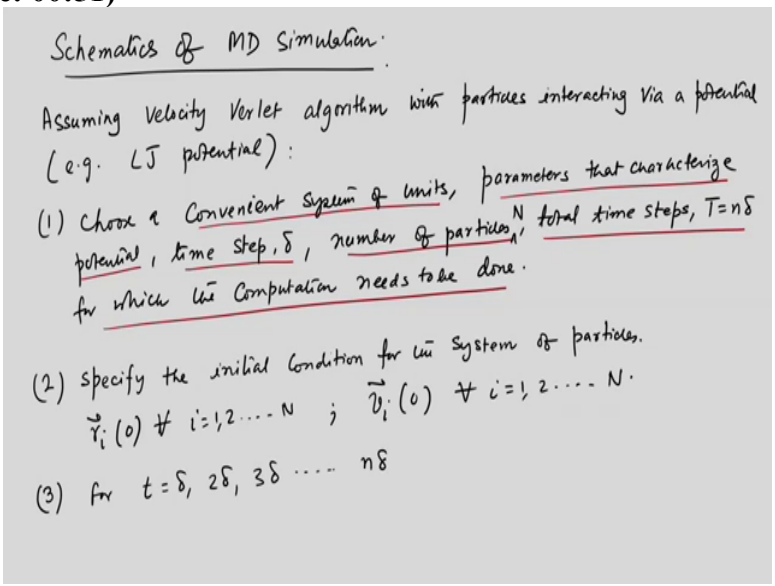
Numerical Methods and Simulation
Techniques for Scientists And Engineers
Saurabh Basu
Department of Physics
Indian Institute of Technology- Guwahati

Lecture 24
Applications of Molecular dynamics

So, we have looked at the velocity Verlet algorithm and here let us lay down the schematics of MD simulation typical MD simulation, as we said that there is a bare basic we are talking about conservative systems and only position dependent potentials and some of them we have already given examples for.

(Refer Slide Time: 00:51)



So, in order to write down the schematics there are several steps. So, let me outline the steps. So, this is assuming velocity Verlet algorithm it is worth mentioning that there are other algorithms such as leapfrog algorithm or the Demons algorithm, but the most commonly used one is the Verlet algorithm, which has this simultaneous recognition relations for the position and the velocity.

And say with some potential with particles interacting via a potential and this potential could be the Leonard Jones potentials, which we have already discussed. The different steps are one is that choose convenient system of unit parameters that characterize potential very importantly the time step delta which we have discussed at length in the previous discussion and number of particles total time for which one was the trajectory.

So, total time steps which could be written as the call to say n delta where n delta is one time step for which the computation needs to be done. So, let me just go through again on this each

one of these we have seen that convenient system of units is necessary, because most of the time these all these in the natural units or rather the regular units, these quantities are either too small or too large, which is inconvenient for the computers to handle.

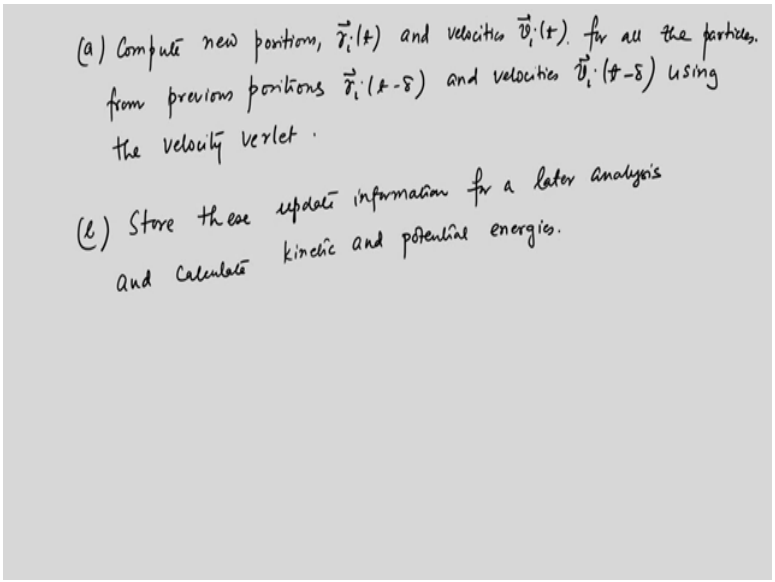
So, in that scenario one actually talks about more like some system of units in which these numerical accuracies do not come into picture the parameters that characterize the potential we have discussed this in two classes before when we had shown that all these noble gases such as you know Argon, krypton and xenon etc. They interact via these Leonard Jones potential. So, the 4 epsilon which is a typical energy scale and the sigma which is a length scale of this and so on those, so these need to be specified in the units that one chooses.

Time step is again very important thing we have seen that the numerical stability and possibilities of getting physical solutions largely depend upon this choice of the time step Delta. In fact, if one chooses a large time step, then it may not be possible to get physically meaningful solution. And of course, the number of particles so, this is a is the number of particles total number of particles, we could choose avegadro of particles or we could choose you know, something like 10 to the 20 number of particles or anything that is sufficient for a given problem.

And the total time over which the computation needs to be done. So, all these things need to be specified at the onset of the computation. Then one needs to specify the initial condition for the system of particles. So, what I mean by that is that one needs to specify ri for i equal to 1 to n so, ri at 0, so, let us just write it. So, ri at 0 that is time equal to 0 and also vi at time t equals 0 for all these particles, 1 2 N.

Now, the simulation is carried on for time steps Delta 2 Delta 3 delta and so, on all the way up to N Delta remember there are two N's which we have not written. So, this is the particle number N which is denoted by a capital N and there is a time step which is denoted by a small m or the total number of timestamps. With this one, let us go to the next page or next slide rather.

(Refer Slide Time: 06:55)

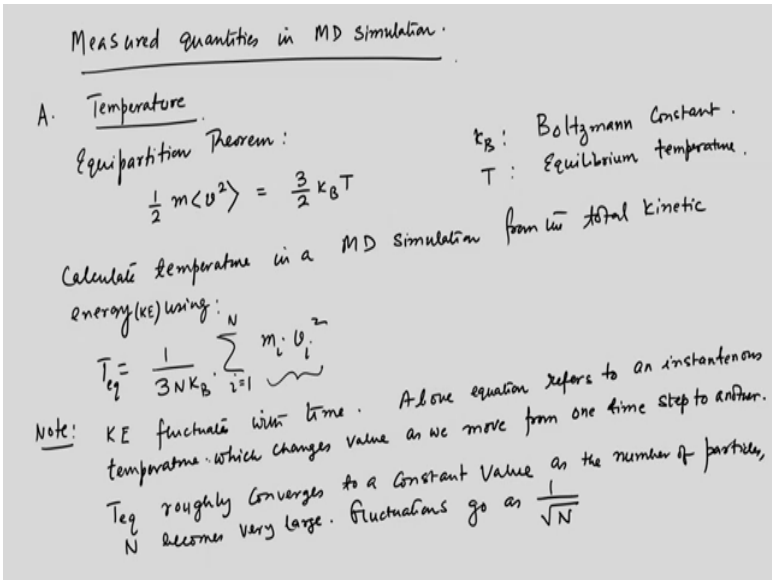


So, what one wants to do with this is that compute at these times the sub sequin times delta to delta etc come to new positions ri t and velocities vi t for all the particles from previous positions ri t - delta and velocities vi t - delta. So, from these previous the knowledge of these positions and velocities at the previous positions, one can calculate the positions and the velocities at the subsequent time steps.

So, this can be done using the velocity varlet the equations that have been told earlier and second is that store these updated information for a later analysis and of course, calculate kinetic and potential energies and remember that this is a micro canonical ensemble in which we actually fix in an E and most of the time we also fix V constant volume, but to fix the constant volume will have to incorporate one more idea which is called as a periodic boundary condition which may lead to violation of the Isotropy of space and hence entropy.

I am not going into details of that, but suppose, we are talking about micro canonical ensemble which is in characterized by NV NE were E being the number of particles we being the volume and E being the energy of the system which all of which remain constant. So, now, having done this MD simulation of what are the ideas or are the what are the quantities to compute for a given system, we would have to get some physical quantities and how these knowledge of these ri's and vi's, they are related to these physical quantities that are of interest.

(Refer Slide Time: 10:10)



So, let us talk about measured quantities in MD simulation. So, what are these measured quantities a number of them let us begin with temperature of the system. So we are not calculating or other we are not having temperature into consideration, but we can of course, calculate temperature by using the quantities that we obtain from the MD simulation. So, a this for a system in equilibrium, at some given temperature T the equal partition states that that the average value of kinetic energy for one particle so this is equal to in three dimension it is a 3 by 2 KT.

So, the statement is that at thermal equilibrium at a temperature T the contribution to each harmonic degree of freedom in a 3 dimensions is 3 by 2 KT. So, which means that if you have one more harmonic degree of freedom as in a harmonic oscillator, we have a half kx square and that would also contribute a 3 by 2 KT and will get added for the average energy and which will give us 3 KT for just a gas of particles not oscillators, it is just half MV square equal to we are talking about free particles, it is 3 by 2 KT Okay.

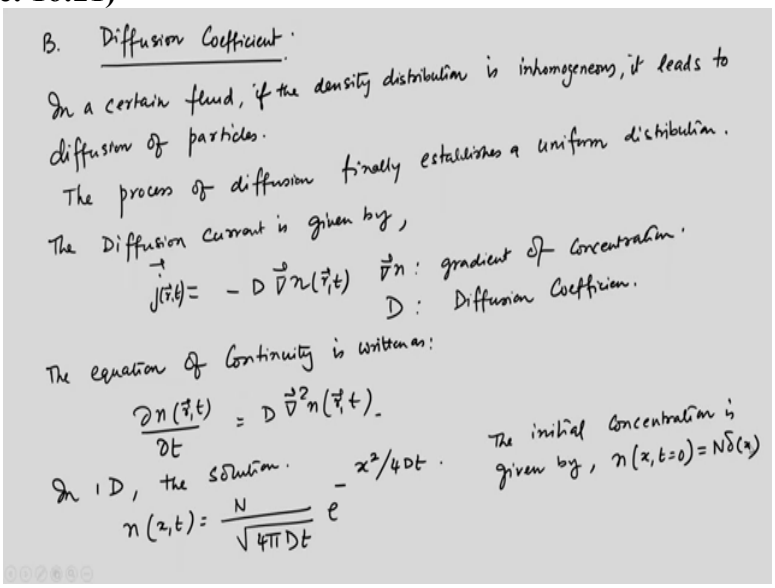
So, calculate temperature in an MD simulation from the total kinetic energy and T is the equilibrium temperatures So this and for N particles, i going from one to N These are mi vi square. So, this is a so, this quantity, you see here this quantity is actually calculated from the MD simulations in one can note that the kinetic energy will write it as KE. So, this is for kinetic energy. The KE fluctuates with time does this equation above equation refers to an instantaneous temperature.

And one can go back and write this as an equilibrium temperature and so, basically; so, this instantaneous temperature is different than the equilibrium temperature and so, this changes value as we move from one time step to another okay. And since this corresponds to the

equilibrium temperature, so, whereas, the equilibrium roughly converges to a constant value as the number of partners particles N become very large because fluctuations in a statistical problem they go as one by root over n and this fluctuations actually vanish and one actually gets TQ which is roughly a constant value.

But there could be some fluctuation even remaining for you know given system with finite number of particles. So, this is the way one calculates temperature even though temperature is not an input to the problem by calculating the velocities and calculating the velocities for these individual particles and then summing it up by according to this equal partition theorem, one can get the temperature of the system and we are fairly convinced that these fluctuations are not large when the number of particles are large. In fact, they are the vanishes 1 by root over N.

(Refer Slide Time: 16:21)



And let us see another one diffusion constant or rather diffusion coefficient called as diffusion coefficient. So, what is the diffusion coefficient? So, in a certain fluid or and the fluid can actually mean gas also, if the density distribution is in homogeneous which means, say for example, this formation of a lump of particles and there the density is much more than any other places in which case, it leads to diffusion of particles.

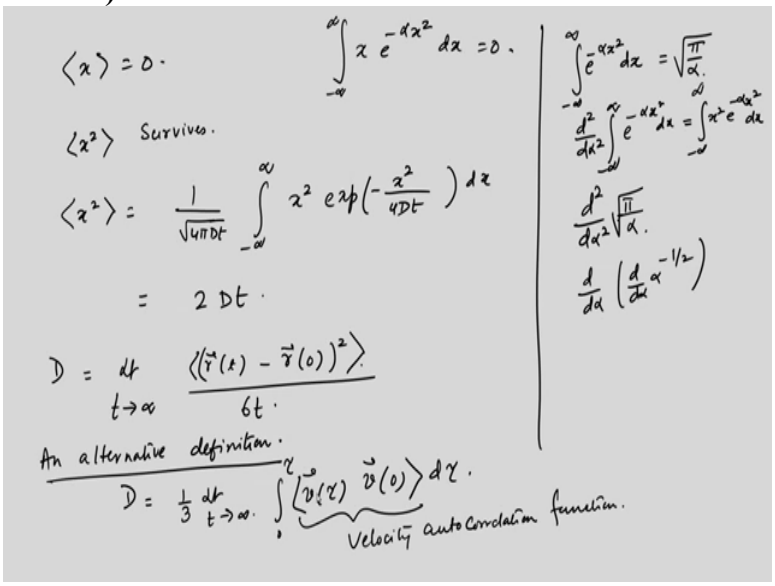
The process of diffusion finally establishes uniform distribution; the diffusion current is given by this j equal to -Dd gradient of n, where these gradient of n is the gradient of density. So, this is a function of r and t, this is a function of r and t of concentration and D is called as the diffusion coefficient. In addition to that the equation of continuity is written as it is a Del n delta t, it is equal to D and Laplacian of the density, okay.

And we will not show the solution because that becomes a problem of mathematical physics, we simply code the result in 1 D, the solution is of the form n xt it is equal to N divided by root

over 4 PI Dt and exponential - x square by 4 Dt. So, from your experience in know that this actually N as a function of x has a Gaussian form, which means that it goes exponential - x squared alpha x square. However, with T it has a slightly more complicated because there is a envelope or rather there is a coefficient which has a form 1 by root over T.

And then it is exponential minus, say, for example, alpha squared over T. So, this is a solution of this equation. And it is given that so this n comes from the initial concentration is given by nx at equal to 0 is n delta x, what it means is that the concentration is at x equal to 0 at elsewhere the concentration is 0. So, when the lump has formed, you talk about that point as rather fix your coordinate system to be 0 there, and then this has a density the total density being n. So, this is how n is defined here. And this has a form with x as a Gaussian form.

(Refer Slide Time: 21:17)



Because it has a Gaussian form, we know that the average value of x has to be 0, because for a Gaussian is a symmetric thing. So, when you say this x and exponential minus alpha x squared, if you sum over all x, then from minus infinity to plus infinity, then this is equal to 0. So, that is why this is the average value of x is equal to 0. But of course, x square survives and x square is given by 1 by root over 4 PIby Dt - infinity to + infinity x squared exponential of - x squared divided by 4 dx.

And if you do the integral, integral is easy to do it is it is a Gaussian integral and so on I keeping the derivation of that, but it is fairly easy to see this is in fact, the double derivative of this thing. So, if you write it as exponential alpha x squared dx from minus infinity to plus infinity has a value which is root over PI by alpha. So this would be some d2 dx2 the alpha 2 exponential - alpha x squared dx.

So, this is that integral, and this looks like x squared exponential - alpha x squared d x maybe with some factor, which one can figure out and so, I ultimately what one needs is to take a 2 d alpha 2 of these 1 of these root over by over alpha, and so, this is like a dd alpha of alpha to the bar minus half, and then you can easily figure out and this has a value which is 2 dt. So, the important point is that in the MD simulations, we can calculate the average value of the position and the square of the average value of the position.

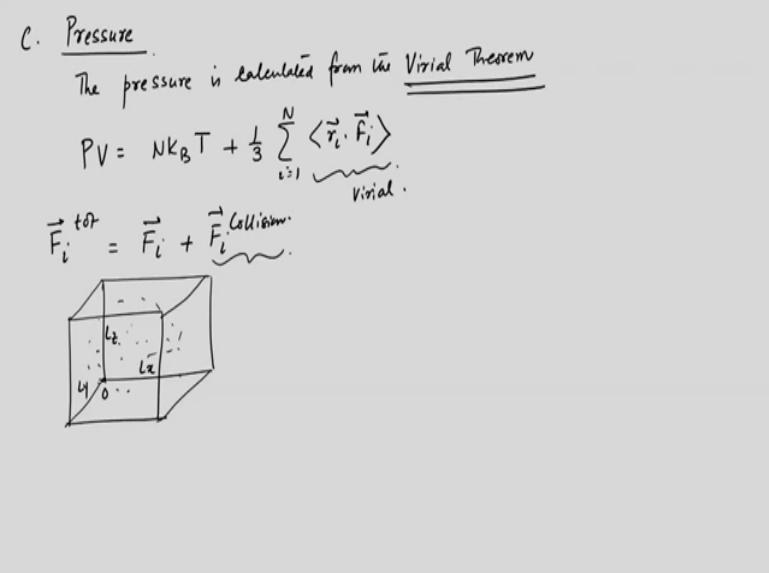
And this comes out in terms of the diffusion coefficient. And hence, because in MD simulations, you can calculate the left hand side you can calculate this right hand side or the rather the diffusion coefficient in this particular form. So, the diffusion coefficient in an MD simulation is actually is of course, a long time trajectory is understood that r t - r 0 and square have that divided by 6t.

See there are I have done it in one dimension. So there will be three such integrals which would you know, sort of give additive contribution or other multiplicative contribution and will get a 6dt. So, this fluctuation in this position of the particle with compared to the 0 is given by this and that is why the diffusion coefficient comes out as that basically, there is also an alternate expression for the diffusion coefficient in terms of the velocities.

And the velocity or auto correlation function is. So, D is equal to 1/3 limit t going to infinity 0 to some tau, tau is just a dummy variable of v at tau and V at 0. So, this and then d tau so this is called as the velocity auto correlation function. Okay, so it is a correlation between velocities at different times. So, that will tell you that how the particle velocity is correlated at a given time tau with that at a given time equal to 0.

So as it as the particle drifts away from its initial position, how the velocities are correlated, one has to take a correlation of this kind and then calculate this. Of we have not discussed in details how these correlations are done, but at least one should know that these velocities at time t etc, they can be computed from the MD simulations and of course, we have 0 has to be specified at the beginning. So these are some of the quantities that one can calculate let us see one more of them.

(Refer Slide Time: 26:22)



And we will do it in sort of bit of a hurry, because a full treatment of the MD simulations in just a week is impossible, but we will sort of see that get a flavour that what MD simulations do, we know that it is a deterministic simulation. So the pressure is calculated from the Virial theorem. This is an important theorem in statistical physics, one may want to look up I am simply giving the result.

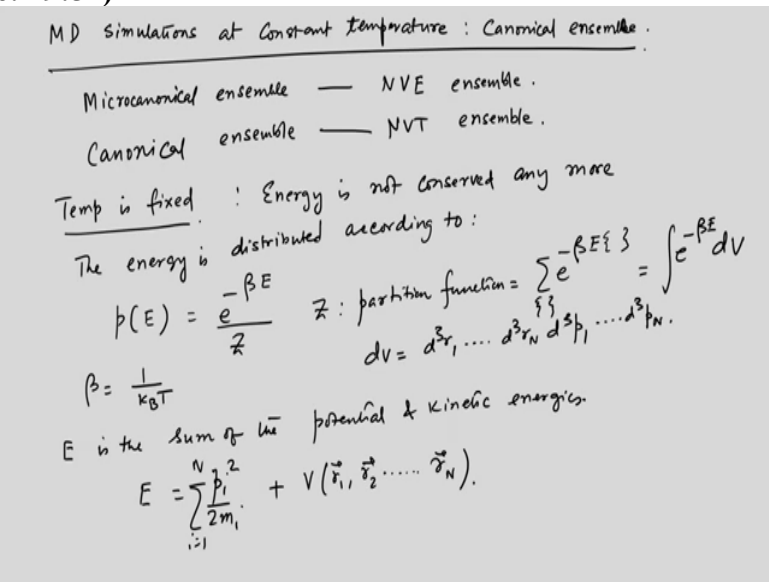
So it is a PV equal to NKT + 1/3, sum over i from 1 to N. And so this is called as the virial this last quantities called as a virial. So this is the virial ok, so the expression for pressure is like this, where this Fi are can be computed from this total force on the ith particle, it is equal to this Fi which is written here, which is due to the interaction which comes arise out of some potential sale, Leonard Jones potential and plus Fi coalition.

Let us write it as just okay, that is right full coalition. And so, this coalition can be calculated by taking a box like this Okay, fixing the origin here and this is Lx This is Ly and this is Lz . So, the coalition this force will originate from the all these the gas is contained in this box, okay. And these gas molecules will collide with the walls of the container, they will reverse the velocities and come back and this corresponds to pressure that is being given by virtue of these Coalition's.

And so, these are one can calculate and you see that these are the things are ri and fi are the things that can be calculated from the MD simulations and using this formula one can calculate the pressure exerted by shear gas a classical ideal gas or for that matter anything in classical ideal gas Of course, they're not interacting particles were these they are no forces being act acting on them and one has only collision the forces which contribute to the or rather the

pressure is only contributed by the molecules colliding with the walls of the container the container that is shown here okay.

(Refer Slide Time: 29:54)



And the last thing that we want to say here is the canonical ensemble, so, MD simulations alright, constant temperature canonical ensemble okay. So, the micro canonical ensemble is called as the NVE ensemble, this is the language of the MD simulations. And of course, in real experiments, you understand that you can do not have a control on the energy whether that can be kept constant instead, what one can do is that one can keep the temperature constant and this is called as a canonical ensemble.

So, a system can be made in contact with a heat bath, that is the system can actually exchange energy with the heat bath and finally come to equilibrium with the that of the bath the temperature of the bath and this is called as the NVT ensemble or this is called canonical ensemble okay, so, these are so, temperature is fixed here and the energy is not conserved anymore. So, that energy is actually distributed according to pE equals to exponential - beta e divided by Z whereas Z is called as the partition function and defined as sum over exponential – beta E.

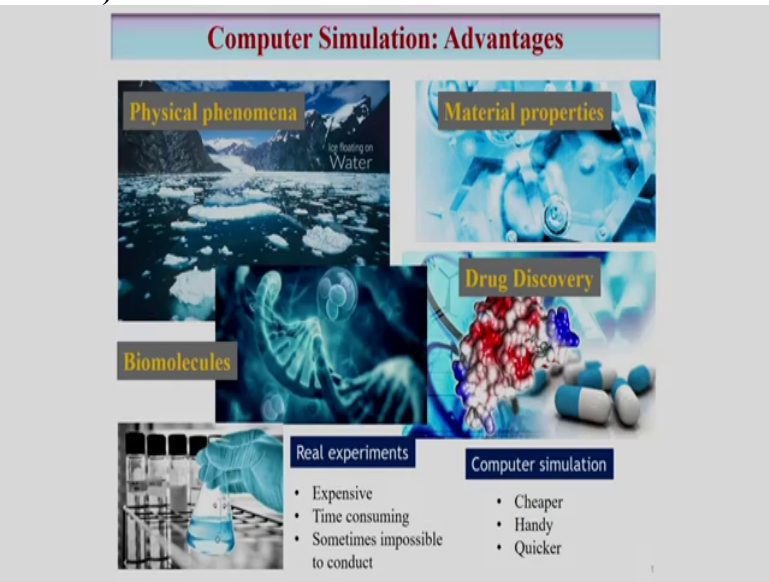
Now, this beta E is actually a functional of certain or other function of certain degrees of freedom. So, we have to sum over all these degrees of freedom or one can write it little more sort of cleanly using an integration sign for a continuum system where dv is actually a phase space Integral it is a 6 dimensional, 6 and a dimensional integral. So, this is d3 rN and d3p1 d3 pn okay. So, this is how it is distributed. So, the energy is distributed according to this at a given temperature rt where beta is called as the inverse temperature and is defined as this okay.

KT is the Boltzmann constant. So, you see that this is dimension less it makes it dimension less and so on. So, the E is the sum of potential and kinetic energy okay. So, E equal to pi squared over 2 mi sum over i to N plus we have a potential which purely depends on the position variables and most of the time we actually consider that they actually do not depend upon the vectors of the positions as vectors, but the magnitude of them are the distance between two particles pairwise distance between two particles and so on.

So, again following this ensemble, just like what we have done in the previous case, we can calculate physical quantities and so on we can calculate pressure temperature or this diffusion coefficient or some other things. And these are the motivation about using molecular dynamics which for us a fluid most importantly, we talking about gases or fluids. And one can do the simulation use the velocity were like scheme and can get the position.

And the velocity is that all subsequent you know times starting with some initial condition for a given interaction potential, which could be you know, most of the time they are pairwise interaction potential and one can calculate this physical quantities.

(Refer Slide Time: 35:34)



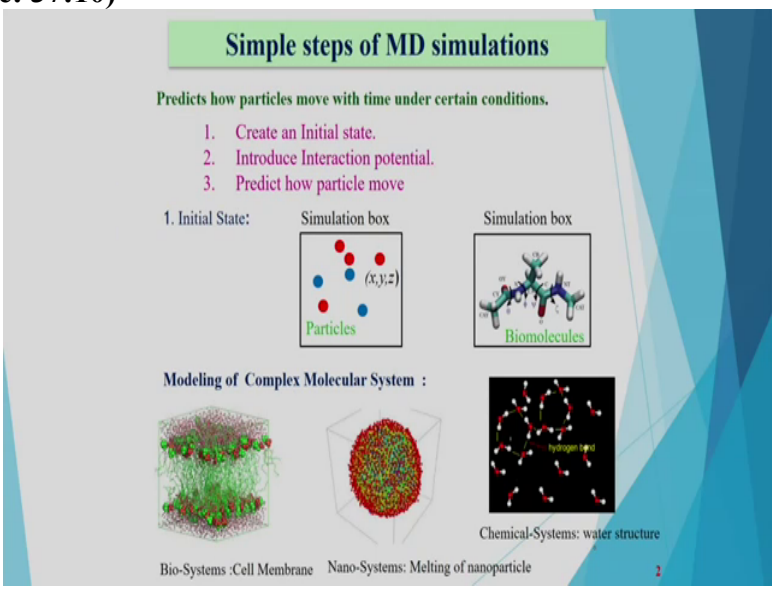
So, let us recapitulate on the salient aspects of the simulation and in particular, the MD simulation and let us see some applications of MD simulations. Some of those applications or other this movie that I am going to show is prepared by Dr. Susmitha Gosh who is a postdoctoral fellow at IMSC Chennai. So, we have seen that computer simulations are very essential and they have a lot of advantages in understanding things.

Such as physical phenomena that is, Ice formation on water or in our understanding of bio-molecules, their movements, material properties, such as, you know, new materials and how they can be transformed into better substances in terms of applications and utilities such as

maybe conduction properties, etc. And then drug discovery, these simulations have played a very important role in discovering new drugs.

And not only their discovery, but how they are delivered to the target sites. We have emphasized this earlier, and we are doing it again, that the real experiments are in principle expensive, and they are time consuming and sometimes they are too difficult to conduct. And, of course, apart from the fact that they could have hazards and safety issues, whereas the computer simulations are cheaper, they are handy and they are quicker.

(Refer Slide Time: 37:10)



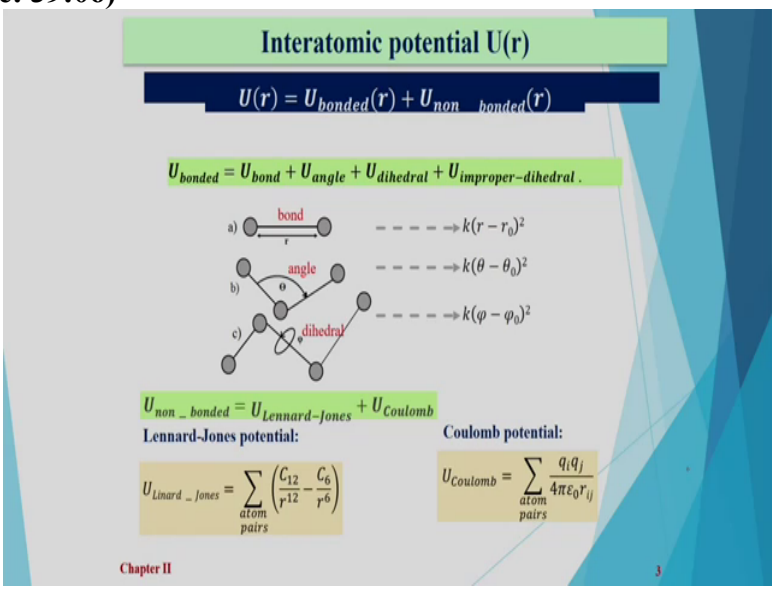
So, in particular, let us focus on the MD simulations that we have been doing and these MD simulations predict how particles move with the passage of time under certain conditions and under the influence of certain kind of potential or the forces that are derived from them. So, what we do is that we create an initial state then we introduce an interaction potential or there could be an external potential as well.

And then we predict how the particles move according to certain algorithm and we have seen that we have used this Varlet algorithm. So, let us say that the initial state they are a variety of situations they are shown as this the simulation box in the first one where there are red particles and blue particles and the coordinates are represented by x, y and z. So, they just particles we are not specifying what kind of particles they are, but they are inside a box.

Now, this could give rise to additional complications such as periodic boundary conditions etc, which we have not discussed. However, there are you know extensions of the algorithm that we have done to incorporate the periodic boundary conditions. Let us talk about the bio-molecules in a similar simulation box then there could be you know, bio-systems such as cell membranes, etc.

Then nano systems, where we could study the melting of the nano-particles or other disintegration of the nano-particles, then there are water, the structure of water is all known to us, it is H2O. So, these are water structures and we can simulate them using MD simulations and see that how they molecules actually move around or the crystal structure that moves around and so on.

(Refer Slide Time: 39:06)

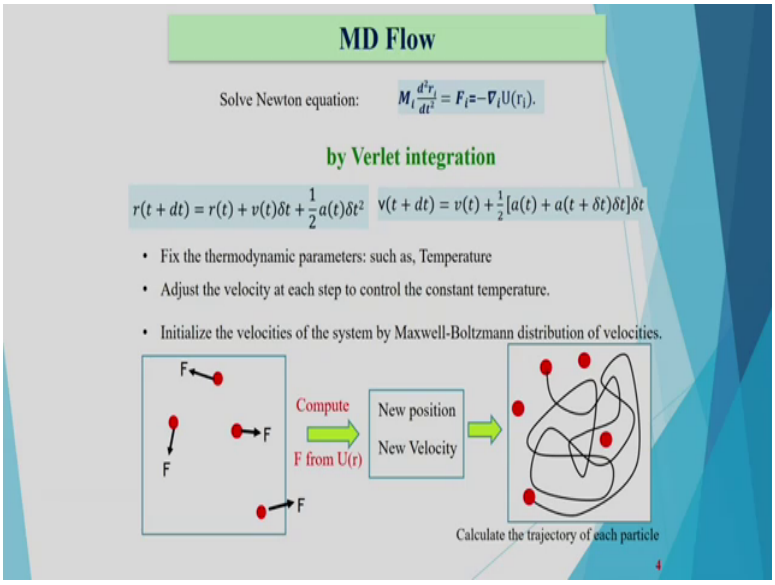


So, we remind ourselves that we have an inter atomic potential which we call as U r, this could be a bonded part and there is a non bonded part and among the bonded part we have bond, which is the equilibrium distance between the two atoms or molecules and which we consider as harmonic it may not be, but let us consider is harmonic to begin with, so, they go as kr - r 0 pole squared, where r 0 is the equilibrium distance.

And they can also go a harmonic function of this angle theta, theta - theta 0 whole square, or there could be a dihedral angle, which is shown in the figure. And they could go as k phi - phi 0 squared. And then the non bonded term is something that we have seen which could be Leonard Jones plus Coulomb. Or more plus Coulomb or any other potential and let us particularly talk about this Leonard Jones potential spelling mistake, Leonard Jones is written as it is written there.

So it is a 6-12 potential which you have been exposed to and then there are Coulomb terms which are looking like this Qi Qj and 4 by epsilon 0 rij, where the depend upon the mutual distances between the ith and the ith molecule.

(Refer Slide Time: 40:41)



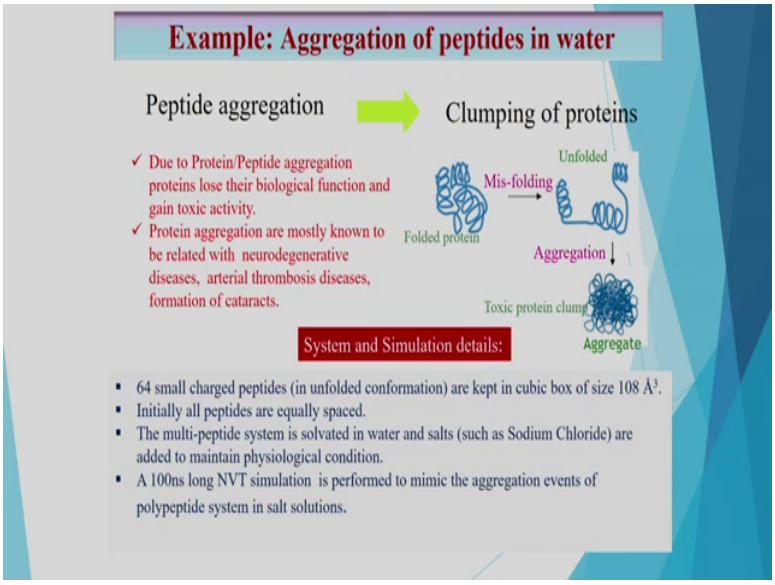
And then in order to study the flow of these particles, we have to solve Newton's equation, which is simple, which is mi d2 ri dt to equal to fi, where fi is derived from these potentials by the negative gradient of these potentials that we have just seen. And then this equation will be solved using the Verlet algorithm, which we have learned earlier that is, it is a simultaneous algorithm for the position and the velocity of the particle and which are you know, written as this we have derived these things.

And now, there are more complicated algorithm as you have seen Demons algorithm etc. But they are a little more complicated to implement, whether this is the basic algorithm that we have seen, then we have to fix the thermodynamic parameters such as temperature or maybe pressure, and then we have to calculate the velocities at each step to control the constant temperature initialize the velocities of the system by Maxwell Boltzmann distribution of velocities, this we have not particularly discussed, but you take it from any distribution, this is a Gaussian distribution.

So, the say the velocity is the initial velocities are according to some distribution taken from a distribution and that distribution is called as the maximal Boltzmann distribution, which is simply a Gaussian distribution. And then of course, using this algorithm, you calculate the new position and the new velocities of these particles. So, these red particles were initially having a configuration like this and then there are, these are the new configurations and then they will go to new configurations for the new configurations and so on.

And this is how the MD flow goes on, where we understand that this is how these are you know, the flow of these particles are computed.

(Refer Slide Time: 42:43)



So, in particular, we are going to focus on aggregation of peptides in liquid for example, the liquid could be water and there are implications in several fields of physics and biophysics, physics and biology on this, because this peptide aggregations could lead to clumping or this conglomeration of proteins and if this happens, then there could be a you know, if it happens in the brain, then there could be a possibilities of stroke and other neural disorder.

So, due to this protein, which a long chain these peptides are long chain protein molecules and if they aggregate then they lose their biological functions. So, the proteins are unable to do their normal biological functions and thus the gain toxic nature or they have toxic acquired toxic properties. And these protein aggregation is known for or rather the origin of many of the neurodegenerative diseases such as arterial thrombosis, formation of cataracts and so on okay.

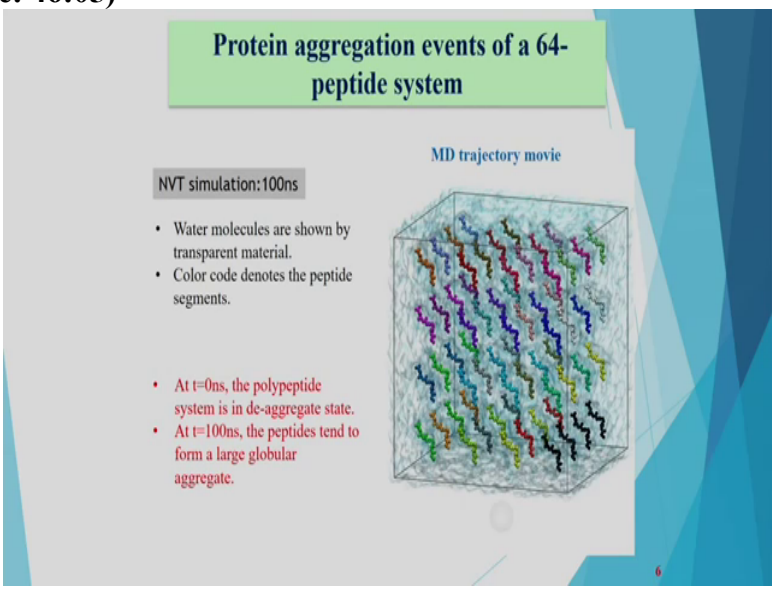
So, this is the healthy protein the folded protein and then if it is miss folded, it gets can get unfolded in this fashion. And then when they aggregate they form these bunching and when they form this bunches, they lose their activities normal biological activities and these are something that we are going to show in our simulation. So, the system and simulation details are that 64 small charged peptides in the unfolded confirmation that is this confirmation that appears here this confirmation that appears here.

So, we take them there in a cubic box of size 108 cubic Angstrom initially all the peptides are equally space this is how we start our simulations and then we run the MD trajectory. And this multi peptide says stem is now solve it in water and salt and this salts could be a various types I mean, we have taken a sodium chloride here in this particular case it is a sodium chloride and it is shown that the in the presence of sodium chloride they have actually form aggregates and if

you use a different salt or a different you know, ionic concentration of if you use a different pH then of course, the results could be different.

Like for example KCL potassium chloride actually prevents can these aggregation to a large extent. Then, it is 100 nanosecond long NVT we have discussed this NVT this is a particular ensemble that we talk about in which number of particles volume and the temperature remain constant NVT simulation is an MD simulation is performed to mimic the aggregation events of the poly peptide system in this solid solutions that we just talked about.

(Refer Slide Time: 46:05)



So, let us see this simulation alright. So, protein aggregation events have a 64 peptide system as we have just said and this is over 100 nanoseconds which we have shown here and these water molecules are shown by the transparent materials. So, this cubic box is filled with water and there is of course, sodium chloride or some salt is being added and the colour code denotes the peptide segments. So, these are the peptide segments okay.

So, at t equals 0 nanosecond that is at the beginning of the simulation, the poly peptide system is in the D aggregate state. So, now they are no aggregation so, they are not touching each other they are all maintaining a distance. And as I run the simulation, the simulation will be for about a minute.

(Video Starts: 46:59)

And you see that they will former aggregate see that they are these are these MD trajectories being computed using the Verlet algorithm. And they are sort of slowly over a period of time, as a time progresses, the time is, is about one minute, I mean the real time, so, it is over 100 nanoseconds. So, let me tell you what I have, what I am showing here is that there are slides which are taken after every 500 microseconds okay.

So, this is the typical timescale of these peptide movement. And then there are something like 200 of these slides are being joined to make a movie okay. So, you could actually have more you know, slides, but that becomes a difficulty for eyes to track. Now, you see that they are forming an aggregate and suppose this happens in the brain and these peptides are nothing other neurons.

(Video Ends: 48:06)

And then these neurons bunching together or aggregating together would result in you know, severe disease that could lead to stroke or it could lead to other neuro disorders in the brain. So this is just a movie to show that this is how MD trajectories are carried out. So once again, just saying that, so because it is 100 nanosecond movie rather, the physical time is about a minute, as I said, but it is a movie that is you know, taken over hundred nanoseconds.

So, there are the MD simulations are carried out after every 500 people seconds, that is the timestamp delta t that we have used. And these slides are made or rather these snapshots are taken after every 500 Pico seconds and 200 of them are being joined together to make 100 nanosecond movie and this is what you see that they are barely bunched an aggregated. So, this one real example of showing the how MD tragically works in real systems, and how they can be relevant for predicting various you know, biological disorder that goes on.

So, normal proteins would like to be un-bunched for their normal biological activity, but for some reason, if it happens, that they are getting bunched in a solution in a salt solution, then of course, that will lead to disorder. As I said that potassium chloride or some other salt actually have revealed that the bunching is almost not there over the period that we have carried out the simulation and so, this is one of the applications of the MD simulations that we want to present here.

So, let us let me give you some closing remarks that during the course of or the other the proceedings of this course, we have learned numerical techniques, remember that these are techniques that are to be implemented in the computer, we have not talked about real codes either in a given computer language, that is the job of the candidates are the students in order to implement these methods that we learn and get results in you know, using a computer mostly for which the analytic results are not possible.

So, we have looked at these roots of these non linear equations, they could be you know, non linear equations of varied complexity and so on. So, there are two main things that we have learned one is called is a bisection method, which is not very efficient, because you have to give

two choices in order to and they should, you know, rather bound route from both sides on the lower side and the upper side.

So, that is not something very efficient and then we have learned Newton Rapson, which requires only one guess to be given. And of course, the convergence and the error involved on all these methods we have discussed and then we have talked about interpolation algorithms. This is something very important for all experimental scientists and engineers to understand that what is the nature or behaviour of a certain experimental data.

If we only present large number of data in terms of numbers, and so, they are not very intuitive for us to understand that in certain region of the parameters space, how they behave, whether they have a singularity, which means they are divergence. And so, for that, one needs to do an interpolation scheme, these interpolations of are quite helpful in understanding the pathological behaviour of certain quantities in the parameters space.

Then we have talked about derivatives integrations, integrations are quite well described by a number of methods and they are quite accurate in case we can make these greed sizes to be small. And then we have talked about the differential equations, which form one of the most important things in any branch of science and engineering, we need to solve equations in order to understand the behaviour of particular system.

And this could be first order differential equations, second order or even higher order, they could be partial differential equations or ordinary differential equations, we have learned some techniques, majorly, it is the Runge-Kutta methods that give a solution to that, then we have talked about some initial and boundary value problems. And finally, towards the last couple of weeks of the course, or maybe a little more, we have talked about simulation techniques.

And these simulation techniques have been shown, or rather emphasize that they are extremely important for and can replace the actual experiments in the lab, which could be extremely time taking or it could be impeded by requirement of large amount of resources. Whereas doing that in the computer would get one to understand the outcome and the pitfalls of this particular system. Its associated you know, advantages and disadvantages, etc.

And in that, we have seen two kinds of simulations one is called as a deterministic simulation The other is called as the stochastic simulation that deterministic simulation is the one that we have learned last of the at the end of the course, which is molecular dynamics and the stochastic one is the Monte Carlo, both are very versatile simulation techniques. So, we have given techniques and some of the examples in real systems where it is applied.

And this makes the course complete. And we have feel that you have been able to learn things that are intended through this course and would be useful for your career. Thank you.