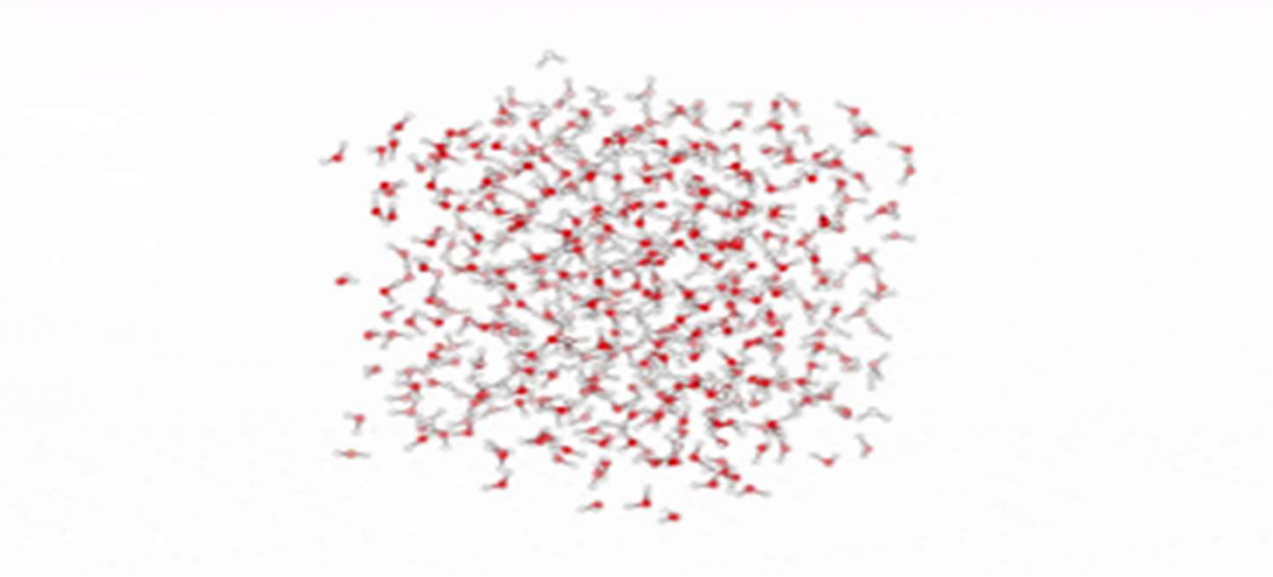
Module - 05



Molecular Dynamics

Syllabus

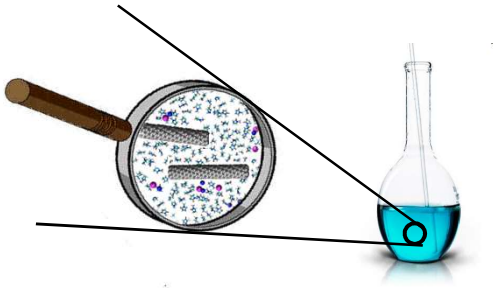
Name of Module	Syllabus
Module – 05 *Molecular Dynamics*	Introduction to ensembles, force fields, integration of Newton's laws of motion, force calculation, energy minimization, periodic boundary conditions, choice of input configuration, velocities, and time-step, applications, and calculation of simple thermodynamic variables.

- ***Molecular Dynamics*** is a computer simulation of physical movements of atoms and molecules. The atoms and molecules are allowed to interact for a fixed period of time giving a view of the dynamic "evolution" of the system.

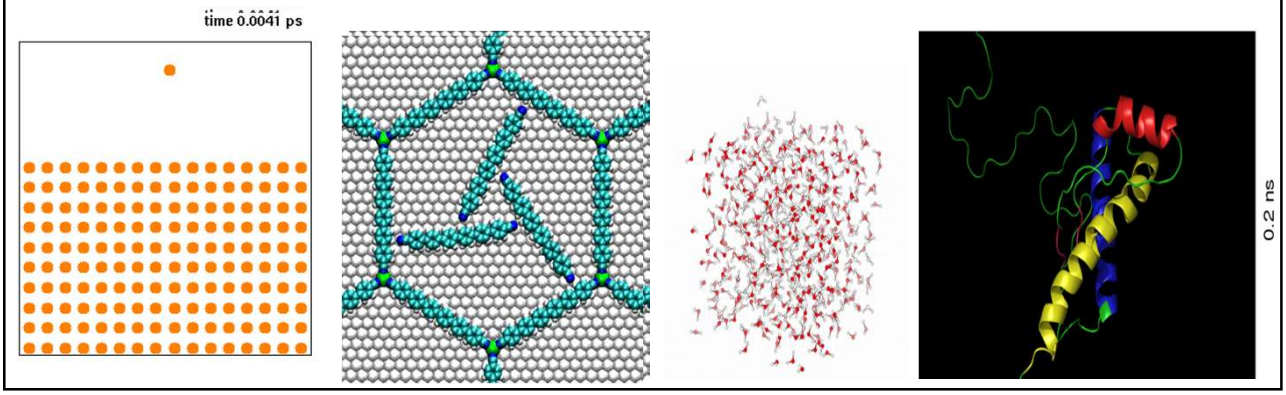
- The trajectories of molecules and atoms are determined by numerically solving the classical equations of motion (Newton's equations) for a system of interacting particles, where forces between the particles and **potential energy** are defined by **molecular mechanics force fields.**

- This was first accomplished in **1957 and 1959** for a system of hard spheres by Adler. Now is applied mostly in **materials science and modeling of biomolecules.**

MD is a computational microscope. Understanding how things work on a detailed *molecular* level ranging from electronic structures to long-time phase behavior of molecules



MD System

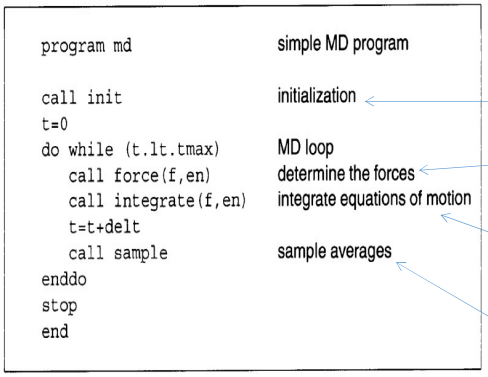


Molecular Dynamics Simulation

A molecular dynamics simulation can be constructed in following manner-

Specify the condition of run / parameters – Initial Temperature, Number of particles in the system, density and time step.

Molecular Dynamics Programme-



1) We read the parameters specifying the conditions of the run : N, T, V, time step,… data input
2) We initialize the system initial (initial position - r_i and velocity - v_i)
3) We compute the forces (f) from the interaction potential (force field, en) forces
4) We integrate the Newton's equations of motion (core of the simulation) and update the trajectories. integrate
5) We compute various quantities by averaging

To start simulation we should assign the initial velocities and position to all the particles in the system. Position of particle should be chosen compatible with the structure that we are aiming to simulate.

The position of particle should not be in overlapping position and this can be avoided by placing the particles on cubic lattice.

In this algorithm we have chosen to start the run from cubic lattice. The values of densities and initial temperature are chosen in such a way that the system remain unstable.

First put the each particle on crystal lattice and then we attribute to each velocity component of every particle a value that is derived from uniform distribution in the interval [-0.5, 0.5] (Maxwellian Distribution).

Subsequently shift all the velocities so that total momentum is zero. The resulting velocities are so adjusted to have desired value of mean kinetic energy.

$$\langle v^2_\alpha \rangle = k_B T / m$$

Where v_α is the α component of the velocity of a particle.

Instantenous temperature at time t, T(t)

$$kB\ T(t) = \sum_{i=1}^{N} \frac{m v^2_{\alpha,\,i}(t)}{Nf}$$

It is clear from the above equation that we can adjust the instantaneous temperature T(t) to match the desired temperature (T) by altering the velocities with factor $[T/T(t)]^{1/2}$.

Ensemble of system

The entire universe is governed by the laws of thermodynamics through the transfer of energy between matter. This process is too complex to consider directly, instead we consider a part of the universe (i.e., the system) separately from the rest of its surroundings. The system houses the process or thermodynamic state under consideration, whereas the surroundings contain everything else.

The system can be described using an ensemble also known as a statistical ensemble, which is an idealization consisting a large number of copies of a system, considered all at once, each of which represents a possible state that the real system might be in.

We can consider different systems with different degrees of separation from their surroundings, from completely isolated (e.g., microcanonical ensemble) to completely open to its surroundings (e.g. grand canonical ensemble). However, for the purposes of molecular simulations only three main ensembles will be considered. Namely, the **microcanonical, canonical and isothermal-isobaric ensembles.**

A collection of a large number of systems which are identical with the system under consideration in a number of aspects (such as Volume, number of particles etc.) is called an ensemble of system.

Ensembles are of three types depending upon the thermodynamic variables kept constant.

1) Microcanonical ensemble
2) Canonical ensemble
3) Grand canonical ensemble

Microcanonical ensemble:

An ensemble of systems in which each system (member) has the same value of number of particle (N), volume (V) and energy (E) is called a microcanonical ensemble. In classical thermodynamic sense each system in microcanonical ensemble is like an isolated system. Hence we can imagine a microcanonical ensemble in which each system (member) is separated by rigid, impermeable, adiabatic walls so that **neither energy nor material particles can flow from one system to the other.**

Canonical ensemble:

If all the members (systems) of an ensemble have the same value of N, V and T, then it is called a canonical ensemble. It can be set up by imagining rigid but conducting walls separating the different systems through which energy can pass but not the particles. Due to the conducting walls, each member of the ensemble has the same temperature, but may not have the same energy. Thus each system in **a canonical ensemble is like a closed system in the thermodynamic sense.**

Grand canonical ensemble:

In a grand canonical ensemble for each system, V, T and μ (chemical potential) for each component is same. In thermodynamic sense **each member of a grand canonical ensemble is an open system** such that matter can flow between the systems and the composition of each member may fluctuate.

Force Calculation

To calculate the average values of the properties of the system, it is necessary to simulate the dynamic behaviour of the system based on the model of the intra- and intermolecular interaction present.

For any arrangement of atoms in the system, the force acting on each atom due to interaction with other atom can be calculated by differentiating potential energy function.

From the force acting on each atom it is possible to determine its acceleration via a Newton's law of motion.

Integration of all equations of motion yield trajectory that describes how the position, velocities and acceleration of particles vary with time and from which average values of properties we can determined through mathematical equation.

The problem which we face is for 'macroscopic' number of atoms or molecules (of order of 10^{23}) it is not possible to determine the initial configuration of the system and we can only integrate the equation of motion and calculate the trajectory.

Gibb's & Boltzmann recognized this problem and developed statistical mechanics in which single system evolving in time is replaced by large numbers of replications of the system that are considered simultaneously.

The ensemble average or expected value is the average value of the property over all the replications of ensembles generated by simulation.

Force Fields

Molecular mechanics is a non-quantum mechanical technique for calculating energies and some properties of molecules. Molecular Mechanics treats the nuclei and electrons in the molecule.

Molecular mechanics is based on a mathematical model of a molecule as a collection of balls (corresponding to the atoms) held together by springs (corresponding to the bonds).

Many problems which we deal in molecular modelling are too large and can not be managed by quantum mechanics.

Quantum mechanical model deals with electrons in the system and if we ignore some of the electrons then also large number of electrons needs to be considered and calculation of properties is time consuming.

Force Field Methods (also known as Molecular Mechanics) do not consider the electronic motion and calculate the energy of the system as a function a function of nuclear positions only.

In this model the energy of the molecule changes with geometry because the springs resist being stretched or bent away from some "natural" length or angle and the balls resist being pushed too closely together.

The principle behind MM is to express the energy of a molecule as a function of its resistance toward bond stretching, bond bending, and atom crowding and to use this energy equation to find the bond lengths, angles and dihedrals corresponding to the minimum-energy geometry – or more precisely to the various possible potential energy surface minima.

In other words MM uses a conceptually mechanical model of a molecule to find its minimum-energy geometry (for flexible molecules, the geometries of the various conformers).

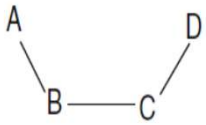
The form of the mathematical expression for the energy and the parameters in MM constitute a force field and molecular mechanics methods are sometimes called force field methods.

The potential energy of a molecule can be written

$$E = \sum_{bonds} E_{Stretch} + \sum_{angles} E_{Bend} + \sum_{dihedrals} E_{Torsion} + \sum_{Pairs} E_{Non\ bond}$$

Where $E_{stretch}$ etc. are contributions from bond stretching, angle bending, torsional motion (rotation) around single bonds and interactions between atoms or groups which are non-bonded (not directly bonded together).

The sums are over all the bonds, all the angles defined by three atoms A–B–C, all the dihedral angles defined by four atoms A–B–C–D, and all pairs of significant non-bonded interactions. The mathematical form of these terms and the parameters in them constitute a particular **force field.**



Molecular mechanics is based on simple model of interaction within the system with contributions from process like bond stretching, Angle bending and rotation about single bond (Torsional angle).

Molecular modelling force field for molecular systems depends on intra- and intermolecular forces acting within the system [Bond Stretching, Angle Bending, Bond Rotation (Torsional Angle) and Non-bonded interactions].

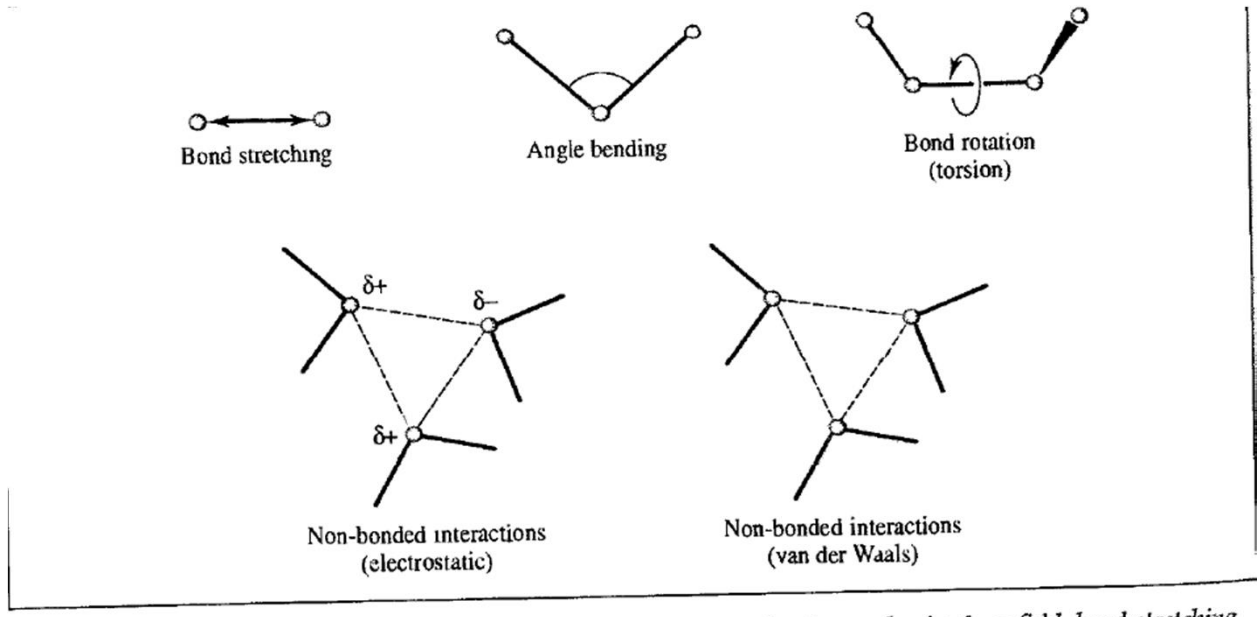


Fig 4 1 Schematic representation of the four key contributions to a molecular mechanics force field bond stretching, angle bending and torsional terms and non-bonded interactions

Integration of Newton's Law of Motion

Molecular dynamics is a method in which the motion of each individual atom or molecule is computed according to Newton's second law. It usually involves a large number of particles, from few tens to a thousand or even several millions of particles.

Macroscopic quantities are extracted from the microscopic trajectories of particles. It is a tool which can be used to understand macroscopic chemistry from an atomic point of view.

Molecular dynamics has many applications—the thermodynamic properties of gas, liquid, and solid; phase transitions; structure and dynamics of macromolecules (e.g., polymers, DNA, proteins); hydro-dynamical fluid flow; plasma and electrons, transport phenomena, etc

Verlet Algorithm

This algorithm is particularly suited for molecular dynamics. It has been widely used in many areas from simulation of liquids and solids to biological molecules.

Newton's Law of Motion is defined as

$$m_i \frac{d^2 r_i}{dt^2} = Fi$$

where r_i is the position vector of the i^{th} particle, m_i is the mass of this particle and F_i is the total force acting on this particle.

Integration of Newton's Law of motion can be obtained by applying Taylor expansion of particle around time 't'

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \frac{\Delta t^3}{3!}r + \Theta(\Delta t^4)$$

A Taylor series is a series expansion of a function about a point.

----------Equation (1)

Similarly

$$r(t - \Delta t) = r(t) - v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 - \frac{\Delta t^3}{3!}r + \Theta(\Delta t^4)$$

----------Equation (2)

Adding equation (1) & (2)

$$r(t+\Delta t) + r(t-\Delta t) = 2r(t) + \frac{f(t)}{m}\Delta t^2 + \Theta(\Delta t^4)$$

$$r(t+\Delta t) = 2r(t) + \frac{f(t)}{m}\Delta t^2 + \Theta(\Delta t^4) - r(t-\Delta t)$$

In above equation Δt^4 is an error and can be ignored

$$\boldsymbol{r(t+\Delta t) = 2r(t) + \frac{f(t)}{m}\Delta t^2 - r(t-\Delta t)}$$

Subtracting equation (2) from (1)

$$r(t+\Delta t) - r(t-\Delta t) = 2v(t)\Delta t + \Theta(\Delta t^3)$$

Above equation is accurate to order Δt^2

$$r(t+\Delta t) - r(t-\Delta t) = 2v(t)\Delta t + \Theta(\Delta t^2)$$

$$\boldsymbol{\frac{r(t+\Delta t) - r(t-\Delta t)}{2\Delta t}) - \theta(\Delta t^2) = v(t)}$$

Above equation can be used to calculate the velocity or kinetic energy using Verlet Algorithm.

Energy Minimization

In molecular modelling the area of interest is the minimum points on the energy surface. Minimum energy arrangements of atoms corresponds to the stable state of the system and any movement away from minimum gives configuration with higher energy.

There can be large number of minima on the energy surface but the minima with lowest energy is known as **global energy minimum.**

To identify those geometries of the system which corresponds to minimum energy on energy surface we use energy minimization algorithm.

Energy minimization algorithm helps in understanding how the system changes from one minimum energy structure to another minimum energy structure, how the relative position of atom changes during a reaction, changes in structure as the molecule changes its conformation, the highest point on the pathway between two minima which represents transaction state (Saddle point).

Both maxima (saddle point) and minima on the energy surface represents the stationary points where the first derivative of energy function is zero.

Energy Minimization: Statement of the problem

Let the function 'f' depends on one / more variables x_1, x_2, x_3……x_i . Find the value of those variables where the value of 'f' is minimum.

At minimum point $\frac{\partial f}{\partial x_i} = 0$ *and* $\frac{\partial^2 f}{\partial xi^2} > 0$

Molecular mechanics minimization are always performed in Cartesian coordinates where energy is a function of 3N variables.

For analytical functions the minimum of the function can be found by finding its derivative using calculus but the same can not be applied to molecular systems as energy varies with coordinates in a complex manner. In molecular system the minima is determined using numerical methods.

Let us understand how various minimisation algorithm operates.

Let us consider a function of two variables 'x' and 'y' such that f(x,y) = $x^2 + 2y^2$. The function is represented as contour diagram.

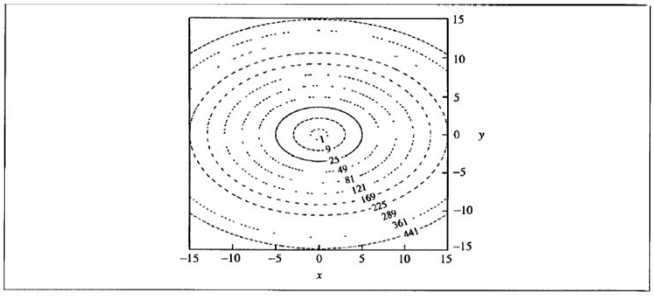


Fig. 5.2. The function $x^2 + 2y^2$.

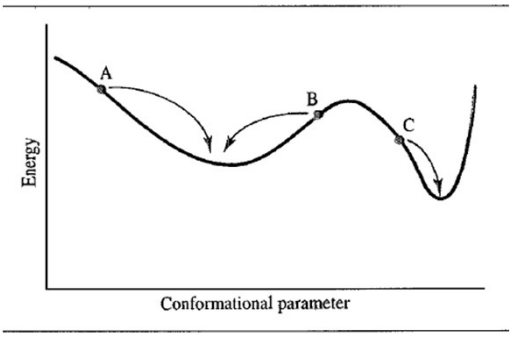
The function $f(x,y) = x^2 + 2y^2$ has one minima located at the origin.

Minimization algorithm can be of two types :
(1) Algorithm using derivative of energy with respect to the coordinates: provides information about the shape of energy surface. (2) Algorithm which do not use derivative.

No single minimization method is best for all the molecular modelling problems and hence most of the software provide choice of methods.

Most of the minimization algothrim only go down hill on energy surface so they can only locate minima near the starting point.

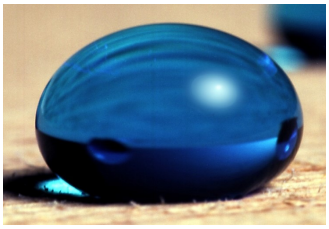
To locate more than one minima or global energy minimum we need to generate different starting point and each of which is minimised but no algothrim is capable of identifying global energy minimum from arbitrary starting point.



The input to minimum energy program consist of set of initial coordinates for the system. The initial coordinates may be obtained from experimental techniques like X-Ray Crystallography or NMR and in other cases theoretical method is applied such as conformational search algothrim.

Geometry and Boundary Conditions

- Most simulations probe the structural and thermodynamic properties of a system of a few thousand particles.
- For such small systems the choice of boundary conditions can strongly affect the measured properties.
- **For example:** In a cubic crystal of 1000 atoms 49% of the atoms are at the surface. For a crystal of 1,000,000 atoms only 6% of them are at the surface.



- What should we do at the boundaries of our simulated system?
- If nothing special is done atoms near the boundary would have less neighbours than atoms inside.
- This causes surface effects in the simulation to be much more important than they are in the real system.
- Finite systems: use open boundary conditions, i.e. no boundaries at all, just N particles in space
- Macroscopic systems: real macroscopic systems have a much larger number of particles ($\sim 10^{23}$) than can be handled in a simulation

Periodic Boundary Conditions

An initial value problem (IVP) is an ordinary differential equation together with an initial condition which specifies the value of the unknown function at a given point in the domain (set of inputs to the function).

Modelling a system require solving an initial value problem and the differential initial value is an equation which specifies how the system evolves with time given the initial conditions of the problem.

In an initial value problem all of the conditions are specified at the same value of the independent variable (and that value is at the lower boundary of the domain, thus the term "initial" value is used.

Boundary value problems are similar to initial value problems. A boundary value problem has conditions specified at the extremes ("boundaries") of the independent variable in the equation.

For example: If the independent variable is time over the domain **[0,1]**, a boundary value problem would specify values for at both and whereas an initial value problem would specify a value of and at time.

For example: Finding the temperature at all points of an iron bar with one end kept at absolute zero and the other end at the freezing point of water would be a boundary value problem.

Example of a boundary value problem (in one spatial dimension)

Let the general second order differential equation is defined as **y"(x) + y(x) = 0** which is to be solved for the unknown function y(x) with the boundary conditions y(0) =0, $y(\frac{\pi}{2}) = 2$
Without the boundary conditions, the general solution to this equation is
y(x) = A sinx + B cosx
From the boundary condition y(0) =0 one obtains
0 = A*0 + B *1
which implies that B=0. From the boundary condition $y(\frac{\pi}{2}) = 2$ one finds ; 2 = A*1

So, A=2 One sees that imposing boundary conditions allowed one to determine a unique solution, which in this case is
y(x) = 2 sinx

Computer simulation is the process of mathematical modelling performed on a computer, which is designed to predict the behaviour of, or the outcome of, a real-world or physical system. The reliability of some mathematical models can be determined by comparing their results to the real-world outcomes they aim to predict.

Macroscopic systems are extremely large and are therefore expensive to compute by molecular simulations. **Hence Periodic Boundary Conditions are used to predict the behaviour of particle in real world by using molecular simulation.**

For example: One gram of water has about 3 x 10^{22} molecules, a number too large to be calculated even on computers.

Boundary condition.

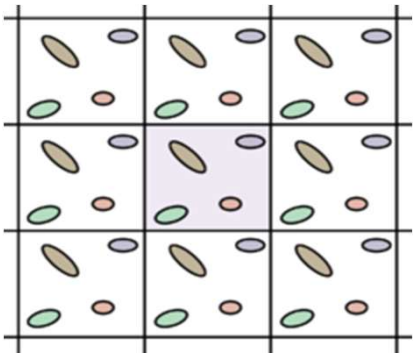
There are two major types of boundary conditions:
Isolated boundary condition **(IBC)**
Periodic boundary condition **(PBC)**

IBC is ideally suited for studying clusters and molecules while PBC is suited for studying bulk liquids and solids.

In IBC, the N-particle system is surrounded by vacuum; these particles interact among themselves but are presumed to be so far away from everything else in the universe that no interactions with the outside occur except perhaps responding to some well-defined "external forcing.

In PBC, one explicitly keeps track of the motion of N particles in the so-called supercell but the supercell is surrounded by infinitely replicated, periodic images of itself. Therefore a particle may interact not only with particles in the same supercell but also with particles in adjacent image supercells

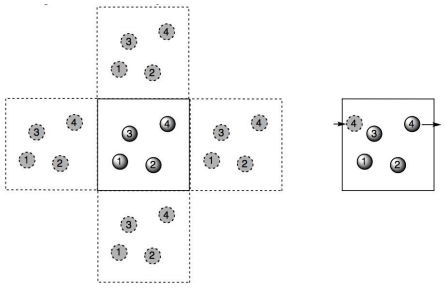
Periodic boundary conditions (PBCs) are a set of boundary conditions which are often chosen for approximating a large (infinite) system by using a small part called a *unit cell*. PBCs are often used in computer simulations and mathematical modelling.



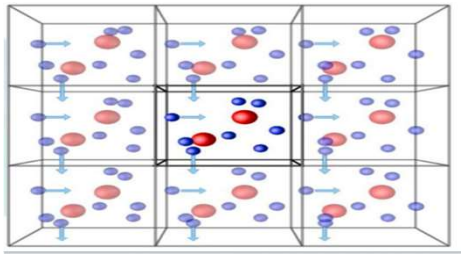
During the simulation particles are free to move in the central (original) cell; therefore, their periodic images of the adjacent cells move in an identical way. This means any particle that crosses one boundary of the cell will reappear on the opposite side.

Periodic boundary conditions enable us to mimic / imitate an infinite system by treating a relatively small part of a system to achieve a reasonable representation of the infinite system. The particles of this small subsystem are controlled by a set of boundary conditions called a unit cell (e.g., a three-dimensional box).

During the simulation, particles are free to move in the central (original) cell; therefore, their periodic images of the adjacent cells move in an identical way. This means when any particle that crosses one boundary through the one side of the unit cell, will reappear on the opposite side with the same velocity.



The large systems approximated by PBCs consist of an infinite number of unit cells. In computer simulations, one of these is the original simulation box and others are copies called *images*.



During the simulation only the properties of the original simulation box need to be recorded and propagated. Each individual particle in the simulation interacts with the closest image of the remaining particles in the system.

Length and Times of MD Simulation

Typical experiment sample contains ~ 10^{23} atoms!

Typical MD simulations (on a single CPU)

a) Can include 1000 – 10,000 atoms (~20-40 Å in size)!

b) run length ~ 1 –10 *ns* (10^{-9} seconds)!

Consequence of system size:

Larger fraction of atoms are on the surface, $\frac{Ns}{N} = \frac{4\pi r^2 dr\ \rho\ /\ m}{\frac{4}{3}\pi r^3 \rho\ /\ m} = 3\frac{dr}{r}$

$$\frac{Ns}{N}(Expt.) \sim 3\frac{(3A)}{10^8 A} \sim 10^{-7}$$

$$\frac{Ns}{N}(MD) \sim 3\frac{(3A)}{20A} \sim 0.45$$

Surface atoms have different environment than bulk atoms!

Calculation of Simple Thermodynamic Properties

A wide variety of thermodynamic properties can be calculated from computer simulation.

Comparison of experimental & calculated values for such properties is important way by which the accuracy of simulation and underlying energy model can be qualified/examined.

Simulation methods also enable predications to be made of thermodynamic property of the system for which no experimental data is available.

Simulation can provide the structural information about the conformational changes in molecules and the distribution of molecules in a system.

Thermodynamic properties which can be calculated are Internal Energy, Heat Capacity, Pressure, Temperature, Radial Distribution Functions.

Internal Energy

The internal energy can easily be obtained from simulation as the ensembles average of the energies of the states that are measured during the course of the simulations.

$$\boldsymbol{E = \frac{1}{M}\sum_{i=1}^{M} E_i}$$

Heat Capacity

At phase transaction the heat capacity shows a characteristic dependence upon the temperature (First order phase transaction is characterised by infinite heat capacity at transaction and in second order phase transaction the heat capacity changes discontinuously).

$$C_v = \left(\frac{\partial U}{\partial T}\right)_V$$

Heat Capacity can be calculated by performing series of simulations at different temperature and differentiating the energy with respect to temperature.

Heat Capacity can be calculated from single simulation by considering the instantaneous change in the energy as follows -

Heat Capacity can be calculated from single simulation by considering the instantaneous change in the energy as follows -

$$C_v = \frac{\{< E^2 > - < E > 2\}}{k_B T^2}$$

$$< (E - < E >)^2 >= \{< E^2 > - < E > 2$$

Given

$$C_v = \frac{\{< E^2 > - < E > 2\}}{k_B T^2}$$

Pressure

In computer simulation the pressure is calculated via a virial theorem of Claucius.

The virial is defined as the expectation value of the sum of the products of the coordinates of the particle and the force acting on them. It is defined as

$$\boldsymbol{W} = \sum \boldsymbol{x_i p_{xi}}$$

Where,
x_i = Coordinate of particle
P_{xi} = Momentum along the coordinate

By definitation of virial theorem, Virial is equal to $-3Nk_BT$.

In ideal gas the only forces are due to the interaction between the gas & and the container and the virial is equal to -3PV. This can also be obtained by $PV = Nk_BT$.

The total virial for real system is equal to sum of ideal gas part and the a contribution due to the interaction between the particles.

$$W = -3PV + \sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} \frac{dv(r_{ij})}{dr_{ij}} = -3NkBT$$

$$P = \frac{1}{V}\left[NkBT - \frac{1}{3}\sum_{i=1}^{N} \sum_{j=i+1}^{N} r_{ij} \frac{dv(r_{ij})}{dr_{ij}} = \right]$$

The forces are calculated as part of a molecular simulation and little effort is required to calculate virial and thus the pressure.

Pressure

In a canonical ensemble the total temperature is constant but in microcanonical ensemble the temperature fluctuate. The temperature is directly related to the kinetic energy of the system as follows

$$k = \left[\sum_{i=1}^{N} \frac{Ip_i I^2}{2mi} = \frac{k_B T}{2}(3N - Nc)\right]$$

Where,
p_i = Total momentum of particle 'i'
m_i = mass of particle 'i'

Radial Distribution Function

Radial distribution functions are the useful ways to describe the structure of a system particularly liquids.

Radial distribution functions can be measured experimentally using X-ray diffraction. The regular arrangement of the atoms in a crystal gives characteristics X-ray diffraction pattern with bright sharp spots.

For the liquids, diffraction patterns has region of high and low intensity but no sharp spots. The X-ray diffraction pattern can also be analysed to calculate an experimental distribution function, which can be compared with the data obtained from simulation.

Thermodynamic properties can be calculated using Radial distribution functions if pairwise additivity of forces is assumed. These properties are usually given as an ideal gas part plus a real gas part.

Limitation of MD

• Quality of the force field

• Size and Time – atomistic simulations can be performed only for systems of a few tenths of angstroms on the length scale and for a few nanoseconds on the time scale

• Conformational freedom of the molecule – the number of possible conformations a molecule can adopt is enormous, growing exponentially with the number or rotatable bonds.

• Only applicable to systems that have been parameterized

• Connectivity of atoms cannot change during dynamics – no chemical reactions.