particular particle in relation to the other particles constituting the system. In such case the force is described the first partial derivative of the potential energy function as given by the second half of equation (2.2) as follows [14]:

$$\mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} \mathcal{U}(\mathbf{r}^N) \qquad (2.10)$$

The potential U defines a force $\mathbf{f}$ at every point $\mathbf{r}$ in space, and hence the set of forces is usually referred to as a force field.

Due to the complexity and unpredictability of particle motion, the above mentioned situation is almost impossible to study analytically. That is why this situation is the one where molecular dynamics is most useful. The first two situations can be studied analytically although it requires a lot of efforts and certain conditions to abide by. The third situation however can only, in most cases, be studied via the application of MD. This particular type of problem shall be fully analyzed in following chapters.

2.3 MOLECULAR DYNAMICS WITH CONTINUOUS POTENTIALS

As we briefly stated earlier, when dealing with realistic (less idealized) systems, the force is dependent on positions, not only of the particle monitored, but also of all the particles with which it interacts. That is every time the position of the particle 'i' changes the force on it shall change, and also every time the position of any of the particles with which this particle 'i' interacts changes, the force on particle 'i' shall change. The situation becomes a many-body-system rather than a one-body-system, and in order to be able to fully analyze it we must study the continuous potential energy function. Under the conditions of this continuous potential the many-body-problem is tackled in the normal manner.

The potential functions for non-bonded energies are obtained by summing over the interactions between particles. The potential energy in that case can be calculated from the sum of contributions of energies between atom pairs. The Lennard-Jones potential is an example of such non-bonded potential [16]. However, in many-body potentials the potential energy does not only depend on the interacting pair. It includes contributions and effects of three or more particles. The potential energy cannot be simply found by a sum over pairs of atoms, since these interactions are calculated 'explicitly' as a combination of higher-order terms. The equations of motion for this particular type of problem cannot be solved by normal methods and hence there are certain techniques that have been especially developed for obtaining solutions for the many-body problem. One of these methods is referred to as the finite difference method, and it is a form of integration that is utilized to solve the equations of motion [17,18]. In its essential form this method provides approximations of the solutions of differential equations using finite difference equations to approximate derivatives.

2.3.1 The Finite Difference Method

As we have explicitly discussed before, when carrying out a molecular dynamics simulation one is trying to arrive at certain results, which in many cases involves the prediction of the positions and velocities of

particles at any point in time. For a many-body system the finite difference technique is utilized to illustrate trajectories with continuous potential models. We shall assume that these trajectories are pair wise additive. The integration of the Newton-Euler equation is broken down into a number of consecutive stages which are separated from each other in time by a certain time interval δt. The total force on any one particle at any time t is calculated using vector sums of the particle's interactions with other particles in the ensemble. Having arrived at a value for the force, which is assumed to be fixed during a time step, one can now calculate the acceleration. Now that we have the positions, velocities and accelerations of the particles at time t, we can use them to calculate the positions and velocities of the particles at time $t+\delta t$. Now the forces and accelerations of the particles at their new positions are calculated, from which we can determine the positions and velocities of the particles at time $t + 2\delta t$. The process is a repetitive pattern which lasts as long as the determined number of time steps. The diagram below describes how the finite difference method relies on discretizing a certain function [19].
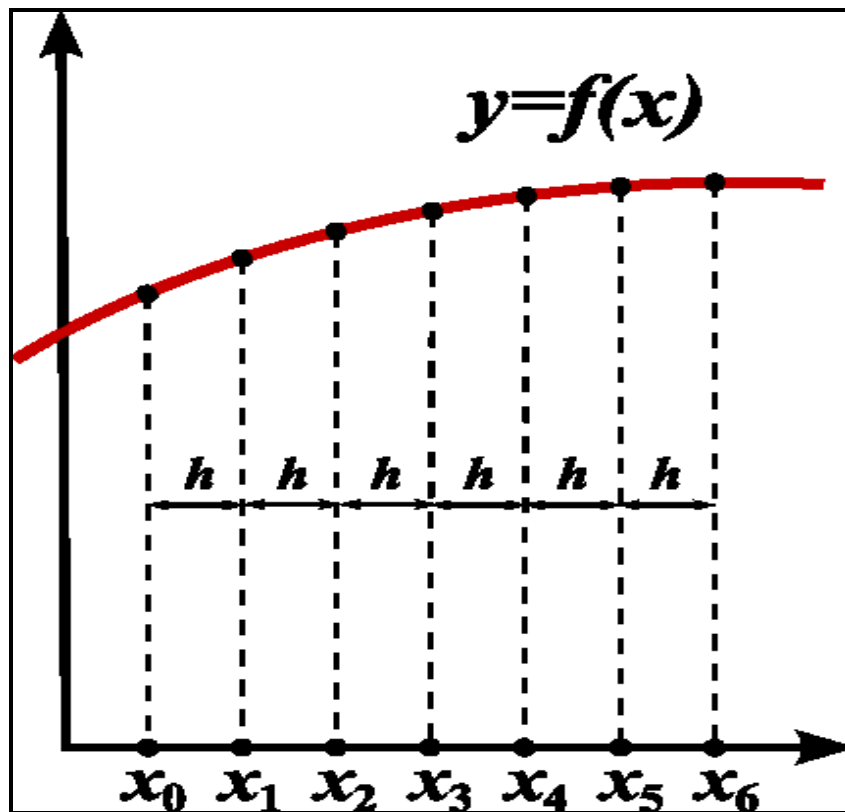
Figure (2.1): The finite difference method relies on discretizing a function on a grid

The first question that comes to one's mind at this point is how we can approximate the derivatives of a known function by finite difference formulas based only on values of the function itself at discrete points. Many algorithms for integrating the equations of motion using the finite difference method have been developed over the years among which some are very often used in MD calculations. The core of these algorithms is the assumption that all dynamic properties of the system can be approximated by utilizing Taylor series expansions. Taylor series expansions of the basic dynamic quantities of a system (position, velocity, acceleration...etc) are shown below [20]:

$$r(t + \delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \frac{1}{24}\delta t^4 c(t) + \dots \qquad (2.11)$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t) + \delta t \mathbf{a}(t) + \frac{1}{2}\delta t^2 \mathbf{b}(t) + \frac{1}{6}\delta t^3 \mathbf{c}(t) + \dots \qquad (2.12)$$

$$\mathbf{a}(t + \delta t) = \mathbf{a}(t) + \delta t b(t) + \frac{1}{2}\delta t^2 c(t) + \dots \qquad (2.13)$$

$$b(t + \delta t) = b(t) + \delta t c(t) + \dots \qquad (2.14)$$

The first derivative of position with respect to time gives the velocity (v), the second derivative with respect to time gives the acceleration (a) and the third and fourth derivatives of position with respect to time are denoted by (b) and (c) respectively.

Having established this point so far, we now have to choose the most suitable integration method to work with.

2.3.2 Different Integration Algorithms

There is a large number of integration methods with which we may proceed to solve the above problem. All of these techniques are sufficient; however some of them require more efforts than others. When choosing the most suitable technique one must take a number of factors into account. First and foremost we must evaluate the computational effort required for that

technique. We must also take the cost effectiveness of the algorithm into consideration. The energy conservation factor is also of great importance and it can be easily established via calculating the root-mean-square fluctuation, which is a statistical measure of the magnitude of a varying quantity, and plotting it against time step. Different algorithms vary in the rate at which the error varies with the time step [16]. The method should also conserve momentum, be time reversible and permit the usage of long time steps. The memory required and the synchronization of positions and velocities are also important factors to be considered when choosing the most suitable algorithm as well as how easy it is to program. It is also important that the algorithm be able to give accurate results (as precise as computationally possible), i.e. same results as an exact analytical trajectory.

Among the Integration techniques which have proved very successful are [7,20,21,22,23,24,27]:

1. **The Verlet algorithm [21]:** This method is the most commonly used in MD. It uses two sets of coordinates namely the positions and accelerations at time t and the accelerations from the previous run $r(t - \delta t)$ to calculate the positions at $(t + \delta t)$ and $r(t + \delta t)$. By addition of simple simultaneous equations we arrive at the relation:

   $r(t + \delta t)=2r(t)- r(t - \delta t)+ \delta t^{2} \boldsymbol{a}(t)$ \hfill (2.15)

   where r is the position and $\boldsymbol{a}$ is the acceleration. The equation does not give the velocities, however they can be calculated using a straight forward relation of the form:

   $\boldsymbol{v}(t) = [r(t+ \delta t)-r(t- \delta t)]/2 \delta t$ \hfill (2.16)

   As we may see, the Verlet algorithm is a direct approach which has proven very successful especially with its modest storage requirements. However it has also proven to have a lower degree of precision due to the fact that the components being added together while calculating positions are not of the same magnitude. That is where we add two relatively large

quantities $2r(t)$ and $r(t-\delta t)$, to a relatively smaller quantity $(\delta t^2 a(t))$. This, as well as the fact that there is no direct way to obtain the velocities is among the major disadvantages of this method. One other disadvantage is that the Verlet technique is not a self starting method. That is acquiring new positions is dependent on current and previous positions.

2. **The Leap-Frog algorithm [24]:** this second technique is a variation of the Verlet algorithm. This method calculates velocities $v(t+1/2\delta t)$ from the velocities calculated at time $(t-1/2\delta t)$ and acceleration at time t. From these calculations, the positions are obtained. That is the velocity at time t is calculated by the relation:

$$v(t)=1/2[v(t+1/2\delta t)+v(t-1/2\delta t)] \qquad (2.17)$$

where:

$$v(t+1/2\delta t)=v(t-1/2\delta t)+\delta t\, a(t) \qquad (2.18)$$

And positions are given by:

$$r(t+\delta t)=r(t)+\delta t\, v(t+1/2\delta t) \qquad (2.19)$$

As we can see this method is straightforward. Unlike the Verlet Algorithm, the leap-frog algorithm explicitly includes the velocities and it does not have calculations including quantities of highly varying magnitudes. However, it too has its disadvantages the most obvious of which is that the positions and velocities are not synchronized which inhibits the ability to acquire a number of important quantities like the Kinetic Energy contribution to total energy.

3. **The Velocity Verlet Method [12,22,24]:** This algorithm is a corrected version of the Verlet technique which allows acquiring positions, velocities and accelerations and at the same time has a higher degree of precision. This method uses the following equations:

$$r(t+\delta t) = r(t) + \delta t\, v(t) + 1/2\, \delta t\, 2a(t) \qquad\qquad (2.20)$$

$$v(t+\delta t) = v(t) + 1/2\, \delta t\, [a(t) + a(t+\delta t)] \qquad\qquad (2.21)$$

4. **The Beeman Integration Scheme [12,21]:** This is one of the more accurate techniques available. It was designed to allow for high numbers of particles in MD simulations. The expression it gives for velocity does not allow for a high degree of error and thus is highly proficient. Having a more accurate description of the velocity allows for better conservation of energy since Kinetic energy is obtained directly rather than through assumptions or estimations. The technique is rather expensive however it is cost effective. There are two variations of the method, the first of which was introduced in 1973 [21] is known as Beeman's method and as mentioned above it is a direct variation of the Verlet method that uses different formulas for velocities. The second variation is one that was introduced later in 1976. It is the direct variant of the third order method in the predictor corrector technique. The Beeman algorithm is related to the Verlet method and uses the following equations [23,24,28]:

$$r(t+\delta t) = r(t) + \delta t v(t) + 2/3 \delta t^2 a(t) - 1/6\, \delta t^2 a(t-\delta t) \qquad\qquad (2.22)$$

$$v(t+\delta t) = v(t) + 1/3 \delta t a(t+\delta t) + 5/6 \delta t a(t) - 1/6\, \delta t a(t-a) \qquad\qquad (2.23)$$

This integration scheme is widely used today with its predictor corrector variation to attain the best, most accurate results. The more complex calculations and equations make this algorithm more expensive than others.

2.3.3 The Predictor Corrector Integration Method:

This method is a more advanced family of algorithms which was developed at a later stage than the Verlet variations. In its core, it somehow depends on the various Verlet techniques, however it has introduced the ability to choose a method that is correct to a given order. As suggested from the name, this algorithm predicts the degree of error in its calculations and is able to provide means of correcting them. It is a highly systematic method

which in its core relies on first calculating a rough approximation of the desired quantity, this is the prediction step. The correction step comes next where the initial approximation is refined using other means. By definition the predictor corrector methods 'proceed by extrapolating a polynomial fit to the derivative from the previous points to the new point (predictor step), then using this to interpolate the derivative (the corrector step)' [25]. In MD, the first step of the method is one where new positions, velocities, accelerations and any other terms of higher order are predicted via the use of a regular Taylor expansion as previously discussed. The second step involves evaluating the forces at the new positions using the potential energies and then deriving the new accelerations $a(t+\delta t)$. Having evaluated these velocities, the next step is to compare these values with those predicted from the Taylor series expansion, $a^c(t+\delta t)$ and then the difference (which is assumed to be the error) is applied, as a corrector, to positions, velocities and other calculated amounts.

The correction stage proceeds in the following order [26,27,30].

$$\Delta a(t+\delta t)=a^c(t+\delta t)- a(t+\delta t) \qquad\qquad (2.24)$$

THEN,

$$r^c(t+\delta t)=r(t+\delta t)+c_0\Delta a(t+\delta t) \qquad\qquad (2.25)$$

$$v^c(t+\delta t)=v(t+\delta t)+c_1\Delta a(t+\delta t) \qquad\qquad (2.26)$$

$$a^c(t+\delta t)/2=a(t+\delta t)/2+c_2\Delta a(t+\delta t) \qquad\qquad (2.27)$$

$$b^c(t+\delta t)/6=b(t+\delta t)/6+c_3\Delta a(t+\delta t) \qquad\qquad (2.28)$$

The best values for the coefficients c_0, c_1, c_2, c_3...etc depend on the order of the Taylor series expansion and hence vary from one situation to the next [29,32].

The Predictor Corrector method is a modification that can be applied to any of the integration techniques discussed earlier. In our research, as shall be discussed in more detail in following chapters, we used the modified

version of the Beeman integration scheme. This particular procedure which was proposed in 1976 by Beeman [30,31] is one of the most accurate algorithms in the Verlet family for calculating velocities. The Beeman algorithm which is based on equations (2.22) and (2.23) ensures that calculations of velocity and position are synchronized. First it evaluates the new positions using the current positions, velocities and accelerations as well as previous accelerations. Then the forces are reevaluated on basis of the new positions. The forces are then used to update the velocities and accelerations. As we see, like most other Verlet methods, this algorithm is not self starting.

The Modified Beeman's Algorithm may only be derived for a fixed time step. It was found to give the most stable computations. The modified Beeman's algorithm calculates velocities and forces self-consistently. In the equations of the modified Beeman algorithm the symbol x is used to denote any dynamic variable (including but not exclusive to: MD cell edge, centre-of-mass co-ordinate, quaternion etc). The symbol $x^{\bullet(p)}$ represents the 'predicted' velocities and $x^{\bullet(c)}$ represents the 'corrected' velocities. The algorithm may be described in the below stated series of equations [31]:

1) $\quad x(t+\delta t) \quad = x(t) + \delta t\,\dot{x}(t) + \dfrac{\delta t^2}{6}[4\,\ddot{x}(t) - \ddot{x} - \delta t]$ \hfill (2.29)

2) $\quad \dot{x}^{(p)}(t+\delta t) = \dot{x}(t) + \dfrac{\delta t}{2}[3\,\ddot{x}(t) - \ddot{x}(t-\delta t)]$ \hfill (2.30)

3) $\quad \ddot{x}(t+\delta t) \quad = F(\{x_i(t+\delta t), \dot{x}_i^{(p)}(t+\delta t)\}, i = 1...\text{n})$ \hfill (2.31)

4) $\quad \dot{x}^{(c)}(t+\delta t) = \dot{x}(t) + \dfrac{\delta t}{6}[2\,\ddot{x}(t+\delta t) + 5\,\ddot{x}(t) - \ddot{x}(t-\delta t)]$ \hfill (2.32)

The final step is to replace $\dot{x}^{(p)}$ with $\dot{x}^{(c)}$ and go to step 2 until it converges. That is we should apply the predictor corrector method to steps 2 till the end of the cycle until the predicted and corrected velocities converge and the error (discrepancy) becomes

minimal. This should not require more than three of four cycles if the initial configuration and conditions are accurate.

Now that we have calculated the velocities and positions, which in this method are dependent on accelerations at time *(t-δt)* and *(t+δt)*, we can deal with the translational aspect of the molecular motion. The following step it to start dealing with the rotational motion aspect which is also of great importance in understanding the dynamics of the system.

Perhaps one of the most commonly used notations for representing orientations and rotational motion of objects in three dimensions is unit quaternions [33]. Quaternions, also referred to as versors, are used to represent an orientation or a certain rotation relative to a reference position. Quaternions provide a simple method to represent such space rotations via the encoding of the axis vector and the angle of rotation which are the components used in the representation. The quaternions encode these axis-angle components in four numbers followed by the application of the corresponding rotation to a position vector that represents a certain point relative to the origin. Unit quaternions are given by [34]:

$$\mathbf{q} = \begin{bmatrix} q_0 & q_1 & q_2 & q_3 \end{bmatrix}^T \tag{2.33}$$

$$|\mathbf{q}|^2 = q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1 \tag{2.34}$$

The quaternions are related to rotations around an axis by the following equations [35]:

$$\mathbf{q_0} = \cos(\alpha/2) \tag{2.35}$$

$$\mathbf{q_1} = \sin(\alpha/2)\cos(\beta_x) \tag{2.36}$$

$$\mathbf{q_2} = \sin(\alpha/2)\cos(\beta_y) \tag{2.37}$$

$$\mathbf{q_3} = \sin(\alpha/2)\cos(\beta_z) \tag{2.38}$$

where α is a rotational angle and $\cos(\beta_x)$, $\cos(\beta_y)$ and $\cos(\beta_z)$ are directional cosines related to Euler's method.

Compared to Euler angles, not only are they easier to compose and simpler algebraically, but they also help arrive at equations of motion which have no singularities. This is an important factor when running an MD simulation since it helps avoid energy and force overflow during the simulation. The same factors apply when comparing quaternions to rotation matrices which may be less numerically stable. There are a number of algorithms and simple computational codes which help convert quaternions to Euler angles format. The following equations relate quaternions to Euler angles [36]:

$$q_0 = \cos(\frac{\varphi + \psi}{2}) \cdot \cos(\frac{\theta}{2}) \qquad (2.39)$$

$$q_1 = \sin(\frac{\varphi - \psi}{2}) \cdot \sin(\frac{\theta}{2}) \qquad (2.40)$$

$$q_2 = \cos(\frac{\varphi - \psi}{2}) \cdot \sin(\frac{\theta}{2}) \qquad (2.41)$$

$$q_3 = \sin(\frac{\varphi + \psi}{2}) \cdot \cos(\frac{\theta}{2}) \qquad (2.42)$$

Given that the vector form may be expressed as follows:

$$q \equiv (q_0, q_1, q_2, q_3) \qquad (2.43)$$

Furthermore, Pawley has defined specific algebraic methods to solve quaternions [35,36]. For example, when multiplying quaternions we deal with each as a four element vector and use a matrix product technique to come up with the solution. So if we have two quaternions denoted by q and r, then the four-vector format of each shall be $q \equiv (q_0; q_1; q_2; q_3)$ and $r \equiv (r_0; r_1; r_2; r_3)$, then their product may be expressed as follows [35]:

$$qr = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{pmatrix} \begin{pmatrix} r_0 \\ r_1 \\ r_2 \\ r_3 \end{pmatrix} \tag{2.44}$$

We may also define the conjugate of the quaternion as $q* = (q_0', q_1', q_2', q_3')$ and the norm $|q| \equiv \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}$

Having established this, and since our work requires that we find the time derivative of the quaternions and weave it into the equation of motion there are a number of important relations we need to understand as follows[37,38]:

$$2\dot{q} = q(0, \omega^p) \tag{2.45}$$

The above equation establishes a relation between the time derivative of a quaternion and the principal frame angular velocity (ω^p). The second derivative is for the above equation is given by[39,41]:

$$2\ddot{q} = q(\frac{-1}{2 \cdot (\omega^p)^2}, \dot{\omega}^p) \tag{2.46}$$

$$= q(-2|\dot{q}|^2, \dot{\omega}^p) \tag{2.47}$$

Given the relations above we can now utilize the quaternion and their derivatives as dynamic variables for the rotational aspect of motion. The components of the quaternion shall be treated independently and solved via the regular algorithm techniques we described previously.

2.4 LIMITATIONS AND CONSTRAINTS OF MD SIMULATIONS:

In order to be able to make proper use of MD simulations, we should be well aware of its limitations. There are a number of limitations which are of more importance than the rest. These limitations are [40]:

1. **Quantum Effects:** Certain dynamical events usually involve quantum effects such as changes in the bonding structure, tunneling and presence of certain intermediaries. These effects cannot be accounted for by straightforward force field simulations. It can be solved via the use of quantum or *ab initio* MD.

2. **Reliability of interatomic potentials:** Results of the simulation will only be realistic if potential energy function is an imitation of the forces experienced by the real atoms. Therefore designing a good force field is a very tricky aspect of the simulation.

3. **Time limitations:** Perhaps the most severe problem of MD simulations is the time limitation. The time scale for some chemical processes may extend over many orders of magnitude. There are always limitations on the time scales that can be investigated. Usually simulation runs are short (in the range of nanoseconds). In a very few cases it may extend to microsecond ranges. Wrong choice of the time step has negative consequences on the results of the simulation runs. Therefore one must test whether or not the simulation has reached equilibrium before we can fully rely on the averages calculated from that run.

The length of the run depends on the size of the system as well as its physical properties. There are no fixed rules for choosing the most suitable time step. If the time step chosen is too small the trajectory will only cover a limited proportion of the phase space. If it is too large it may give rise to high energy overlaps between atoms and hence may result in instabilities in the integration algorithm. This shall increase the error and thus give inaccurate result which has discrepancies with the analytical results since the energy and linear momentum conservation would be violated. Furthermore if the violation is too high it may result in a total failure of the program resulting from numerical overflow. The diagram below shows an example of the two extremes. If the time step is very small (as on the left) the phase space is covered rather slowly. If the time step is too large (as in the middle) this gives rise to instabilities. With the

appropriate time step (as on the right) the phase space is efficiently covered which helps increase the efficiency of the simulation run and thus provide more accurate results [42].
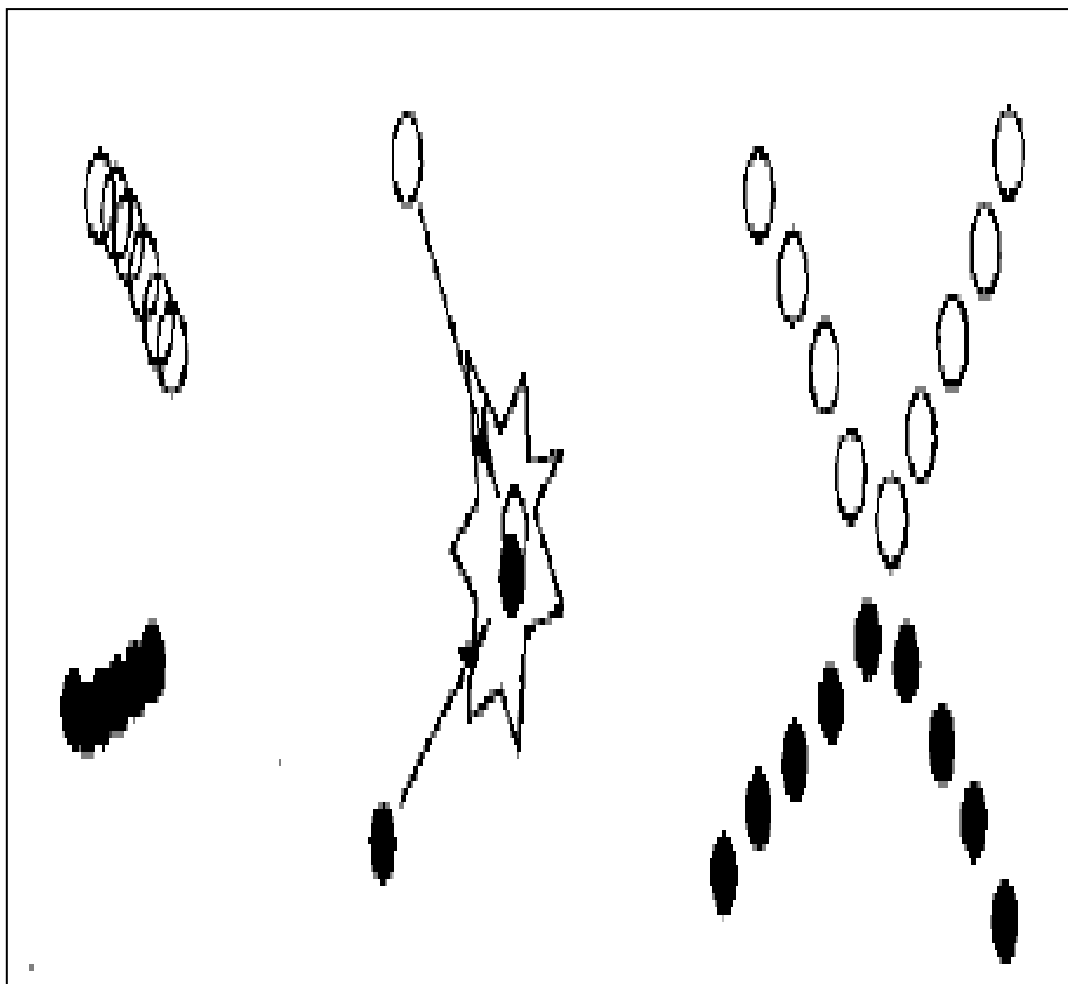
Figure (2.2): A very small time step (left) phase is covered very slowly; a large time step (middle) gives instabilities. An appropriate time step (right) phase space is covered efficiently and collisions occur smoothly.

In most MD problems, it is very important to choose a time step that is well balanced with the correct trajectory. That is for a small time step much more computer time shall be needed, and for a rather large time step we give rise to the largest errors to arise, furthermore a large time step can result in blowing up the trajectory (taking it out of proportion). This has placed a very severe limitation on the time step: the time step must be approximately one order of magnitude smaller than the shortest motion. Furthermore experience has

given us researchers a chance to estimate appropriate time steps for certain systems. The table below briefs this point [42]:

System	Types of motion	Suggested time step (seconds)
Atoms	Translation	10^{-14}
Rigid Molecules	Translation and rotation	5×10^{-15}
Flexible molecules, rigid bonds	Translation, rotation, torsion	2×10^{-15}
Flexible molecules, flexible bonds	Translation, rotation, torsion, vibration	10^{-15} or 5×10^{-16}

Table (2.1): Suggested time steps for different types in motion present in various systems.

4. **Surface Effects:** The system we are dealing with, particularly if it is in the gaseous or liquid phase, shall in most cases be contained in any sort of container which itself is made up of particles. This means that it is really difficult to isolate the system we are studying from interactions with the particles in the walls of the container. This problem presents itself in a more obvious way when the number of molecules used in the simulation is small.

5. **Size Limitations:** The rules which apply to choosing the time step also apply to choosing the size of the MD ensemble. Furthermore, the length of the time step places certain limitations on the size of the problem. For large systems the evaluation of forces is computationally more challenging and therefore each integration step takes a longer time period. The simulation box size is of as much importance as the simulation time step. In short, the MD box size must be large enough to avoid boundary

condition effects. Boundary conditions are usually used when the sample size is small and the surface effects are not of particular interest. The size of the box must be large enough in order to prevent periodic artifacts from occurring as a result of unphysical topology of the simulation. If the box is too small, a macromolecule may interact with its own image in a neighboring box which results in highly unphysical dynamics in most macromolecules. The appropriate box size in relation to the size of the molecules depends on a number of factors, namely the length of the simulation (time step), the desired accuracy as well as the anticipated dynamics.

A number of these limitations, particularly the last two, can be avoided (to a certain extent) via the use of a technique referred to as the Periodic Boundary Conditions.

2.4.1 Periodic Boundary Conditions:

A simple definition of Periodic Boundary Conditions (PBC) is that it is a tool which makes it possible for MD simulations to use a relatively small number of molecules to study the properties and behavior of the overall system. In any MD box we may describe it a fast substitution system which leaves the number of molecules within the unit cell constant. That is the moment one molecule moves out of the central cell another molecule instantly moves in its position from the opposite side of a second cell. The total number of molecules inside the central cell remains constant.

So in short, periodic boundary conditions are those boundary conditions that are' often used to simulate a large system by modeling a small part that is far from its edge' [12,44]. In a more up-to-date sense, these conditions represent topologies of some video games. In practice, a unit cell or simulation box is set to be suitable for three dimensional tiling. When a certain object passes through one face of the unit cell, it is reflected on the opposite face with the same velocity. The atoms in a computational unit cell are replicated throughout the space to form an infinite lattice. The idea of PBC can be best illustrated by the diagram below [43]:
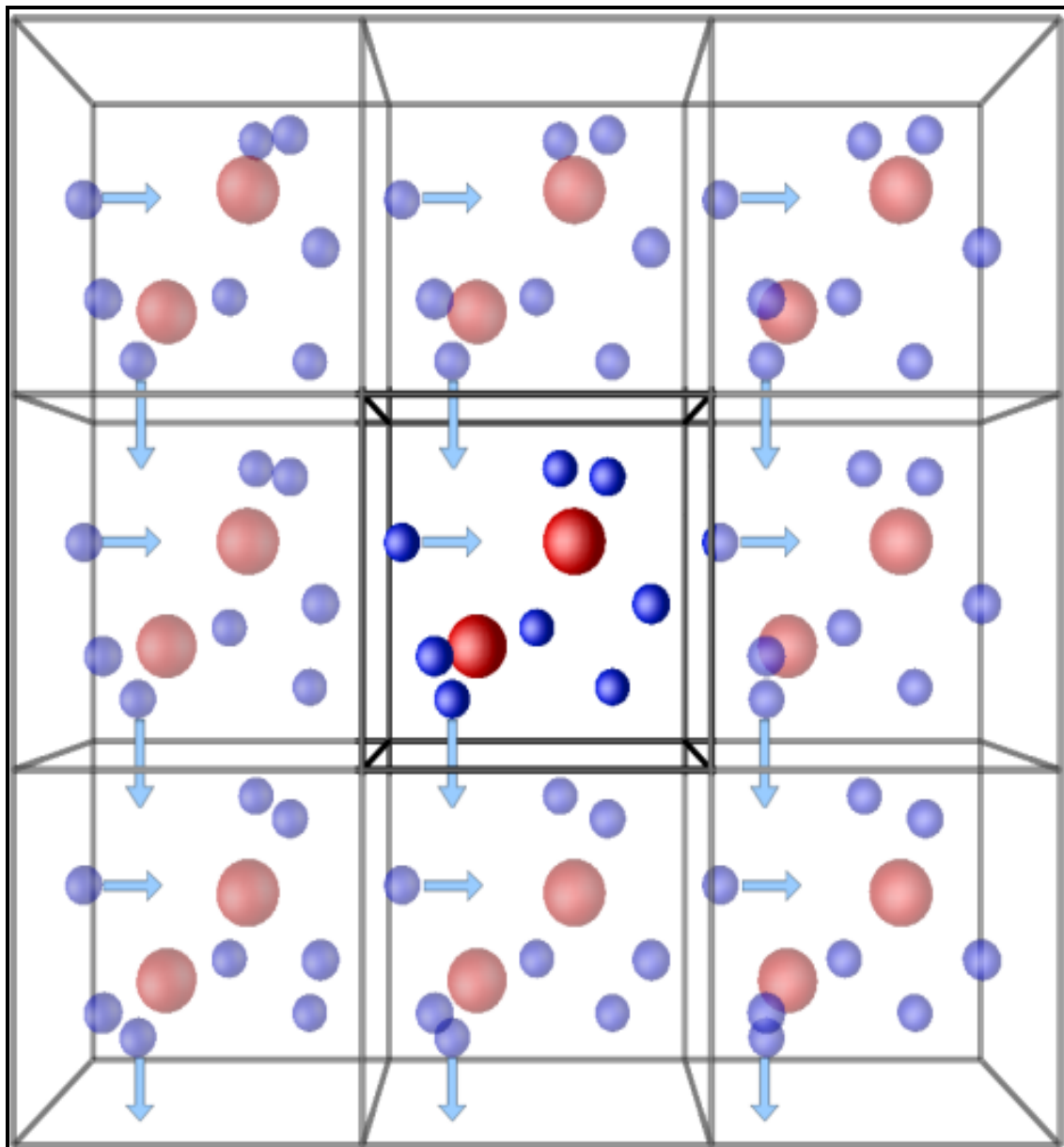
Figure (2.3): Schematic representation of the idea of periodic boundary conditions

The periodic boundary conditions are usually applied to simulate bulk gases, liquids, crystals or mixtures. They are particularly useful for simulating parts of a bulk system with no surface present. In many cases it may be used with Ewald summation methods to account for electrostatic forces in the system.

Periodic boundary conditions are also used to truncate and modify the strain field arising from inhomogenuity of solid systems. In ionic systems on the other hand it may be used in conjunction with other factors to avoid summing to an infinite charge, i.e. the net electrostatic charge must be zero. Under the PBC, linear momentum is conserved whilst angular momentum is not. In certain cases it is not best to use the PBC since it may alter the sampling of the simulations. Such cases are when we are dealing with microcanonical ensembles where particle number, volume and energy are constant, also referred to as the MD ensemble. In such case PBC shall alter the sampling due to the conservation of linear momentum as well as the position of the center of mass. The conservation of such quantities introduce certain artifacts to a number of important factors such as the statistical mechanical definition of temperature, the departure of velocity from its mean Boltzmann distribution as well as the violation of the equipartition for systems containing heterogeneous masses.

Periodic boundary conditions also require the unit cell to be in shape to be perfectly tiled into a 3D crystal. This makes the cubic or rectangular prism most intuitive and a highly common choice along with the hexagonal prism, the elongated dodecahedron, the rhombic dodecahedron and the truncated octahedron. On the other hand it makes spherical and elliptical droplets very difficult to use. However, in certain cases non-cubic shapes are used for example when trying to eliminate the influence of the cubic symmetry on a shape of crystal nucleus in certain liquids. The figure below shows, a rhombic dodecahedron, an elongated dodecahedron, a hexagonal prism, a truncated octahedron and a Parapellepiped unit cell [45,46,47].
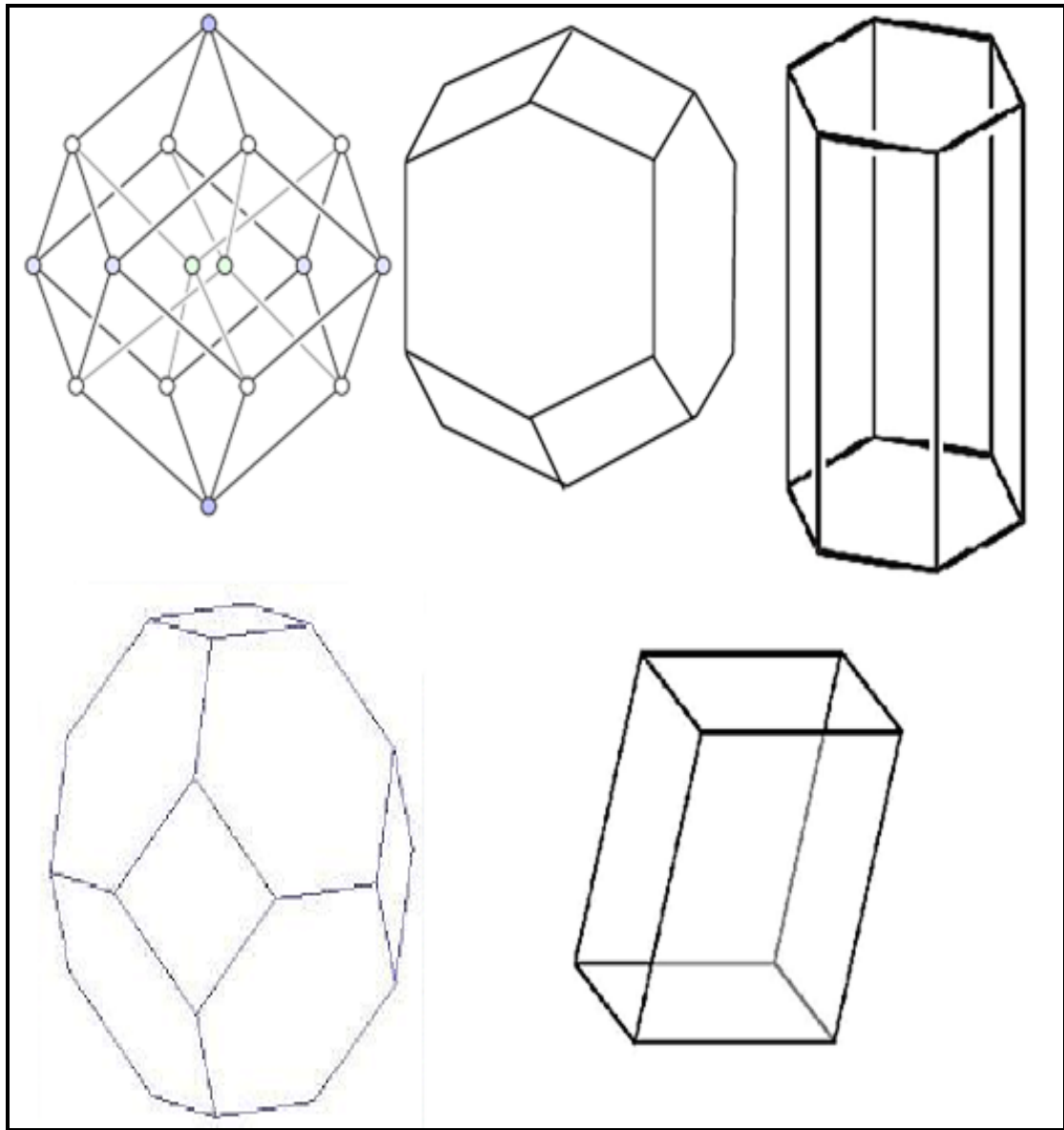
Figure (2.4):Different periodic cells used in computer simulations. From the left (top) a rhombic dodecahedron, an elongated dodecahedron, a hexagonal prism, (bottom) a truncated octahedron and a Parapellepiped unit cell

2.5 TRUNCATING POTENTIALS

Perhaps the most time-consuming part of a simulation is the calculation of the non-bonded forces and energies at every time step. The time required for this part of the simulation run for a pair wise potential energy model is directly proportional to the square of the number of molecules in the ensemble.

There are a number of methods that can be used to reduce such effects. When calculating the potential energy of a system with a certain configuration we first assume that the potential is pair wise additive whilst the effect of three-body or higher level many body interactions are ignored. Let's assume we are studying the system in the following diagram [48]:
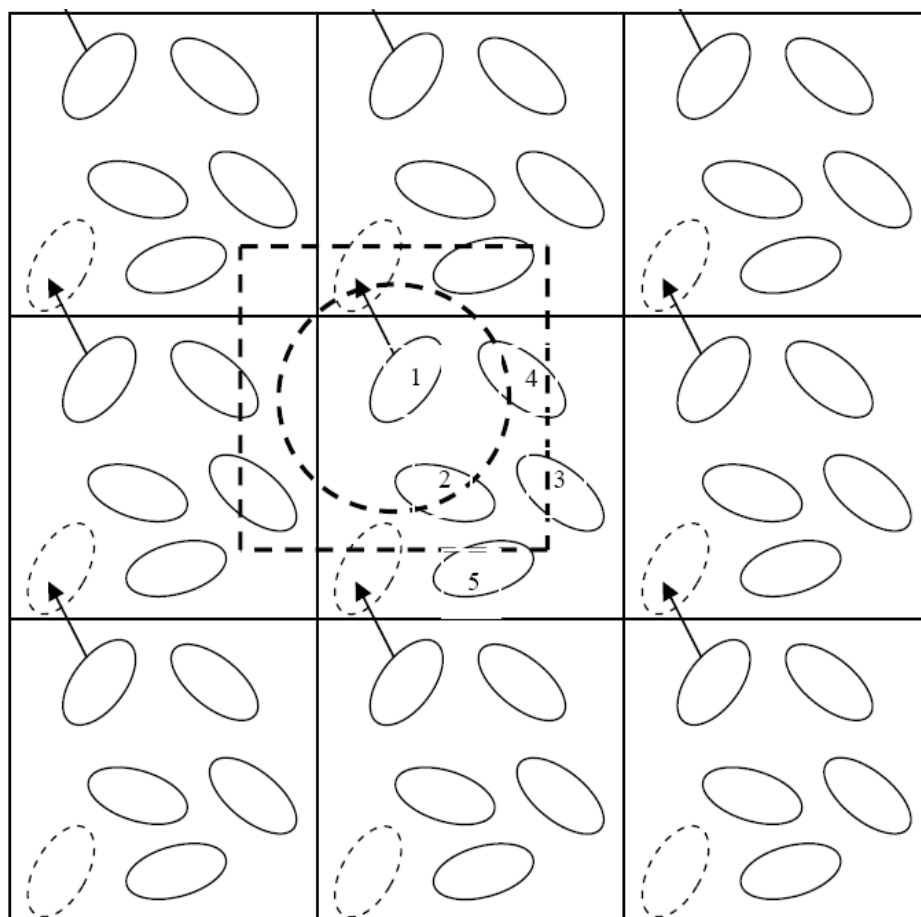
Figure (2.5): Periodic boundary conditions: a 2D periodic lattice where the unit cell contains five particles

In order to calculate the force and the potential energy involving molecule '1'we shall follow our previous assumption that the potential energy is pair wise additive and ignore other interactions. Now we start by taking the interactions of molecule '1' with all other molecules in the MD box. While proceeding to add these interactions we will find that we must take into account the interactions of molecule '1' and the images in the replica boxes. This will lead to an infinite number of interactions which are impossible to calculate in practice. Let's take the Lennard-Jones potential as an example. The L-J potential is a 'mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules'.[49,50] It is most commonly expressed by the following relations [51]:

$$V_{LJ} = 4\varepsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]$$
$$= \varepsilon \left[\left(\frac{r_m}{r}\right)^{12} - 2\left(\frac{r_m}{r}\right)^{6} \right] \qquad (2.48)$$

This particular potential is not the most accurate, it is however one of the simplest. That is why it is most commonly used in computer simulations. As previously discussed it is practically impossible to calculate the desired quantities due to the infinite number of interactions present. This gave rise to the truncated potentials. This method was established to save computational time. The idea of truncation is based on the fact that foe short range potentials, main contributions to energy and force are from nearby particles. We can therefore choose to truncate the potential at some certain distance which we refer to as the cutoff radius (r_c). We then assume that the potential contributions at any distance beyond (larger than) r_c is zero. The L.J potential for example is often truncated at a cutoff radius of $r_c = 2.5\sigma$ or 3.2 σ [51] where:

$$V_{LJ}(r_c) = V_{LJ}(2.5\sigma) = 4\varepsilon \left[\left(\frac{\sigma}{2.5\sigma}\right)^{12} - \left(\frac{\sigma}{2.5\sigma}\right)^{6} \right] \approx -0.0163\varepsilon \qquad (2.49)$$

That is, at the cutoff radius, the L-J potential V_{LJ} is about 1/60th of its minimum value. However, a simple truncation of potential may result in a new problem: whenever a particle pair ``crosses'' the cutoff radius, the energy

makes a small jump. When these events are repeated a large number of times, it is likely to spoil the energy conservation in a simulation. The L-J potential method takes into account this discontinuity that may result from truncation. To avoid the described leap discontinuity at the cutoff radius, L-J potential is shifted upward by a minimal amount so that at the cutoff radius the truncated potential is exactly equal to zero. That is, the potential is often shifted in order to vanish at the cutoff radius. So, given the L-J potential relation mentioned above, we can deduce the truncated L-J potential which is expressed as follows [50,51]:

$$V_{LJ_{trunc}}(r) := \begin{cases} V_{LJ}(r) - V_{LJ}(r_c) & \text{for } r \leq r_c \\ 0 & \text{for } r > r_c. \end{cases} \qquad (2.50)$$

Truncation does of course have an effect on the physical quantities, yet it is not easy to estimate truncation effects particularly for geometries with low symmetries. However truncation effects are usually large for quantities like surface energy.

2.6 MINIMUM IMAGE CONVENTION AND LONG RANGE CORRECTION:

As previously mentioned, since the number of interactions involved when calculating the force and energy are infinite, this makes it impossible to calculate them in practice. In order to overcome this issue an approximation method has been introduced. The 'minimum image convention' approximation method in essence relies on the placing a restriction on the force range. That is in short-ranged force, the only interactions involved in the calculations are those which take place among the particles closest to particle 1. We may assume that particle 1 is the center and then construct a smaller box with the same size and shape as the original one. This box is now referred to as the minimum image convention box. Now we will only consider the interactions which take place between particle 1 and all other particles whose centers are inside the minimum image convention box. Now

the summations which involve particle 1 can be limited to a number of N-1 which facilitates matters to a very large degree [16].

The minimum image convention limits the terms involved in calculating energy and force to 1/2 N(N-1). This is a reasonable range for simulations using a relatively small number of molecules. For simulations with a number of molecules within the range of say several thousands, the simulation shall still be computationally intensive. This is where we use the truncation method in conjunction with the minimum image convention method. The cutoff radius is chosen, now further limiting the molecules with which particle 1 is assumed to interact. In order to keep the cutoff radius consistent with the minimum image convention, it must not be larger than half the box length and in order not to change the properties of the material extensively it must not be too short [52].

Until now we have dealt with all the interactions that may be referred to as short ranged interactions/forces. These are the forces or interactions whose effects vanish at any distance beyond the cutoff radius. In most cases, a potential is defined to be short range if and when it decreases r faster than or equal to $r^{-(d-1)}$ where d is the dimensionality of the system. However if a potential does not fit the above description it cannot be assumed as short range. Now we need to deal with the long range forces, which may in some cases, be of crucial importance. Such cases include [52,53]:

- Coulomb interactions: when ions are involved.

- Dipole-dipole interactions: when particles are neutral but polarized.

- Vander Waal's attraction: when there are induced dipole moments.

Under these conditions, we cannot estimate a cutoff radius since it is essential that we take into account the interaction of each particle with all other particles within the MD box. We cannot also use the minimum image convention method since the particle would also interact with its own image. The interactions coming from far

away particles are not negligible. Although this type of interaction decreases with distance, the number of interaction increases so that the total contribution of all far away particles can in certain cases have a weight or contribution to the total interactions as that of the neighboring particles. This is where a technique referred to as the long range correction method presents itself. This technique is usually used in simulations of bulk fluids, as well as those involving gas-solid interactions. It is a method which poses an assumption that the radial distribution function for pair wise interactions is unity beyond a certain cutoff radius. There are a number of methods used to deal with the long range forces among which the most commonly used are the Ewald summation method which we shall discuss in detail as well as the Particle-Mesh Ewald method and the reaction field method. Less commonly used is the cell multiple method [16,54].

2.7 THE EWALD SUMMATION:

The Ewald summation method is one of the most commonly used methods when it comes to estimating long range interactions. In definition it is 'a technique for evaluating the potential, subject to periodic boundary conditions, due to a lattice point charge, plus a screening background'. It was named after Paul Peter Ewald, who first introduced it in 1921 [48]. It may be treated as a special case of the Poisson formula. The procedure by which the Ewald summation solves the long range interaction problem is that it decomposes the interaction potential into two components: a short range component that is summed in real space, along with a long term component summed in Fourier space [55,56].

The diagram below describes how the Ewald summation method replicates the simulation box to convergence in a small system versus the use of the radial cutoff method [55].
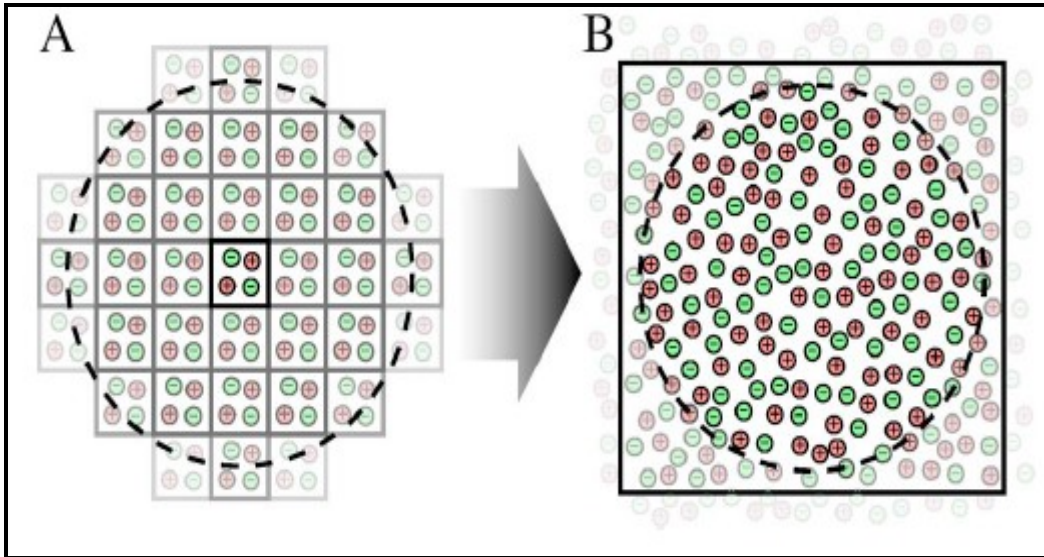
Figure (2.6): (A) A small system and Ewald sum replicates the simulation box to convergence. (B) Radial cutoff method showing a larger system reaching convergence

This diagram below describes the Ewald summation in terms of the Gaussian distribution where the set of charges in real space are cancelled out by those in the Fourier (reciprocal) space (2.8)[48].
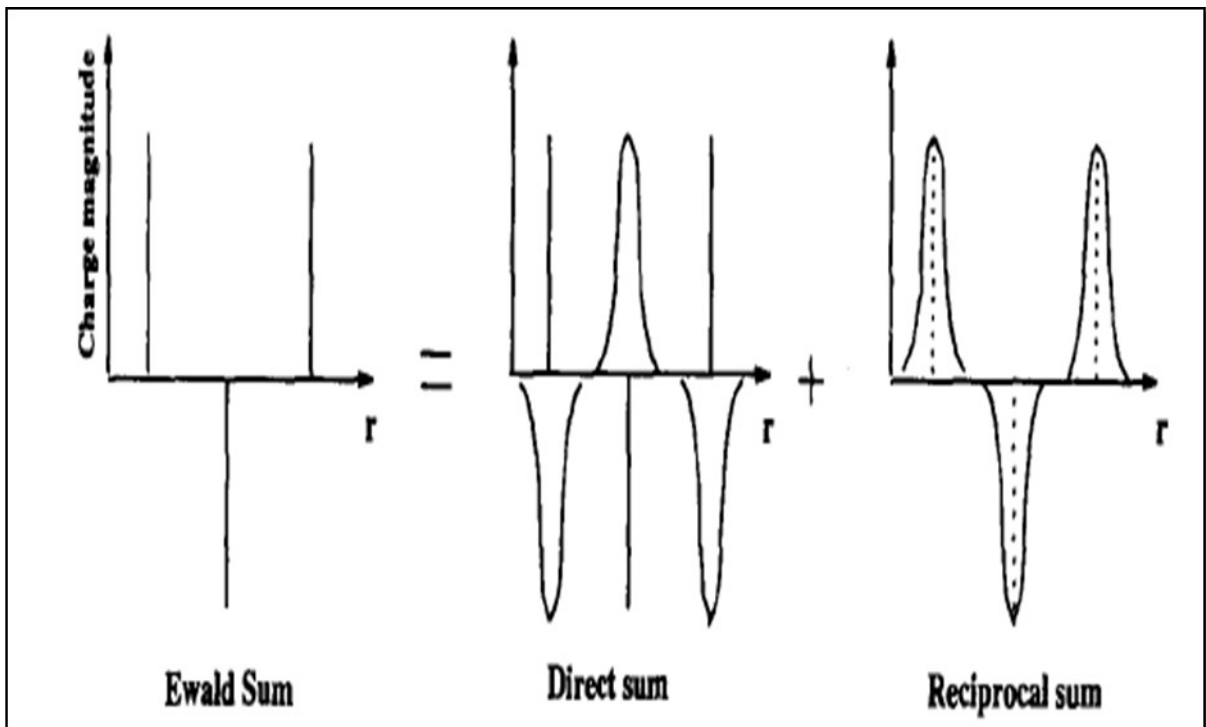
Figure (2.7): The Ewald sum components of a one-dimensional point charge system. The vertical lines are (+ / -) unit charges and the Gaussians are also normalized to unity.

Under PBC, the total Coulomb energy (Ewald Sum) of a system of N particles in a cubic box of size L and their infinite replicas with no external field is given by:

$$V_{Coulomb} = V_R + V_F - V_S + V_{B(cr)} \tag{2.51}$$

The Ewald summation method transforms the 'conditionally convergent sum' into four components: the real space term (V_R), the Fourier (reciprocal) space term (V_F), the self interaction term (V_S) and the boundary term $V_{B(cr)}$ which is a function of the dipole moment of the simulation cell and the dielectric constant surrounding the media of the periodically replicated cells. The first three components are given by [16]:

$$V_R = \frac{1}{4\pi\epsilon_0} \sum_{r_{ij} < r_{cut}} \frac{q_i q_j}{r_{ij}} \operatorname{erfc}(\alpha r_{ij}) \tag{2.52}$$

$$V_F = \frac{1}{2V\epsilon_0} \sum_{k \neq 0} \frac{\exp\left(-|k|^2/4\alpha^2\right)}{|k|^2} Q_k \overline{Q}_k \tag{2.53}$$

$$V_S = \frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2. \tag{2.54}$$

where r_{ij} is the distance between charges i and j, *erfc* is the complementary error function and Q_k is the density squared at wave vector k and is given by [16]:

$$Q_k = \sum_i q_i \exp(\imath k \cdot r_i), \qquad k = \frac{2\pi}{L}(k_1, k_2, k_3), \ k_i \in Z, \ i = 1, 2, 3; \tag{2.55}$$

given that L is the length of the simulation box and $V=L^3$.

If we assume that the MD box is spherical and is divided into cubic cells all of which have the same dimensions (length of side =L) and all of which are surrounding a central cell, then the position of any of the cubic cells can be described by a certain vector operation with integral components that are multiples of L. The

charge interactions of the particles can be described by the following relation [16,48]:

$$V = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$ (2.56)

If the box is three dimensional there will be six nearest neighboring cells to the central cell whose positions may be given by the coordinates (0,0,L), (0,0,-L), (0,L,0), (0,-L,0), (L,0,0) and (-L,0,0). These *replica* cells are separated from the central cell by a distance r_{box}. Thus the interactions between the charges in the central cell and those in the replicas can be given by[16,54]:

$$V = \frac{1}{2} \sum_{nbox=1}^{6} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 |\mathbf{r}_{ij} + \mathbf{r}_{box}|}$$ (2.57)

Now if we accumulate the charge-charge interactions within the central cell with the charge-charge interactions between particles in the central cells and those in the six nearest replicas surrounding it we arrive at the following relation:

$$V = \frac{1}{2} \sum_{|\mathbf{n}|=0}^{\infty} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0 |\mathbf{r}_{ij} + \mathbf{n}|}$$ (2.58)

We then need to consider the contribution of interactions between the central cell and surrounding medium including points outside the MD box. A correction term is introduced which depends on the properties of the surrounding medium. Permittivity is amongst the properties which highly affects the definition of the correction term. That is if the permittivity is infinite the effect of the surrounding medium is negligible and we can ignore the correction term. That is the case when the surrounding material is a conductor. If however the surrounding medium has a relative permittivity of 1 it is considered to be a vacuum space and this correction term must be introduced [56]:

$$V_{correction} = \frac{2\pi}{3L^3} \left| \sum_{i=1}^{N} \frac{q_i^2}{4\pi\varepsilon_0} \mathbf{r}_i \right|^2 \qquad (2.59)$$

given that r_i is the distance between the particle of interest and the border of the unit cell.

For a convergent series, which converges rather slowly, like the one introduced in equation (2.58) above, the time it would take to solve this equation may be very long and thus affect the results of our simulation. Increasing the time boundaries is not always the best method since it has its negative aspects as we have discussed earlier. So another method was introduced to serve as an intermediary for solving this matter. The series is divided into two convergent series, the first summed in the real space and the other summed in the Fourier space. This was fully described at the beginning of this section. However, it may also be carried out via the usage of Gaussian distributions for each charge. We shall assume that each charge involved in the system is surrounded by a Gaussian distribution of charges of equal magnitude and opposite direction (thus neutralizing the system). We shall design two Gaussian distribution the first of which shall include a group of positively and negatively charged particles distributed in proportion around the real system. This will result in a contribution to the potential energy which corresponds to the real space. The second Gaussian distribution shall represent the Fourier (mirror) space and shall counteract the first distribution.

If the Gaussian distribution is given by [16,57]:

$$\rho_i(\mathbf{r}) = \frac{q_i \alpha^3}{\pi^{\frac{3}{2}}} e^{-\alpha^2 r^2} \qquad (2.60)$$

then we can convert the sum representing the charged particles into another representing all interactions including the neutralizing factor. Now we may write an equation denoting the interactions of the central MD cell as well as the Gaussian distribution for real space as follows[16,54,55,56]:

$$V = \frac{1}{2} \sum_{|\mathrm{n}|=0}^{\infty} {}^{\backslash} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{erfc\left(\alpha\left|\mathrm{r}_{ij} + \mathrm{n}\right|\right)}{\left|\mathrm{r}_{ij} + \mathrm{n}\right|}$$
(2.61)

where α is the damping or convergence parameter with units of $Å^{-1}$ and *erfc* (the error function) is given by:

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-t^2} dt$$
(2.62)

Equation (59) is a rapidly convergent series representing real space, whilst the following series (also a rapidly converging series) shall represent the reciprocal space:

$$V = \frac{1}{2} \sum_{k\neq 0} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{1}{\pi L^3} \frac{q_i q_j}{4\pi\varepsilon_0} \frac{4\pi^3}{k^2} e^{-\frac{k^2}{4\alpha^2}} \cos(\mathbf{k} \cdot \mathbf{r}_{ij})$$
(2.63)

where *k* represents reciprocal vectors and are given by $2\pi\mathbf{n}/L^2$.

There are three parameters which control the convergence of the sums given above. These parameters are:

- n_{max}: an integer representing the range of the real space and controlling its maximum number of vectors

- m_{max}: an integer representing the summation range in the reciprocal space as well as its maximum number of vectors

- The Ewald damping (convergence) parameter (α): an amount representing the relative rate at which the real and reciprocal sums converge.

As we see in equation (2.61), the damping parameter α appears in the numerator and thus the larger its value the higher the accuracy of the calculation of the interaction contribution of the central cell. In equation (2.63) on the other hand the damping parameter appears in the denominator thus inversing the effect since we would require a smaller value as possible

for α to include a smaller number of terms as possible. It is very important that we choose the optimal value for the damping parameter. It must be a value that balances the effect of the parameter in the two equations thus not heavily affecting either. That is it cannot be too big as to interfere with equation (2.63) yet not too small to reduce the efficiency of equation (2.61). There are many ways we can choose the optimal value for α which includes comparisons with either experimental data or previous results. In our field of study the optimal value for the damping parameter which gives the best results for simulations was found to be in the range of ($\alpha = 5/L$).

To summarize this, there are a number of factors we must consider when choosing Ewald parameters. These factors are [48]:

1. **System size:** A larger system requires a large damping parameter and/or cutoff radius in order to limit the number of pair-wise interactions such that the real-space sum converges faster.

2. **Accuracy desired:** Choosing a larger cutoff radius, nmax, or mmax usually yields more accurate results. However sometimes this may be inefficient.

3. **CPU time consumed:** The larger the damping parameter the less the work that shall be done in the real sum which is usually the most time consuming part.

4. **Cutoff radius:** The smaller the cutoff radius the larger we must make the damping parameter so that the real sum can converge rapidly with a reasonable number of vectors.

5. **the reciprocal sum:** Usually the reciprocal sum is the once which is calculated more efficiently and hence it would be better to choose the damping parameter to minimize the real sum.

6. **Large systems:** For systems with N more than 10,000 Rycerz and Jacobs suggested a value for the damping parameter given by [48]:

$$\alpha = \tfrac{1}{2}\left(N^{1/3} + \frac{\beta L}{2R_{\text{cutoff}}}\right)$$

(2.64)

Having arrived at the best terms for the sums representing the real space as well as the reciprocal space we must now introduce the self interaction component which is given by:

$$V = -\frac{\alpha}{\sqrt{\pi}}\sum_{k=1}^{N}\frac{q_k^2}{4\pi\varepsilon_0}$$

(2.65)

Now adding the three components we arrive at:

$$V = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\begin{array}{l}\displaystyle\sum_{|\mathbf{n}|=0}^{\infty}{}^{\backslash}\frac{q_iq_j}{4\pi\varepsilon_0}\frac{erfc(\alpha|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|}\\[2ex]\displaystyle +\sum_{k\neq0}\frac{1}{\pi L^3}\frac{q_iq_j}{4\pi\varepsilon_0}\frac{4\pi^3}{k^2}e^{-\frac{k^2}{4\alpha^2}}\cos(\mathbf{k}.\mathbf{r}_{ij})\\[2ex]\displaystyle -\frac{\alpha}{\sqrt{\pi}}\sum_{k=1}^{N}\frac{q_k^2}{4\pi\varepsilon_0}\\[2ex]\displaystyle +\frac{2\pi}{3L^3}\left|\sum_{i=1}^{N}\frac{q_i^2}{4\pi\varepsilon_0}\mathbf{r}_i\right|^2\end{array}\right]$$

(2.66)

The Ewald summation method is applied to calculate other quantities such as force, total energy and other.

CHAPTER 3: TETRAHEDRAL MOLECULES AND TETRAFLUOROMETHANE

3.1: MOLECULAR GEOMETRIES

When studying a certain molecule, whether solid liquid or gas, it is very important to understand the geometry and orientation of that molecule. Molecular geometry is usually defined as 'the three dimensional arrangement of atoms that constitute a molecule'[57]. It determines a number of properties of the substance including polarity, reactivity, phase, phase transitions, color, magnetic properties, point groups, biological activity, and electron affinity as well as a number of other properties which are all of high importance in understanding the nature of that substance. The type of bonding as well as electronic configuration of the molecule is also of great importance. Both these properties greatly affect the electronegativity of the molecule and thus its reactivity towards other substance of the same or different kind. Furthermore, the nature of the chemical bond determines the position of each atom relative to other atoms. Bond lengths, bond angles, and torsion angles also play an important role in describing the geometry of the molecule, its orientation and reactivity.

MD simulations are one of the very important tools used to study geometries and related characteristics of the atoms. Knowing the geometry of a molecule and the other factors described above then facilitate the modeling stage of simulations aimed towards studying other aspects of the molecule including phase transitions. They also give a brief overview of the range within which we may expect our results to turn out. They are also important when deciding on the boundaries of the simulation. The bulkier the molecule, the longer the time needed for the system to equilibrate. The bulkier the molecule, the harder it would be to include a large number N in the simulation. So it has a role in deciding the time step as well as the number of molecules. It also has a rule in deciding the optimum conditions to set for the simulation such as temperature, pressure, cutoff radius and other factors.

The geometry of the molecules differs vastly. It may be determined by various methods including a number of spectroscopic methods along with a range of

diffraction methods. The vibrational and rotational motion of the molecule allows techniques like Infra Red spectroscopy, Raman Spectroscopy and Microwave Spectroscopy to determine the geometries. For solids, x-ray crystallography and electron diffraction are among the most commonly used methods for determining the geometry of the molecule. The first of these two methods uses the distance between the nuclei whilst the second one uses the concentration of electron density. For small molecules in the gas phase we may use Gas electron diffraction to determine the geometry.

There are a huge number of geometries and orientations that molecules adapt, all of which are in basis related to one of the more common geometries. The most common geometries, according to the Valence Shell Electron Pair Repulsion theory are: linear, trigonal planar, bent, tetrahedral, octahedral, pyramidal and square planar. The table below shows the shapes of the basic classes of Molecular Geometries [57,58]:

Geometry	*Shape*
Linear	
Trigonal Planar	
Bent	
Tetrahedral	
Octahedral	
Pyramidal	
Square Planar	

Table (3.1): Molecular geometries and their shapes

3.2: TETRAHEDRAL MOLECULES

A Tetrahedral Molecular Geometry is one of the less complicated molecular geometries present. Tetrahedral molecules have four pairs of electrons around the central atom. Each electron pair forms a covalent bond with one of the corner atoms. This will leave the central atom with no lone pairs of electrons and four bonding pairs of electrons. The bonding pairs of electrons repel each other as much as possible which results in the tetrahedral orientation. In three dimensional space a tetrahedral molecule will have four identical faces (identical in all aspects only when the four corner atoms are the same). Each of the faces is an equilateral triangle making the molecule perfectly symmetrical. The diagram on the left shows the equilateral triangular faces of a tetrahedral molecule while the one on the right shows the Lewis structure of a tetrahedral molecule [59,60].
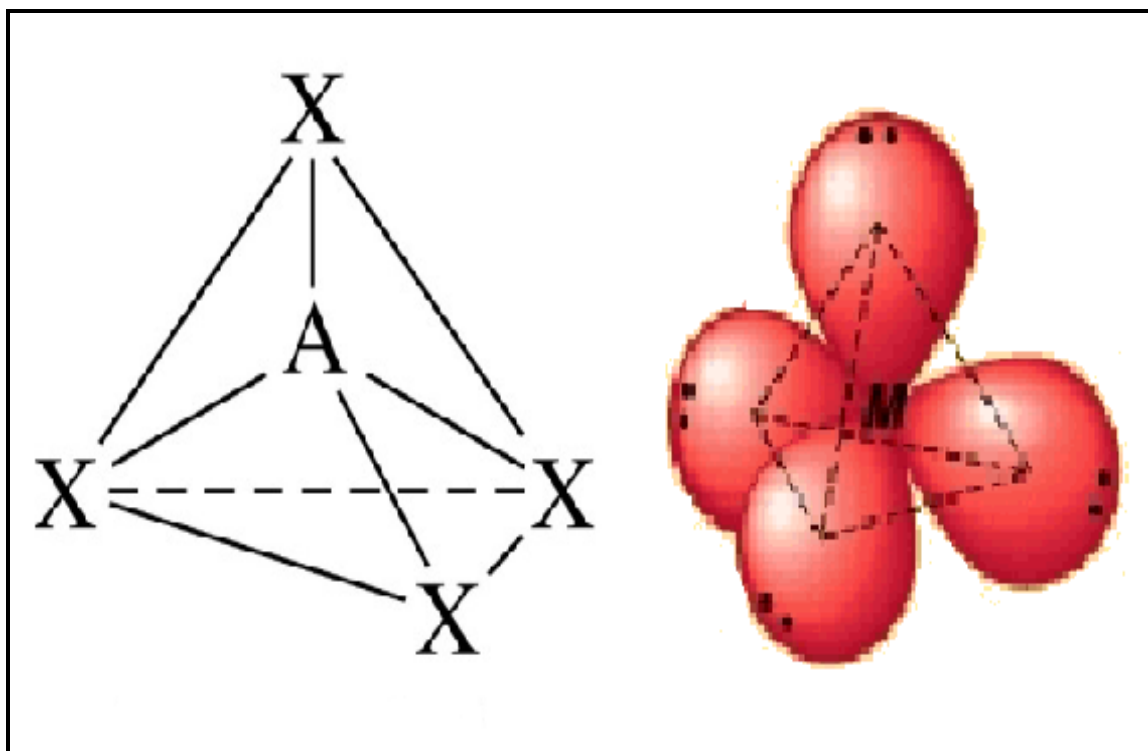
Figure (3.1): Left: equilateral triangular faces of a tetrahedral molecule. Right Lewis structure and atomic orbital orientations of a tetrahedral molecule

In organic compounds the central atom is a Carbon atom (C). In the simplest hydrocarbon, methane (CH_4), the four corner atoms are Hydrogen. In other hydrocarbons, the central C atom is surrounded by four substituent

atoms. A substituent is an atom or group of atoms substituted in place of a Hydrogen atom on the parent chain of a hydrocarbon at the corners of a tetrahedron [60,61].
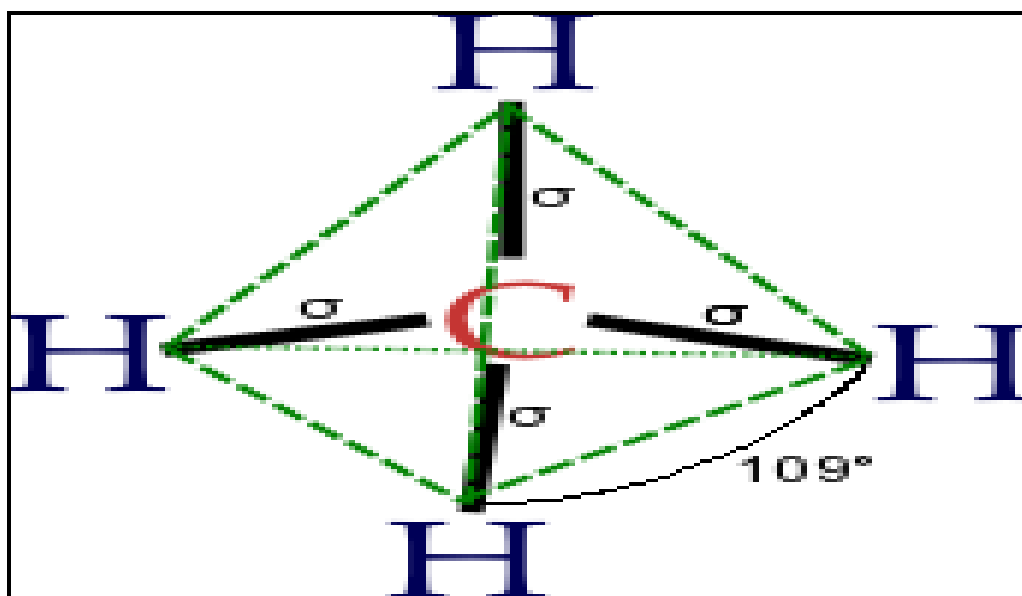
Figure (3.2) Tetrahedral methane molecule with the central C atom surrounded by 4 H atoms at the corners of a tetrahedron with bond angles of 109º

So in such geometry where there is one atom located at the center with four other atoms located at the corners of a tetrahedron the bond angles, when all four of the corner atoms are the same, are equal to $\cos^{-1}(-1/3)$ which is equivalent to an angle of approximately 109.5º. The tetrahedral geometry is most common in the first half of the periodic table and more so in organic compounds [60].

However, not all molecules where the central atom has four lone pairs of electrons are tetrahedral. In cases where one lone pair is not shared the resulting geometry of the molecule is triangular pyramidal. The triangular pyramidal is also a tetrahedron; however it differs from a tetrahedral molecule in the location of their central atom. In a pyramidal shape molecule the central atom is located at the apex of the pyramid whilst in a tetrahedral molecule the central atom is at the center of the tetrahedron. An example of a molecule with pyramidal geometry is ammonia (NH_3). The images below

show the difference between a tetrahedral molecule and a pyramidal molecule [62].
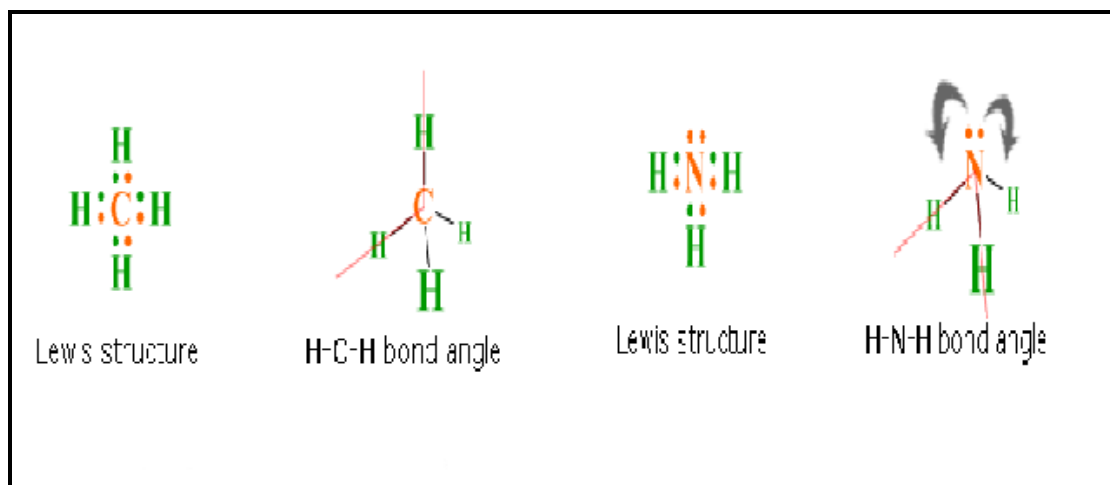
Figure (3.3) From the left: Lewis structure of methane (CH₄) -bond angles in tetrahedral (CH₄) - Lewis structure of Ammonia (NH₃)- bond angle in pyramidal (NH₃)

In the figure below, the black cloud on the top vertex of the tetrahedron in a pyramidal geometry is the electron cloud formed due to the presence of a non-bonded lone pair of electrons[62].
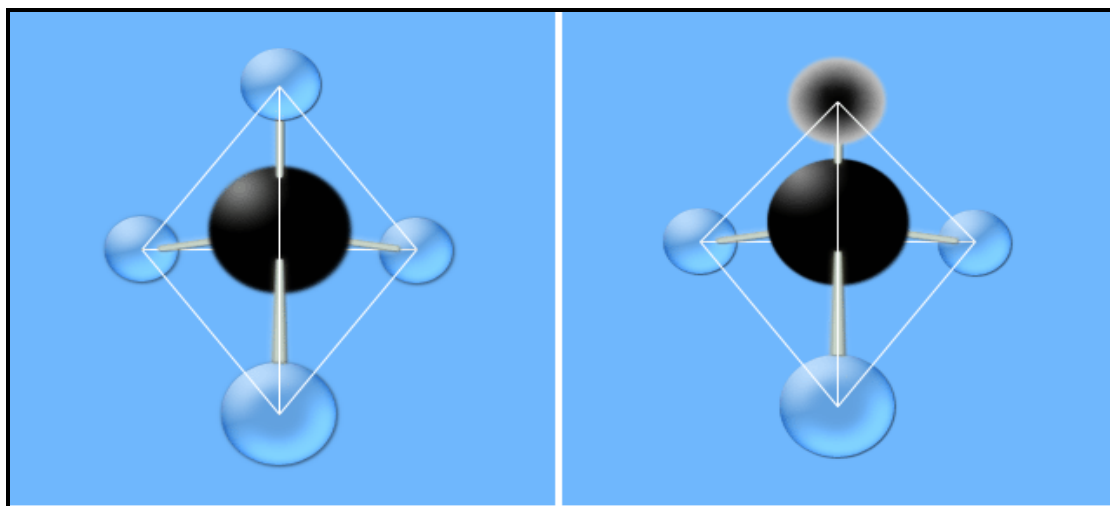
Figure (3.4): The difference in geometries between a tetrahedral molecule (left) and a pyramidal molecule (right) due to the presence of a non-bonded (lone) pair of electrons in pyramidal molecules

We must note that not all tetrahedral molecules are perfectly symmetric. They can be chiral. That is 'a type of molecule that has a non-superposable mirror image' [57,63]. In more simple terms, it is a molecule where the four faces are not identical. In Tetrahedral

molecules this is the case when all four substituent atoms are different. The difference in nature, electronegativity and other factors alter the tetrahedron orientations, bond angles and thus symmetry of the molecules. This is reflected in the point groups assigned to these molecules.

3.3: CRYSTAL STRUCTURE

The crystal structure of a molecule is defined as the 'unique arrangement of atoms or molecules in a crystalline liquid or solid' [63]. It is composed of a pattern where the atoms are arranged in a definitive manner and the lattice is highly ordered. The lattice, also known as the Bravais Lattice is a repetitive pattern generated by a set of discrete translation operations and vectors. The points form identical unit cells which fill the entire lattice. The lattice is described by a set of parameters known as the lattice parameters which consist of the lengths of the sides of the unit cells as well as the angles between them. The unit cells are therefore considered the simplest repeating unit in a crystal.

The positions of the atoms inside a unit cell are described in terms of (xi,yi,zi). The entire structure of the crystal is described in terms of its unit cells. The unit cell itself is defined in terms of the lattice points or all the points in space about which the particles are free to vibrate. August Bravais showed that crystals can be divided into 14 unit cells.[64] These unit cells have certain criteria among which the most important are: the opposite faces of a unit cell are parallel and every edge of the unit cell connects equivalent points. There is a total of 14 unit cells which fall into seven categories. The figure below shows the 14 Bravais unit cells [65]:
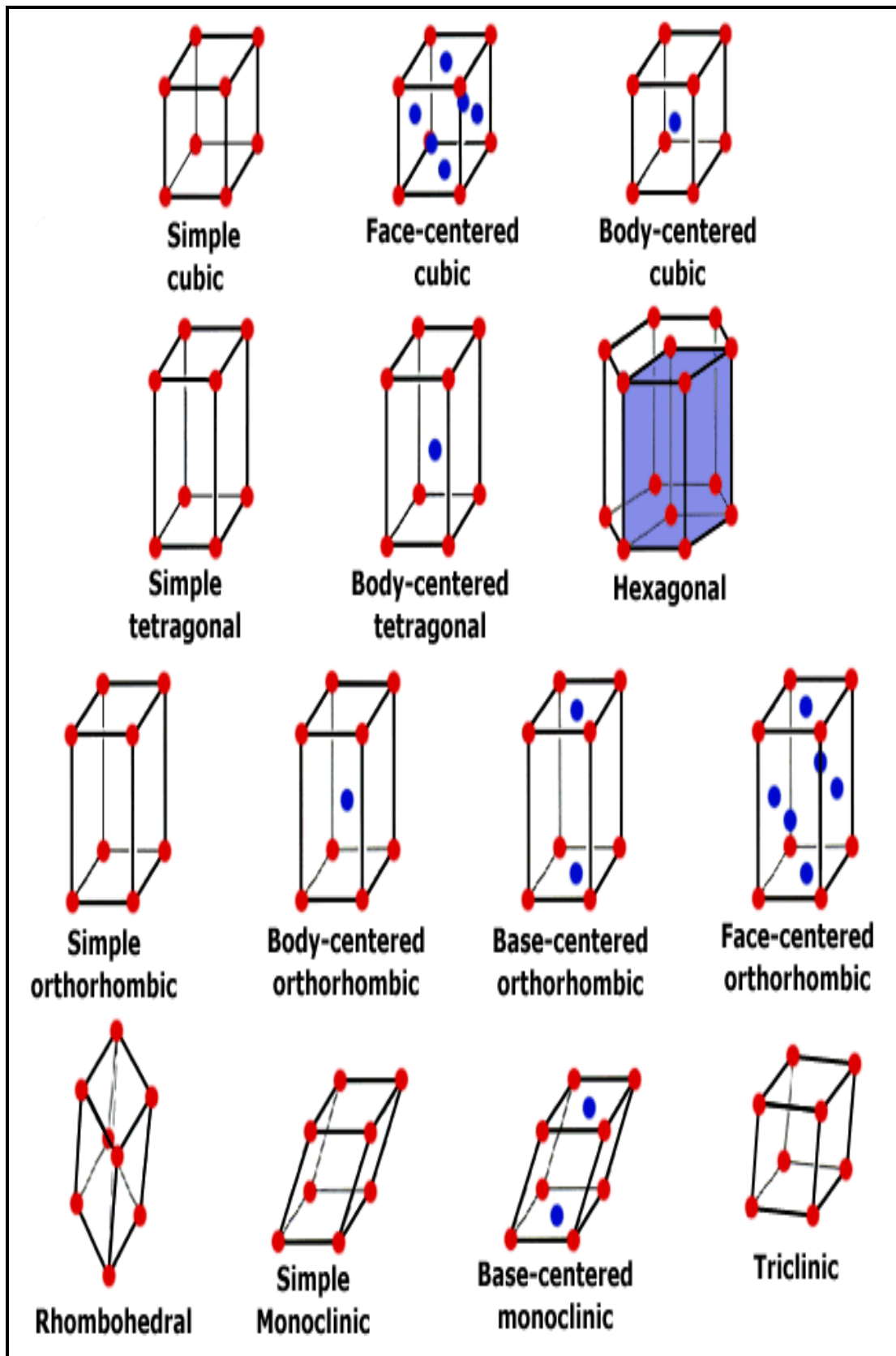
Figure (3.5): 14 Bravais Lattice Systems

The 14 unit cells are subcategorized into 7 groups which differ in the lengths of the sides of the unit cells (a,b,c) as well as the three internal angles (α,β,γ). The 7 groups are described in table (3.2).

Lattice Type	Side Lengths	Internal Angles
Cubic	a=b=c	$\alpha=\beta=\gamma=90^o$
Tetragonal	a=b≠c	$\alpha=\beta=\gamma=90^o$
Monoclinic	a≠b≠c	$\beta=\gamma=90^o\neq\alpha$
Orthorhombic	a≠b≠c	$\alpha=\beta=\gamma=90^o$
Rhombohedral	a=b=c	$\alpha=\beta=\gamma\neq90^o$
Hexagonal	a=b≠c	$\beta=\gamma=90^o\neq\alpha$ ($\alpha=120^o$)
Triclinic	a≠b≠c	$\alpha\neq\beta\neq\gamma\neq90^o$

Table (3.2): 7 Bravais Lattice Groups with relative side lengths and bond angles

The figure below shows the difference between the three cubic unit cells: simple cubic, body-centered-cubic (BCC) and face-centered-cubic (FCC) [66,67]:
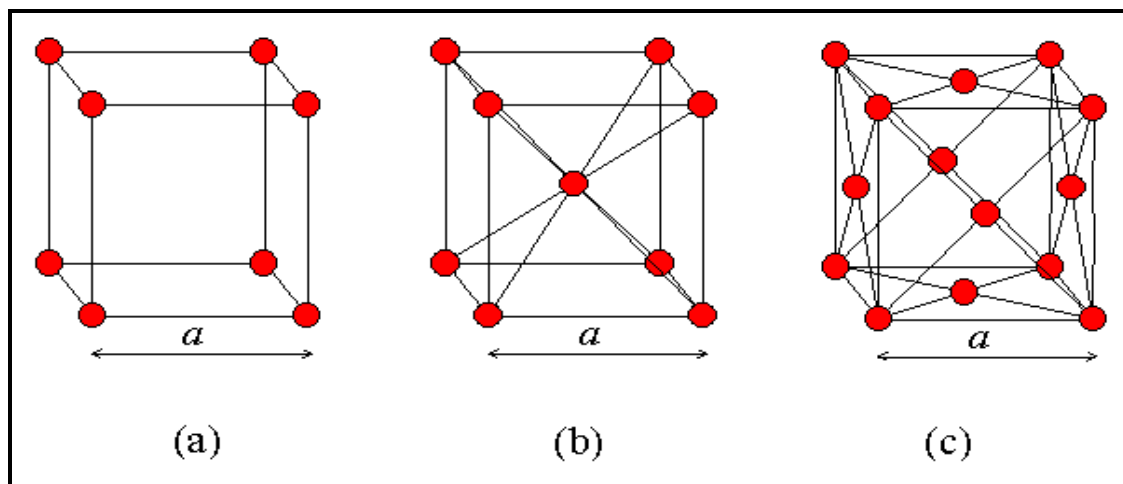
Figure (3.6): The difference between a simple cubic unit cell (a), a body-centered cubic unit cell (b) and a face-centered cubic (c) unit cell

- In the simple cubic cell, also referred to as primitive cubic, there is one lattice point at each corner of the cube. This means that each of the atoms placed at one of the vertices of the unit cell is shared equally between eight adjacent unit cells and thus the total contribution of the atom towards one unit cell is 1/8. Multiplying (1/8 * 8) we get 1 atom per unit cell in this particular structure. The side length in terms of the radius of the atom is given by *a=2r*. The packing fraction of the unit cell (volume of sphere/volume of unit cell) is 52%.

- For a BCC unit cell there is one lattice point at the center of the unit cell. This atom fully belongs to the unit cell and is not share by any adjacent cells and thus its contribution to the unit cell is 1. There is also the one lattice point at each corner also shared by eight adjacent unit cells and contributing to the unit cell by 1/8. Therefore the number of lattice points (or atoms) per unit cell is equal to (1+ (1/8 * 8))= 2. The side length in terms of *r* is given by: *a=4r/√3*. The packing fraction of the unit cell is 68.01%.

- An FCC system has one lattice point at each corner as well as one lattice point on each of the six faces of the cube. As explained before the points located at the vertices each contributes by 1/8. Each of the faces of the cube is shared by 2 adjacent cells thus contributing to each of them by 1/2. The total number of lattice points from the faces is therefore (1/2 * 6 = 3), which makes the total number of lattice points (or atoms) in an FCC = ((1/8*8) + (1/2*6)) = 4. The side length in terms of *r* is given by: *a=4r/√3*. The packing fraction of the unit cell is 74.05%.

The figure below emphasizes the points stated above. It shows the packing style of each of the three cubic unit cells [68].
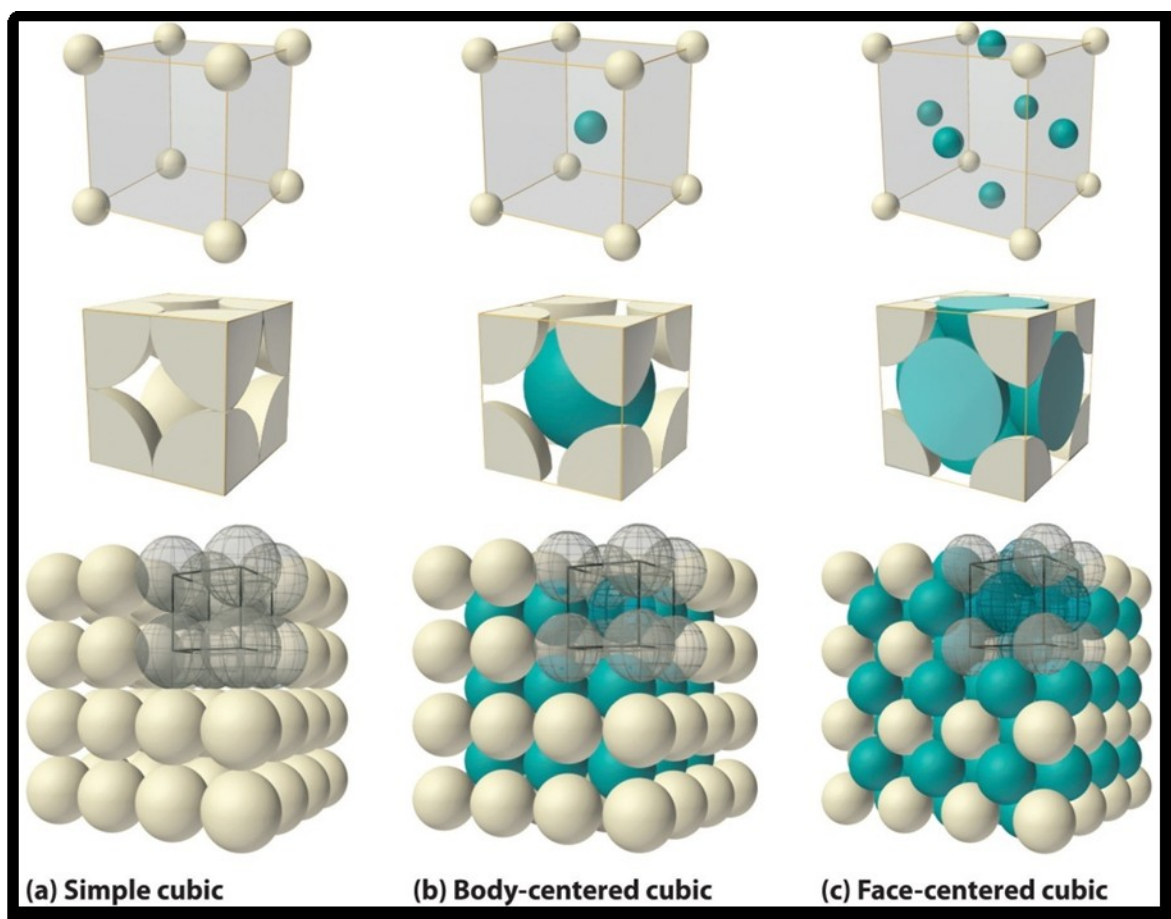
Figure (3.7): A ball and stick model(top), a space filling cutaway model that shows the portion of each atom that lies within the unit cell (middle) and an aggregate of several unit cells (bottom) for the three types of cubic unit cells

3.4: MOLECULAR SYMMETRY

Molecular symmetry describes the symmetry present in molecules as well as the classification of these molecules on basis of their symmetry. It helps predict a molecule's chemical properties including dipole moment, spectroscopic transitions, phase transitions, reactivity and other properties of equal importance.

Therefore it is very important when studying the properties and behavior of certain molecules, that we understand the symmetry of that molecule. Molecular symmetry is a very important field whether in chemistry, physics or biology. A small change in the symmetry of a molecule may result in changing its properties

completely. It is perhaps enough to say that carbon nanotubes, coal and diamonds are all made up of exactly the same material and the only difference between them is their crystallography which in its essence depends on the symmetry of the molecule and the phase within which it exists. Certain cyclic molecules differ vastly in their properties on bases of the position of the functional group. A molecule with a functional group in the meta positions for example may have a melting point, boiling point, color, odor or reactivity which differs completely from the exact same molecule with the exact same functional group in the ortho or para positions.

A molecule's symmetry may be determined via a number of methods. There are a number of practical techniques which include X-ray crystallography and a number of of spectroscopic techniques such as IR and Raman spectroscopy. Other frameworks for studying Molecular symmetry include the point group approach which shall be discussed in detail later.

Some objects are more symmetrical than others; some are highly symmetric while others are completely asymmetric. A sphere for example is more symmetric than a cube, since it looks exactly the same after rotation around any angle about the diameter. A cube on the other hand only looks exactly the same of rotated through certain angles and around certain axis. For example, a cube would look exactly the same after a rotation of 90° around an axis passing through the center of any of its opposite faces. Similarly, it would look the same if rotated by an angle of 120° around an axis passing through any of its opposite corners.

This approach provides another definition for symmetry which may be expressed as a molecules possession of a number of indistinguishable configurations. The following diagram shows the rotation of the trigonal planar BF_3 by an angle of 120° leaving it indistinguishable from its original orientation [69,70].
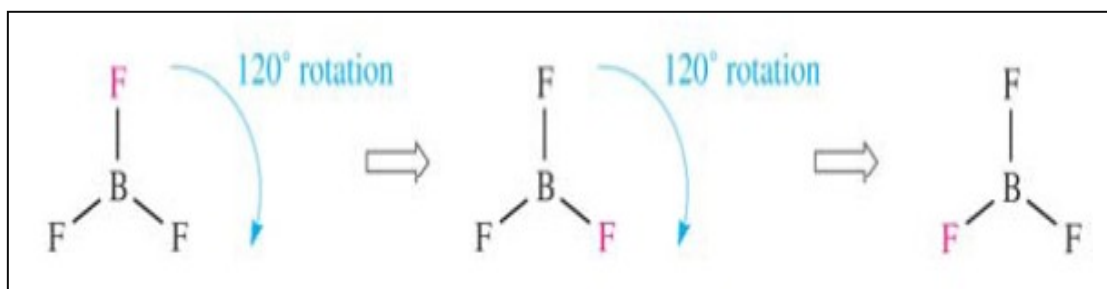
Figure (3.8): Symmetry operation performed on a BF3 molecule leaving it indistinguishable from its original orientation

Generally, any action or move which leaves the molecule looking exactly the same after the transformation as it did before is referred to as a symmetry operation. There are a lot of symmetry operations among which the most applicable and most widely used are rotations, reflections and inversions. Each of these operations is associated with what we refer to as a symmetry element which is either a point, line (axis) or plane with respect to which the symmetry operation is performed. The most important symmetry operations and their corresponding symmetry elements are described as follows [69,70,71,79,80]:

- **Identity (E):** This operation is about not moving the molecule at all. The entire object may be treated as a symmetry element. All molecules possess this operation. Those which *only* have the Identity operation are the least ordered or the most asymmetric. Examples may include $C_3H_6O_3$, CHClBrF, and lysergic acid denoted by the formula below:
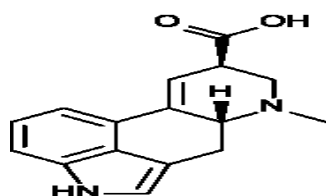
- **n-fold rotations (C_n):** This operation is a rotation about an n-fold axis of symmetry. C_n is describes as a rotation by an angle of *360/n*. That is C_1 is a rotation through 360° (same as E) whilst C_2 is a rotation through 180°. Water, (H_2O) for example possesses this symmetry

operation at an angle of 180°. That is it possesses one twofold axis, C_2. Ammonia (NH3) molecules have one threefold axis C_3. Hence an ammonia molecule has two rotations the first one being C_3 at an angle of 120o and the other $C_3^{'2}$ at an angle of -120° (or240°). The benzene ring on the other hand, C_6H_6 has one six fold axis C_6 and six twofold axes C_2. The largest C_n is referred to as the principal rotation axis. Linear molecules have an infinite number of axis C_∞. The following figures describe the process of the proper rotation around a C_3 axis, [72] and the rotational axes of the Benzene ring [73]..
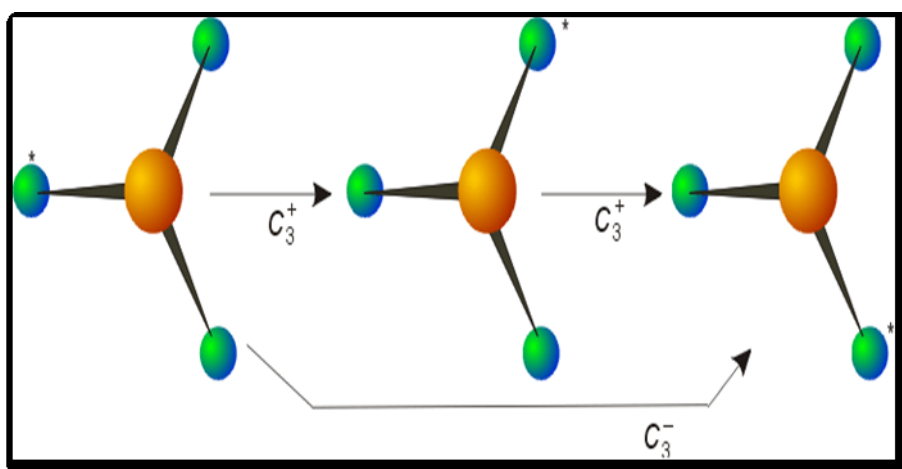
Figure (3.9): A C_3 axis has two symmetry operations associated with it; rotation by 120° in a clockwise or counterclockwise direction providing two different orientations
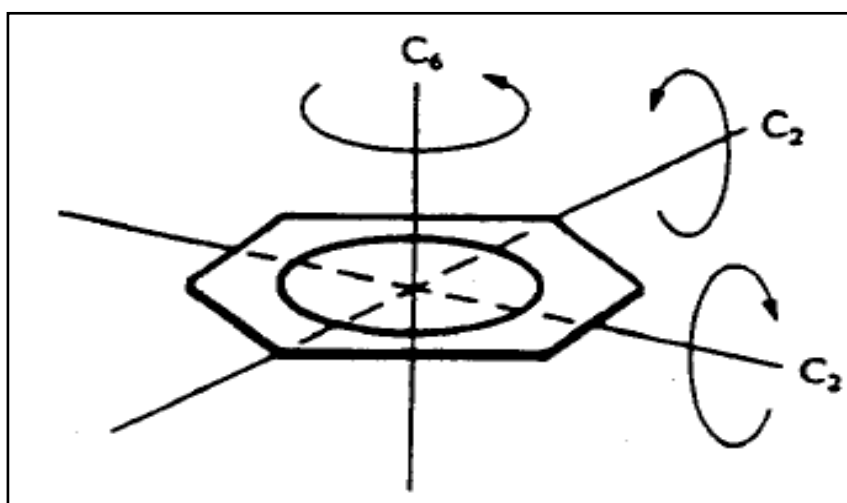
Figure (3.10): axes of rotation in a benzene ring

- **Horizontal Mirror Plane** σ_h: Mirror plane perpendicular to the principal axis of rotation. The Benzene ring for example has a C6 principal axis and a horizontal mirror plane.

- **Vertical Mirror Plane** σ_v: Mirror plane that is parallel to and encompasses the principal rotation axis. A Dihedral mirror plane σ_d is a vertical mirror plane that bisects the angle between two C_2 axes. Water for example, H_2O, has two vertical planes of symmetry. The representation in the figure below describes a number of the mirror planes present [72]:
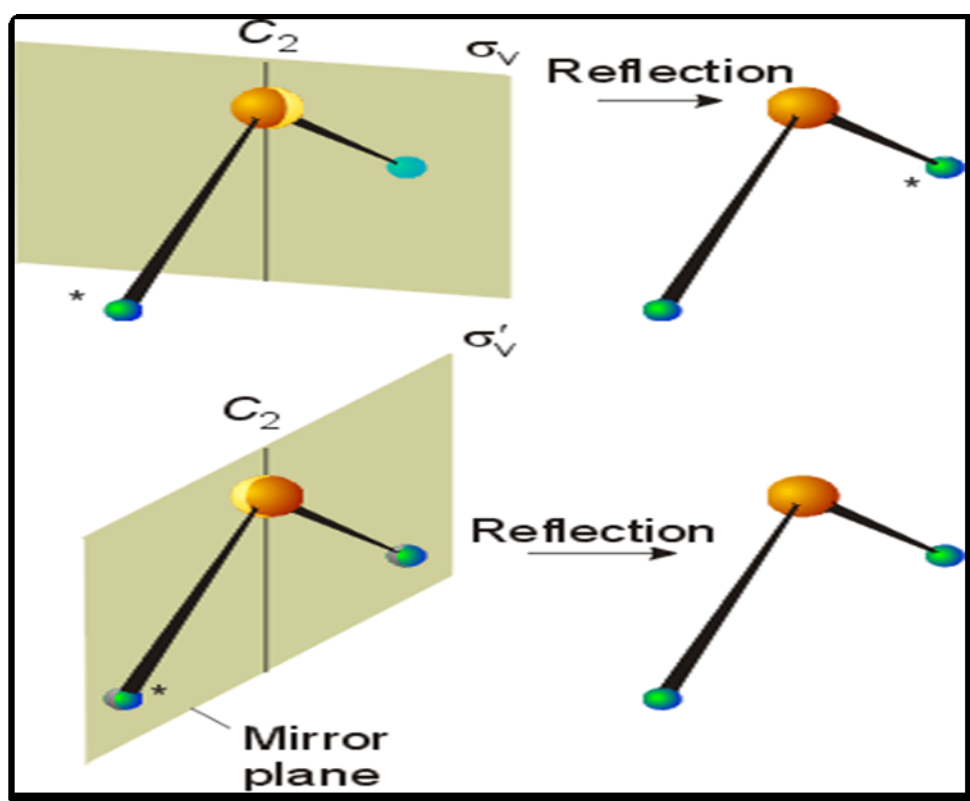
Figure (3.11): Mirror Planes in a water molecule

The Benzene ring has three vertical mirror planes one σ_v plane and two σ_d associated with the three C_2 axes and containing the principal axis C_6 The diagrams below shows the rotation axes as well as the mirror planes for the Benzene ring (3.13) [74]:
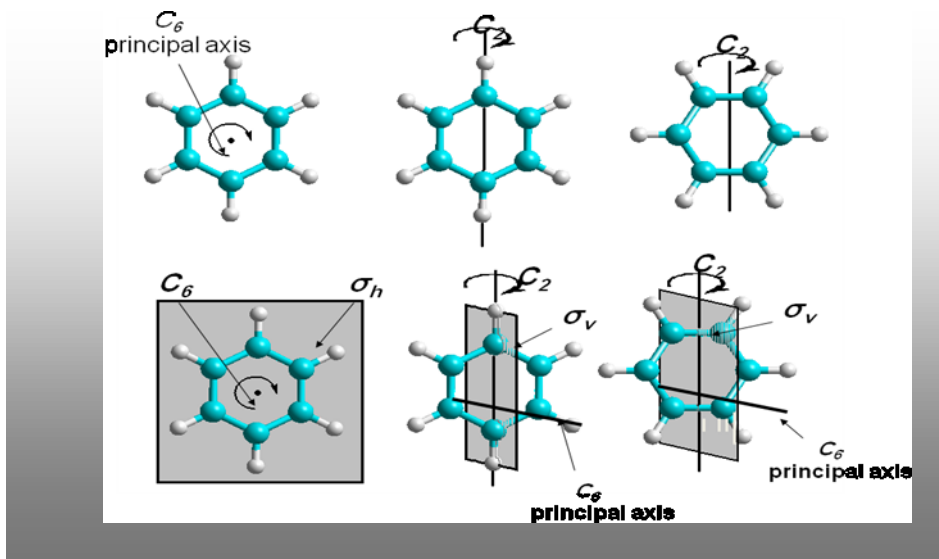
Figure (3.12): Rotational axes (top) and mirror planes (bottom) of the Benzene ring

- **Inversion (*i*) through the center of symmetry:** This operation transforms all coordinates of the object according to the rule *(x,y,z)* *(-x.-y,-z).* In other words, the inversion center, also referred to as the center of symmetry exists in a molecule if for any atom in the molecule there is an identical atom which is diametrically opposite the center of inversion and at n equal distance from it. The benzene ring has a center of inversion exactly at the center of the ring. Inversion may be described by the image below [72].
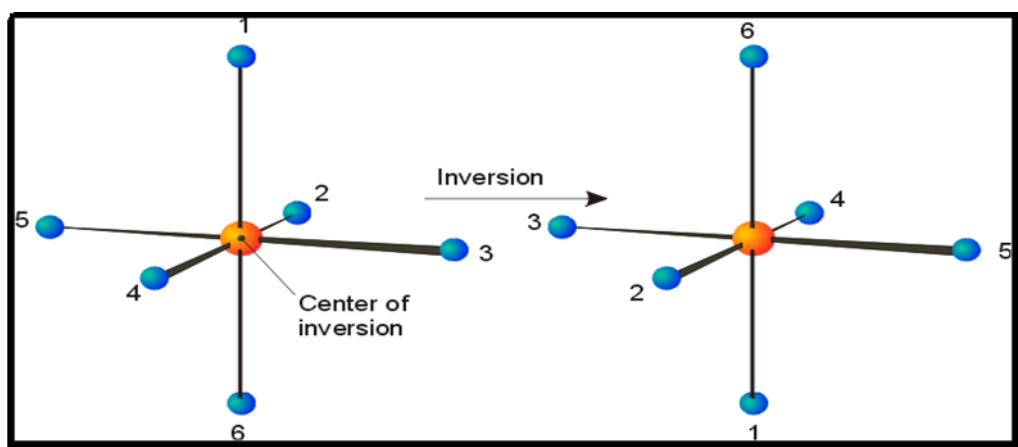
Figure (3.13): Inversion in a tetrahedral molecule

- **n-fold Improper rotation S_n:** The improper rotation can be described as a combination of two successive symmetry operations. The first is a rotation by an angle of 360/n and the second is a reflection through the plane perpendicular to the axis of rotation. Neither of these operations needs to be a symmetry operation on a stand-alone basis. It is the end results which we are considering. The CH4 for example, has three S_4 axes. Each of the three consists of a rotation of 90o about the axis followed by a reflection through a mirror plane perpendicular to the axis. The improper rotation may be the hardest to visualize, however the following figure of the S_4 rotation of methane (a tetrahedral molecule) might make it a bit easier[72,73].
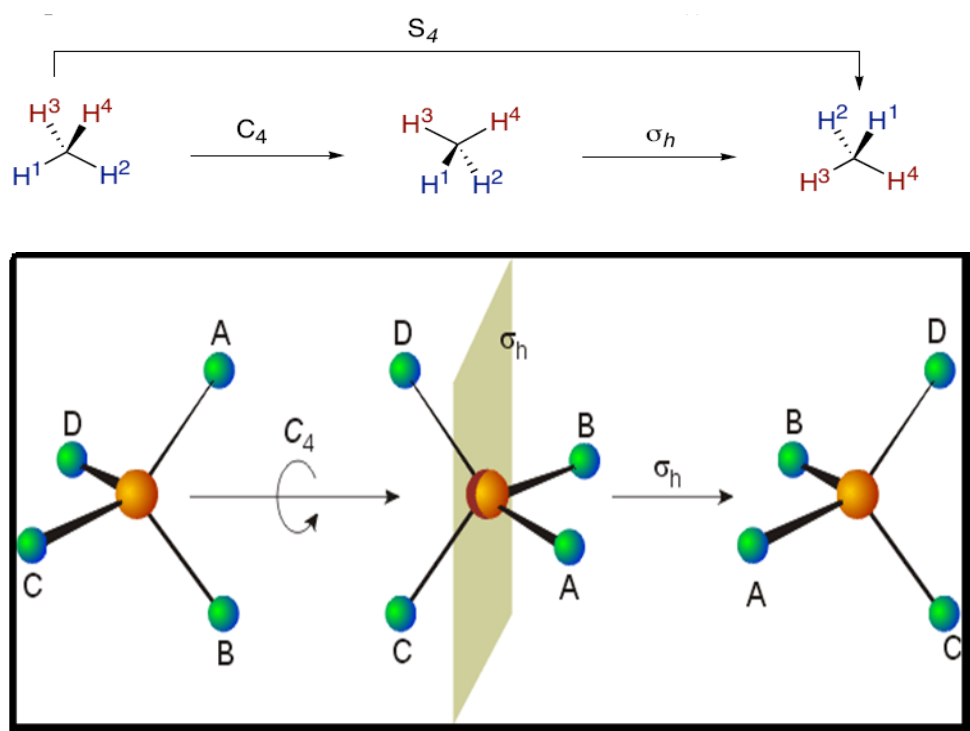
Figure (3.14): Improper rotation S4 around a C4 axis in a tetrahedral molecule

3.5: POINT GROUP

As we have seen in the previous section, the symmetry of a molecule which is located on a symmetry axis, cut by symmetry planes or centered on an inversion center is referred to as point symmetry. Having identified the symmetry operations and elements of a certain molecule, one can now group

certain molecules according to the similarities in their symmetries. In other words we can say that the collection of symmetry elements present in a molecule form a group typically referred to as a point group. The name 'point' group comes from the fact that all symmetry elements in a molecule (whether points, lines or planes) will intersect at one single 'point'. The Group theory has made it much easier to classify molecules according to their symmetries. According to the theory, certain symmetry operations may be members of a group if they satisfy the following group axioms [75,80]:

1. The successive application of a set of operations (two or more) shall give an end result equivalent to that of the application of a member of the group. So in mathematical terms we may say that if operations A, B and C are all operations belonging to the same group then we can say that A.B=C (generally A.B is not equivalent to B.A)

2. All groups have the Identity operation E. Therefore we may say that A.E=E.A=A

3. The reciprocal of each operation in the group is also an operation in the group. That is A^{-1}=B. It is important to note that $A.A^{-1}=A^{-1}.A$=E

4. The multiplication of two or more symmetry operations is associative. That is A.B.C= (A.B).C= A.(B.C).

Following the rules above one can easily identify the elements belonging to a certain group which in turn shall make it easy to assign a point group to any molecule. This describes one of the approaches used to identify the point group of a molecule which depends on character tables. A character table is a matrix showing all the symmetry operation belonging to a certain point group. These tables give all the information required on the symmetry operations of a certain group as well as its irreducible representations. We can simply find all the symmetry elements and then refer to the character tables to in order to find the group to which the molecule belongs. Each point group has a unique character table organized into its own matrix. The column headings are the symmetry operations which are grouped into classes. The

first column from the left identifies the name of the point group and then the Symmetry representation labels which also provide information about the degeneracies. Rows are called irreducible representations of the point group. The top row describes the symmetry operations present. The last two columns are referred to as symmetry functions and they provide information about vectors and atomic orbitals. The main body of the table consists of characters (numbers) labeling the symmetry properties of the Molecular Orbitals or modes of molecular vibrations [76,77].

C_{2v}	E	$C_2(z)$	$\sigma_v(xz)$	$\sigma_v(yz)$	Linear rotations	Quadratic
A1	1	1	1	1	Z	x^2, y^2, z^2
A2	1	1	-1	-1	R_z	Xy
B1	1	-1	1	-1	x, R_y	Xz
B2	1	-1	-1	1	y, R_x	Yz

Table (3.3): C_{2v} Character table

The representations are labeled according to certain criteria. Those are:

1. A: if the rotation around a principal axis is symmetrical (degenerate)

2. B: if the rotation around the principal axis in asymmetrical (degenerate)

3. E and T: are doubly and triply degenerate representations respectively

4. The value 1: the symmetry operation leaves the sign or phase of the vector or orbital unchanged (symmetric)

5. The value -1: The symmetry operation results in a change in sign or phase of the vector or orbital (asymmetric)

6. For point groups with an inversion center: the subscript *g* (*gerade*) denotes an even operation with no change in sign.

7. For point groups with an inversion center: the subscript u (*ungerade*) denotes an uneven operation which results in a change in sign.

8. For the groups $C_{\infty v}$ and $D_{\infty h}$ the symbols describe the angular momentum

The water molecule H_2O, for example, belongs to the C_{2v} point group [78].
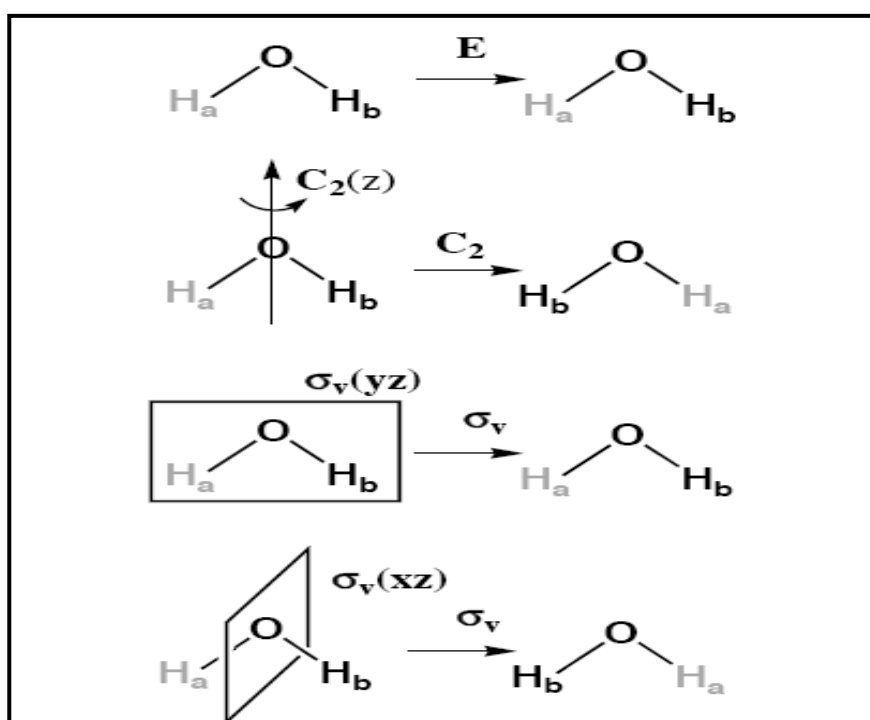
Figure (3.15): Symmetry operations in a water molecule which belongs to the C_{2v} point group

As seen from the diagram above, a water molecule has the following Symmetry elements:

- Identity E

- One rotation about a C2(z) axis

- One reflection in the σv(yz) plane

- One reflection in the σv(xz) plane

To make it even easier, assigning the point group to a certain molecule can be done via the use of the flow chart depicted below [72]:
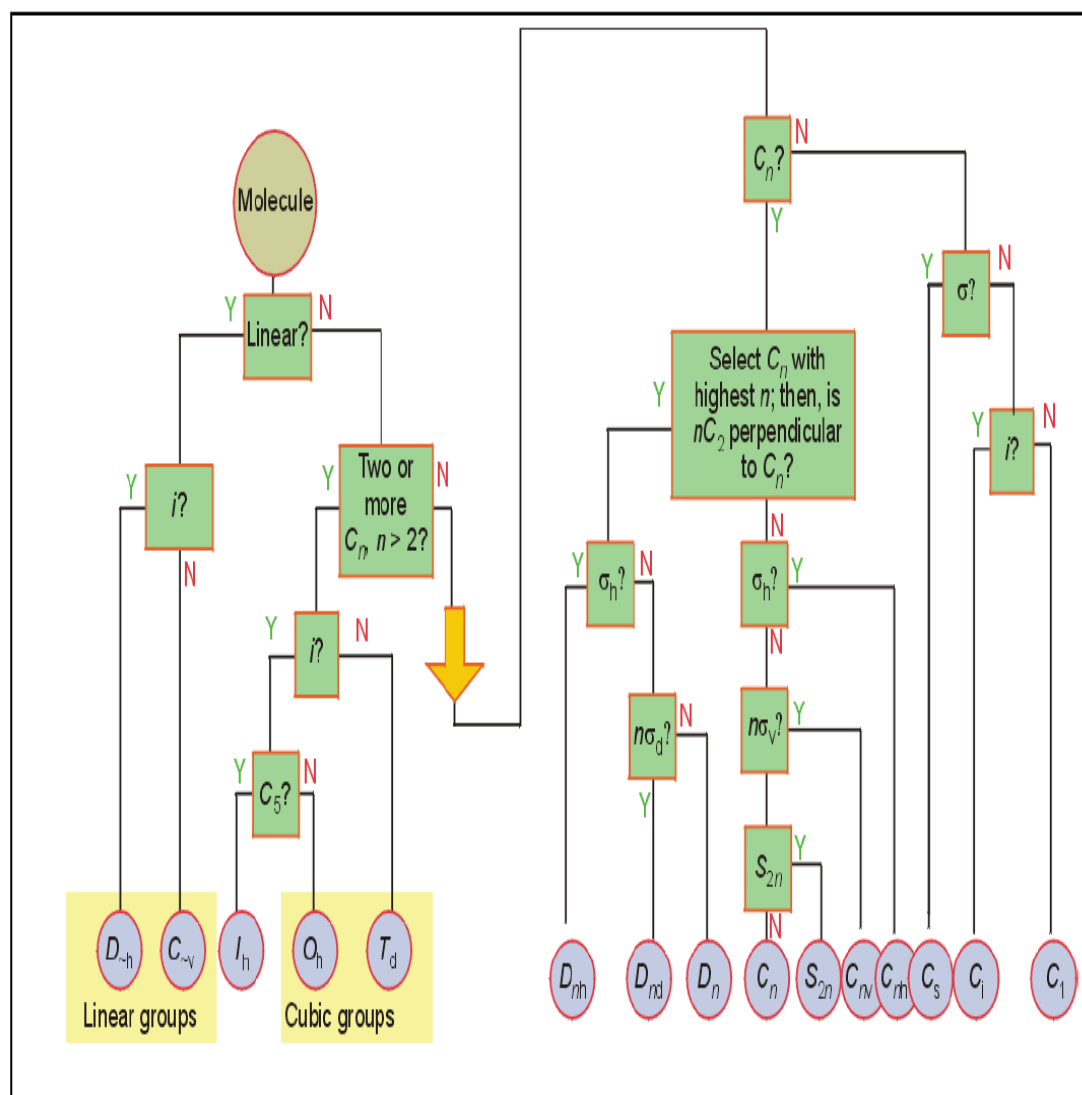
Figure (3.16): Point group decision tree

It is possible with the above flow chart and careful examination of drawings and three dimensional models of a molecule to assign the point group with reasonable efficiency after some practice.[81] A list of point groups has been developed to further simplify matters. The table below [82] has been developed to show the basic characteristics of every point group. Molecules containing these characteristics can be automatically placed under that group. The table however assumes that all ligands (atoms or groups of atoms) attached to the central atom are identical [81].

Point Group		Symmetry Elements Present
C_1		E
C_s		E, σ_h
C_i		E, i
C_n		E, C_n
D_n	n = odd	E, C_n, n$\perp C_2$
D_n	n = even	E, C_n, $^{n/2}\perp C_2'$, $^{n/2}\perp C_2''$
C_{nv}	n = odd	E, C_n, nσ_v
C_{nv}	n = even	E, C_n, $^{n/2}\sigma_v$, $^{n/2}\sigma_d$
C_{nh}	n = odd	E, C_n, σ_h, S_n
C_{nh}	n = even	E, C_n, σ_h, S_n, i
D_{nh}	n = odd	E, C_n, σ_h, n$\perp C_2$, S_n, nσ_v
D_{nh}	n = even	E, C_n, σ_h, $^{n/2}\perp C_2'$, $^{n/2}\perp C_2''$, S_n, $^{n/2}\sigma_v$, $^{n/2}\sigma_d$, i
D_{nd}	n = odd	E, C_n, n$\perp C_2$, i, S_{2n}, nσ_d
D_{nd}	n = even	E, C_n, nC_2', S_{2n}, nσ_d
S_n	n = even only	E, S_n, $C_{n/2}$ and i if $^{n}/2$ odd
T		E, $4C_3$, $3C_2$
T_h		E, $4C_3$, $3C_2$, $4S_{2n}$, i, $3\sigma_h$
T_d		E, $4C_3$, $3C_2$, $3S_4$, $6\sigma_d$
O		E, $3C_4$, $4C_3$, $6C_2$
O_h		E, $3C_4$, $4C_3$, $6C_2$, $4S_6$, $3S_4$, i, $3\sigma_h$, $6\sigma_d$
I		E, $6C_5$, $10C_3$, $15C_2$
I_h		E, $6C_5$, $10C_3$, $15C_2$, i, $6S_{10}$, $10S_6$, 15σ
K_h		E, infinite numbers of all symmetry elements

Table (3.4): Point groups and the symmetry operations associated with each

For a molecule with different ligands around the central atom are not all the same then the symmetry will be lower and the molecule would belong to a lower symmetry point group.

Further description of the symmetry of a molecule at this level is done via Molecular Orbital diagrams, Walsh diagrams as well as Tanabe Sugano diagrams. However, for the purpose of our study assigning point groups and there from space groups which we shall discuss in detail below, is sufficient.

3.6: SPACE GROUPS

However, there is yet another level where we can take this. The symmetry of a molecule may be further investigated. Another known identification method which has existed for centuries is the pane symmetry group [83]. This was a form of mathematical classification of two-dimensional repetitive patterns on basis of the symmetries on the pattern.

What was first introduced as 17 groups was developed over the centuries by Schonflies (1891), Fyodorov (1891), Barlow (1894) and Burkhardt (1967) [84] to become one of the most important classification methods for crystal structures in three dimensions [84]. They are now referred to as space groups and can be described as a blend between the 14 Bravais Lattices and the 32 Crystallographic point groups. The resulting description becomes some sort of combination of the translational symmetry of a unit cell including lattice centering, the point group symmetry operations of reflection, rotation and rotinversion as well as the screw axis (rotation about an axis followed by a translation along the direction of the axis) and glide plane (a reflection in a plane followed by a parallel translation) symmetry operations [85].

These various combinations have resulted in a total of 230 different space groups describing all possible crystal symmetries [85]. The following table shows the names and symbols given to the 14 Bravais lattices [86,87,89].

Symbol	Point Symmetry	Name
aP	I	Primitive triclinic
mP	2/m	Primitive monoclinic
mC	2/m	One-face-centered monoclinic
oP	mmm	Primitive orthorhombic
oC	mmm	one-face-centered orthorhombic
oI	mmm	Body-centered orthorhombic
oF	mmm	(all)face-centered orthorhombic
tP	4/mmm	primitive tetragonal
tI	4/mmm	Body-centered tetragonal
hP	6/mmm	Primitive hexagonal
hR	3m	Rhombohedral
cP	m3m	Primitive cubic
cI	m3m	Body-centered cubic
cF	m3m	Face-centered cubic

Table (3.5): Point Symmetry and space groups

A full table of the 230 available space groups is available in many chemistry and physics books on symmetry.

3.7: SYMMETRY, POINT GROUPS AND SPACE GROUPS OF TETRAHEDRAL MOLECULES

As previously mentioned, tetrahedral molecules is made up of a central atom surrounded by four ligands. The perfectly symmetrical tetrahedral molecule (tetrahedron), where all four ligands are the same, belongs to the point group T_d. All tetrahedral molecules with the formula MX_4 are perfectly symmetrical and all have a T_d point group. As seen in the above table, the T_d group has an order of 24; it has the following symmetry elements: E, $4C_3$, $3C_2$, $3S_4$ and $6\sigma_d$. The symmetry operations for this group are: {E, $8C_3$, $3C_2$, $6S_4$ and $6\sigma_d$} [88].

However not all tetrahedral molecules are perfectly symmetrical since they do not always have four similar ligands around the central atom, and furthermore, they do not always have four ligands around the central atom. the broader definition of a tetrahedral molecule states that it is a molecule with a central atom and four surrounding lone pairs of electrons. If all four pairs of electrons are covalently bonded to four identical ligands we have a perfectly symmetrical tetrahedral molecule with a Td point group.

If all four lone pairs are bonded to a combination of different atoms the symmetry changes. For example the molecule $CHCl_3$ belongs to the C_{3v} point group, the molecule CH_2F_2 belongs to the C_{2v} point group and the molecule CHFClBr belongs to the C_1 point group. Similarly, If only three pairs of electrons are bonded and a lone pair remains the molecule is tetrahedral however the point group is not Td. An example is Ammonia (NH_3), which can be classified as tetrahedral if we consider the lone pair a fourth ligand. Ammonia belongs to the C3v point group. Water (H_2O) can be placed under tetrahedral molecules if we assume Oxygen is the central atom, with two pairs of electrons bonded to the two Hydrogen atoms and two pairs left unbounded. Water belongs to the C_{2v} point group.

Since space groups are combinations of point group symmetries and Bravais lattice structures, the variations in either of these quantities shall cause a variation in the space group of the molecule. So a tetrahedral

molecule in the Td group in FCC shall have a space group different from the same molecule in the same Td group but in BCC. Similarly 2 molecules in the FCC lattice structure with different point groups shall have different space groups. The alternations are many and that is why we have almost 230 space groups and more than 15 describing T_d molecules alone [89]. Methane for example is a tetrahedral molecule belonging to the Td point group. It exists in a number of lattice structures among which FCC, BCC and HCP (hexagonal close packed) are the most common. Henceforth, methane exists in different forms each belonging to a certain space group including *Fm3c* and *Fm3m* [90].

3.8: TETRAFLUOROMETHANE CF$_4$

Carbon Tetraflouride, Tetrafluoro Carbon, Methane tetrafluoride, Tetrafluoromethane are all names given to the molecule with the formula CF_4. Tetrafluoromethane is the IUPAC name, however carbon tetrafluoride is the name most commonly used and therefore we will be using it throughout this thesis. CF_4 belongs to the haloalkane family (specifically halomethane); it is the simplest fluorocarbon. Carbon tetrafluoride is one of the strongest bonded haloalkanes due to the nature of its constituent atoms. The carbon and four fluorine atoms are bonded together by means of a polar covalent bond as seen below [91].
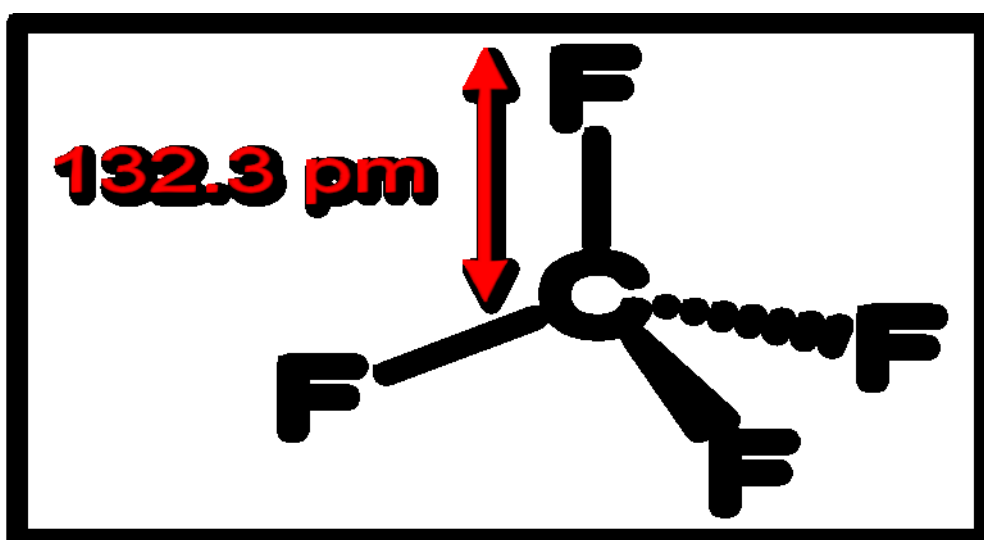
Figure (3.17) Tetrafluoromethane

As a matter of fact fluorine is the most electronegative atom in the periodic table. It has an electronegativity of 4.0 on the Pauling scale [92,93]. Carbon on the other hand has an electronegativity of 2.55 on the Pauling scale [92,93]. Electronegativity is defined as an atom's tendency to attract the bonded pair of electrons towards itself. Thus the central carbon has a significant positive charge due to the displacement of the bonded electrons towards the Fluorine atom. The difference in electronegativities of the atoms and the displacement of partial positive and negative charges towards carbon and fluorine respectively provide an ionic character to the molecule which shortens and highly strengthens the four C-F bonds. The diagram below describes to point stated above [94]:
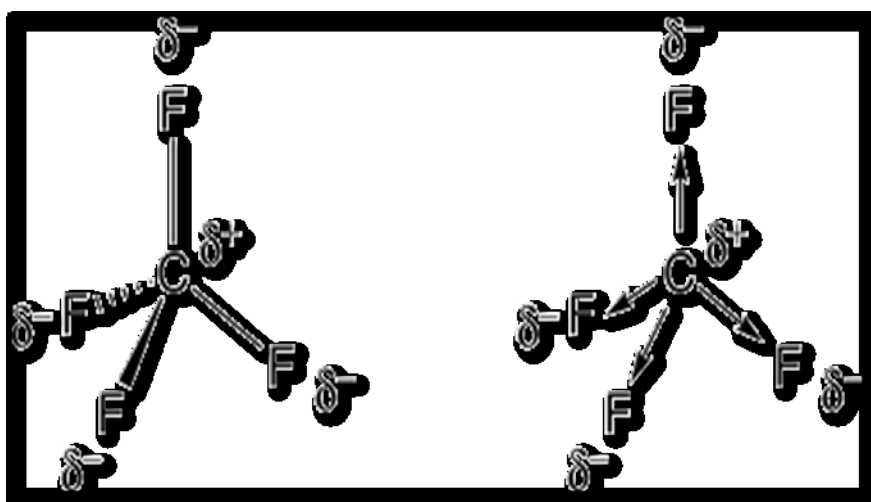
Figure (3.18): Partial positive charge on the central Carbon atom as a result of the displacement of the bonded electrons towards the highly electronegative Fluorine atoms

Although the large difference in the electronegativities results in partial polarity of each of the constituent atoms, the molecule itself is nonpolar. This is due to the symmetric shape of the molecule and the tetrahedral arrangement of the atoms which evenly distributes the partial charges thus avoiding a displaced polarity. The high stability of CF_4 due to the strength and shortness of the C-F bonds (bond energy of 515 kJ/mole) has resulted in the inertness of the molecule towards most acids and hydroxides. It is however strongly, and sometimes explosively reactive towards alkali metals. Combustion of CF_4 produces toxic gases including hydrogen fluoride,

carbonyl fluoride and carbon monoxide. CF_4 is insoluble in water under normal conditions, however it is miscible in most organic solvents.

In nature, CF_4 is a gas, which is odorless, colorless, inflammable and highly inert. It is a greenhouse gas too, that is it absorbs and emits radiation within the thermal infrared range being one of the many gases which contribute to the greenhouse effect. Carbon tetrafluoride is produced when any carbon compound or pure carbon is burned in a fluorine atmosphere. It is prepared via a number of methods among which the most commonly used are the fluorination of carbon dioxide, carbon monoxide or phosgene. Industrial manufacture of CF_4 is usually prepared from either dichlorodifluoromethane or chlorotrifluoromethane. Other methods for the preparation of CF_4 include the following [95]:

1. Treatment of metal carbides with Fluorine at 20° C for 90 min or with CoF_3 at 440^oC for 9 hours. The yield of CF_4 in this process is between 68% and 90%.

$$MnCm + F_2 \xrightarrow{20^oC,\ 90\ min} CF_4 + C_2F_6 + C_3F_8 + C_4F_{10}$$

2. Fluorination by the use of BrF3 of tetrachloromethane or tetraiodomethane under available conditions or tetrachloromethane at yields around 60% CF_4

$$Cl_4 + BrF_3 \xrightarrow{RT} CBr_2F_2 + CF_4$$

3. Electrolysis of concentrated trifluoroacetic acid solutions with platinum electrodes.

$$CF_3CO_2H \xrightarrow{electrolysis} CF_4 + C_2F_6 + CO_2 + O_2$$

4. Electrochemical fluorination of dimethyl sulfide in the presence of HF yields 60% CF_4

$$SMe2 \xrightarrow{HF\ electrolysis} CF_4 + (CF_3)_2SF_4 + CF_3SF_3 + SF_6$$
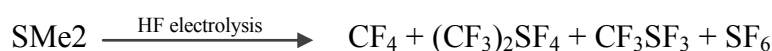
5. Oxidation of tetrafluoroethylene with dioxygen

$$C_2F_4 \xrightarrow{\text{Oxygen}} CF_4 + CO_2$$
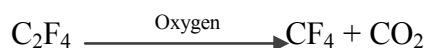
Carbon tetrafluoride is used in many applications. It is utilized as a refrigerant at low temperatures. Due to the reactivity of the fluorine atom, and the mixed nature of the molecule itself (ionic, polar and neutral) it is one of the most commonly used compounds in a number of plasma assisted electronic chip manufacture processes. In semiconductor processing and preparation as a plasma etchant for silicon, silicon dioxide and silicon nitride and a number of other films. Another important process in semiconductor processing is strip resistance which again is one of the applications of CF_4. Furthermore CF_4 is mixed with oxygen and used for plasma cleaning of Chemical Vapor Deposition (CVD) reactors. Perhaps it is for this reason that carbon tetrafluoride has gained all this wide interest. It can, under certain conditions, be used in semiconductor manufacture which due to its availability and relatively reasonable price makes it an ideal tool. Much research is being carried out today in order to regulate the use of tetrafluoro carbon in semiconductor processes. The fact that it is a greenhouse gas with high greenhouse warming potential is alerting. a number of available recovery/recycle technologies have been introduced to exhaust gases containing CF_4 (and perfluorocarbons (PFCs) in general) from semiconductor manufacturing. Such techniques include cryogenic condensation/distillation, pressure swing absorption and membrane separation.

The molar mass of CF_4 is 88.0043 g/mol. Its melting point is at 90K and boiling point is at 145 K. CF_4 is a tetrahedral molecule (diagram 3.17), with a C-F bond length of 1.32 +/- 0.02A [96]. Carbon tetrafluoride is known to exist in a number of phases thus existing in a number of different crystal lattice structures including monoclinic, FCC and BCC. It belongs to the T_d point group and with the variations in its crystal structure its space group varies too. It is assigned a number of space groups including $cP2_1/c$ and C2/c. This shall be discussed and analyzed in detail through our simulation results [97].

CHAPTER 4: MOLECULAR DYNAMICS SIMULATION: METHODOLOGY, SET UP AND ANALYSIS

There is a wide range of molecular dynamics software packages that one can use to carry out a simulation. Different software has different features and their own merits. There are certain criteria one must satisfy when choosing the appropriate software to use in their work. First and foremost the program must be user friendly. It is preferable that the program be suitable to use with more than one operating system. It should be designed to provide output data which includes all or most of the information we are intending to study in order to avoid using more than one program at a time. Preferably the program should be as general as possible so as to treat any system within its designed PBC.

For the purpose of our study we have chosen to use Keith Refson's software named Moldy (Molecular Dynamics) [41]. Moldy is a 'highly portable C program for performing molecular-dynamics simulations of solids and liquids using periodic boundary conditions' [41,98]. Its code is reasonably easy to understand and utilize. It has also been optimized as to give the highest possible performance whether used in serial mode on a conventional workstation or on parallel system via interfacing. The Program has the following specifications [41,98], most of which we have already discussed in previous chapters:

1. It uses a standardized system whose information is fully detailed in an input file which also specifies the run time as well as the constitution of the system which can be made of any number of atoms, ions, molecules or mixtures of these constituents in any ratio.

2. The input file, referred to as system specification file, includes information on the nature of the system, the type and number of each species, the interaction potential used as well as the initial configuration.

3. Information needed to initiate and control a run is given in a file named control file. The control file should include all parameters governing the simulation run including but not limited to the number of time steps, the length of each time step, the frequency of the output, the name of the output file, the name of the specification file, as well as any other files to be read from or written to.

4. The program does not support many-body forces, polarizable atoms, point dipoles or higher multipoles.

5. Molecules are treated in the rigid molecule approximation and the system is assumed to contain atoms and molecules interacting via pair wise potentials.

6. Rotational motion is modeled using quaternion method.

7. MD cell does not necessarily have to be cubic which means the program can simulate both solids and liquids.

8. Equations of motion are integrated using the modified Beeman algorithm.

9. All simulations are performed in NVE ensembles or isobaric / isothermal ensembles.

10. It supports L-J potential functions; however the minimum image convention method is not used. Instead the program takes into account interactions of all images of a particle present within the cutoff radius.

11. The Ewald summation method is used to calculate long range electrostatic forces.

12. Moldy provides a number of output files. Most important are the periodic output file, which contains instantaneous values of a number of thermodynamic quantities, the restart file for state information and the dump files which store permanent record of the particle trajectory.

Having chosen the most appropriate and compatible program, and having also decided on the system and the problem being studied one may now proceed with the simulation.

4.1 THE INITIAL CONFIGURATION OF THE SYSTEM

It is quite important when starting a simulation to provide as much information as possible about the system. As previously mentioned, we need to specify the type and number of molecules in the system. Furthermore, the initial configuration of the system must be specified too. That is, we must specify the initial positions and initial velocities of each and every atom as well as a number of other quantities depending on the type of simulation we shall be performing, whether it shall be constant temperature, constant pressure, both or neither. However, first we need to establish the main structure of the system with its positions and velocities at the beginning to the MD run [98].

4.1.1 The Initial positions and velocities

Positions of atoms and their orientations within a unit cell may be chosen randomly. As the simulation run proceeds the system will rearrange itself and shall adapt the equilibrium position, which we can find through the analysis of the dump file. For further runs we may either choose to place the atoms randomly as we had done before or we may use the equilibrium positions which had been provided by the previous run, a process known as the skew start method [98]. Furthermore, positions of atoms within the unit cell in relation to one another may be attained from previous experimental or analytical data. The figure below [99] shows the placement of atoms at the lattice points in the initial configuration of a Molecular Dynamics system.
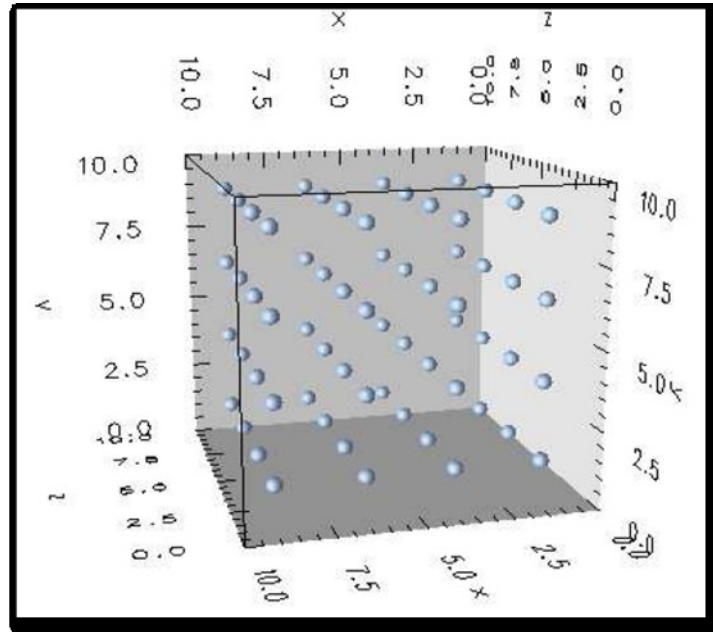
Figure (4.1): Placement of atoms at lattice points in the initial configuration structure adopted during a Molecular Dynamics Simulation

Usually, one of the well known Bravais lattice structures is used to describe the initial configuration of the system. When it is one of the possible structures of the system, the Face Centered Cubic (also referred to as cubic close packed) unit cell structure is the one most commonly used. As a matter of fact this is the structure we shall be using an initial configuration throughout our work.

In FCC, the system contains a number of molecules equal to $4M^3$, where M={2,3,4,5,...}. This is why the system must consist of a number of molecules equal to {32, 108, 256, 500, 864....}. In our work we used systems containing 108, 256 and 500 molecules.

The standard orientations of the molecules within this type of unit cell may be used as the initial orientations of the molecules in the system. Otherwise, we may use random orientations or we may randomize the standard orientations of a given structure. In our work we randomized the orientations suggested by the FCC lattice structural parameters. However, it is very important to know that the closer the orientations are to the real orientations of the atoms and molecules, the faster the simulation shall

proceed. It is always best to choose those orientations which somehow reflect the expected orientations which the molecules shall adopt as the system equilibrates.

As in all molecules in the FCC structure, the packing percentage is somewhere in the range of 74%. relating this information to the relative density of the molecule can be useful when deciding upon the system size.

The figure below shows the CF4 crystal in FCC/ccp arrangement [97].
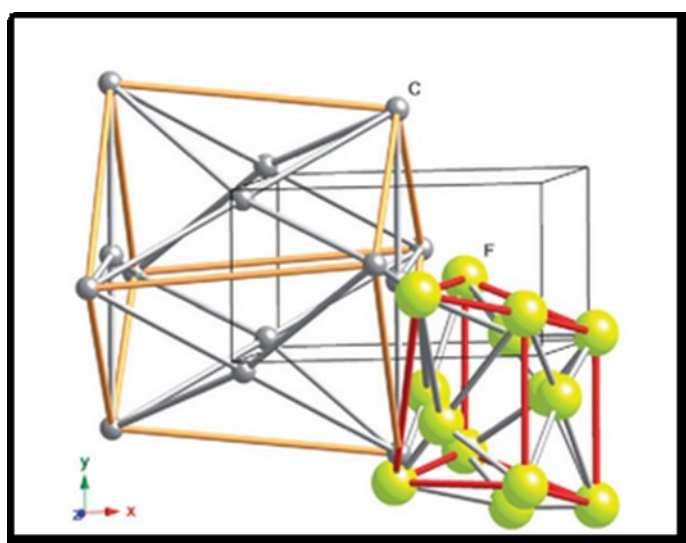
Figure (4.2): Carbon Tetrafluoride in a Face-Centered Cubic lattice structure

Up until this point things seem rather easy. However, when it comes to setting the initial velocities of the atoms in the system, things are a little more complicated. Whether the positions have been chosen randomly or on basis of a previous run, or via the use of the skew start, the initial center of mass velocities shall be defined using the Maxwell-Boltzmann distribution at the temperature which has been specified for the simulation. In other words, initial velocities shall be chosen from a set of random numbers belonging to a Gaussian distribution, and shall then be normalized to achieve a probability density $p(v)$ of x,y,z components of the velocity of molecule k. That is we will be measuring the probability that an atom will have a certain velocity when that atom of given mass is in a certain position at a fixed temperature. This relation for the probability distribution is given by [98]:

$$p(v_{ik}) = \left(\frac{m_k}{2\pi k_B T}\right)^{1/2} \exp(-\frac{m_k v_{ik}^2}{2k_B T})$$

(4.1)

With a random number generator one can now generate several random numbers R_{ik} with unit variance. Using the random number R_{ik} we arrive at the following equation to describe each component of velocity [98]:

$$v_{ik} = \sqrt{\frac{k_B T}{m_k}} R_{ik}$$

(4.2)

The 'k' in this equation specifies the direction vector for x,y or z components.

Having specified the velocities of each point in the system, we must now find the angular velocities which measures the rate of change of angular displacement and which is particularly important in our work [100]. The same procedure used to assign the velocities is now used to assign the angular velocities. First we have the probability distribution function given by [98,101]:

$$p(\omega_{ik}^P) = \left(\frac{I_{ik}}{2\pi k_B T}\right)^{1/2} \exp(-\frac{I_{ik}(\omega_{ik}^P)^2}{2k_B T})$$

(4.3)

The quantity I_{ik} is the moment of inertia of molecule i around the k direction.

Second, a random number generator is used to assign a random number velocity given by the following relation [98]:

$$\omega_{ik}^P = \sqrt{\frac{k_B T}{I_{ik}}} R_{ik}$$

(4.4)

Moldy uses quaternions to describe angular coordinates. The quaternions equations which we have previously described are treated to give us the following relations for the first and second derivatives:

$$\dot{\mathbf{q}} = \mathbf{q}(0, \omega^P/2)$$

(4.5)

$$\ddot{\mathbf{q}} = -\frac{1}{4}(\omega^P)^2\mathbf{q}$$

(4.6)

It is worth mentioning that if and when a molecule has less than three degrees of freedom, their corresponding angular velocities are set to zero.

Now that we have decided upon the initial positions, velocities and angular velocities of the system, we need to decide on the quantities which we shall keep fixed or constant throughout the molecular dynamics run. Usually the total energy 'E' of the system is conserved throughout the time of the run. That is, at each time step the sum of the kinetic energy and potential energy is zero. A number of constraints may be applied in order to minimize the fluctuations in total energy during a run. Furthermore, the total linear momentum 'P' of the system is set at a value of zero in order to avoid the movement of the molecular dynamics lattice (box) during the run. This is simply achieved via simple arithmetic steps involving the division of total linear momentum in each direction by the mass of the system thus giving velocities of the particles which is then utilized to adjust the momentum to zero. This is also repeated at each time step to ensure conservation of the set quantities.

4.1.2: Temperatures and Pressure

The initial temperature and pressure of the MD simulation are very important. Therefore they need to be thoroughly understood in order for us to be able to assign the most suitable values for both quantities at the beginning of each run. The trajectories produced for each particle are quite sensitive to small perturbations in the initial conditions provided, which makes precision a must. The temperature and pressure of a system are among the most important equilibrium properties and assigning them correctly shall help greatly in producing correct results. Some MD simulations are performed under conditions which conserve energy, volume and momentum. The regular most commonly used ensemble is one where number of molecules N,

volume, V and energy E are kept constant. Such microcanonical system is commonly referred to as an NVE ensemble. However, In certain situations it is desirable to perform simulations of a fluid for particular values of temperature and/or pressure or under conditions which allow energy and volume to fluctuate. The ensemble thus becomes an NVT or NPT rather than NVE. There are a number of methods which involve performing MD simulations under constant temperature and/or pressure. In these methods, the time averages of properties are equal to averages over either the isoenthalpic-isobaric, canonical, or isothermal-isobaric ensembles. These methods are particularly important when studying certain problems such as pressure induced phase transitions, temperature induced phase transitions, and isothermal-isobaric experimental ensembles.

4.1.2.1 MD simulations at constant Temperature

At constant temperature, the energy of a system of N particle fluctuates. In order to perform simulation on such a system a mechanism for introducing energy fluctuations must be devised.

The relation between temperature and K.E of the system (calculated from the K.E for every degree of freedom) of an unconstrained system may be given by [101]:

$$\langle K \rangle_{NVT} = \frac{3}{2} N k_B T \qquad (4.7)$$

In order to perform MD simulations at constant temperature one should simply keep the kinetic energy of the ensemble constant by scaling the velocities at each time step. However, this when performing this simulation method, one cannot be sure that the configurations produced belong to the canonical ensemble. However, a constraint method has been introduced by Hoover *et al* where an additional velocity dependent term is inserted into the relation and summed up with the force to maintain a constant total kinetic energy [102]. Thus, the produced canonical ensemble shall only involve potential energy terms for the kinetic energy fluctuations shall be suppressed.

So if we use the method of velocity scaling, the scaled velocities are given by [101]:

$$v_i(t+\Delta t/2) = v_i(t + \Delta t/2)\beta + F_i(t)\,\Delta t/m \qquad (4.8)$$

Where: $\beta^2 = (3(N-1)kT_r/m) \Big/ \sum_{i=1}^{N} v_i^2(t - \Delta t/2)$ \qquad (4.9)

The removal of three degrees of freedom by placing a constraint of zero to the total linear momentum is what results in the introduction of the factor N-1 rather than N in the above relation.

Now that the velocity is scaled for each particle, the change in the temperature of the system from the temperature T(t) at time *t* when the velocities have been scaled by a factor λ is given by:

$$\Delta T = \frac{1}{2}\sum_{i=1}^{N}\frac{2}{3}\frac{m_i(\lambda v_i)^2}{Nk_B} - \frac{1}{2}\sum_{i=1}^{N}\frac{2}{3}\frac{m_i v_i^2}{Nk_B} \qquad (4.10)$$

$$\Delta T = (\lambda^2 - 1)T(t) \qquad (4.11)$$

$$\lambda = \sqrt{\frac{T_{new}}{T(t)}} \qquad (4.12)$$

Having arrived at a simple factor linking the new temperatures (the required one) with the current temperature T(t), one can easily control the temperature by simply multiplying the velocity at every time step by the factor given by equation (4.12).

Another method to keep control over the temperature is to bring the system into contact with an external heat bath, also referred to as a thermostat [102]. The temperature of the thermostat is fixed prior to the experiment in order to provide or remove thermal energy from the system as per our simulation requirements. Under these conditions, the probability of finding the system in a given energy state is simply given by the Boltzmann distribution. This constant temperature method was first proposed by

Andersen in hope to achieve a certain overall temperature for the system. The velocities, as before, shall be scaled at each time step.

The rate at which the temperature changes is proportional to the difference between the temperature of the system and that of the external thermostat. This relation below is what we shall use to scale the velocities accordingly. Assuming τ is the coupling parameter, then:

$$\frac{dT(t)}{dt} = \frac{1}{\tau}\left(T_{bath} - T(t)\right) \tag{4.13}$$

The change in temperature after every time step is given by:

$$\Delta T = \frac{\delta t}{\tau}\left(T_{bath} - T(t)\right) \tag{4.14}$$

which results in a velocity scaling factor given by:

$$\lambda^2 = 1 + \frac{\delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right) \tag{4.15}$$

However, the previously mentioned methods cannot be applied to all ensembles. They are particularly not suitable in some cases to the extent that the results acquired may be so far from the real values. This is where Anderson's advanced coupling method proved very useful. This approach, also referred to as the stochastic collisions method couples the system to an external heat bath. The coupling may be represented by stochastic impulsive forces that act occasionally on randomly selected particles. [102,103] The collisions of the system with the heat bath transports the system from one energy shell to another. Between these stochastic collisions, the system evolves at constant energy. The collisions ensure that all accessible constant-energy shells are visited according to their weight in the Boltzmann distribution. To perform the simulation, two terms must be decided upon; the first is the desired temperature 'T' and the second is the mean rate at which each particle is subjected to stochastic collisions 'v'. The probability that a particular particle is suffers a stochastic collision in any small time interval is

given by '$v\Delta t$'. Furthermore, a random number generator is used to decide upon the time intervals between successive collisions of each particle. The intervals are described according to:

$$P(t) = \nu\, e^{-\nu t}$$

(4.16)

Having chosen the initial set of positions and velocities, the Hamiltonian equation of motion is integrated until the time of the first stochastic collision. If the particle suffering from this first collision is given the name 'i', the value of the velocity of particle 'i' is chosen at random from a Boltzmann distribution at temperature T. The same is applied to the momentum of particle 'i' [103]. The changes to velocity and momentum takes place instantly and then the equation of motion is solved for all particles until the occurrence of the next stochastic collision and so on. The results obtained by this procedure is a trajectory specified by positions, velocities and/or momenta which is then used to calculate time averages for any other function. The time average for any of the functions that we calculate from this resulting trajectory is equal to the ensemble time average of the function for the canonical ensemble at temperature **T**.

The effect of the overall process described by the stochastic collisions method is analogous to a series of microcanonical simulations set at a series of different energy levels represented by a Gaussian function. However, a more capable method was later introduced by Nose and developed by Hoover [104]. The method is referred to as the Nose-Hoover thermostat method and it is a deterministic method. The main idea behind this method is that it considers the thermostat or heat bath as an integral part of the system. Furthermore, it fixes the average temperature of the system under simulation but meanwhile allowing for a fluctuation of the temperature with a distribution of canonical nature. The heat bath used is introduced with an extra degree of freedom (s). The potential energy of the reservoir is described by $(f+1)k_B T \ln(s)$, with f being the number of degrees of freedom and **T** the desired temperature. Furthermore the

velocities of the real atoms in the system (excluding those posed by the external thermostat) are given by :

$$\mathbf{v}_i = s \frac{d\mathbf{r}_i}{dt} \quad \text{where } r_i \text{ is the position of particle i.} \tag{4.17}$$

4.1.2.2 MD simulations at constant Pressure

At constant pressure the volume of the system of N particles is allowed to fluctuate. Constant pressure simulations require PBC. The pressure is controlled via dynamically adjusting the size of the unit cell and rescaling atomic coordinates (other than those of fixed atoms) during the simulation [103]. The change in the simulation cell can be isotropic where the shape of the cell remains unchanged, or it can be anisotropic where the shape of the cell changes.

To describe the fluctuations in the volume of the system under constant pressure, a molecular dynamics method is devised where the volume of the system is treated as a variable rather than a constant. So for example we may have a simulation of solid-state phase transition which allows for changes in the size or symmetry of the unit cell. This is particularly important for our work, for it allows us to know the new structure, symmetry, point group and space group of the system after a phase transition since we had allowed the size and symmetry of the cell to fluctuate.

Anderson's introduction of the constant pressure MD method was a breakthrough in the MD simulation field [103]. Anderson introduced a modification to the MD method where the average volume was to be determined by the ratio between internal and externally set pressure. Having determined the volume, it is then allowed to fluctuate. Anderson's ensemble was referred to as an HPN ensemble since it required enthalpy conservation. Anderson's ensemble was developed by Rahman and Parrinello to allow for changes in the shape of the MD cell [103]. Today, the Extended Lagrangian Method is the most commonly used for simulations at constant pressure. According to this method, the pressure tensor shall depend on the chosen

coupling scheme. The virial is the product of the positions and the derivative of the potential energy function. According to the virial theorem the atomic and molecular pressure of a system of N molecules are given by [105]:

$$P_{\text{atm}} = \langle \Pi_{\text{atm}} \rangle = \left\langle \frac{1}{3V} \sum_{\alpha}^{N} \sum_{i}^{N\alpha} \left(\frac{\mathbf{p}_{\alpha i}^{2}}{m_{\alpha i}} + \mathbf{r}_{\alpha i} \cdot \mathbf{f}_{\alpha i} \right) \right\rangle \qquad (4.18)$$

$$P_{\text{mol}} = \langle \Pi_{\text{mol}} \rangle = \left\langle \frac{1}{3V} \sum_{\alpha}^{N} \left(\frac{\mathbf{P}_{\alpha}^{2}}{M_{\alpha}} + \mathbf{R}_{\alpha} \cdot \mathbf{F}_{\alpha} \right) \right\rangle \qquad (4.19)$$

However, there are a number of simpler arithmetic to be used when dealing with constant pressure simulations. Allowing the volume of the system to fluctuate we can then relate the relative fluctuation to a factor referred to as the isothermal compressibility 'β' where:

$$K = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_{T} \qquad (4.20)$$

The isothermal compressibility factor is a measure of the relative change in the volume of the system as a response to a change in pressure. The compressibility factor is very important. at constant pressure it is directly proportional to changes in volume, whilst at constant volume it is inversely proportional to changes in volume. Furthermore, for an isobaric system, a variation in volume may be induced by either a one dimensional or a three dimensional volume change. The mean square volume displacement is related to K by [101]:

$$K = \frac{1}{k_{B}T} \frac{\langle V^{2} \rangle - \langle V \rangle^{2}}{\langle V^{2} \rangle} \qquad (4.21)$$

There are a number of methods that are used to control the pressure during a simulation. Among these methods the most widely used are: the Berendsen Method, the Andersen Method and the Parrinello-Rahman method [102,106]. The Berendsen et la method is the one we shall focus on. It relies on the weak coupling to an external bath. The Langevin equation of motion is

then modified, eliminating the stochastic force and replacing it with a constant friction term with a variable friction proportional to the constraint. The equations utilized in this particular method are simple thus allowing for easy implementation. According to this method the rate of change of pressure with time is described by the following relation:

$$\frac{dP(t)}{dt} = \frac{1}{\tau_p}\left(P_{bath} - P(t)\right) \tag{4.22}$$

given that τ_p is the coupling constant also referred to as the relaxation constant, P_{bath} is the pressure of the designed bath, and $P(t)$ is the actual pressure at time t. Now we may proceed to introduce a scaling factor 'λ' to the volume of the system

$$\lambda = 1 - K\frac{\delta t}{\tau_P}\left(P - P_{bath}\right) \tag{4.23}$$

The same factor is applied to scale the coordinates by a factor $\lambda^{1/3}$. The new positions are now calculated using the following equation:

$$r_i^{\backslash} = \lambda^{1/3}r_i \tag{4.24}$$

We may experience difficulty when it comes to determining the ensemble on which these results are based. Anderson was the first to introduce a new method to overcome such problem. The extended system method of Andersen introduced an additional degree of freedom which leads to mass dependent oscillations in both volume and pressure. Andersen preserved the shape of the cell while allowing the volume of the cell to change isotropically. The basic idea of the system is to treat the volume of the cell as a dynamic variable. Then the Lagrangian of the system is modified so that it contains a term in K.E with a user defined mass and a potential term which is derived from an external pressure acting on the volume of the system like a piston. In this case we have a (NTP) ensemble where the MD cell is considered a cube of length $V^{1/3}$ with real coordinates r_i being expressed as:

$$r_i = V^{1/3}x_i \qquad (4.25)$$

where x_i is a scaled coordinate with values of its components being limited to a range from 0 to 1. Similarly the velocity is expressed by:

$$\dot{r}_i = V^{1/3}\dot{x}_i \qquad (4.26)$$

The coupled Newtonian equations of motion for this type of system is given by:

$$\ddot{r}_i = \mathbf{F}_i/mV^{1/3} - (2/3)\dot{r}_i\dot{V}/V \qquad (4.27)$$

$$\ddot{V} = (P - P_\mathrm{E})/M \qquad (4.28)$$

Moldy utilizes the 'extended-Lagrangian' methods of Nose and Hoover as well as Parinnello and Rahman to implement the isobaric and isothermal ensembles when required.

4.2 RUNNING THE SIMULATION

Now that we have decided upon the initial conditions as well as the initial positions, velocities and ensemble type we shall be using, we are ready to run the program. The control as well as the system specification files is written, the first to provide the parameters governing the run and the second to define the system itself. Moldy is given the code to start the run. If for any reason the run does not immediately start one must revise the input file.

As the simulation run proceeds, a number of trajectories are being calculated at each time step. The potential function at each time step is very important. Integration on the function provides the force on each atom at every time step. For atoms governed by the Lennard-Jones potential, the force on atom i by atom j is given by [101]:

$$f_{ij} = \frac{r_{ij}}{|r_{ij}|}\frac{24\varepsilon}{\sigma}\left[2\left(\frac{\sigma}{r_{ij}}\right)^{13} - \left(\frac{\sigma}{r_{ij}}\right)^{7}\right] \qquad (4.29)$$

Knowing that the force on atom *i* by atom *j* is equal in magnitude and opposite in direction, thus the force on atom *j* is equal to the negative of the force on atom *i*. That is: *fij* = - *fji*. This simplifies matters for it enables us to calculate the forces between any pair of atoms only once.

4.2.1 Phases of the MD Run

Once it has begun, the MD run shall smoothly continue unless given a termination order. The system automatically adopts the ensemble which has been designed. Integration algorithms are carried out on basis of the method we have previously chosen and the type of potential used is the Lennard Jones potential. As the run continues it passes through a number of stages each having its own properties which when compiled with those of the other stages provides means of simplifying the output files and clarifying it as much as possible.

As suggested by the chart below, the molecular dynamics run may be segregated into two distinct phases: the equilibration phase and the production phase [107].
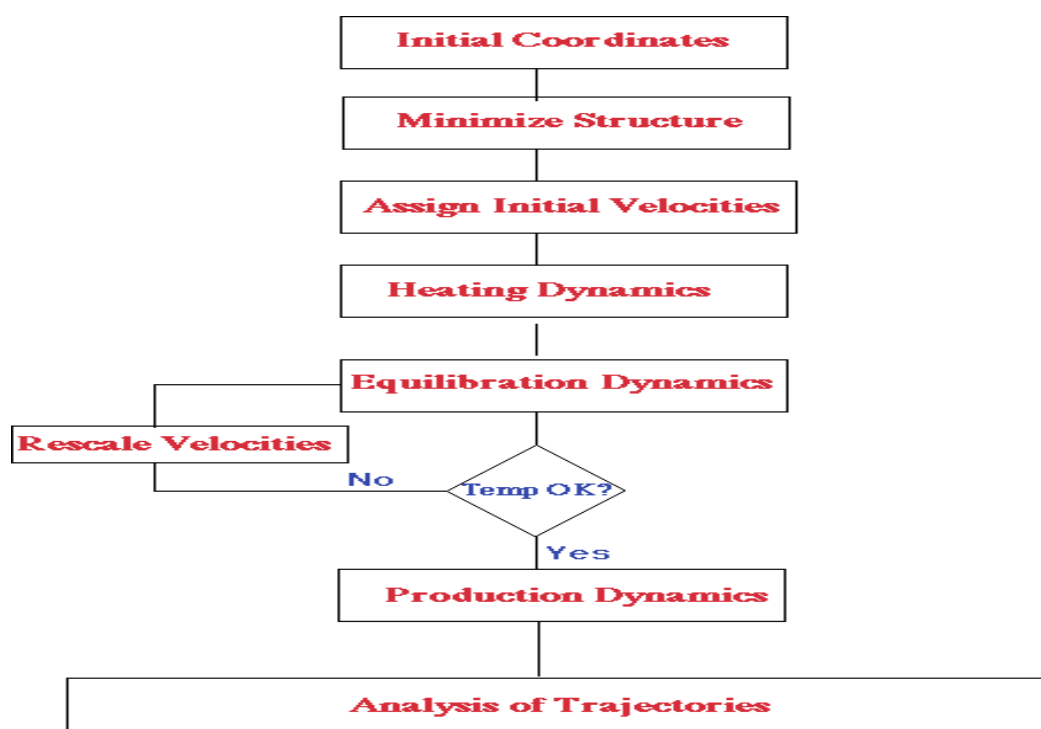
Figure (4.3): Phases (stages) of a molecular dynamics run

4.2.1.1 The Equilibration phase

During the equilibration phase the system is driven from the initial state to the equilibrium state. Once the desired temperature has been reached, the simulation system continues through which time several properties are mentioned, particularly the structure, the pressure, the temperature, the velocity and the energy. The aim of the equilibration phase is to run the simulation until these properties have stabilized with respect to time.

Once the desired temperatures and pressures are reached then the run can be terminated. During the equilibration phase, and on basis of the ensemble type we have used, we must monitor certain quantities very closely. In the case of a microcanonical ensemble for example, we must monitor the total energy at every time step. Once we have reached a stage where the variations in total energy can be disregarded, (i.e. within one in10^4) we can say that the system has reached equilibrium since we have now reached a point where we have conservation of the total energy. For an NVT ensemble we must monitor the temperature and the kinetic energy at every time step, similarly with the NPT ensemble we may need to monitor the changes in volume and/or structure of the unit cell.

To sum it all up, it is very important to remember that the equilibration phase is the most important part of the simulation. In order for us to make the best use of this phase one must be very careful while deciding when to terminate the equilibrium phase and start the production phase. In order to do so, a number of properties are to be closely monitored during the time of the simulation particularly those quantities which we expect to conserve or drive to a certain value during the simulation. Such quantities include, as we have mentioned before, the total energy for a microcanonical ensemble and the temperature for an isothermal ensemble. Besides these quantities, we must also monitor certain parameters which describe the overall structure of the system. Such structural parameters are referred to as 'order parameters'.

Order parameters give us an idea about the degree of order/disorder present in a certain system at a certain point in time. It is usually defined as a set of observable physical quantities which are able to distinguish between two distinct phases [108]. It is usually zero in one phase (usually above the critical point) and non- zero in the other thus allowing us to differentiate between the two. Furthermore, it characterizes the onset of order at the phase transition. There are a number of order parameters including, the translational order parameter, the rotational order parameter, the tetrahedral order parameter and the bond orientational order parameter. among these the most commonly used, and particularly in our work, are the translational and rotational order parameters which may be described as follows[109]:

- *Translational order parameter:* Used to distinguish between the solid and liquid phase of a system. In our work we have decided to start the simulation in the solid phase (FCC structure), while we expect that in a number of runs the simulation shall terminate with the system in the liquid phase. To measure the translational order parameter of a system initially having an ordered FCC structure Verlet introduced the following relation:

$$\lambda = \frac{1}{3} \left[\lambda_x + \lambda_y + \lambda_z \right] \qquad (4.30)$$

Where *a* is the length of the side of a side of the unit cell and:

$$\lambda_x = \frac{1}{N} \sum_{i=1}^{N} \cos(\frac{4\pi x_i}{a}) \qquad (4.31)$$

At the beginning of the simulation the Cartesian coordinates are all multiples of a/2 and the translational order parameter is unity. As the simulation time advances it comes to a point when all positions are randomized, at which time the translational order parameter gradually decreases to zero. As the system reaches equilibrium the translational order parameter fluctuates in inverse proportion to the square of the size of the unit cell i.e. proportional to $1/\sqrt{N}$.

- *Rotational order parameter*: Used to determine the orientation state of the system and degree of order exhibited. Viellard was able through his work with Baron to suggest a relation for the rotational order parameter which is now commonly used in most MD simulations. The rotational order parameter is given by:

$$P_l = \frac{1}{N}\sum_{i=1}^{N}\cos(\gamma_i) \hspace{3cm} (4.32)$$

where γ_i is the angle between the current and original direction of the molecular axis of molecule *i*. When the molecules are perfectly aligned the value of Pi is unity. Complete rotational disorder is characterized by a value of zero. Fluctuations about the average is also proportional to $1/\sqrt{N}$.

4.2.1.2 The Production Phase

Once we have reached this state of dynamic equilibrium the runs can be terminated. The following phase of the MD run is referred to as the production phase during which thermodynamic parameters as well as other information is calculated. During this phase, yet another set of factors must be checked and monitored in order to ensure that the results we have obtained are indeed sampled from the ensemble we had used. In addition to the conservation of energy, we must also ensure that the resulting velocities describe a Boltzmann distribution, and similarly the kinetic energy must be equally distributed among the three Cartesian directions. During the production phase a number of temporary output files are produced at regular intervals including such information as the positions, energies, velocities and other dynamical quantities of the system at certain times intervals. These folders are important when defining certain quantities. The positions for example are extracted from one of the imtermediate files written during the production phase, namely known as the Dump file. In addition to the above, if we need to start a run from the last configuration of a previous run then the dump file is utilized as an initial configuration for the new run.

Furthermore, it is during this production phase that a number of critically important parameters are calculated. That is simply because the production phase is the only period in the term of the simulation run when the system is physically stable/consistent.

Molecular Dynamics properties are functions of positions and velocities both being function of time. When we need to measure a physical quantity during the simulation, a time average of this quantity is performed, given that the quantity is averaged over the total time of the production phase rather than the entire simulation time for the same reasons mentioned above. In numerical language, a certain quantity A can be expressed in terms of coordinates and velocities of the particles in the system as follows [98, 101]:

$$A(t) = f\left(r_1(t),......r_N(t), v_1(t),......v_N(t)\right) \qquad (4.33)$$

With the fact that the coordinates and velocities are functions of time, the average for quantity A may be given by:

$$\langle A \rangle = \frac{1}{N_T} \sum_{t=1}^{N_T} A(t) \qquad (4.34)$$

where t denotes an index which extends over the time steps from 1 to N_T (total number of steps). The average of quantity A can be calculated via one of two methods. Both methods are equivalently efficient but the first one is easier.

1. The quantity A at time t is calculated at every time step throughout the MD run. Having the $A(t)$ we then simply find the sum of $A(t)$, $\sum_t A(t)$, after each time step, which is also automatically updated. At the end of the simulation run the average is directly obtained from dividing the $\sum_t A(t)$ by the number of time steps.

2. Quantities which are periodically being dumped in one of the intermediate production files like the dump file may not be automatically averaged at the end of each time step or at the end of the simulation.

Another program may be used in conjunction with the main program to read the dump files and calculate the desired quantities.

Regardless of the method we use, there is a number of important quantities that are calculated during the production phase of the run. These quantities are described in the following sections.

4.2.1.2(a) Potential Energy

For any given system, the total potential energy is given by:

$$U = \sum_i \sum_{j>i} \sum_\alpha \sum_\beta \phi_{i\alpha j\beta}(r_{i\alpha j\beta}).$$

(4.35)

The potential energy function denoted by $\phi_{i\alpha j\beta}(r_{i\alpha j\beta})$ is the key to deriving the forces which in turn are important in determining the dynamics of the system. The average potential energy is computed using summations of instantaneous values. For example the two-body interactions in a system can be described by [98]:

$$\Phi(t) = \sum_i \sum_{j>i} \phi\left(\left|r_i(t) - r_j(t)\right|\right)$$

(4.36)

4.2.1.2(b) Kinetic Energy

The kinetic energy of a system at any instant in time (t) is given by[98,101]:

$$K(t) = \frac{1}{2} \sum_i m_i \left[v_i(t)\right]^2$$

(4.37)

This is used to calculate the total kinetic energy of the system. The kinetic energy is then used to calculate a number of important quantities including for example the effective temperature of the system which is given by the ensemble average of its kinetic energy.

The total energy is the sum of potential and kinetic energies. It can be simply calculated via the relation $H= K + V$. Furthermore, it is a conserved quantity in Newtonian dynamics and thus does not require to be calculated at every time step. However it is common practice to do the calculation at each time step in order to make sure that it is indeed conserved. The total energy must remain constant throughout the run despite any fluctuations in its components. The potential and kinetic energies are sort of exchanged throughout the simulation so that the overall fluctuation in their sum remains almost zero.

There is no way to ensure that the sum in fluctuations of K(t) and V(t) will equal zero. Practically the total fluctuations may not be zero and thus the total energy may fluctuate throughout the simulation. However these fluctuations should be almost negligible, that is they must be less than a certain value beyond which one can say that there is something incorrect with the conditions of the system thus resulting in these fluctuations and making the energy non conserved. The typical amount of fluctuations in total energy that we may accept in a simulation is in the range of $1/10^4$ or less (one part in 10^4). If the fluctuations exceed this limit it cannot be ignored. It may of course be caused by a number of factors among which is the incorrect (usually longer) value for the time step. In such case, the fluctuations in total energy are simply reduced by reducing the time step. [110] Energy conservation may also be violated if the cutoff method chosen is not sufficiently good or if the computer itself in numerically limited. The following plot illustrates the conservation of total energy throughout a simulation run [111]:
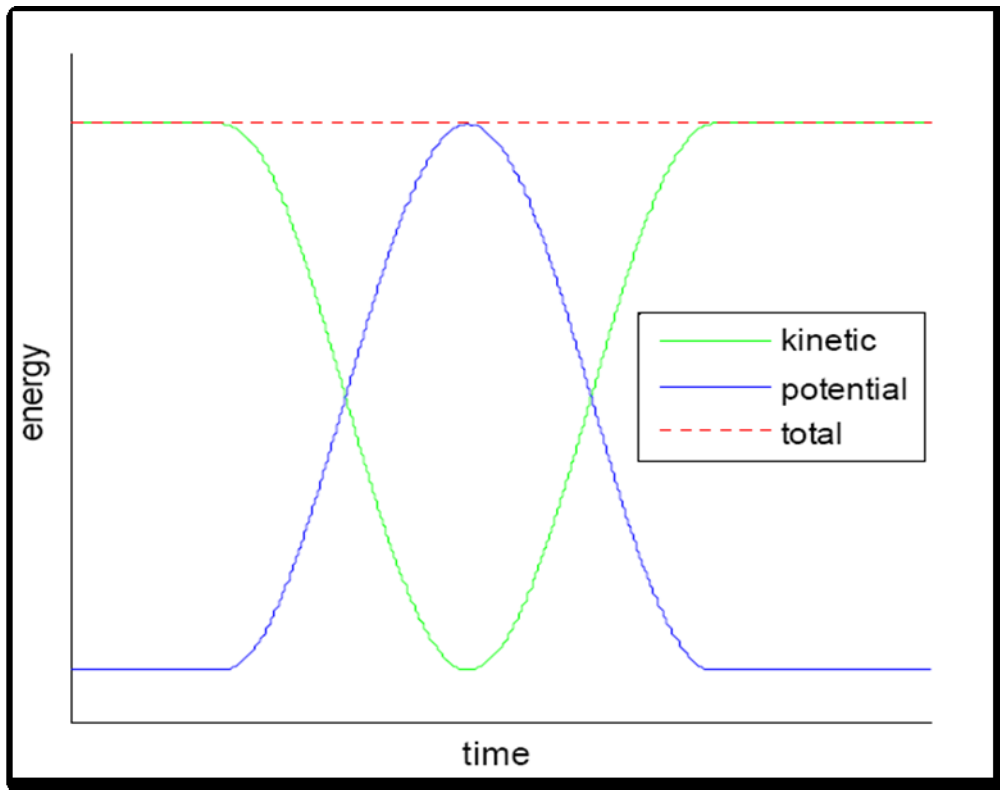
Figure (4.4): Energy conservation in a MD simulation run

Furthermore, the figure below [42] shows an example of the results concerning energy measurements throughout a MD run.
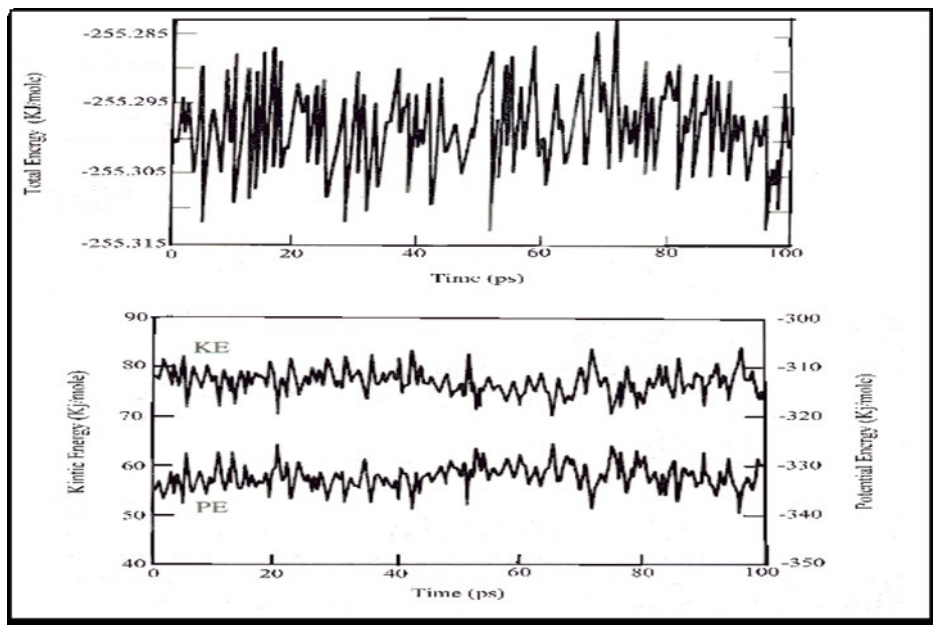

Figure (4.5) Variation in total energy versus time for the production phase of a molecular dynamics simulation of 256 argon atoms at a temperature of 100K and a density of 1.396

4.2.1.2(d) Pressure

The approach usually adapted to calculate the pressure P is a molecular simulation involves an ensemble average of the instantaneous microscopic pressures. Yet there are a number of methods to do so. The virial theorem of Clausius is the most commonly used. [111][102] The virial can be defined as the expectation value of the sum of the products of coordinates (x_i) of the particles and the first derivative of the momentum along that coordinate ($\dot{p}_{x_i}$) which according th Newton's second Law is equivalent to the forces acting on them.

The virial theorem states that the virial W is equal to $(-3Nk_BT)$. That is[101]:

$$W = \sum x_i \, \dot{p}_{x_i} = -3NK_BT \tag{4.38}$$

For an ideal gas, the only forces acting on a certain particle is a result of the interactions between the gas particle and the container within which it is placed. In such case the virial is simply $-3PV$, which corresponds to the value obtained from the relation $PV = Nk_BT$

For a real gas, or a liquid, the forces between the particles among themselves have an effect of the value of the virial. The total virial in such cases is equal to the sum of the virial of the ideal gas and a term corresponding to the contribution by the interactions between the particles. This gives the following relation:

$$W = -3PV + \sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} f_{ij} = -3NK_BT \tag{4.39}$$

This gives a direct relation for the pressure of the system given by:

$$P = \frac{1}{V} \left[NK_BT - \frac{1}{3K_BT} \sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} f_{ij} \right] \tag{4.40}$$

4.2.1.2(e) Temperature

In a canonical ensemble the total temperature of the system is constant. However, for a microcanonical ensemble the temperature fluctuates and is directly proportional to the kinetic energy. The relation between temperature and kinetic energy is given by [98][101]:

$$K = \sum_{i=1}^{N} \frac{|\mathbf{p_i}|^2}{2m_i} = \frac{k_B T}{2}(3N - N_c) \qquad (4.41)$$

where $\mathbf{p_i}$ is the total momentum and m_i is the mass of a particle.

According to the theorem of equipartition of energy each degree of freedom contributes $k_B T/2$ to the kinetic energy. Therefore, if there are N particles in a system, each having three degrees of freedom, and then the kinetic energy will be given by: $3Nk_B T/2$. The term N_c given in the equation (105), represents the number of constraints on the system. During the simulation, the total linear momentum of the system is usually constrained to a value of zero. This is equivalent to removing three degrees of freedom from the system and hence we may substitute 3 for N_c. A value of N_c is used for every type of system according to the constraints it is placed under.

4.2.1.2(f) Radial Distribution Function

The radial distribution function (RDF), denoted by g(r), and also referred to as the pair distribution function, is a very effective way of describing the average structure of disordered molecular systems such as liquids, as well as some disordered and ordered solids. In simple terms, the radial distribution function describes how density varies as a function of distance from a reference particle. It describes the probability that an atom will be found at a distance r from another atom with reference to the ideal gas distribution. The general algorithm depicted in this method is to determine how many particles are found within a distance r and $r+dr$ from another particle. This is usually achieved via the calculation of the distances between all particle pairs.

The RDF for a certain crystal will show an infinite number of sharp peaks. The heights of these peaks as well as the separations amongst them is characteristic of the lattice structure of the crystal. On the other hand, the radial distribution function for a liquid is characterized by a small number of peaks, with short distance intervals between them which decays to a constant value as we move farther away from the central atom. The radial distribution function is zero for distances which are shorter than the atomic diameter. This is a direct result of the strong repulsive forces present within that distance [98,101,111]. A typical RDF is shown below (4.5) [112]:
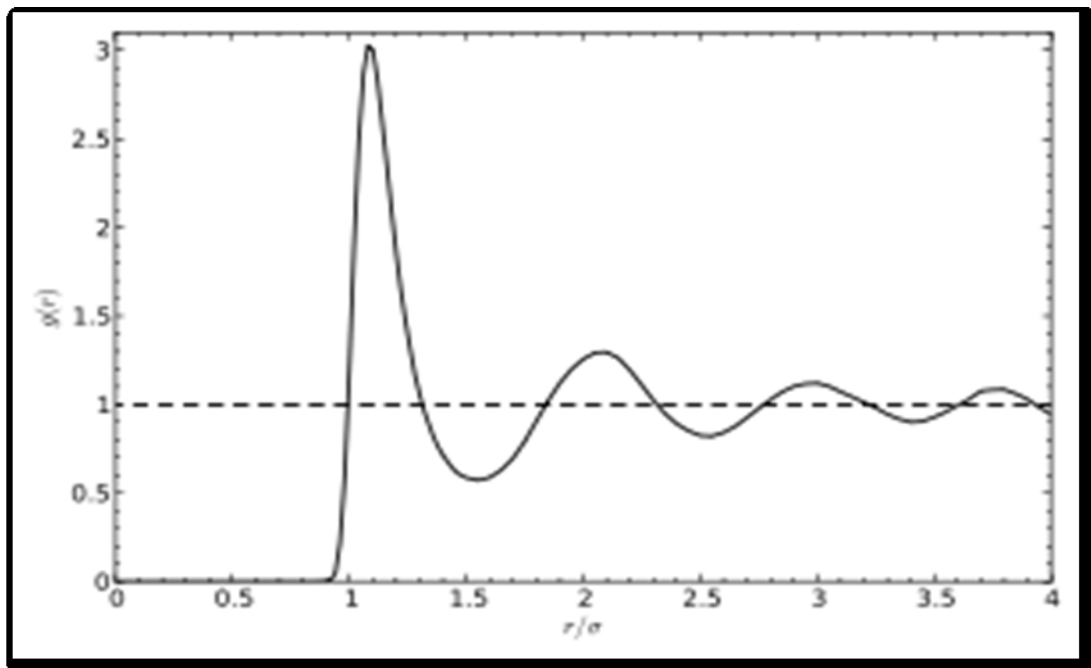
Figure (4.6) Radial distribution function determined from a 100 ps molecular dynamics simulation of liquid argon at a temperature of 100 K and a density of 1.396

In order to calculate the RDF from a simulation, the neighbors around each atom or molecule shall be categorized into distance histograms. The number of neighbors within every histogram is the averaged over the entire simulation. To further describe the process, let us consider a spherical shell of thickness δr at a distance r from a certain atom. The shell is shown below (4.6) [101]:
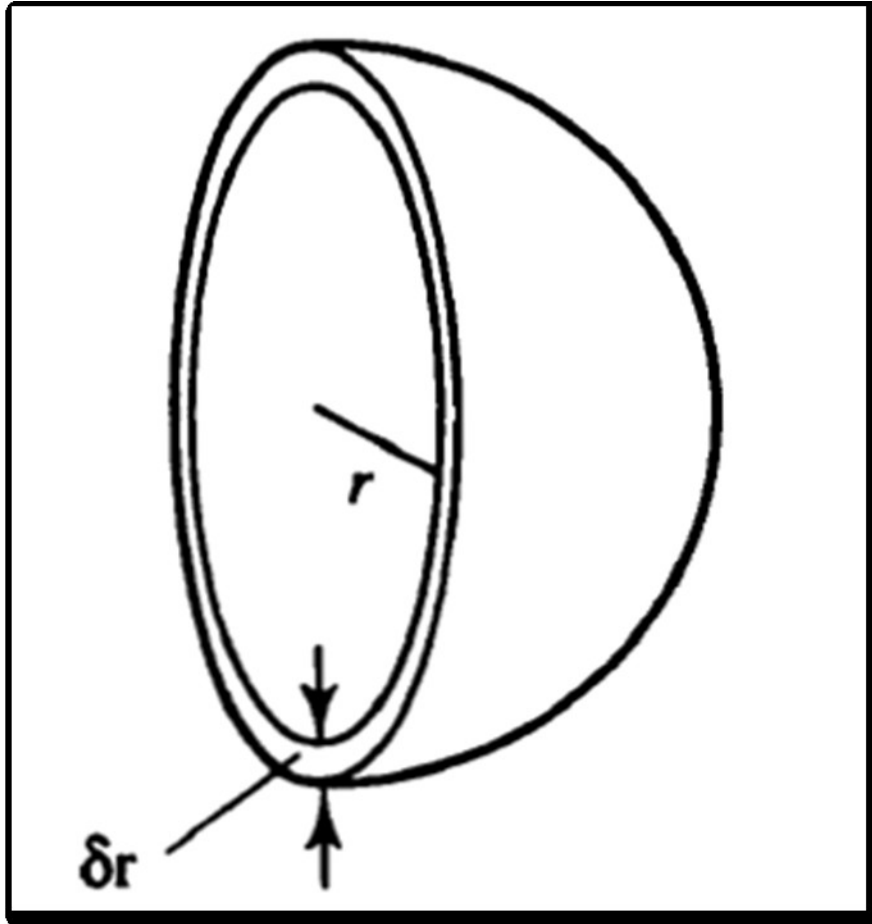
Figure (4.7): : Radial distribution functions use a spherical shell of thickness

The volume of the shell is given by:

$$V = \frac{4}{3}\pi(r + \delta r)^3 - \frac{4}{3}\pi r^3$$
$$= 4\pi r^2 \delta r + 4\pi r \delta r^2 + \frac{4}{3}\pi \delta r^3 \approx 4\pi r^2 \delta r$$

(106)

ρ is the number of particles per unit volume. Therefore we can find the total number of atoms in the shell which shall be given by $4\pi\rho r^2\delta r$. If the entire space around the atom is segmented into cells each of which having a thickness of δr and each having a number of atoms equal to δr, the number of atoms inside each shell can be calculated and compared with ideal gas distribution. Furthermore, if the number of atoms in every cell is given by n(r) then the local density $\rho(r)$ will be $\rho(r)$ = n(r)/4 πr^2dr. In molecules with more than one type of atom, the radial distribution function is used to

calculate a number of site-site distribution functions. For CF_4 we have three distinct functions namely: g(C-C), g(F-F) and g(C-F). The information provided by these functions is used to describe the orientations of the CF_4 molecule.

4.3 ANALYSIS OF RESULTS

Now that the run has terminated and all the data produced has been extracted we have arrived to the most critical point in the series of steps of carrying out the computer simulation; that is analyzing the trajectories. In general, there are two types of parameters that may be computed during the molecular dynamics run. The first include all long term averages of quantities like energies, temperatures, cell vectors, pressure, viral, mean square forces, mean square torque and dipole moments. The second set of parameters includes the radial distribution functions, as well as other values which are not potential oriented. Yet a third set of parameters is calculated separately after the completion of the MD run. These may include correlation function as well as complex thermodynamic averages.

There are a number of parameters that are of high importance to our work. Combined together, these parameters allow us to study the system thoroughly and be able to detect any change in the structure which would mean that the system has experience a phase transition. The changes that may take place are detected via a number of indicators which work together to monitor the state of the system during a series of molecular dynamics runs. These indicators are particularly important for us since we are primarily interested in the phase transitions of the system at given pressures and temperatures. We can summarize these indicators as follows:

4.3.1 The Caloric Curve

The caloric curve is one of the most important indicators of a phase transition. The caloric curve depicts the relation between the total energy of the system and its temperature or pressure. Once the total energy H, and the temperature T or pressure P are measured in different runs corresponding to different thermodynamic states, the caloric curve H(T) or H(P) can be constructed. A jump or sharp and sudden change in the H(T) caloric curve implies a first order phase transition, whilst a jump in the derivative of the H(T) implies a second or higher order phase transition.

The most common firs-order transition which can be observed via this method is melting. That is a phase transition from an ordered crystal solid phase to a partially disordered liquid phase. As the system abandons the crystalline structure and becomes disordered, a sharp change is observed on the H(T) curve. The sharp change shall occur at the latent heat of fusion of the system. In simple terms, at this point the system absorbs heat but no change occurs to its overall temperature. This results in the sharp and sudden change in the slope of the caloric curve which to us is an indication of the process which had just taken place. Usually such change takes place at a temperature higher than the theoretical melting value of the system. This is justified by the fact that at this particular moment the system undergoes hysteresis effects due to the necessity to wait for a seed of the liquid phase to appear. Once the liquid seed containing a few atoms is formed, the liquid phase starts to grow as the solid phase disintegrates simultaneously. Typically in a molecular dynamics run, one must 'overshoot' and set the temperature 20-30% above the real melting temperature in order to be able to actually observe melting [113]. The graph below is a typical caloric curve for the MD simulation of 256 molecule of methane at constant pressure (4.7) [101].
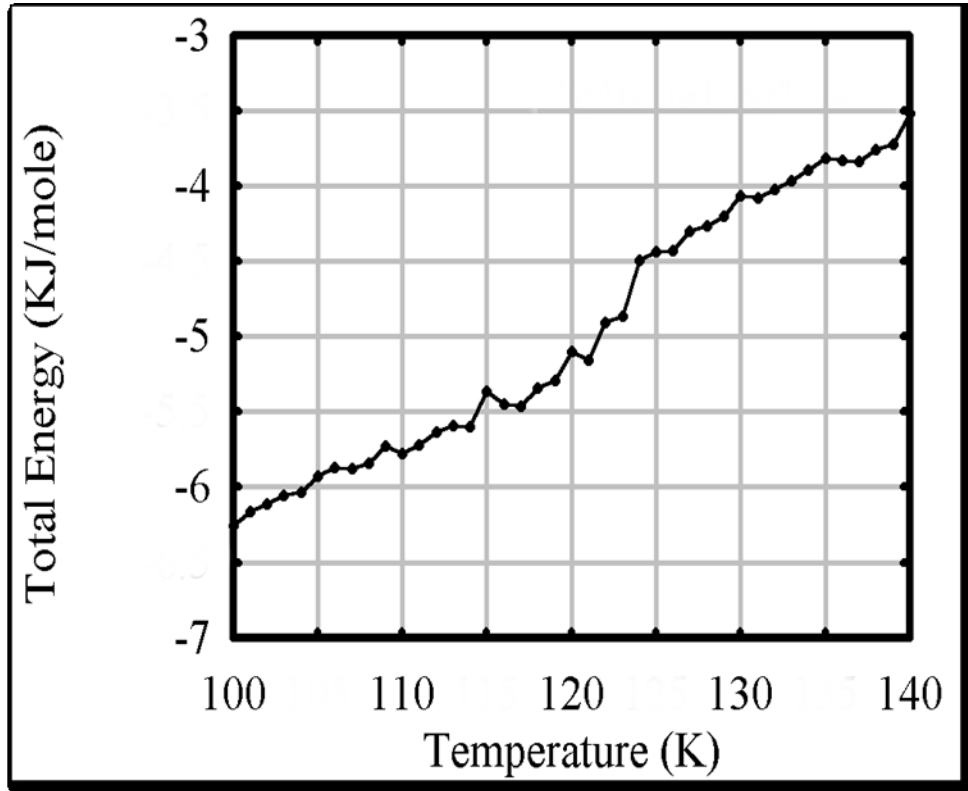
Figure (4.8): The caloric for the MD simulation of 256 molecules of methane at constant pressure

4.3.2 The Binder Fourth Order Cumulant

The analysis of the behavior of the fourth order cumulant of certain physical quantities like order parameters of energy is a well used approach to characterize the order of a phase transition. The fourth order cumulant of energy is given by [114]:

$$\Gamma = 1 - \frac{\left\langle H^4 \right\rangle}{3 \left\langle H^2 \right\rangle^2} \qquad (106)$$

where H is the total energy of the system. For two identical systems in every aspect except the number of molecules in each, an intersection point on the curve of their fourth cumulants indicates a phase transition. The graph below shows a typical fourth order cumulant intersection plot [101]:
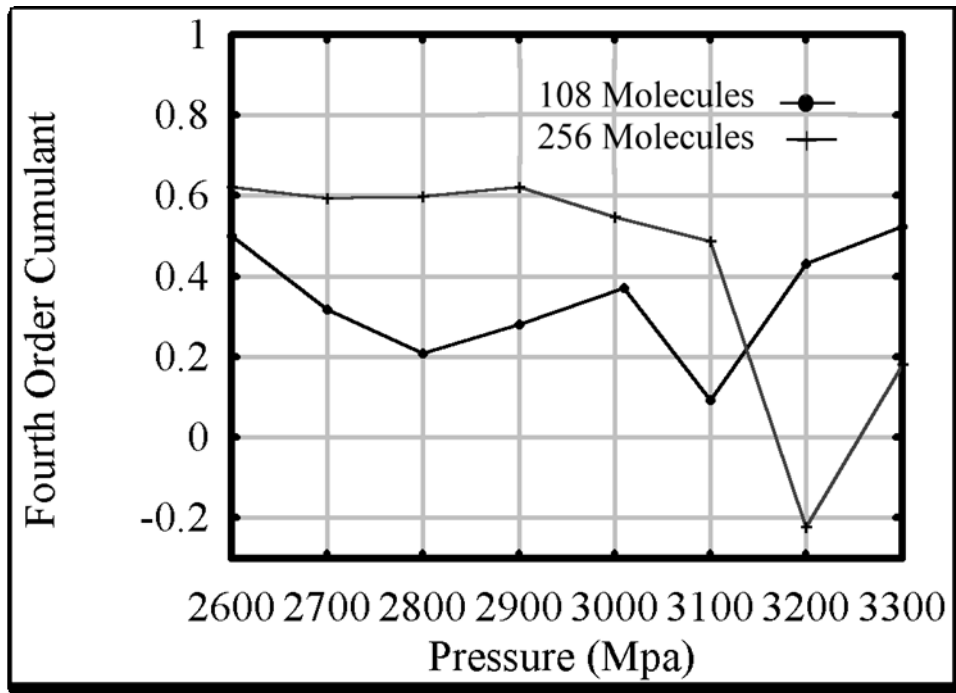
Figure (4.9): Binder fourth order cumulant curves for 2 methane models (108 molecules and 256 molecules) simulated at constant pressure of 100 MPa

4.3.3 The Translational and Rotational Order Parameters

A full account of these quantities has been discussed in previous sections. In summary the translational order parameter is used to distinguish between a solid and liquid phase of the system, whilst the rotational order parameter is used to determine the orientation of the system that is it determines the degree to which the system is ordered/disordered.

A graph of the translational order parameter versus temperature or pressure for a series of MD run can provide information about the melting point of the system. On the other hand, a plot of the rotational order parameter versus temperature or pressure can determine any change in the orientation of the molecules in the system. Below are typical translational and rotational order parameter plots from the MD simulations [42,101].

Figure (4.10): Translational order parameter of MD simulation of methane at P=100 MPa, and N=256 molecules.

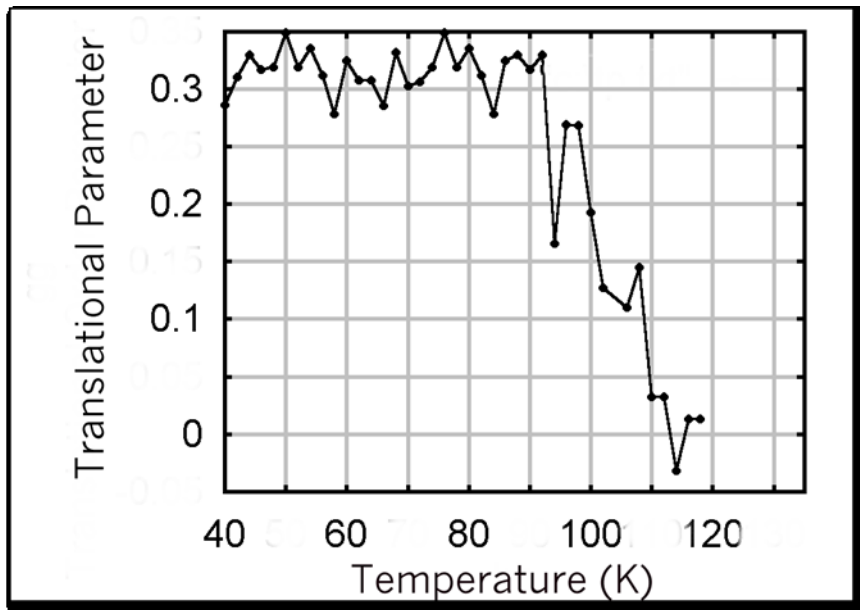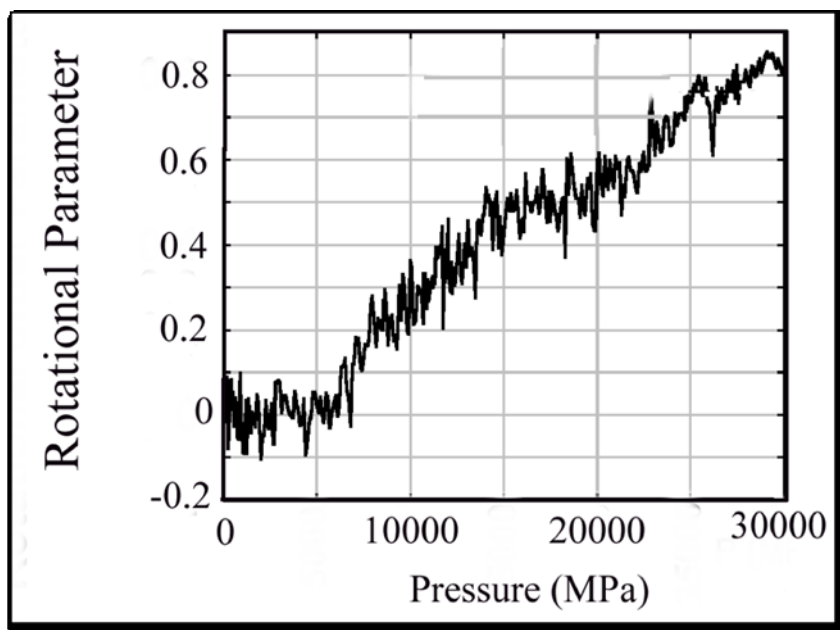*Figure (4.11): Rotational order parameter of MD simulation of methane at T=295 K and N=256 molecules.*

CHAPTER 5: PHASE TRANSITIONS OF CARBON TETRAFLUORIDE: RESULTS

Carbon Tetrafluoride is one of the major sources of fluorine or carbon-fluorine free radicals used in a number of wafer-etch processes in the manufacture of semiconductors. It is used with oxygen to etch polysilicon, silicon dioxide and silicon nitride. The interest in chemical dry etching has considerably increased in the recent years due to its advantages over other etching techniques. Due to the spatial separation of the excitation and reaction volume, the surface processes taking place do not in any stage involve charged particles. Instead they are totally controlled by chemical reactions of neutral particles. As a result of the reduction of energetic particles, charge induced effects are almost negligible. In addition to that, dry etching also reduces the handling and disposal of toxic chemicals such as hot phosphoric which is one of the chemicals used in the process. Carbon Tetrafluoride is relatively inert under standard conditions. High purity CF_4 gives way to better dimensional and profile control of the etching process. Furthermore, experimental studies have shown that the etch rate in silicon based compounds by Carbon Tetrafluoride can be manipulated via the introduction of N_2 into the process. This has caused a large industrial demand on CF_4 as one of the most preferably used gases in the etching process.

For the reason above, as well as the incorporation of CF_4 in other areas of the semiconductor manufacturing processes such as chamber cleaning, MEMS etching applications and Flat Panel Display etching applications, Carbon Tetrafluoride has attracted the attention of a lot of researchers. The aim of most research was to fully understand its properties and characteristics, particularly those related to the ability to work as an etchant. Furthermore, CF_4 is also widely used today in electronic components surface cleaning, solar battery production, laser technology, space technology, metal smelting, meteorological insulation, low temperature refrigeration and many other applications.

The use of CF_4 in etching processes remains the most important of the applications of Carbon Tetrafluoride. This became an incentive that much research be carried on the crystalline structure of CF_4. The crystal lattice, orientations, point groups and space groups of the different phases are particularly important when studying the physisorption of CF_4 onto graphite as well as determining the thickness of the CF_4 to be used under each set of work conditions. However, studying the structural phase transitions of CF_4 proved to be a complicated matter. Despite the closeness in structure between CF_4 and other simple tetrahedral molecules including methane, the phase diagram of Carbon Tetrafluoride is not as simple as that of others. There have been a number of researches performed to probe the interatomic interactions of the molecule at high pressure. Unfortunately the results of these various researches have not been in agreement. Similarly, the phase diagram of CF_4 at low temperatures is still a vague area whereas it is rather important to investigate the structure of the molecule within low temperature ranges. This is due to the fact that such conditions of low temperatures are usually imposed in certain industrial processes where higher temperatures may interfere with the chemical reaction rate and mechanism. In the etching process for example, the relational between etching rate and temperature is not always linear, which means that in order to optimize the results we should be able to use the etchant at any temperature needed.

Furthermore, the crystalline structures which CF_4 adapt are not fully agreed upon. Many high pressure studies have been carried out to test whether the structural arrangements shall vary and although it did prove that the positions of the molecules are considerably displaced under conditions of high pressure, the crystalline structure itself, and particularly the space groups have yet to be resolved. This controversial nature of the crystalline structure and phase transitions of CF_4 is what encouraged us to extend the research in order to probe this issue under conditions of low temperatures and high pressures.

5.1 PREVIOUS WORK

When validating our model, the comparisons with experimental and previous analytical results was rather important. For this cause we have utilized experimental and analytical results that have been agreed upon and validated by a number of credible research papers and scientific journals. Such results have given us details on the structure and phase diagram of CF_4 as follows:

According to the latest Raman Spectroscopic results, at room temperature CF_4 undergoes the following transitions [114,115]:

- From a liquid phase to the first solid phase (I) at a pressure of 1.6±0.30 Gpa.
- From solid phase (I) to solid phase (II) at a pressure of 2.83±0.25 GPa.
- From solid phase (II) to solid phase (III) at a pressure of 3.45±0.15 GPa.

Furthermore, a number of other studies using both Raman spectroscopy and X-ray diffraction methods have given the following information which has not yet been supported by simulation methods or other experimental data[115,116,117,118]:

- A transition to phase (III) at a pressure of 9GPa was obtained by Shindo et al [117] and verified by Nakahata et al [118].
- A transition from the solid phase (III) to solid phase (IV) at a pressure of 13.6GPa was also obtained in the work of Shindo et al [117] and Nakahata et al [118]. However, particularly this transition was not observed in any other works. Instead Lorenzana et al [116] in their work have observed what they assumed to be a solid-solid transition from phase (III) to phase (IV) at a pressure of 16GPa.

In our work we have tried to revisit the particulars on phase transitions mentioned above particularly numbers 4 and 5. We have also tried

to simulate the system at pressures higher than 16GPa to study the behavior of CF_4 molecules and whether or not they undergo any phase changes. Furthermore we have also worked on a range of constant pressures to observe phase transitions at low temperature. This shall be discussed in full detail in the following section.

Another important aspect of our work was to study the changes which the crystal structure undergoes as subject to changes in the temperatures and/or pressure. The crystal structure, lattice structure and space groups of the solid phases of the CF_4 is also a controversial matter. Particularly phase (II) which is referred to as α-CF_4 is surrounded by some ambiguity. IR and Raman spectroscopic results have suggested that α-CF_4 is a well ordered tetragonal lattice with S_4 symmetry [116]. However this was not in agreement with the X-Ray diffraction results which suggested that α-CF_4 is a well ordered monoclinic lattice belonging to either a C_c [117,120] or a $P_{21/c}$ [118,121] space group. Later studies conducted by Satay et al [122] suggested that the space group of α-CF_4 is $C_{2/2}$ which is deduced from the $P_{21/c}$ previously suggested by Bol'shutkin et al [120]. Space groups of phase III have also been quite controversial, however it was finally agreed by a number of researches that the space group of that phase is a $P_{21/c}$ [115]. The actual symmetry type, lattice structure and space group of CF_4 in all its phases and particularly the α-CF_4 still needs further investigation in order to remove the ambiguities which have been surrounding it so far. This too is part of the work we have carried out.

Given the above information, the phase diagram below was suggested for CF4 [116].
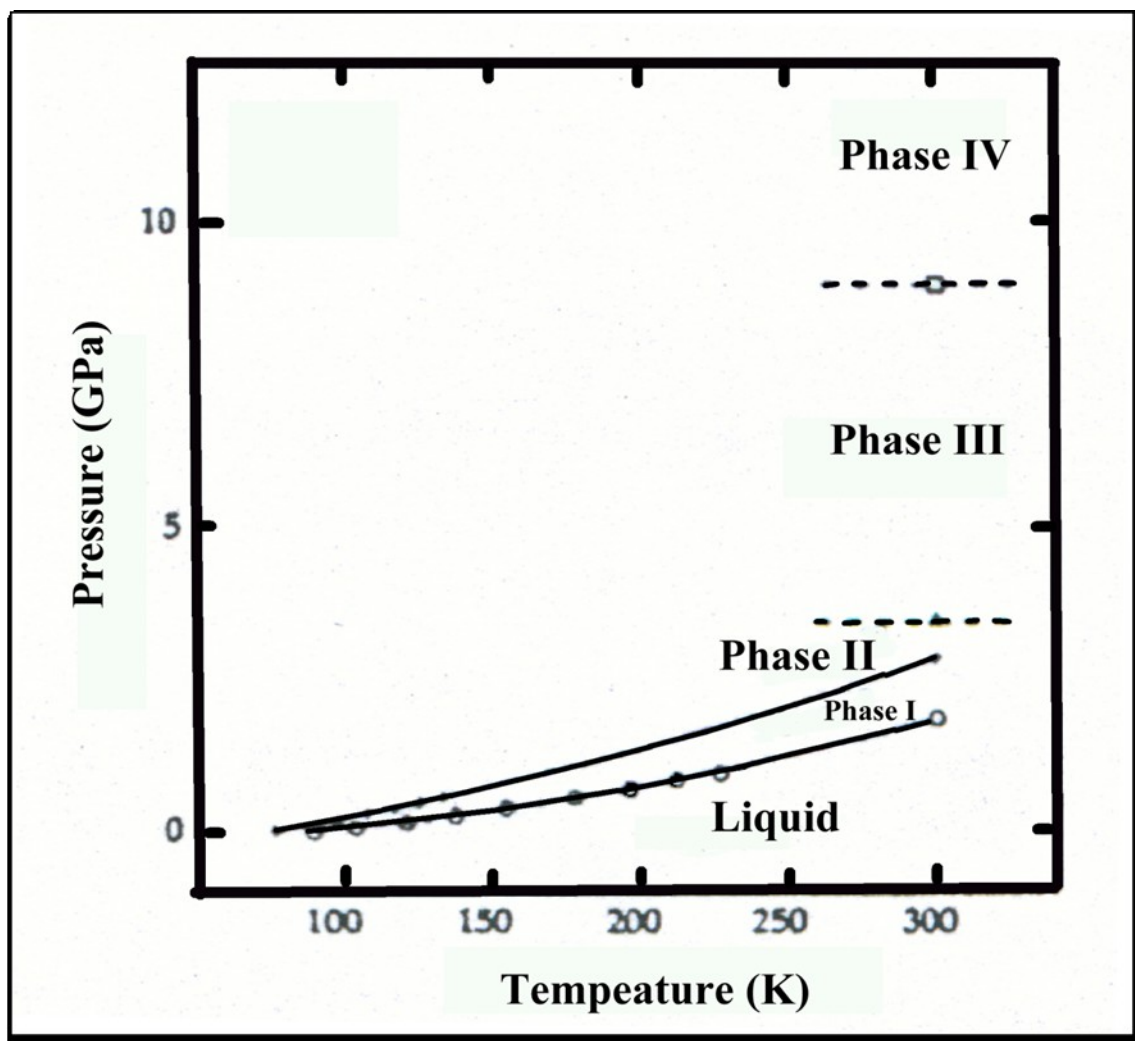
Figure (5.1): Proposed phase diagram for CF₄

Unfortunately, the lack of credible information, the discrepancies in results and the ambiguity surrounding certain phases has resulted in the above phase diagram which is clearly not a complete one. As we see in the proposed phase diagram (5.1) there isn't enough information on the structure of CF_4 at high pressures neither is there enough information about the structure at low temperatures. That is why we have decided to focus our work on those two areas in order to come up with as much clarification as possible as to the structure of CF_4 at high pressures in the ranges between 10 and 20 GPa and also at low temperatures in the ranges from 0-100 degrees Kelvin.

5.2 PROPOSED MODEL

As thoroughly discussed in previous chapters, Carbon Tetrafluoride is a tetrahedral molecule, which in its simplest state belongs to the T_d point group as shown below [123]:
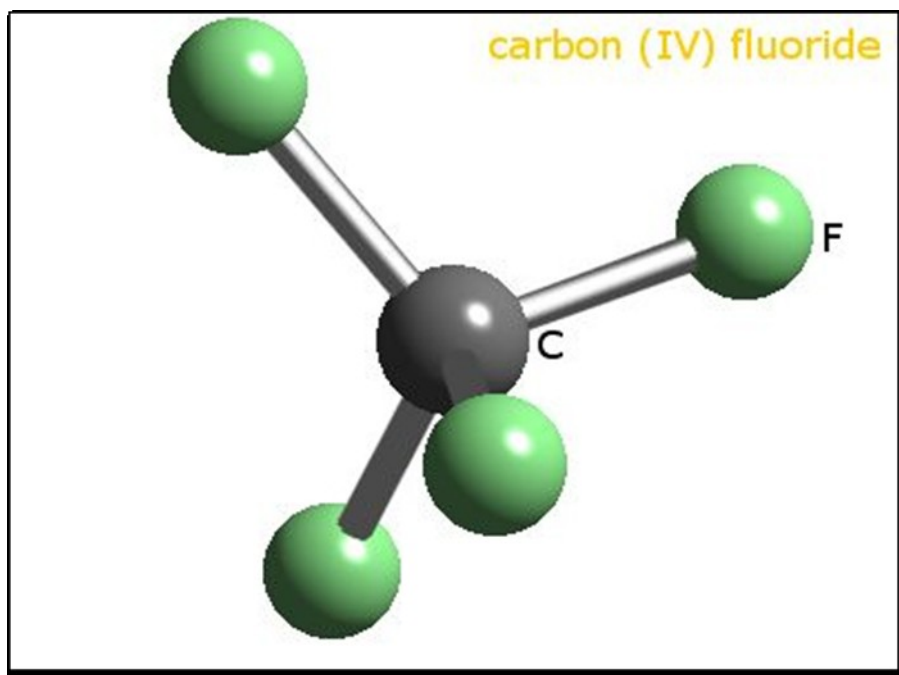
Figure (5.2): Tetrahedral Carbon Tetrafluoride

The gaseous state of CF_4 is one of the easiest states to describe. It is a standard phase where particles are easily dealt with as hard spheres and the spherical shell model is sufficient to describe interactions and potentials. However, when dealing with liquid and solid phases of CF_4 the situation is not all that simple. The potential model utilized must be sufficient to take into account all particle-particle interactions both short and long range. A site-site interatomic potential becomes necessary, In our work we followed the footsteps of Nose and Klein [124] and used the interatomic Lennard-Jones potential however, unlike them we explicitly included the electrostatic force. The relation used is as follows:

$$\Phi_{\alpha\beta}(r_{ij}) = 4\varepsilon_{\alpha\beta}\left[\left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{\alpha\beta}}{r_{ij}}\right)^{6}\right] + \frac{q_{i\alpha}q_{j\beta}}{4\pi\varepsilon_{o}r_{ij}} \qquad (5.1)$$

In the relation we point out the distance between any atoms in the term r_{ij} and we also point out the nature of the atom (C or F) referring to them by α and β respectively. Furthermore, the collision diameters is $\sigma_{\alpha\beta}$ and depth of the potential well is $\varepsilon_{\alpha\beta}$ where both are LJ parameters. The relation between atoms of the same nature is described by the parameters used by *Singer et al* and *Murad and Gubbins* in their work [125,126]. Other parameters are obtained using Lorentz-Berthelot mixing rules as follows [109]:

$$\sigma_{CF} = \frac{1}{2}\left[\sigma_{CC} + \sigma_{FF}\right] \qquad (5.2)$$

and

$$\varepsilon_{CF} = \sqrt{\varepsilon_{CC}\varepsilon_{FF}} \qquad (109)$$

Altogether, the set of LJ parameters is given in the table below:

Parameter	ε_{CC}/K_B	ε_{FF}/K_B	ε_{CF}/K_B	σ_{CC}	σ_{CC}	σ_{CC}
Value	51.2 K	52.8 K	51.99 K	3.35 Å	2.83 Å	3.09 Å

Table (5.1) Lennard Jones Parameters on Carbon and Fluorine atoms

Atomic partial charges may easily be obtained using the partial equalization of orbital electronegativities. This method is one of the most precise approaches which provides results usually consistent with experimental values. The masses and partial charges on the Carbon and Fluorine atoms are given below:

Atom	Partial charge (e)	Mass (amu)
Carbon	0.5617	12.01
Fluorine	-0.1404	18.9984

Table (5.2): Partial charges and atomic masses of Carbon and Fluorine atoms

5.3 SIMULATION SETUP

It was not easy deciding upon the most appropriate conditions under which we should run our simulations. Knowing the importance of choosing the correct number of molecules, time step, as well as other factors like pressure, temperature...etc we had to be very precise with our choices. Finally we decided that in order to closely simulate the system we need to at least three MD boxes. So we chose to use MD boxes containing 108, 256 and 500 molecules. The runs were initiated using the FCC as our initial configuration since this particular lattice is most suitable for CF_4. As discussed earlier, we used the software 'Moldy' to carry out our runs and calculations.

As previously explained, Moldy allows the use of boundary conditions, which we have chose to impose in all directions. Furthermore, we allowed the variation in size shape and symmetry of the MD box in order for the system to fully equilibrate particularly in the solid phases. The integration algorithm utilized to provide the solutions of the Newtonian equations of motion is the Beeman Algorithm which we have discussed in detail in previous chapters. We have used the isobaric-isothermal ensemble which allows us to provide definite values for the temperature and pressure at the beginning of each simulation run. in an isobaric-isothermal ensemble simulation, the molecules experience energy and volume fluctuations driving the system to equilibration. Throughout the runs we were closely monitoring the equilibration to make sure we have allowed enough time for the system to

reach equilibrium and for the trajectory to cover the entire phase space yet not too long so that we do not allow a rise to high energy overlaps between atoms and hence instabilities in the integration algorithm [42].

5.4 OUTPUT ANALYSIS

Termination of the simulation runs is only the beginning of what may be considered the hardest stage of the entire process. This stage involves the analysis of the data that we have extracted from each simulation run. We then proceed to produce caloric curves, Binder fourth cumulant curves, RDF's, and order parameters curves in order to describe the system at various conditions. Furthermore, careful analysis of the given results and produced curves was carried out to allow us to isolate temperatures and pressures where there might have been a phase transition and to identify the type of transitions as well as the structure and symmetry of the system at each phase. These calculations were either entirely carried out using Moldy, or the data was extracted from Moldy and then dealt with separately either by the use of another computational program like MATLAB or by simply utilizing an excel spreadsheet. The process by which data was extracted and analyzed may be summarized as follows:

1. Data is extracted and used to calculate the fourth order cumulant. For constant temperature simulations the fourth order cumulant is plotted against pressure and in constant pressure simulations it is plotted against temperature for two system sizes (either 108 and 256 or 256 and 500 or 108 and 500). The points at which the curves for the two systems intersect indicate phase changes.

2. Data is obtained and utilized to draw caloric curves at specified constant temperatures (Total Energy Vs. Pressure) or at specified constant pressures (Total Energy Vs. Temperature). A discontinuity or sudden change in the slope of the caloric curve indicates a phase change.

3. The translational order parameter is plotted to distinguish between a

- 124 -

solid and a liquid phase.

4. The rotational order parameter is plotted to specify the rotational order/disorder.

5. Lattice constants, both positions and angles (a, b, c, α, β, γ) are extracted in order to identify the lattice structure and space group.

6. Finally the RDF is obtained and the number of nearest neighbors and next neighbors is directly extracted from the output file.

5.5 RESULTS OF CONSTANT PRESSURE SIMULATIONS

As discussed earlier, it was rather important to us to revisit previously probed areas as well as try to explore the areas which have not yet been fully developed. We carried out a series of runs over the three MS box sizes at constant pressure. We started with runs at 100 MPa, 1 GPa, 6 GPa, 15 GPa and 35 GPa. Under each of these pressures the temperature was varied from T = 0 K to T = 450 K at intervals of 2 degrees Kelvin.

The calculations discussed in the previous section were carried out for each set of coordinates and results were tabulated. It will be extremely difficult to present all the calculations and graphs for the runs carried out which at the interval we used were around 175 runs at every pressure value. Therefore in the following pages we shall present a sample of out result analysis, particularly for areas where a phase transition was determined. We may consider for example the runs carried out at P=100 MPa and the runs carried out at 6 GPa

5.5.1 Results at P = 100 MPa and T = 0 K to 450K

First, by plotting the caloric curve for the systems N = 256 and N = 500 molecules at P = 100MPa we were able to observe two distinct changes in the slope of the graphs. Segmenting the Caloric curves makes it rather easier to examine the place and magnitude of such discontinuity. The two overall graphs are seen below:
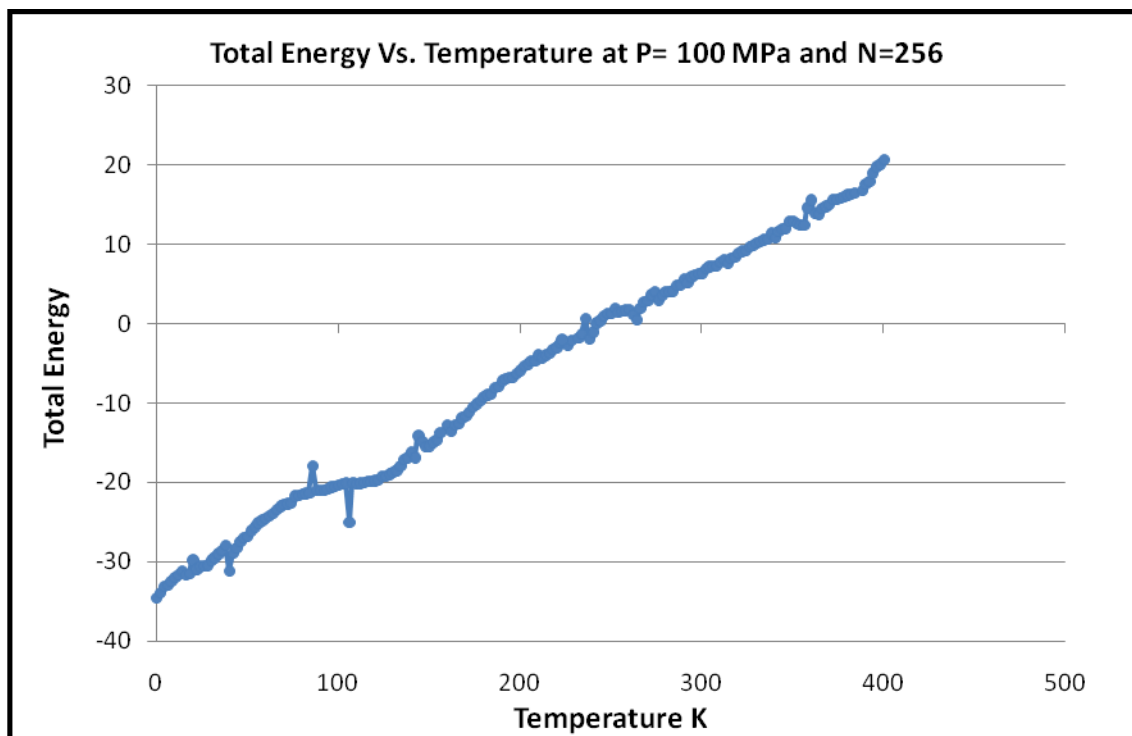
Figure (5.3): Caloric curve for simulation of 256 molecules of CF_4 at P=100 MPa
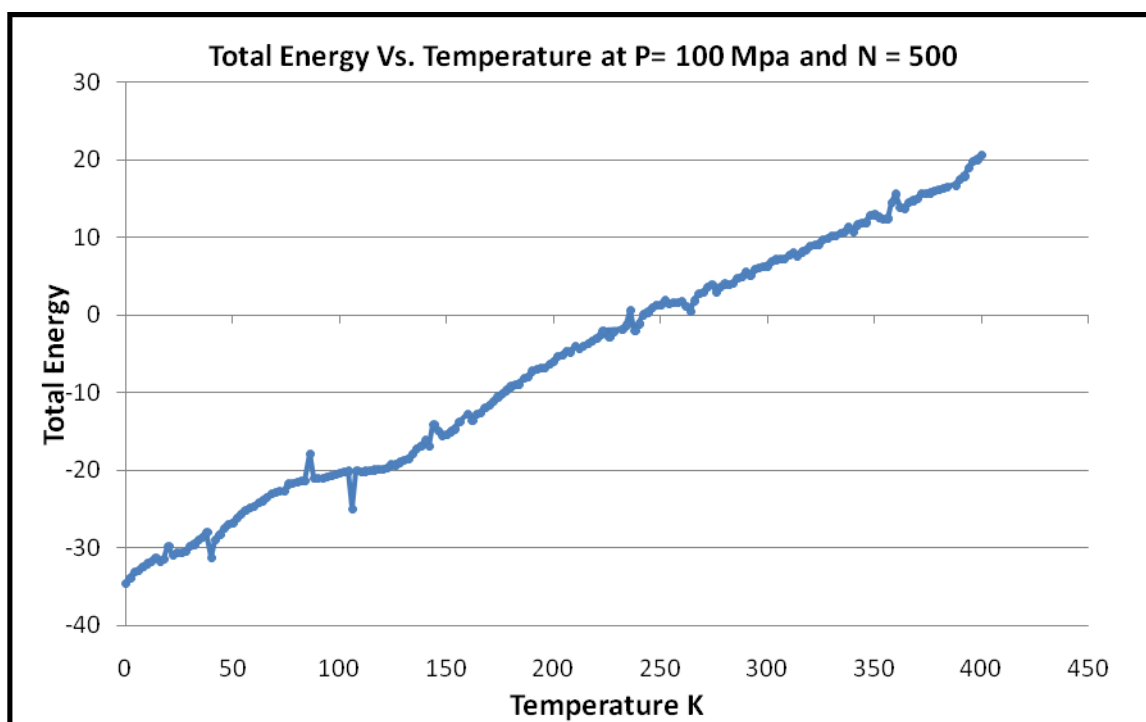

Figure (5.4): Caloric curve for simulation of 500 molecules of CF_4 at P=100 MPa

Closer studies performed on the caloric curve of the system with N=256, and segmentation at areas where there appears to be a sudden change in the slope gave us the following results:
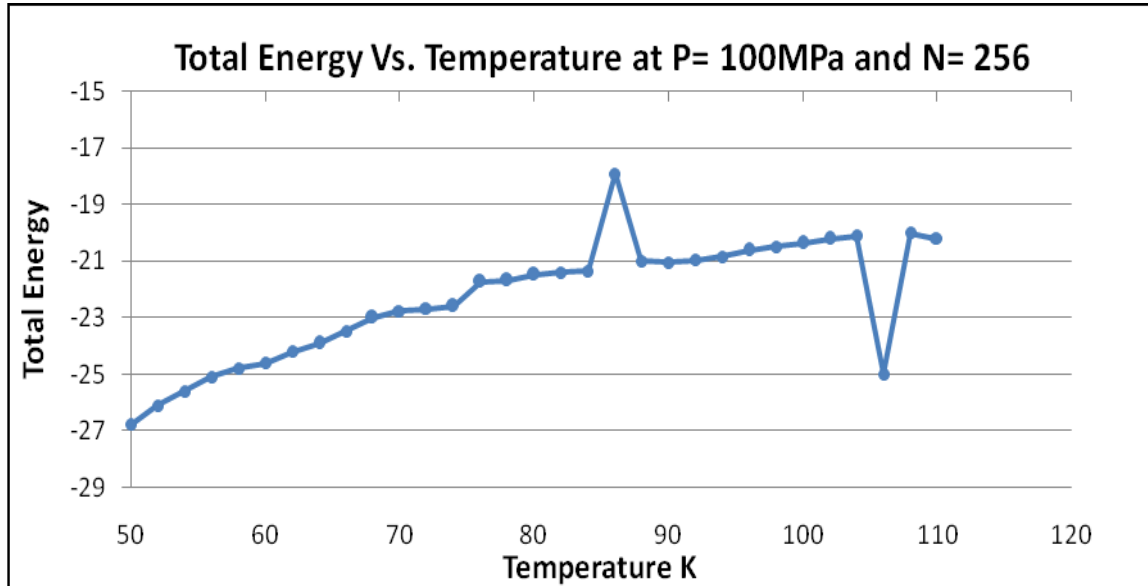

Figure (5.5): Caloric curve for simulation of 256 molecules of CF4 at P=100 MPa

Furthermore, the fourth cumulants of the systems where N=108 and N=256 where plotted to check for areas of intersection. The diagram below only shows the areas of intersection:
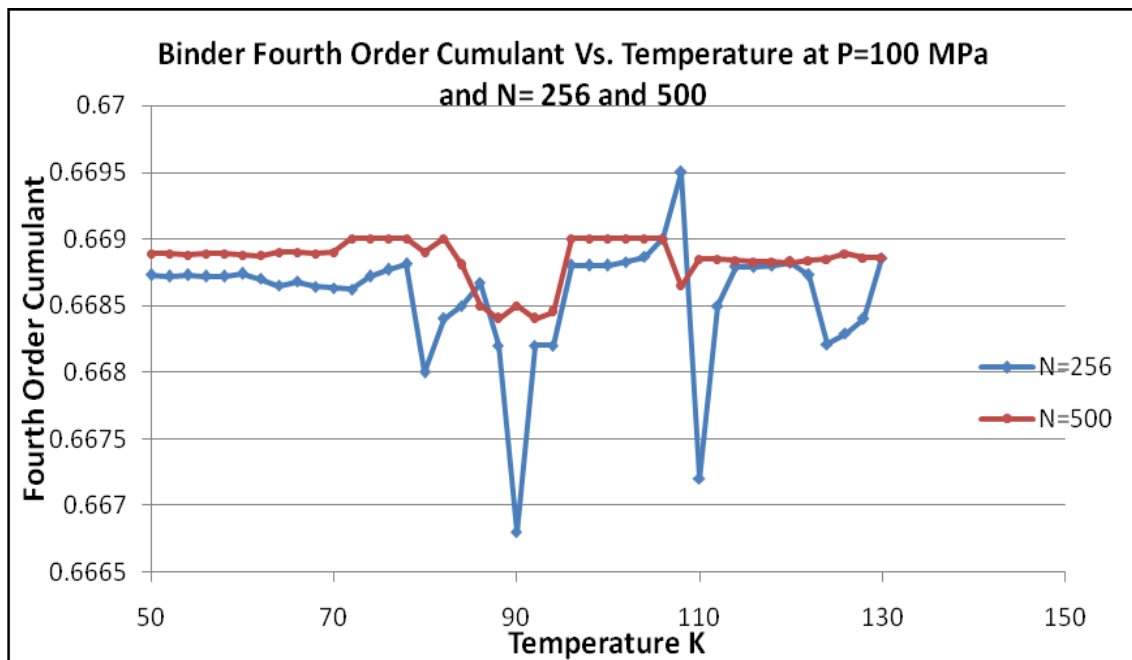

Figure (5.6): Binder fourth order cumulant curves for 2 CF$_4$ models (108 molecules and 256 molecules) simulated at constant pressure of 100 MPa

Now that we have the caloric curves as well as the fourth order cumulant, it is quite clear that Carbon Tetrafluoride undergoes two distinct phase transitions at P=100MPa, the first of which occurs at a temperature of 86±0.5 and 106±0.4.

In order to study the nature of the transition, the translational order parameter is plotted against temperature. An order parameter close to the value zero indicates a transition to a liquid (disordered) structure and as we see in the plot below, the order parameter fluctuates with increasing temperature and does in fact approach zero at the temperature range of 106± 0.4 which indicates that the transition is from the liquid state to the solid state thus indicating also that the phase transition occurring at T=86±0.5 is a solid-solid phase transition.

Figure (5.7): Translational Order Parameter curve for CF₄ model with N=500 simulated at constant pressure of 100 MPa

Furthermore, since we happen to have two solid phases so far among which the first transition at T=86 ± 0.5K takes place, the Rotational Order Parameter is obtained to distinguish between the structures of both phases. The curve below indicates that at the lower side of the temperature of the transition the system is well ordered and is transitioned into a less ordered

state as the temperature increases. That is the rotational order parameter remains well above zero within temperatures below 86 K, beyond which the rotational order parameter drops to zero and maintains a position within that range until the second temperature at which the transition takes place.
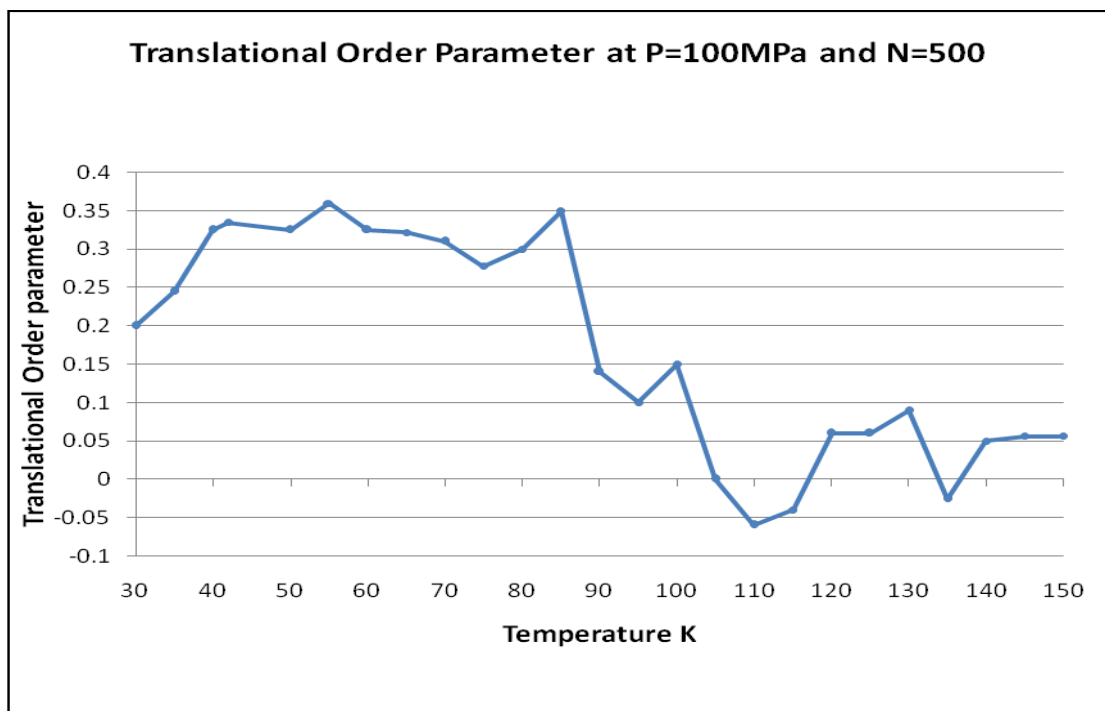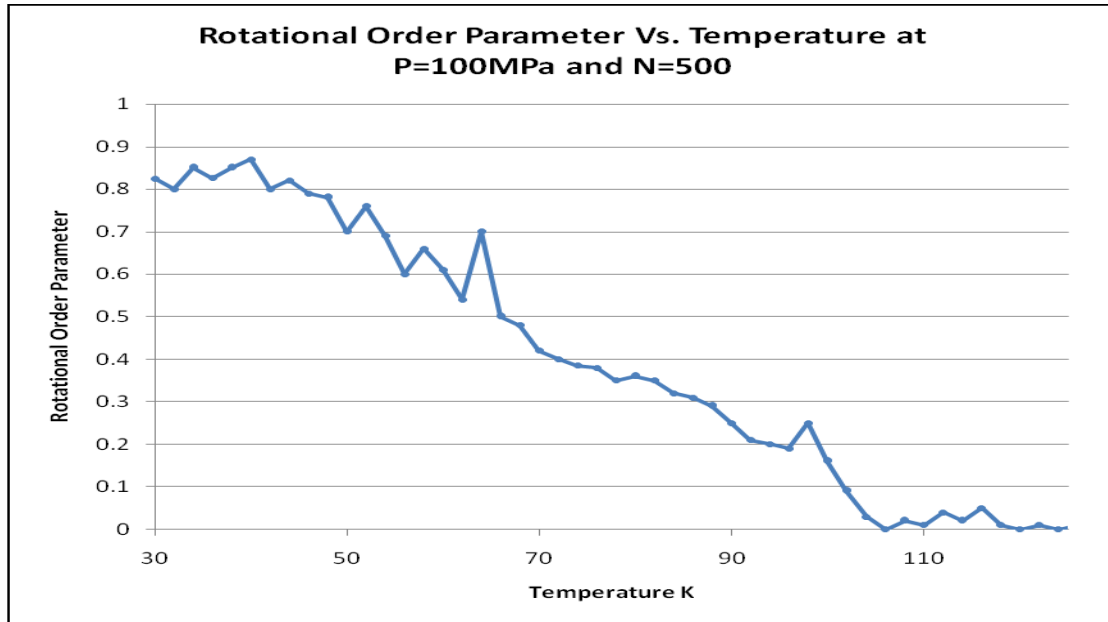

Figure (5.8): Translational Order Parameter curve for CF$_4$ model with N=500 simulated at constant pressure of 100 MPa

The last stage of our analysis for this particular pressure was to obtain the Radial Distribution Function (RDF) for the two starting lattice structures. The RDF was obtained at a temperature T= 95 K (intermediate between the two transitional temperatures) and our constant pressure of 100 MPa. As we see from our curve below, the structure of the cell and the simulation does not depend on the initial configuration of the model.
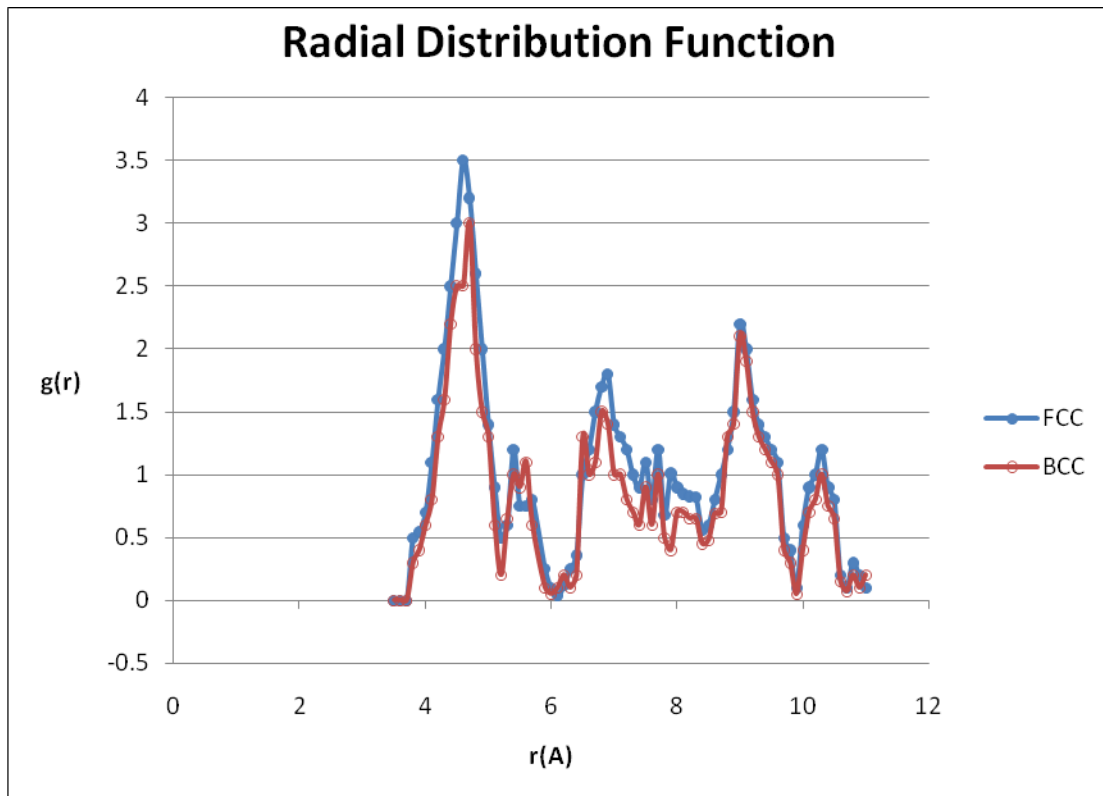
Figure (5.9): RDF curves for CF₄ model with N=500 simulated at constant pressure of 100 MPa and 2 starting configurations (FCC) and (BCC)

Although the caloric curve and the fourth cumulant intersection at P = 1 GPa did not show any other phase transitions in the system containing 256 molecules, a signal was detected in the Caloric curve for the 500 molecule system at a low temperature in the area of 45 K. However, the signal was not strong enough under the pressure conditions. The above procedure was repeated with other pressures, namely 1 GPa, 6GPa, 15 GPa and 35 GPa. The results obtained from the data acquired via these simulations did in fact support the findings above and also provided other information. The series of runs which were carried out at 20 GPa did actually clearly suggest a solid-solid phase transition at a temperature 42±0.4 K. Furthermore, runs carried out at a pressure of 35 GPa confirmed the transition at 42±0.4 and also showed another phase transition at 430±2 K. The tables and figures below summarize the results for all simulations at constant pressure.

5.5.2 Results at P = 100 MPa, 1 GPa, 6GPa, 15 GPa, 20 GPa and 35 GPa and T = 0 K to 450K

a. P = 100 MPa

Nature of Phase Transition	Temperature in K at which transition occurs	
	Current Simulation work	Previous Experimental work
Solid Phase II to Solid Phase I	86± 0.6	87 [126]
Solid Phase I to Liquid Phase (Melting)	107± 0.2	108 [126]

Table (5.3): Phase transitions at P= 100 MPa

b. P = 1 GPa

Nature of Phase Transition	Temperature in K at which transition occurs	
	Current Simulation work	Previous Experimental work
Solid Phase III to Solid Phase II	76± 0.4	76.2 [126]
Solid Phase V to Solid Phase IV	164± 0.5	NA

Table (5.4): Phase transitions at P= 1 GPa

c. _P = 6 GPa_

Nature of Phase Transition	Temperature in K at which transition occurs	
	Current Simulation work	Previous Experimental work
Solid Phase IV to Solid Phase III	297 ± 0.6	NA
Solid Phase V to Solid Phase IV	163± 0.2	NA

Table (5.5): Phase transitions at P= 6 GPa

d. _P = 15 GPa_

Nature of Phase Transition	Temperature in K at which transition occurs	
	Current Simulation work	Previous Experimental work
Solid Phase V to Solid Phase IV	163± 0.6	NA
Solid Phase VI to Solid Phase III	207± 0.6	203 [127]
Solid-Solid Phase Transition	42 ± 0.4	NA

Table (5.6): Phase transitions at P= 15 GPa

e. *P = 20 GPa*

Nature of Phase Transition	Temperature in K at which transition occurs	
	Current Simulation work	Previous Experimental work
Solid Phase VI to Solid Phase III	207± 0.6	203 [127]
Solid-Solid Phase Transition	42 ± 0.4	NA

Table (5.7): Phase transitions at P= 20 GPa

f. *P = 35 GPa*

Nature of Phase Transition	Temperature in K at which transition occurs	
	Current Simulation work	Previous Experimental work
Solid Phase V to Solid Phase IV	163± 0.6	NA
Solid Phase VI to Solid Phase III	204± 0.6	203 [127]
Solid Phase II to Solid Phase I	42 ± 0.4	NA
Phase Transition	420±2	NA

Table (5.8): Phase transitions at P= 35 GPa
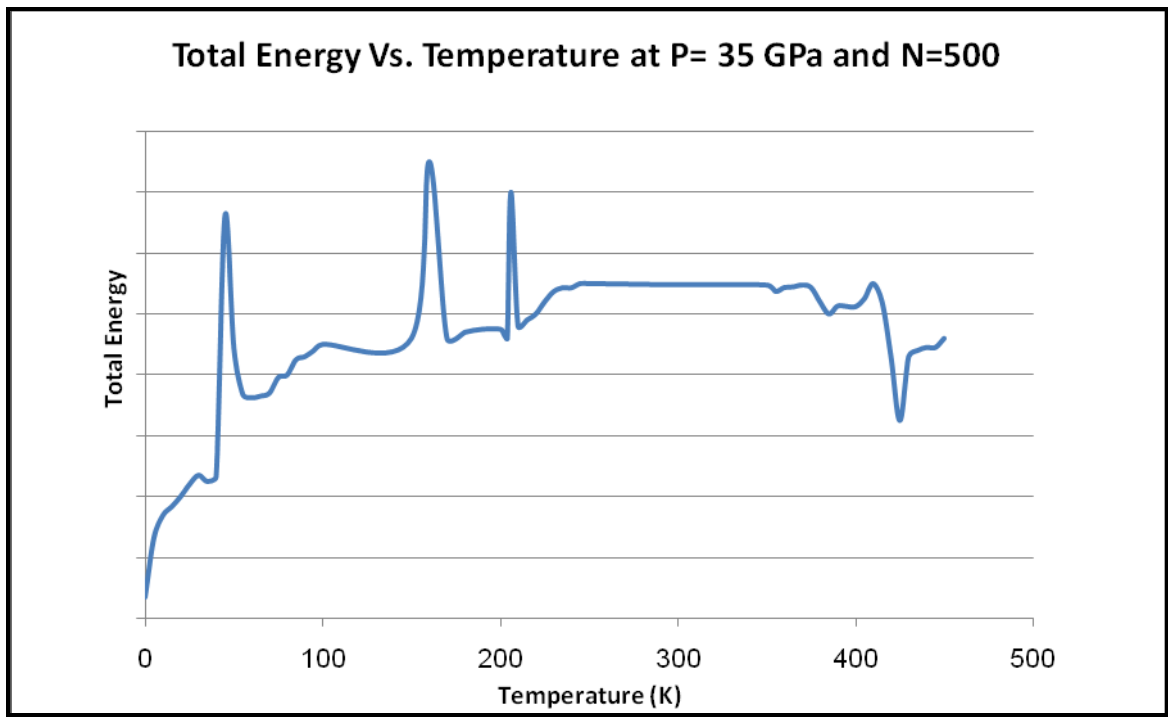
Figure (5.10): Caloric curve for simulation of 500 molecules of CF_4 at P=35 GPa

Figure (5.11): Binder fourth order cumulant curves for simulation of CF_4 at P=35 GPa and N= 256
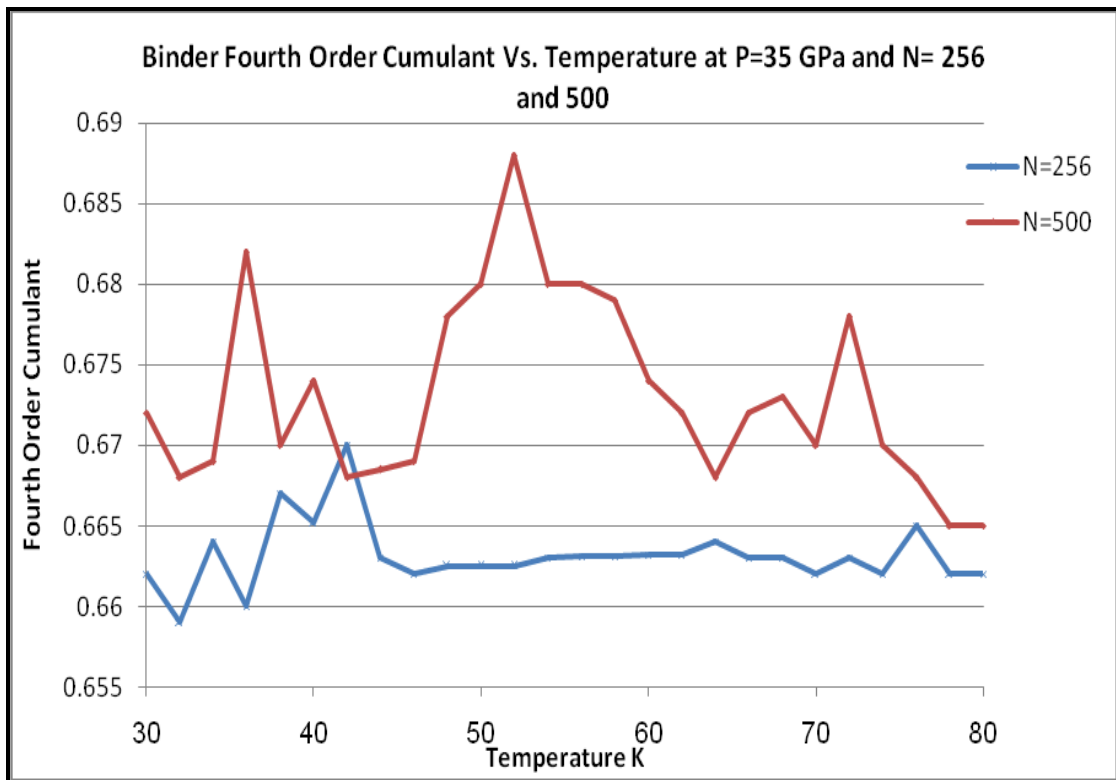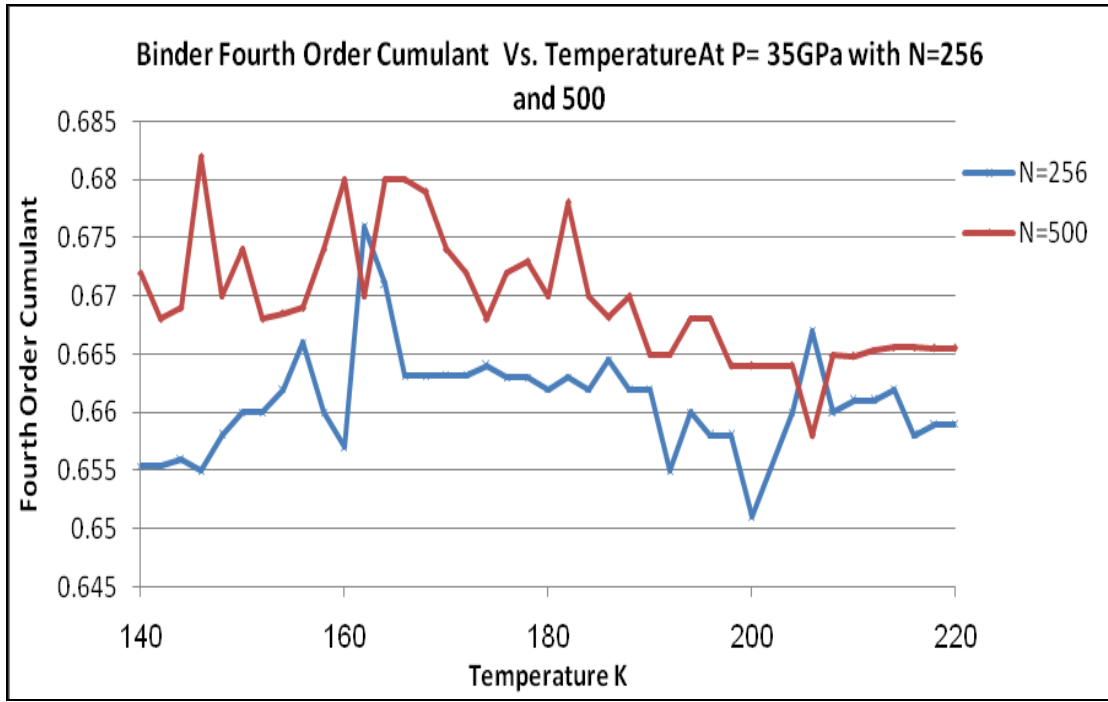*and 500 molecules*

Figure (5.12): Binder fourth order cumulant curves for simulation of CF₄ at P= 35 GPa and N= 256 and 500 molecules
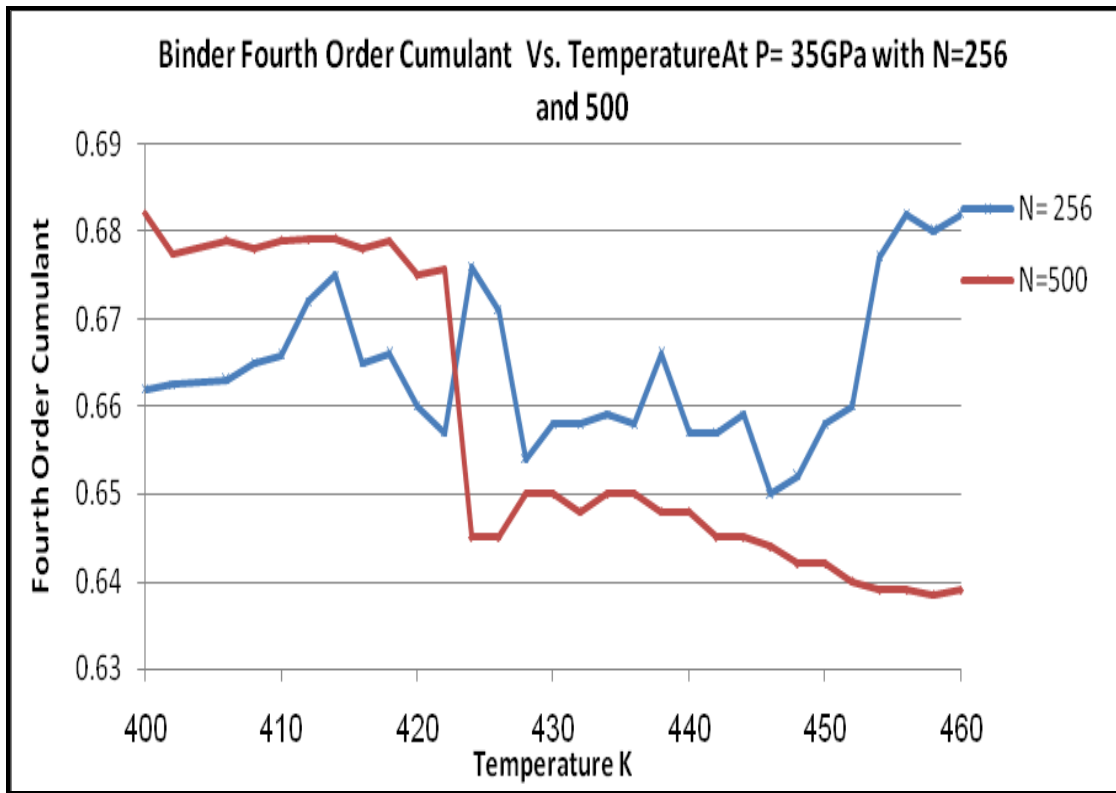

Figure (5.13): Binder fourth order cumulant curves for simulation of CF₄ at P= 35 GPa and N= 256 and 500 molecules

As shown in table (5.8) as well as figures (5.10), (5.11), (5.12) and (5.13), phase transitions were isolated during the P= 35 GPa runs at temperatures which had not previously given such results. The phase transition occurring at 42± 0.4 K has been shown for the first time via simulation techniques. Although it had not possible previously to obtain such results at low temperature ranges, however during our work we had used a computer cluster as well as a large system thus allowing us to simulate the system at these conditions and obtaining results which had not been previously distinguishable. Further, studies have given us information about the type of transitions. The transition occurring at 163± 0.6 K is a solid phase V to solid phase IV transition. Furthermore, the one occurring at 42 ± 0.4 K is a Solid Phase II to Solid Phase I transition.

5.6 RESULTS OF CONSTANT TEMPERATURE SIMULATIONS

The second stage of our work involves carrying out a number of runs at constant temperatures while varying the pressure. The exact same procedure was carried out at temperatures T= 20K, T= 40K, T=106 K, T= 160K, T= 200K, T= 250K and T=300K, as the pressure was varied from 0 GPa to 25 GPa. After terminating the runs the same calculations and observations were made. The results were analyzed in the same manner as was done in the previous section.

First It is worth mentioning that we were able to obtain a distinct phase transition at T=20 30 ± 2.5 MPa. This is one of those transitions at low temperature which are characterized by highly organized structures. Now let's take for example the runs carried out at T=40K. A series of runs were carried out at this low temperature with variations in pressure starting from 0 GPa to 25 GPa. The system with N= 108 molecules did not actually provide any sufficient results. The Caloric curves obtained had no sharp changes and the fourth cumulant curve barely touched the curve for the system with N=256. The system with N= 500 molecules on the other hand had a distinct discontinuity is the slope of the caloric curve corresponding to a pressure of 72 ± 6 MPa as well as an intersection with the Fourth cumulant curve of the N=256 system confirming the above observation.

Furthermore, we shall also point out the results at T=300 K since it is one of those temperatures which has been studied several times and we have several experimental values which we can compare to our work. At 300K, the fourth order cumulant curves of the systems N=256 and N=500 gave four intersection areas corresponding to four areas of discontinuity in the caloric curve. The translational order parameter confirmed the hypothesis we were testing that a transition taking place at a Pressure in the area of 1600 ± 90 MPa is a solid-liquid transition. Furthermore it deduces that all the three other phase transitions are solid-solid transitions. The rotational order parameter in this case was not truly necessary since the results it provided are just factual results that as the pressure increases on a certain system the order of that system increases too. The higher the pressure the higher the degree of the order in the system.

In the following section we have used the runs corresponding to T= 300K for the sake demonstration. All results corresponding to the different temperatures we have tested are listed in later sections.

5.6.1 Results at T= 300K

The caloric curve for the 500 molecules simulation model at T= 300 K was plotted to give the figure below.
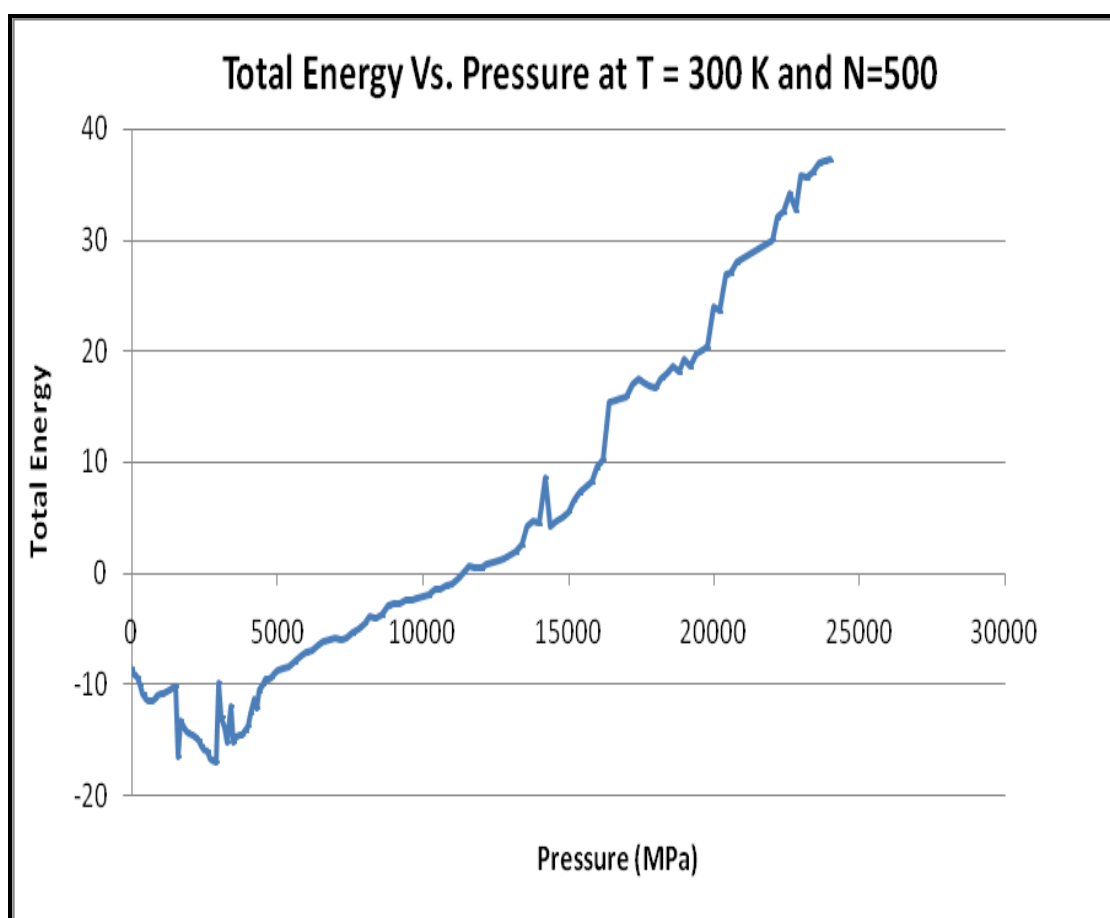
Figure (5.14): Caloric curve for simulation of 500 molecules of CF4 at T=300 K

Closer analysis and segmentation of the caloric curve above gave us the following resulting figures. We have only chosen to present those values where a phase transition is witnessed:
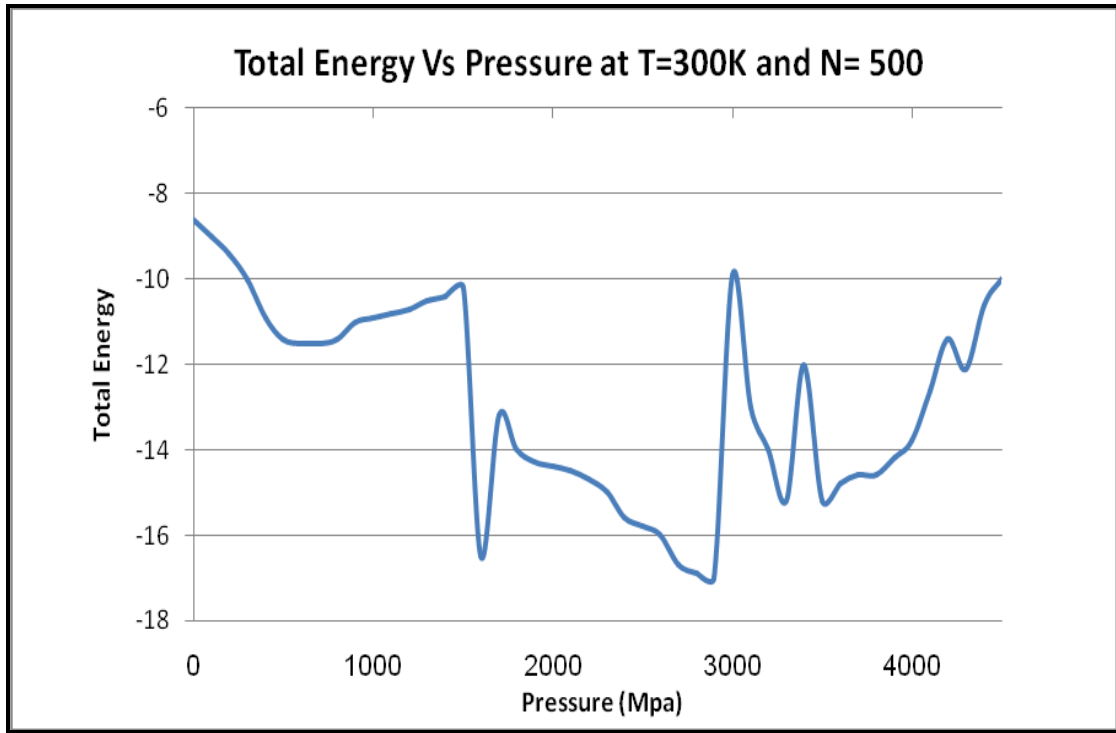
Figure (5.15): Caloric curve for simulation of 500 molecules of CF_4 at T=300 K (for P=0 to P=4500 MPa)
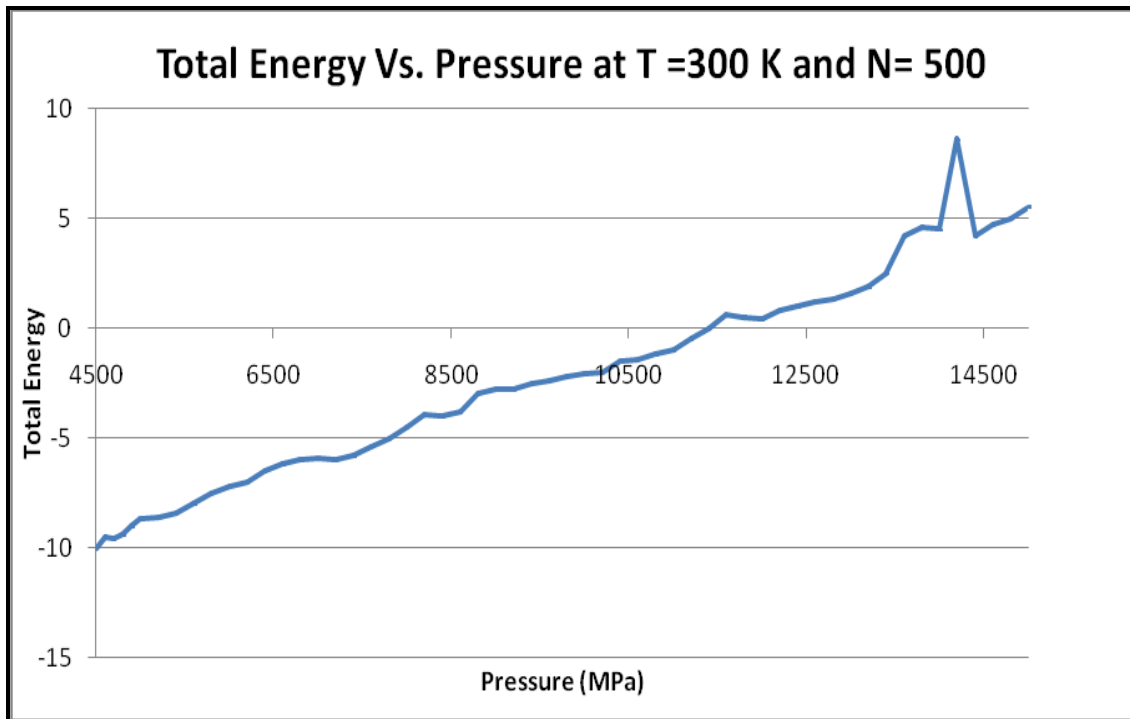
Figure (5.16): Caloric curve for simulation of 500 molecules of CF_4 at T=300 K (for P=4500 to P=15000 MPa)

As seen from figures (5.15) and (5.16), the phase transitions at T= 300K can be summarized as follows:

- Phase transition at 1600 ± 90

- Phase transition at 3000 ± 165

- Phase transition at 3400 ± 132

- Phase transition at 7900 ± 200

- Phase transition at 14200 ± 185

The Binder fourth order cumulants for the N=256 and N=500 models were plotted against each other and the intersections which were isolated confirmed the above stated results.
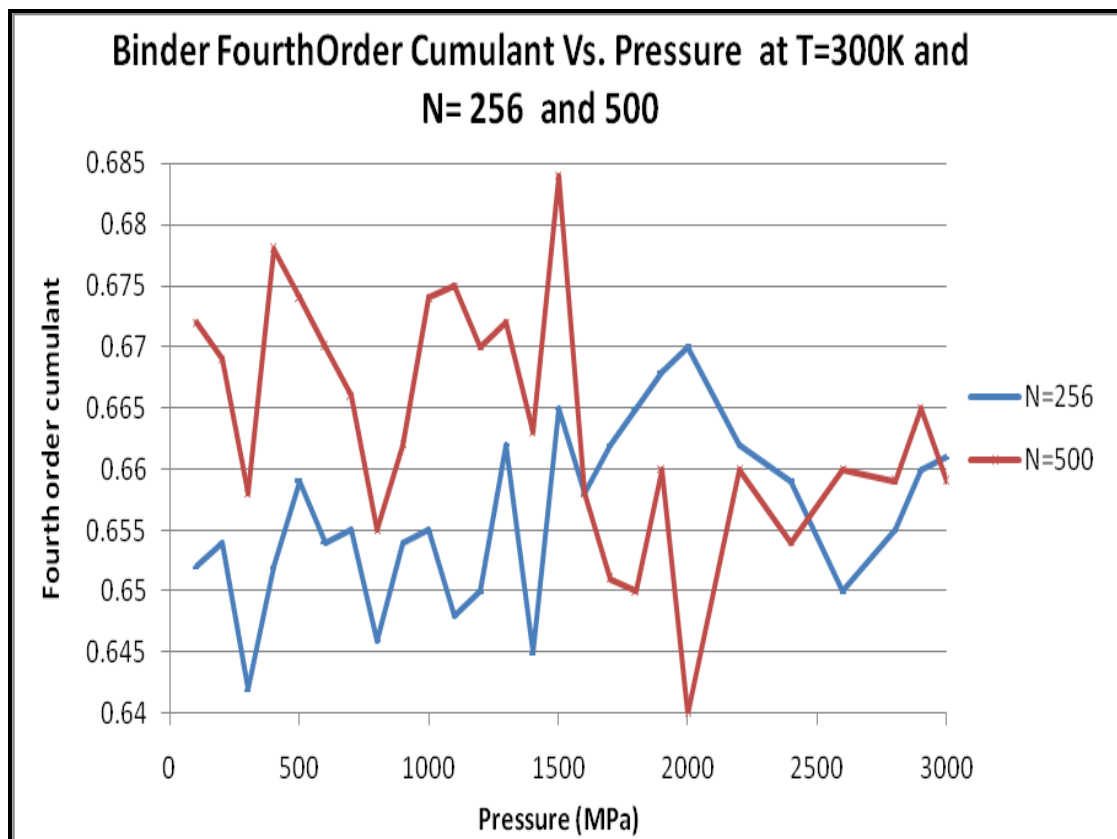
Figure (5.17): Binder fourth order cumulant curves for simulation of CF$_4$ at T = 300 K and N = 256 and N = 500 molecules (P = 0 MPa to P = 3000 MPa)

Figure (5.18): Binder fourth order cumulant curves for simulation of CF$_4$ at T = 300 K and N = 256 and N = 500 molecules (P = 2600 MPa to P = 5800 MPa)
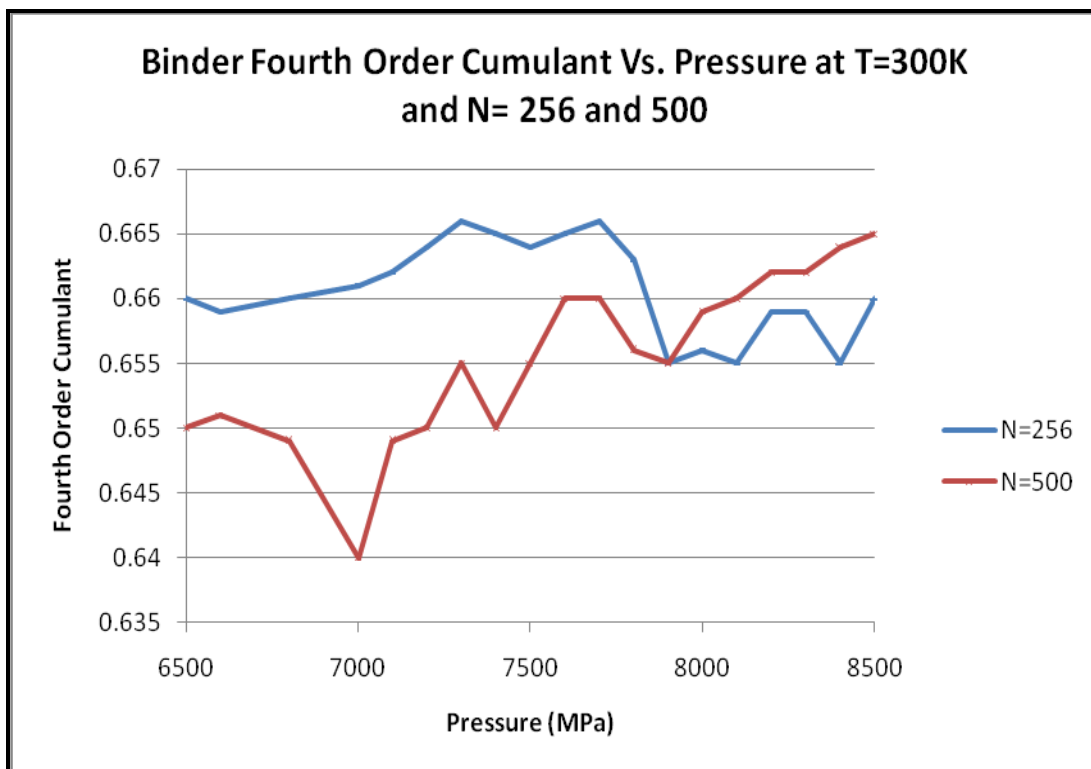
Figure (5.19): Binder fourth order cumulant curves for simulation of CF$_4$ at T = 300 K and N = 256 and N = 500 molecules (P = 6500 MPa to P = 8500 MPa)
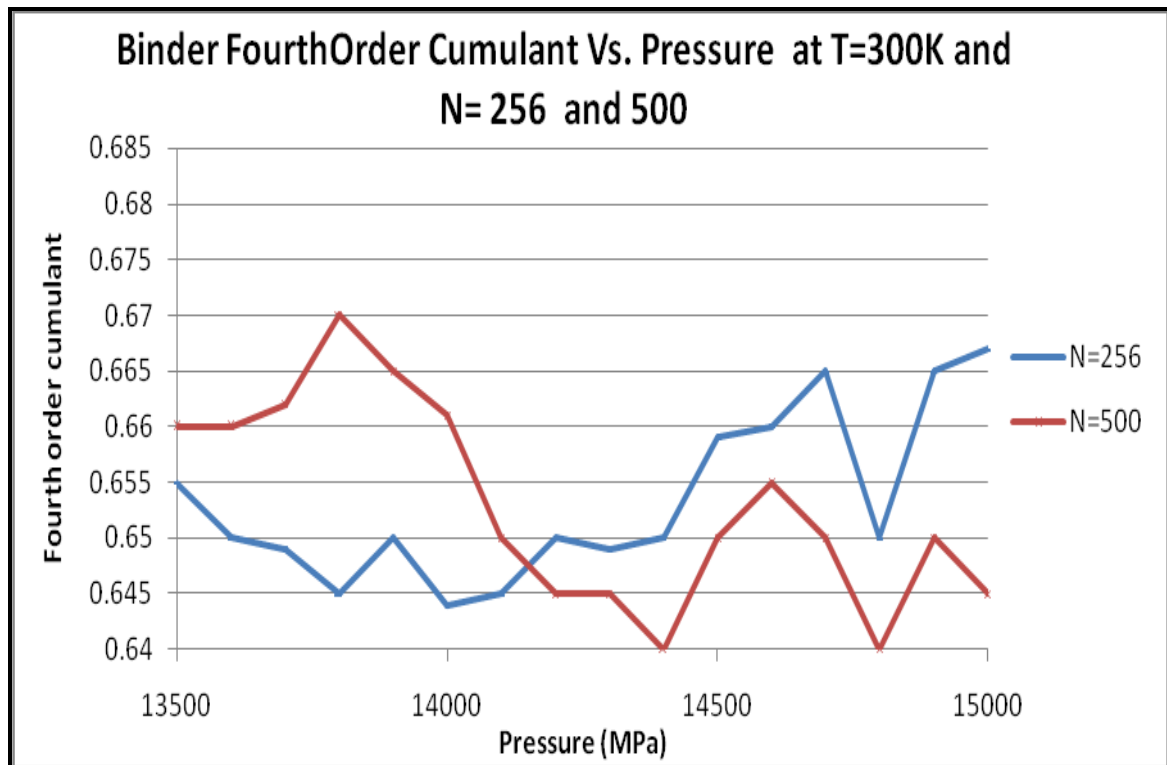
Figure (5.20): Binder fourth order cumulant curves for simulation of CF$_4$ at T = 300 K and N = 256 and N = 500 molecules (P = 13500 MPa to P = 15000 MPa)
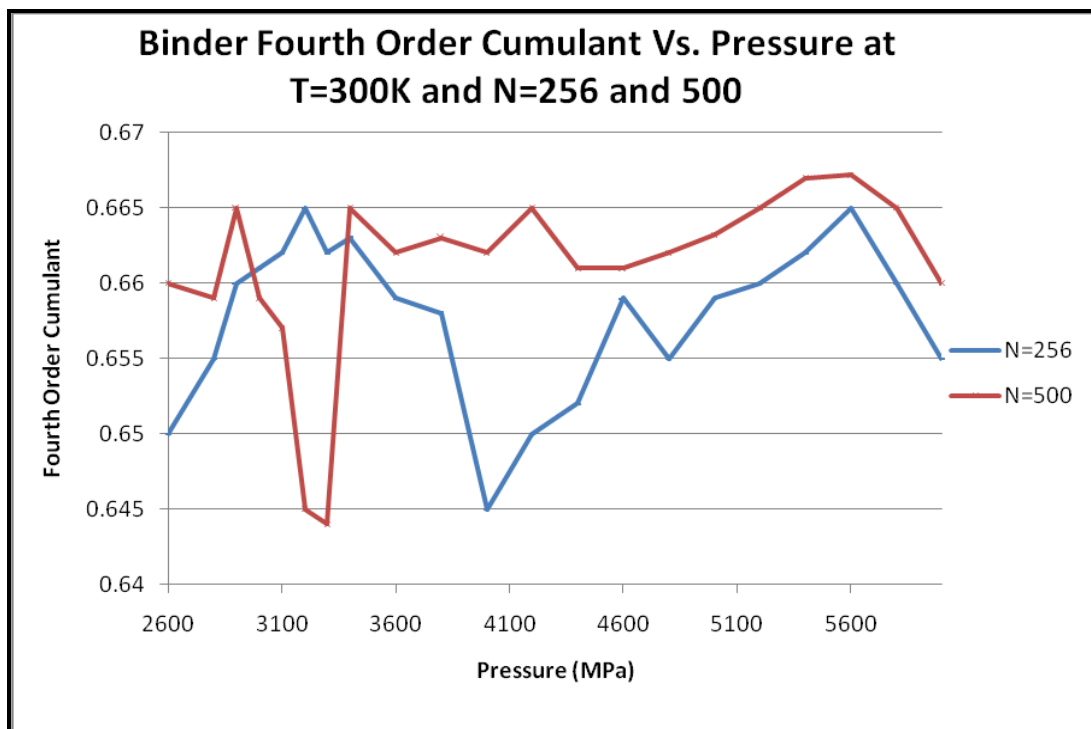
To distinguish between the different transitions we plotted the translational and rotational order parameters. The translational order parameter at T=300 K, figure (5.21), shows that as the pressure increases the translational order parameter tends to zero. At a pressure of P=1600 ± 90 the order parameter reaches zero which indicates that the transition taking place at this particular pressure moves the system from a state of disorder to a highly ordered state, thus it is a liquid-solid phase transition. Consequentially, other phase transitions taking place at higher pressures must be solid-solid phase transitions.

Furthermore, since we have more than one solid-solid phase transitions, and on order to distinguish between the different solid phases, the rotational order parameter is plotted against pressure. As seen in figure (5.22), as the pressure is increased the rotational order parameter tends towards unity thus indicating a highly ordered phase. This implies that high pressure solid phases are more ordered than those exhibited at lower pressure. The figures below illustrate these findings.
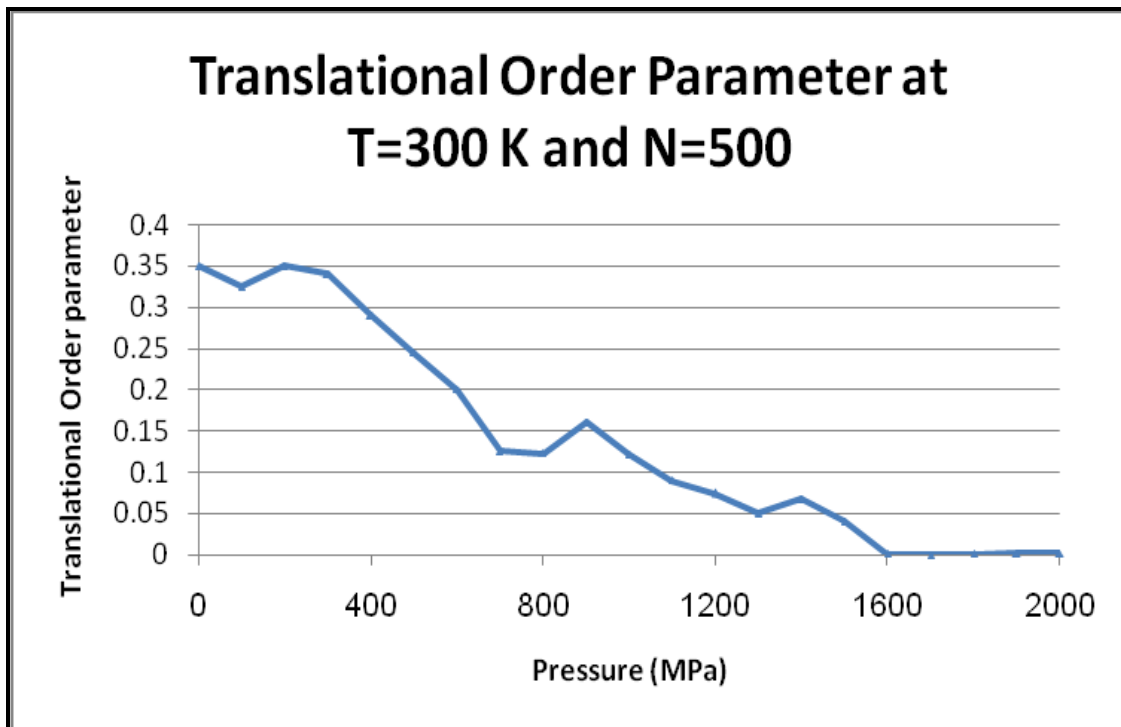
Figure (5.21): Translational Order Parameter curve for CF$_4$ model with N=500 simulated at constant temperature of 300 K
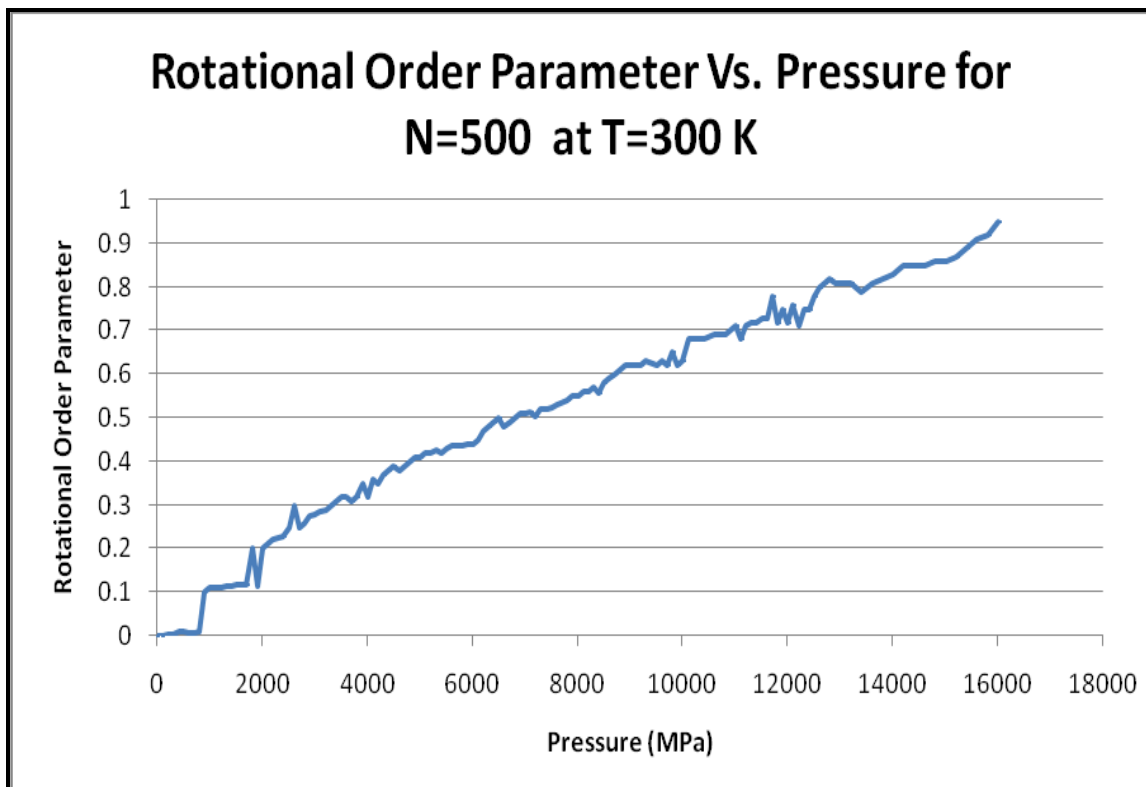
Figure (5.22): Rotational Order Parameter curve for CF$_4$ model with N=500 simulated at constant temperature of 300 K

As previously discussed, the radial distribution function (RDF) is plotted for the two starting lattice structures in order to ensure that the structure of the unit cell as well as all other simulation results are not dependent on the initial configuration.

The results which we have obtained at T=300 K may be summarized in the following table.

Nature of Phase Transition	Pressure in MPa in at which transition occurs	
	Current Simulation work	Previous Experimental work
Liquid to Solid Phase I	1600 ± 90	1860 [128]
Solid Phase I to II	3000 ± 165	2840 [128]
Solid Phase II to III	3400 ± 132	3500 [128]
Solid Phase III to IV	7900 ± 200	8600 [128]
Solid Phase IV to IV	14200 ± 185	13600 [128]

Table (5.9): Phase transitions at T= 300 K

Similarly, the same procedure described above was followed at, T=40K, T=106 K, T= 160K, T= 200K and T= 250K. The results obtained are summarized as follows:

All results corresponding to the different temperatures we have tested are listed below.

5.6.2 Results at T= 40K, 106K, 160K, 200K AND 250 K

Nature of Phase Transition	Pressure in MPa in at which transition occurs	
	Current Simulation work	Previous Experimental work
Liquid to Solid Phase I	72± 6	NA
Solid Phase I to II	140± 3	NA
Solid Phase II to III	Not found	NA
Solid Phase III to IV	Not found	NA
Solid Phase IV to V	Not Found	NA

Table (5.10): Phase transitions at T= 40 K

Nature of Phase Transition	Pressure in MPa in at which transition occurs	
	Current Simulation work	Previous Experimental work
Liquid to Solid Phase I	96 ± 12	91 [128]
Solid Phase I to II	286 ± 25	279 [128]
Solid-Solid Transition	470 ± 32	NA
Solid-Solid Transition	1600 ± 132	NA

Table (5.11): Phase transitions at T= 106 K

Nature of Phase Transition	Pressure in MPa in at which transition occurs	
	Current Simulation work	Previous Experimental work
Liquid to Solid Phase I	370 ± 55	442 [128]
Solid Phase I to II	1100 ±150	905 [128]
Solid-Solid Transition	3900± 170	NA
Solid-Solid Transition	9600 ± 110	NA

Table (5.12): Phase transitions at T= 160 K

Nature of Phase Transition	Pressure in MPa in at which transition occurs	
	Current Simulation work	Previous Experimental work
Liquid to Solid Phase I	800 ± 10	761 [128]
Solid Phase I to II	1580 ±140	1460 [128]
Solid-Solid Transition	9050± 150	NA
Solid-Solid Transition	15300 ± 120	NA
Solid-Solid Transition	18600 ± 200	NA

Table (5.13): Phase transitions at T= 200 K

Nature of Phase Transition	Pressure in MPa in at which transition occurs	
	Current Simulation work	Previous Experimental work
Liquid to Solid Phase I	1300 ± 10	1255 [128]
Solid Phase I to II	2300 ±100	2230 [128]
Solid-Solid Transition	2900± 50	NA
Solid-Solid Transition	11000 ± 200	NA

Table (5.14): Phase transitions at T= 250 K

5.6.3 Crystal Structure, Lattice parameters and Space Groups of the different phases of Carbon Tetrafluoride.

According to the data we have acquired and the calculations carried out, we were able to deduce the following structures for the different phases of Carbon Tetrafluoride.

Phase	a(Å)	b(Å)	c(Å)	α(°)	β(°)	γ(°)	Space Group
I	13.73	12.79	13.36	90	93.6	90	$C_{2/m}$
II	8.21	8.27	6.34	90	87.6	90	C_m
III	8.89	8.86	6.75	90	95.2	90	C_m
IV	7.89	7.87	6.09	90	86.9	90	C_{mm}
V	7.69	7.61	6.32	90	90	90	$C_{mm/2}$
Solid phase at 20K	8.82	4.13	15.35	90	151.01	90	C_c
Solid Phase at 40K	8.71	4.16	15.27	90	150.7	90	C_c
Solid Phase at OK	8.36	4.21	8.13	90	119	90	$C_{2/c}$

Table (5.15): Lattice Parameters and Space Groups of different phases of Carbon Tetrafluoride

CHAPTER 6: CONCLUSIONS & DSCUSSIONS

Molecular Dynamic Simulation is one of the most important simulation methods today. It has a wide range of applications which have given it a broad scope of utilization in most applied sciences.

In this thesis we have worked to simulate a system of CF_4 molecules as an attempt to describe its phase transitions as well as its crystal structure in the solid phase. Most of the results acquired were within good agreement with experimental values most of which have been acquired via the use of spectroscopic techniques such as IR spectroscopy, Raman spectroscopy and X-Ray diffraction.

Furthermore, we were able to probe areas which have yet to be experimented upon such as high pressure areas and low temperature areas. The results which we have acquired can decrease the ambiguity surrounding phase transitions of CF_4. Phase transitions at temperatures 42 ± 0.4 K and 420 ± 200 K were isolated. These particular transitions may have been identified experimentally but had never been isolated using MD simulation. In addition to that a few high pressure transitions were also isolated at high temperatures, particularly 250 K. For example a solid-solid transition was isolated at 11000 ± 200 MPa, and as previously discussed it was rather distinct that the crystal lattice structures of higher pressure solid phases are of more ordered than those of low pressure phases.

These findings could be utilized to suggest a plausible and more comprehensive phase diagram for Carbon Tetrafluoride. As a matter of fact we were able to suggest such phase diagram and are hoping that future experimental and/or simulation work can confirm it.

First, it might help to study a phase diagram that had been recently developed via the utilization of MD simulation. The following figure was proposed in 2006 [101]:
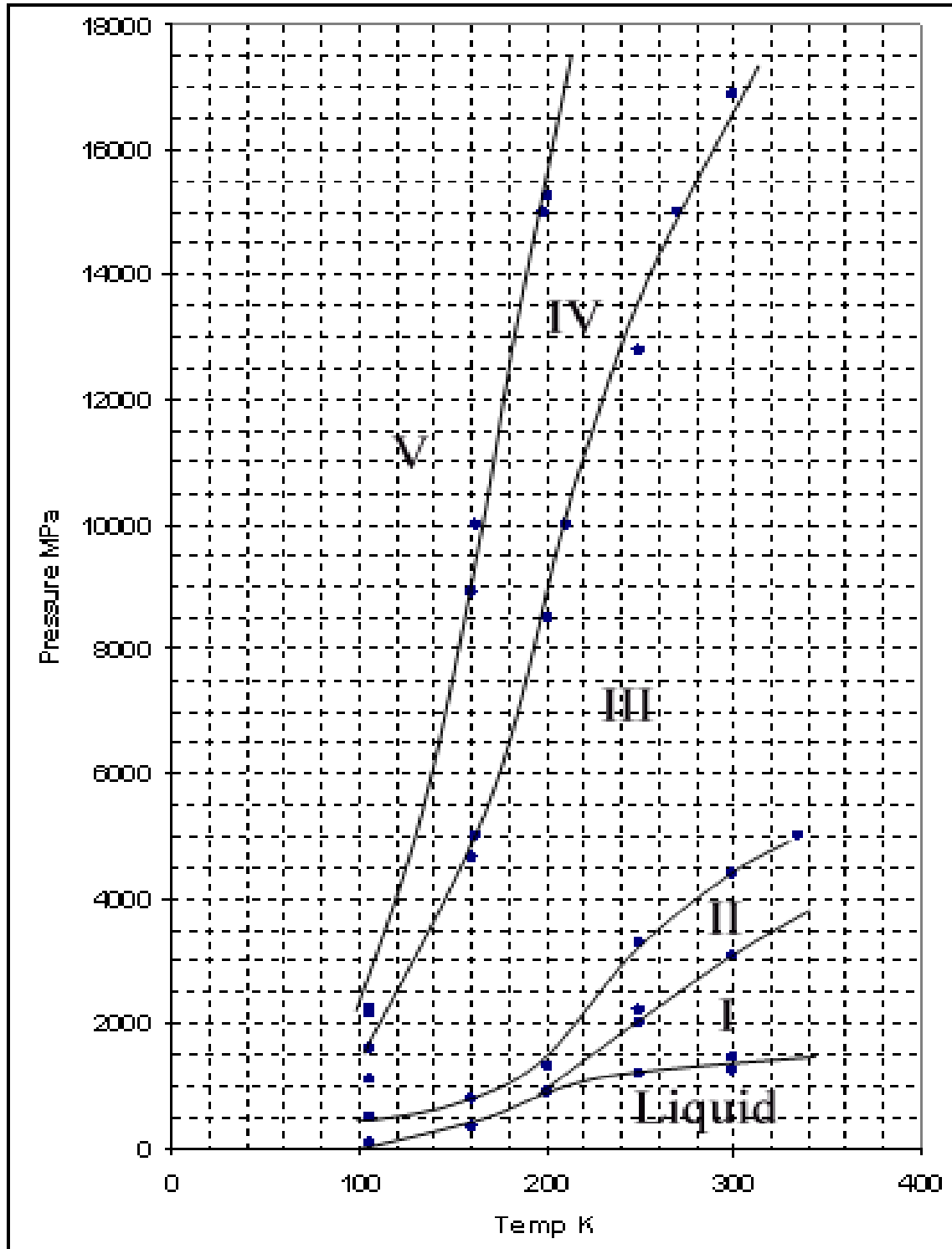
Figure (6.1): Phase Diagram of CF₄ proposed by S. El Sheikh and K. Barakat [101]

In our work, we were able to confirm the findings in the phase diagram proposed by S. El Sheikh and K. Barakat [101]. We were also able to extend the phase diagram to areas of low temperatures and high temperatres which had not been identified in the phase diagram above. The results and findings which we had been able to conclude in this work may be presented as follows:
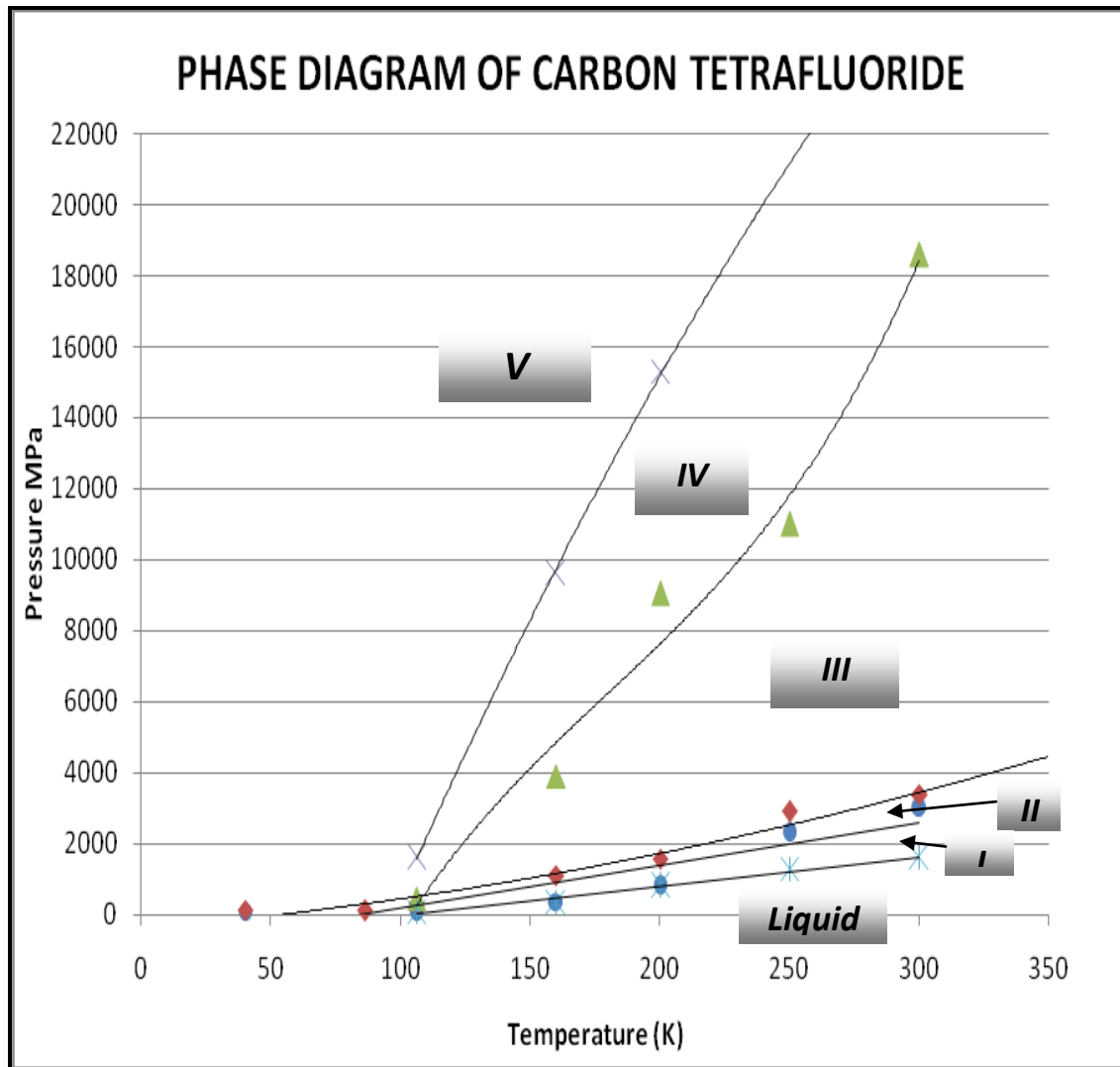
Figure (6.2): Proposed phase diagram of CF$_4$

Having arrived at these results, one can say that MD simulations provide successful means of studying phase transitions and lattice structures of matter. In the past this was only possible as long as the temperatures at which the simulations are carried out are not too low. Low temperatures were

studied by means of quantum (ab-initio) molecular dynamics only. Classical MD simulations had not been able to present accurate results at low temperatures. However, today with the availability of large computer clusters and high technology software packages we were able to simulate larger systems (N=500) and thus classical MD was successful in proposing accurate results at low temperatures as well as high temperature. The larger the simulated system the more accurate the results since we are bringing the simulated system as closely as possible to the structure of the real system. Hence, if the system is large enough, the software is capable and the cluster capacity is high we may confidently say that MD is a successful tool for simulations, even at low temperatures.

Furthermore, one cannot ignore the fact that MD simulations are amongst the most important statistical mechanics methods which enables us to explore the phase space of any system with acceptably accurate results which are comparable to experimental findings. As long as a proper model is utilized, one can consider MD simulations to be a very reliable and relatively simple tool to model, control and observe the behavior of a very large number of systems.

Future work may include simulations of larger bulks of molecules with different initial configurations. For example, one may wish to extend the study carried out for the purpose of this thesis to a system of 864 molecules (FCC $4M^3$, where M= 6) and try to simulate it at temperatures lower than 20K and higher than 450 K. It may also be extended to probing areas of high pressures (50 GPa for example), which despite the fact that it wouldn't be used in practical applications, however it can provide useful information about the validity of the model being utilized. It is true that simulating a larger system is usually quite difficult and requires larger computing facilities, such as a cluster as well as a highly specialized, capable and time effective software package. However the results obtained during these simulations are usually very accurate and much closer to experimental values.

In addition to that, the variations in the size of the MD system as well as the time step allowed, may help us arrive at a certain relation between the two factors, which if proven correct shall facilitate future work. For example, while simulating the system containing 500 molecules much time was consumed while deciding on the most appropriate time step as well as the most suitable number of steps. It was very important that we arrive at the correct conditions, which ensure that the system has reached equilibrium, and yet the simulation does not continue for long after that point. We have noticed during out trial-and-error phase of deciding on the time step, that a larger system requires a shorter time step and a larger number of steps to reach equilibrium. Similarly, we have also noticed that higher pressure and lower temperature simulations gave the comparably more accurate results when we used fewer and longer time steps than when we used more and shorter time steps. Such findings may be tested during future work and if confirmed shall be very useful and shall no doubt save us a lot of time and efforts that are consumed during the initial stages of preparing the model and the various input files.

Furthermore, future work may also include the utilization of quantum (ab-initio) molecular dynamics to further investigate regions of very low temperatures as well as areas of very high temperatures. Ab-initio molecular dynamic techniques require high specification computing systems and thus had not been available to us in the past. However, now we have the facilities which may allow us to utilize ab-initio molecular dynamics simulations to study and analyze crystal structure, phase transitions and last but not least it would allow us to study mechanisms of chemical reactions which cannot be monitored or investigated by experimental techniques.

REFERENCES

1) G. E. P. Box and M. E. Muller (1958). Ann. math. Stat. 29, 610

2) J. B. Gibson, A. N. Goland, M. Milgram, and G. H. Vineyard, (1960) Phys. Rev. 120, 1229.

3) H. Perros. Computer Simulation Techniques: The definitive introduction. Raleigh, NY, 2009

4) http://upload.wikimedia.org/wikipedia/commons/thumb/9/9b/Molecular_simulation_process.svg/733px-Molecular_simulation_process.svg.png

5) H.Muller- Krumbhaar and K. Binder. Dynamic properties of Monte-Carlo method in statistical mechanics. J. Stat. Phys., 8:1-24, 1973

6) M. P. Allen and D. J. Tildesley, Computer Simulations of Liquids (Clarendown, Oxford, 1994).

7) D.C. Rapaport. The Art of Molecular Dynamics Simulation. Second Edition. Cambridge University Press. 2004

8) M. P. Allen. Introduction to Molecular Dynamics Simulation. NIC series, Vol. 23. 2004

9) K. Binder, J. Horbach, W. Kob and W. Paul. Molecular Dynamics Simulation. Institut Fur Physik. Germany. 2003

10) N. Metropolis, A. W Rosenbluth, M. N. Teller, (1953). J. chem. Phys. 21, 087

11) N. Metropolis. And S. Ulam. (1949). J. Am. stat. Ass. 44, 335

12) R. J. Sadus. Molecular Simulation of Fluids: Theory, Algorithms and Object-Orientation. Elsevier Science B.V. Netherlands, 1999.

13) R. H. Stote. Introduction to Molecular Dynamics Simulations. Institut de Chimie LC3-UMR, Franc

14) H. Goldstein Classical mechanics (2nd edn). Addison-Wesley, Reading, MA. 1980

15) J. H. Noggle, Physical Chemistry, 2nd edition, Scott Foresman& Co, 1996

16) F. Hedman. Algorithms for Molecular Dynamics Simulations; Advancing the computational Horizon. Stockholms Universitet. 200

17) C.F. Gerald and P. O. Wheatley. Applied Numerical Analysis. Addison Wesley Longman Inc.1997

18) D. Frenkel and B. Smith. Understanding Molecular Simulation. Second Edition. San Diego Academic Press.2002

19) http://en.wikipedia.org/wiki/Finite_difference_method

20) R D. Skeel. Integration Schemes for Molecular Dynamics and Related Applications. University of Illinois. Urbana USA. 1998

21) W.F. Van Gunsteren and H. J. C. Berendsen. Computer simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry. VCH Germany. 2003

22) D. Frenkel and B. Smit. Understanding Molecular Simulation From Algorithms to Applications. Academic Press. USA. 2002

23) P. Mark Rodger. On the Accuracy of some common Molecular Dynamics Algorithms. Molecular Simulation. Vol. 3 Iss.5-6. 1989

24) W.F Van Gunsteren and H. J. C Berendsen. A Leap-frog Algorithm for Stochastic Dynamics. Molecular Simulation. Vol. 1, Iss. 3, 1988

25) K. Diethelm, N. J. Ford and A. D. Freed. A predictor-Corrector Approach for Numerical Solution of Fractional Differential Equations. Nonlinear Dynamics 29: 3-22. Netherlands 2002

26) C. W.Gear, (1971). Prentice-Hall, Englewood Cliffs, NJ.

27) W C Swope, H C Anderson, P. H Berens and K R Wilson (1982). J. chem. Phys. **76**:637

28) M. J. Sangster, and M. Dixon (1976). Adv. Phvs. **25**, 247

29) R. W. Hockney, and J. W. East wood, (1981). McGraw-Hill, New York.

30) K. Refson, (1985) Physica **131B**, 256–266.

31) D. Beeman. Some multistep methods for use in molecular dynamics calculations. Journal of Computational Physics **20** . (1976)

32) C. W. Gear (1966).. Report ANL 7126, Argonne National Laboratory

33) P. M. Rodger, (1989) Mol. Simul. **3**, 263.

34) D. J. Evans, (1977) Mol. Phys. **34**, no. 2, 317–325.

35) D. J. Evans and S. Murad, (1977)Mol. Phys. **34**, no. 2, 327

36) G. S. Pawley and M. T. Dove, (1985) Mol.Phys. **55**, no. 5, 1147

37) G. S. Pawley, (1981) Mol. Phys. **43**, no. 6,1321–1330.

38) K. Refson and G. S. Pawley, (1987) Mol. Phys. **61**, no. 3, 669

39) .J. G. Powles,W. A. B. Evans, E. McGrath, K. E. Gubbins, and S. Murad, (1979) Mol. Phys. **38**, 893

40) F. A. Bornemann and C. Schutte. A Mathematical Approach to Smoothed Molecular Dynamics: Concerning Potentials for Freezing Bond Angles. ZIB. 1995

41) K. Refson, Moldy: a portable molecular dynamics simulation program for serial and parallel computers., Computer Physics Communications, 126(3):309 (2000) and http://www.earth.ox.ac.uk/~keithr/moldy.html.

42) A. R. Leach, Molecular Modeling: Principles and Applications, Longman (1996).

43) http://isaacs.sourceforge.net/phys/pbc.html

44) R. B. Shirt, S.R. Burt, A. M. Johnson. Periodic boundary condition induced breakdown of the equipartition principle and other kinetic effects of finite sample size in classical hard-sphere molecular dynamics simulation. J Chem Phys 125(16):164102. 2006

45) R. Sonnenschein, (1985)J. Comp. Phys. **59**, no. 2, 347

46) S. K. Joshi and R. A. Mashelkar. Solid State Chemistry: Selected Papers of C. N. R. Rao.World Scientific Publishing Co. Pte. Ltd. 1995

47) http://wanglab.bu.edu/DLPOLY2/node263.html Peter Atkins and Julio de Paula. Atkins Physical Chemistry. 8th Edition W.H Freeman.

48) A. Y. Toukmaji, J. A. Board Jr. Ewald Summation Techniques in perspective: a survey. Duke University 1995.

49) L. Zhigilei. Introduction to Atomistic Simulations. University of Virginia

50) Collection of review articles on interatomic potentials. Feb. 1996 issue of MRS Bulletin.

51) http://en.wikipedia.org/wiki/LennardJones_potential#Truncated_Lennard-Jones_potential

52) B. Smith Phase diagrams of Lennard-Jones Fluids. Journal of Chemical Physics 96 (11) 1992

53) D. W. Brenner. The art and science of an analytical potential. Phys. Stat. Sol. B 217, 23-40 (2000)

54) B.J Garrison and B. Srivastava. Potential energy surfaces for chemical reactions at solid surfaces. Ahhu Rev. Phys. Chem. 46 1995

55) C. J. Fennell and J. D. Gezelter. Is the Ewald Summation still necessary? Journal of Chemical Physics 124, 234104. 2006

56) P. Ewald (1921). Ann. Phys. 64:253

57) http://en.wikipedia.org/wiki/Molecular_geometry

58) http://library.thinkquest.org/C004970/atoms/shapes.htm

59) R. J. Gillespie. Molecular Geometry. Van Nostrand Reinhold. 1972

60) Ronald James Gillespie and Paul L a Popelier. Chemical Bonding and Molecular Geometry: From Lewis to Electron Densities. Oxford University Press. 2001

61) R. F. W. Bader. Atoms in Molecules: A Quantum Theory. Clarendon Press. 1994

62) http://chemistrypartner.blogspot.com/

63) R. S. Cahn and Sir Christopher Ingold. Specifications of Molecular Chirality. Verlag Chemie. GmbH, Germany 2003

64) C. Hammond. The basics of Crystallography and Diffraction. Oxford Press. 1997

65) http://chemed.chem.purdue.edu/genchem/topicreview/bp/ch13/unitcell.php

66) http://www.ee.ccny.cuny.edu/www/web/crouse/I3600/Lectures/Crystalline%20State%20Review.htm

67) T. L. Brown. Chemistry: the Central Science. 10th ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2006

68) http://catalog.flatworldknowledge.com/bookhub/4309?e=averill_1.0-ch12_s02

69) D. J. Willock. Molecular Symmetry. John WIley and Sons Limited. 2009

70) C. E. Housecroft and A. G. Sharpe: Inorganic Chemistry, Pearson Education 2005

71) J. H. Noggle. Physical Chemistry 2nd edition. Foresman & Co, Glenview, IL, 1996

72) M. Gerken.classes.uleth.ca/.../3830%20lecture%20notes%20part...

73) D. C. Harris and M. D. Bertolucci. Symmetry and Spectroscopy: An Introduction to Vibrational and Electronic Spectroscopy. Oxford University. 1978

74) A. Alavi. Symmetry and Perturbation Theory. University of Cambridge. 2009

75) L. Yu Kiang. Axioms and algebraic systems. University of singapore. 1998

76) D. M. Bishop. Group Theory and Chemistry. 1973

77) http://www.webqc.org/printable-symmetrypointgroup-ct-c2v.html

78) http://www.staff.ncl.ac.uk/j.p.goss/symmetry/Water/water0.html

79) M. Ladd; forwarded by Lord Lewis, Harwood. Symmetry and Group Theory in Chemistry. Chemical Science Series, 1998

80) D. S. Schonland, Van Nostrand. 1965. Molecular Symmetry; an introduction to Group theory and Its Uses in Chemistry,

81) G. A. Lawrance. Introduction to Coordination Chemistry. John Wiley and Sons, Ltd. 2010

82) G. Burns. Introduction to Group Theory with Applications; Academic Press: New York. 1977

83) http://en.wikipedia.org/wiki/Space_group

84) S.R. Hall. Space-Group Notation with an Explicit Origin. Acta Cryst. A37: 517-525

85) Plesken, Wilhelm and W. Hanrath. The lattices of six dimensional space. Math. Comp. 43 (168). 1984

86) D. B. Litvin. Tables of crystallographic properties of magnetic space groups. Acta Cryst. A 64 (Pt 3) 2008

87) E. A. Three dimensional Space Groups. Chapter 3 58-98. Arizona University.

88) http://chemwiki.ucdavis.edu/@api/deki/pages/1946/pdf

89) R. Mirman. Point Groups, Space Groups, Crystals, Molecules. World Scientific Publishing Co. 1999.

90) W. Press. Structure and Phase Transition of Solid Heavy Methane (CD_4). J. Chem. Phys. 56, 2597. 1972.

91) http://en.wikipedia.org/wiki/File:Carbon-tetrafluoride-2D-dimensions.png

92) A. L. Allred. Electronegativity values from thermocemical data.Journal of Inorganic and Nuclear Chemistry. Vol. 17 issues 3-4. 1961

93) http://webbook.nist.gov/cgi/cbook.cgi?ID=C75730&Mask=4#Refs

94) http://www.chem.uwec.edu/chem150_s07/elaborations/unit1/unit1-e-polarity/unit1-e-try-this-answer.html

95) O. Meth, A.R. Katritzky, C. Wayne, T. L Gilchrist. Comprehensive Organic Functional Group Transformations; Synthesis: Carbon with Three or Four Attached Heteroatoms. Pergamon 1995.

96) C.W.W Hoffman and R.L. Livingston. The molecular structure of Carbon Tetrafluoride. Indiana. 1953

97) A. K. Eolf, J. Glinnemann and M. Schmidt. Packing of Tetrahedral EX_4 molecules. CrystEngComm. 2008

98) K. Refson. Moldy User's Manual. Parks Road. Oxford. 2001

99) N. Ogbonna. Supervisor: Kristian Muller-Nedebock. Molecular Dynamics Simulation. African Institute for Mathematical Sciences. 2004

100) http://en.wikipedia.org/wiki/Angular_velocity

101) S. M. El-Sheikh, K. Barakat and N. M. Salem "Phase transitions of methane using molecular dynamics simulations" Journal of Chemical Physics 124, 124517, 2006

102) H. C. Andersen. Molecular Dynamics Simulations at constant pressure and/or temperature. Stanford University. 1979

103) S. Nose. A molecular dynamics method for simulations in the canonical ensemble. Division of Chemistry, National Research Council Canada. 1983

104) http://en.wikipedia.org/wiki/Nos%C3%A9%E2%80%93Hoover_thermosta

105) H.C. Anderson. J. Chem. Phys, 72, 2384. 1980.

106) M. Parrinello and A. Rahman, J. Appl. Phys. 52, 7182 (1981).

107) http://www.ch.embnet.org/MD_tutorial/pages/MD.Part3.html

108) L.H Hall. Group Theory and Symmetry in Chemistry. McGraw-Hill. New York. 1969.

109) M. P. Allen and D. J. Tildesley, Computer Simulation of Liquids, Clarendown, Oxford (1994).

110) M. Bosch. The Molecular Dynamic Simulation of Neutral Argon Particles.2008

111) http://www.fisica.uniud.it/~ercolessi/md/md/node37.html

112) http://w3.iams.sinica.edu.tw/lab/jlli/thesis_andy/node14.html

113) http://www.fisica.uniud.it/~ercolessi/md/md/node35.html

114) S-Ho Tsai. Fourth-Order Cumulants to Characterize the Phase Transitions of a Spin-1 Ising Model. State of NY. 1996

115) S. Sasaki, Y. Ikeda and H. Shimizu. J. Phys. Soc. Japan **60** 1560. 1991

116) H. E. Lorenzana, M. J. Lipp, W. J. Evans, and N. Hemmi, J. Low Temp. Phys. **122**, 279 (2001).

117) D. Shindo, J. Phys. Condens. Matter 14 (2002) 10653-10656

118) T. Schlick. Molecular Modeling and Simulation. Springer-Verlag, Berlin, Germany. 2002

119) S. C. Greer and L. Meyer, J. Chem. Phys. **51**, 4583 (1969).

120) D. N. Bol'shutkin, V. M. Gasan, A. I. Prokhvatilov, and A. I. Erenberg, Acta Crystallogr. **B28**, 3542 (1972).

121) Y. A. Sataty, A. Ron, and F. H. Herbstein, J. Chem. Phys. **62**, 1094 (1975).

122) http://www.webelements.com/compounds/carbon/carbon_tetrafluoride.html

123) S. Nosé and M. L. Klein, J. Chem. Phys. 78, 6928 (1983).

124) K. Singer, A. Taylor, and J. V. L. Singer, Mol. Phys. 33, 1757 (1977).

125) S. Murad, in Computer Modeling of Matter, edited by P. Lykos (ACS Symp., Ser. 86, p.62).

126) Nakahata, N. Matsui, Y. Akahama, M. Kobayashi and H. Kawamura. Phase transition of solid CF4 under high Pressure. Rev High Pressure Sci. Technol. Vol.7 1998.

127) S. Nose and M. L. Klein. A study of Solid and Liquid Carbon TetraFluoride ysng constant pressure molecular dynamics technique.

128) D. Van der Putten. J. Phys C. Solid State Phys,20 . 1987

129) L.Verlet. Computer' Experiments' on Classical fluids. I Thermodynamics Properties. 1967

130) http://www.staff.ncl.ac.uk/michael.north/booknet/chapter6/c6q13d.htm

131) http://www.pci.tu-bs.de/aggericke/PC4e_osv/Molecular-Symmetry.pdf

132) http://www.b-u.ac.in/sde_book/msc_phychem.pdf

133) http://www.helsinki.fi/~rummukai/simu/fss.pdf