Basic Statistical Mechanics Prof. Biman Bagchi Department of Chemistry Indian Institute of Technology Bombay Indian Institute of Science, Bangalore

Lecture no- 59 Computer Simulation Methods in Statistical Mechanics Part 1

Welcome back to our online lecture course on statistical mechanics, chemistry and material science. Today we shall start a very important topic which is the computer simulation methods in statistical mechanics. As we have seen, we have gone through many aspects of statistical mechanics from the beginning of the postulates and then on different ensembles and applications.

And one thing that you must have realized that it is a lot of analytical work, a lot of very difficult mathematical methods that are being implemented and these mathematical methods and the requirement of mathematics becomes more and more demanding as you go to the real systems. Actually, at one point, it becomes quite difficult and so one can starts wondering that can one really push this ambitious project of statistical mechanics.

Of explaining the observed variable of crystals phase transitions, biological systems from first principles mean that from an intermolecular potential particularly by the intermolecular potential. So this was the condition state of the art in 1960's and 70's when people are still pursuing analytical work. By end of 1970s, it became quite clear that it might not be feasible to solve these analytical theories.

For example Percus-Yevick equations of the partition functions to get the required observed properties that was a very generous beginning to dawn on people. Fortunately, at the very time the computers were developed to a great extent and some very smart people realized that computer solution can come to rescue in a big way and that was developed some artful techniques and now computer simulation essentially 90% of the applications of statistical mechanics that we do is through computer simulation.

However, the methods of computer simulations are firmly based on the means of advances of statistical mechanics even now. So it seems like that statistical mechanics developed all the principles, the methods and kind of hand it over the equation applies to real systems, because

in the analytical methods solving of integral equation theories or as you see in case Mayer cluster expansion we could get the phase transition.

It is some about the cluster size distribution but even then, the calculation of virial coefficients is extremely difficult for any kind of realistic potential.

(Refer Slide Time: 03:52)

Analytical Methods: Drawbacks

- Simple model (potential) based calculations
- Inadequate results, inefficient predictive power
- Examples: cell theories of liquids, lattice models, mean-field theories, Landau theory etc.
- Approximations required for >2 body interactions

۲

So we are going to kind of just state what I have been saying that simple models based calculations can be done and in certain simple theories like cell theories, lattice models like Ising models, Landau theory Mean Field theory they are wonderful theories they have done wonderful job in explaining and providing us insight what is going on. However, they cannot be successfully applied quantitative prediction for the real systems, like water, liquid water, ethanol, proteins and then DNA.

So, you want to understand the lipid bilayer so all these systems let us stick to basically with water which is raised to a very simple thing so even in water analytical theories did not quite starts, despite tremendous effort of very smart people. Well, we can do some amount of analytical thing to trying to understand liquid to steam transition but even when it comes to freezing it is still has not been successful.

So we faced with these things as then when people realized that, however, is bottleneck are the big wall presented by the analytical theories which cannot be complex in some molecular potential. Intermolecular potential is complex in the sense is more than Lennard-Jones, it has oxygen atom through hydrogen and the end to end charges at moment but then at that time competition came to be a big risk.

(Refer Slide Time: 05:52)



And the amount of potentials which are discussed like Lennard-Jones in a big way and this is just radial potential between two atoms and water is the far cry from these things.

(Refer Slide Time: 06:05)

Intermolecular potentials for real molecular liquids like water, ethanol are far more complex

- Analytical theories cannot be pursued or solved for numbers.
- · This was frustrating
- · Computers came to rescue

So intermolecular potential for real molecular liquids like water, ethanol, dimethyl sulphoxide and many other things are far more complex. So analytical theory did not work and that is what computer simulation came to race to an end as I say 90% of the stat-mac based what now gone with using computer simulations.

(Refer Slide Time: 06:26)

)

of creativity to develop the tools and programmes to be able to mimic structure and dynamics of liquids.

Statistical Mechanics provided the rest of the input, in addition to provide the general framework.

So in fact it required no less ingenuity for less creativity to develop the computer simulation techniques. Here are basic principles and basic methodologies that the things that we need to calculate like radial distribution function, specific heat from energy fluctuations and many other quantities when indeed was the formalism and way to think and go about and equations which are given by statistical mechanics but they are implemented now through computer simulations.

So that is what he said as if the methods of statistical mechanics was developed and handed over to the next generation take it as by using computer simulations.

(Refer Slide Time: 07:14)



geesur Rahman (1927 - 1987)

So computer simulation as I have saying is very smart work and they start it essentially with a very famous people I will cite the paper in the next slide while Metropolis, Rosenbluth Teller

and Teller and they did what is called Monte Carlo simulation that is equilibrium statistical simulations. And few years later in 1950's, both Alder-Wainright did this vision in a time plane by molecular dynamic simulation.

And both treated these hard sphere kind of interactions which was simple to do and got results which was tested against the theoretical results which fortunately well known quite a bit from analytical theories because hard sphere is something analytic theories made some advances. So we then did computer simulations and tested the computer simulation against the analytical theories is solved both the purpose we could taste the analytical theories as I said cannot be done exactly.

But fairly accurately for hard spheres and we can check computer simulations and this is extremely important, that means this verifying computer simulation by statistical mechanics and verifying statistical mechanics by computer simulation. Where you start from first principles I give you intermolecular potentials and you do computer simulations molecular dynamics simulations.

And you give me a bunch of numbers which now can be tasted teory that flows directly from statistical mechanics without going through the computer simulation. So this is something which not sufficiently emphasized that these two things are extremely important and extremely useful in both the ways. Alder and Rosenbluth with the hard spheres the first continuous potential and that was the part of one can see realistic systems was done by an Indian man is Rahman who studied in Hyderabad and then he went to United States.

And made enormous contribution later teamed up with Stillinger did the calculations of water that those work were went on all whole of 1970's I think they written ten papers very nice papers which is done through these computer simulations, then that became available in 1970's in increasing capacity. I should also mention Karplus, Warshell and David were given Nobel Prize in 2014 for their applications of these computer simulation techniques or computer methods to study of biology.

And that was the celebration of the reach of our computer simulation before that Nobel Prize on computer simulation was given a lot of Nobel prizes and the commission give him to analytical theories developed by statistical mechanics.

(Refer Slide Time: 10:26)

Some pioneering works

- N. A. Metropolis, W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller. Equations of State Calculations by Fast Computing Machines J. Chem. Phys. 21(6), 1087–1092 (1953).
- B. J. Alder, T. E. Wainwright. Phase Transition for a Hard Sphere System J. Chem. Phys. 27, 1208–1209 (1957).
- A. Rahman. Correlations in the Motion of Atoms in Liquid Argon Phys. Rev. 136, A405 (1964).
- F. H. Stillinger, A. Rahman. Improved simulation of liquid water by molecular dynamics J. Chem. Phys. 60, 1545 (1974).
- ()

These are the papers. The first paper should be Metropolis, W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, the equations of state calculation of fast computing machines 1953, then they did a hard sphere system, then Barney Alder and Wainwright did the hard sphere and later hard disk and first paper of Metropolis, Rosenbluth with the Rosenbluth was done by Monte Carlo simulations well all that done is molecular dynamic simulation.

So there are two different techniques but the results agreed well with each other and the theories I have been telling. Then really fast continuous potential was done by Rahman is a similar paper and very famous paper in 1964 which is highly cited and highly respected and then Rahman teamed up with Stillinger and did a series of work on water and that time was increase the reach of computer simulation to a significant degree.

(Refer Slide Time: 11:35)



1

Now let us start on main work which is the classification of simulation and computer simulation divided into two groups, one is molecular dynamics which is propagated in times domain and the other Monte Carlo. So remember when you are doing statistical mechanics in the beginning the postulates we are telling there are two postulates of statistical mechanics, one is that equal a priori probability and another that is time average is equal to ensemble average.

Now, the basic idea of statistical mechanics is that you have the phase space of N particles the 6N dimensional space called Γ space and the system is executing a walk in this trajectory, in this multi dimensional phase space and the properties that you observe as a result of this trajectory as the motion of the system on this phase space. So if you want to do dynamics how the correlated time and also structure then we have to really let the system propagate on the 6N dimensional space by moving each atom and molecule, because the movement in phase space is essentially due to the movement of each atoms and molecules moving and rotating. So one point in phase space to another point in phase space means change of configuration all this N particle system or N molecule system or N atomic system.

Monte Carlo on the other hand is gives on the ensemble, so molecular dynamics is time average, Monte Carlo in the ensemble average. There is no time in the Monte Carlo simulation what we start with initial configuration then we follow certain rules you generate new configurations which are allowed according to Boltzmann distribution and some technique like that.

But there is a lot of creativity in creating this Monte Carlo. Monte Carlo fixed which is a huge gambling down we got so casinos in France and Swiss Monte Carlo is a method which is depends on repeated use of random number and which allows us to push the system from one point in the configuration to another configuration space and there are rules so that the systems that we generate follow the basic principles of statistical mechanics.

So the sampling is very important in Monte Carlo so to summarize again molecular dynamics is a time averaging while Monte Carlo simulation is an ensemble based things. So the molecular dynamics has time and Monte Carlo has no time.

(Refer Slide Time: 14:28)



So as I stated repeatedly in this course that is very important to get an overall qualitative picture because statistical mechanics involves lot of techniques but soon when you start doing work or research you tend to get lost into the details especially in computer simulation you start of what kind of numerical technique I used to propagate it what kind of sampling, what kind of random number generator, so there are the many details.

But it is always makes sense to have the big picture in mind so that you know from where you are starting and where to go and you can figure out if you are making mistakes to get an overall understanding of what is expected, what kind of results I need to compare with the experiments are my results correct. So these things are essentially to keep you on the correct state as over the picture is required.

So the computer simulation always has these four steps, any computer simulation has these four steps, we start prepare with initial configuration I will talk about preparing initial configuration example. Then the initial configuration obtained is very far from the system configuration that you want to study. So then that from the initial configuration you have to run the system for a long time on the Monte Carlo over you have to make many moves, you equilibrate the system that you need the temperature, pressure and density condition.

This equilibration is very important thing because the equilibration brings this system to the microscopic state. So you start with a microscopic state initial configuration then you keep on evolving the system, keep on changing the system, so that you ultimately come to a state which I can call this system is in equilibrium so that I can start now measuring the properties. So once it reaches equilibrium the production run, this is the run where we calculate all the quantities.

And along the trajectory or along this sample run and then we use those to get the numbers. So the production not generates the trajectory that to use in analysis to obtain the experimentally observable results, obtain the equilibration has to be longer than the production. So once equilibrium is done production is reliable, so often the equilibration is 5 to 10 times longer than production run. So many times, for example, a simulation of 1000 water molecules then your production run maybe about 20 or 30 nanosecond but the equilibration might be more than that, now might be 50 nanosecond. So this is a very important thing to remember that, these four steps are always involved in.

(Refer Slide Time: 17:33)



So about the initial configuration I told you for example I want to simulate the water. Then in order to generate the configuration of liquid water which is random with random positions of water molecules random orientations. There are two ways to generate the molecules to put the water molecules one by one into the box, other one which is easier to visualize you start with the ice. Although you might be interested in knowing water at $25^{\circ}C$ at ambient condition.

But you can start with the structure of the ice or below and so that ice allows you, so you put up in box around the ice, now increase the temperature as we increase the temperature the ice melts, this is what I was saying, now so ice served as the initial configuration. Now melting and running for a long time is the equilibration so that after a long production done the liquid water that you get has forgotten that it came from ice.

Then we are in a position to start making calculations, so after the calculations and as I told you a thousand or a molecule sometimes the production running from ice can be more than 50 nanosecond or 100 nanosecond so that the final equilibrium state has forgotten the ice and is a equilibrium state of the liquid water. Then we do maybe 10 nanosecond of the production run and keep track of all the positions and momentum now and they use them energy, position and momentum orientation to calculate the properties.

The other is example seeing is the argon let me start with argon which is a crystal and then I build the ice and let us say I create interference as the along the line of the liquid solid. Then you can see that on the left hand side you have the argon fcc crystal on the right hand side we have the random liquid and the very sharp boundary between them. Similarly, if you want to consider a protein then you start with protein data bank which gives me a configuration of all the amino acids.

And I take the excess crystal structure from protein data bank which is all the positions of all the water molecules are there, if I want to study it in water then I dip it in water and run again for a long time so that I get equilibrated structure. Once I get the equilibrated structure I start studying the properties of the protein same we do in DNA or the binary mixtures. So this kind of explaining the steps of initial configuration, equilibration, production and then the analysis of the trajectory.

(Refer Slide Time: 20:36)



Trajectories of two particles from simulation

Here I am going directly to molecular characterization and there are two particles which are not very close to each other, the positions of the particles are projected on 3 dimensional space where we can visualize them and you can see that these each of them seems to be a random haphazard motion, that is what we call a random walk, the trajectories. Reason it is so random because it contains the effects of interactions with all the surrounding molecules. And this is very important because this lot of information is contained into these trajectories.

(Refer Slide Time: 21:15)

)

Basics of Molecular Simulations

- Periodic boundary condition
- Minimum image convention
- Truncation of intermolecular interaction
- Random number generators
- Markov chains
- Force field parameters.

Now we will go on the very basic techniques of computer simulation. So there are the following techniques, one is the periodic boundary condition and minimum image convention they are all barriers closely correlated the range of intermolecular interaction, which has to be

finite otherwise it is very hard to simulated and then something generate random number as molecular simulation. This random number generator has to form a Markov chain.

So these are all connected that we are saying here all the 5 things that contain as I told you a huge amount of highly creative thinking and an enormous amount of effort has gone into thinking this area of computer simulation which is an active area of research in the present state.

(Refer Slide Time: 22:16)

Periodic Boundary Condition (PBC)				
System s surface m	ize dependence: Bull olecules due to unbala	k phase simulation nced forces	s encounter artefacts from	
	Number of particles in bulk system	Number of particles on surface	% of particles on surface	
	216	216 ^{2/3} = 36	16.5	
	1000	10002/3 = 100	10	
	1023	1023×2/3	-106	

(macroscopic)

This drawback is overcome by replacing the boundaries of the system with periodic images of the particles in the system: Periodic Boundary Condition

Let us talk about a periodic boundary condition, it is as you understand the real system is Avogadro number of water molecules 10^{23} but, however, we cannot do that because we are going to solve the interacting systems it turns out we can typically consider it for some more thing like a few thousand water molecules, what really for demanding cases some time to resolve an issue we might do larger systems.

But routinely we will do a few thousand molecules and here we show that, it is one huge problem with a small system that the fraction of molecules on the surface scales us $N^{2/3}$, where n is the number of particles in the system. So in a microscopic system that is very, small, so one in 10⁸ particles is in the system, however, for a small system like 216 which was the initial simulation of water molecules must be done with 216 water molecules.

That is 16.5% of the particles on the surface, 1000 is still 10. So 216 is 36 and 1000 is 100 particles and this percentage on the last column is what the percentage of particles on the surface and you do not want that because this creates a serious problem because we do not want to start in surface effect that has a different goal and different methodologies we want to study the bulk properties but it will influence so much by the surface we cannot really trust it.

So this was the other limitations or drawbacks of finite simulations was circumvented a very carefully implementing this condition is called periodic boundary condition, which creates kind of boots repeated of the same or the original system but removes the boundary.



(Refer Slide Time: 24:29)

Done here by a real system or the red arrows up down up down, this is my Ising model but now these two vertical dashed lines on this red line defines the boundary of the system but I do not want it at the boundary. So what I do is I replicate this on the left-hand line, now see on the left most spin, up spin now has a neighbour on the left hand side because it is down spin, similarly down spin on right side has the neighbour which is on the left.

So this is you may consider these little bit artificial but it serves the purpose we come to the artificiality, it serves the huge purpose with the boundary. So the system is no longer boundaries, it does not have any surface it is extended to infinite systems. And then you can see that what are the limitations further how we go about it but it is true that it has certain correlations will be compromised.

But when the system is sufficiently large, maybe a thousand or few thousand then the particle future on the surface, they fill the correlations but if the correlation between one ends of the surface to another of the surface, it is not long range, then for many purposes the periodic boundary condition. So this is the continuity, it is made at the surface and you know we had one stop we removed the boundary.

(Refer Slide Time: 26:13)



And now in Ising model systems are fixed in digits but in this case the real liquid molecules move around then there is one very beautiful thing that has been invented that if a molecule goes out of the box, like you want on my top left in my shaded area, that is my original box, then other one which is an identical position place to the periodic image moving. So in this way the number of particles in the system is conserved.

This is done for every image. So every image retains the number of particles. So the particle leaves a central box image entering into the exactly opposite side. This is again a very nice technique but on the other hand, it has limitations as you can realize but as I am saying that these limitations are not too drastic.

(Refer Slide Time: 27:19)

Periodic Boundary Condition

Limitations

- Suppresses correlations that are larger in length than the size of the system. This becomes an issue near a phase transition where large length scale correlated fluctuations are important.
- Computer simulations often face difficulties in simulating behavior of the systems near phase transitions.

16

 Statistical mechanical techniques like finite size scaling can be employed to mitigate these limitations.

Then the periodic boundary condition there suppresses correlations for example, dealing in phase transitions particular near the critical phenomena, this periodic boundary condition creates the problem because you are creating an image of a system when correlations go hard beyond. So you are cutting off most long-range correlations, that is why study of phase transitions is very difficult but there are techniques again to circumvent them called finite size scaling but we are not going into that.

(Refer Slide Time: 27:47)

۲

Minimum Image Convention (MIC)

- PBC => Infinite number of interacting pairs => Difficult force calculation
- · Define cutoff region around each particle depending on the type of potential
- Less than or equal to half of the original box.
- · The minimum image region may contain both periodic and real images
- · One particle must not see its own periodic



Next, technique is the minimum image convention (MIC) and this is another beauty again when you do the periodic boundary conditions you are creating a huge number of images, the system has become virtually infinite. Then two things come to be: a) if you are going to include interaction with the spins then all the interaction with all these spins are not required, interaction of one molecule with all the other molecules that you created is not required. And because of the finite range of the interaction potential and more importantly you do not want to interact with your own image. Here is the red dashed line for minimum image convention you might see that show in the red box that I must cut out the interaction of my spin that is the elliptical one that should not interact with itself. So the main system is showing here in this by a double arrow of the bottom and then you draw this thing.

So within that now you are interacting with the central spin is interacting with these two spins on the left and two spins on the right but it does not interact with its own image. So that is the minimum image contains both periodic and real images and the particle must not see its own periodic repetition.

(Refer Slide Time: 29:27)

Truncation of Intermolecular Interactions (TIMI)

- The idea of MIC leads to the discussion of truncation of intermolecular interaction
- A particle in one corner of the box does not interact (or interact negligibly weakly) with particles that are far-off in another corner.
- This is particularly true when the interaction potential is short ranged
- Therefore, we need not consider interactions between all the particles to obtain the total energy of the system.
- Long range corrections (LRC) required in cases when the interaction potential is indeed long ranged: Columbic (-1/r), ion-dipole (-1/r²) or dipole-dipole (-1/r³) interactions.
- 🛞 Ewald, particle mesh Ewald etc.

Then another thing of the minimum image convention that works very well with MIC is the truncation of intermolecular potential. Many times we actually do not have to go all the way to the length of the minimum image convention because intermolecular potentials are often short range and like Lennard-Jones potential we might take up to a third neighbour or so. After that intermolecular potential becomes negligibly small.

So the interaction potential is range then we need not consider interactions between all the particles, then we can have it kind of a box within which we consider the interaction. So this is then combined with the minimum image convention to take care of the finite range of the intermolecular interaction. So as I told you there are many interesting techniques developed.

(Refer Slide Time: 30:34)



And then random number generators and particularly Monte Carlo simulation what we need to do? We need to do many random moves because we have to sample the configuration space and we cannot let it move in a correlated way then many other configurations so far from my initial configuration, we get adenine unexplored. So the basic idea then is to push the system into different directions randomly so that my system samples a maximum number of configurations.

And this is not by using a random number generated so that the positions and the velocities are changed of each atom or molecule are change randomly by small amount but then we, of course, have guidelines to see that there we accept that configuration or not and that is sampling that was developed by Metropolis and that is called Metropolis sampling, but we coming to that in the next lecture.

But we need that for every large random number and one important thing that the random numbers should not be correlated means random number should not repeat itself and this is called when you have a sequence of random numbers where one is uncorrelated to the other. (Refer Slide Time: 32:04)

Markov Chain

- A sequence (or chain) of random transition events that lead the system from one microscopic state to another such that the rate of transition from one state to another is not affected by the past events.
- Markovian process: Memoryless
- Example: Random walk
- Transition from one state to another is independent of any past history

)

So basically we developed a Markovian process that has no correlation, so each random number and the previous one next one, next one, next one, they are all completely uncorrelated each other and that is very important so that I get to sample much of the configuration space. The transition from one microscopic state to another microscopic state has to be as much uncorrelated as possible so that I sample everything.

20

Of course, that is not possible it just push you and put it away from the configuration space, but you allow it to move in all directions. So while the Markov chain is independent but the system is not independent but we push it in different directions so that everything is.



Computer simulation is done in series types in my class because of the universality that allows it $\rho^* = \rho \sigma^3$, where σ is the molecular diameter, ρ is the number density,

 $T^* = k_B T / \varepsilon$, these dimensionless unit allows the transferability that I can now compare methanol with ethanol but then I have to put them in the same dimensionless ρ^* of ethanol will have different ρ because σ is different. Similarly, methanol will be different because of the interaction potential ε . So it is very important that the dimensionally read as I discuss the law of corresponding states allows you to explore certain aspects to be universality, it is a very important thing.

(Refer Slide Time: 33:50)



Final issue I discussed with the lecture is the force field which is like interaction potential as I say Lennard-Jones is very simple even that was generated because Lennard-Jones did it by using the temperature dependent of the virial coefficient from the equation of state which was measured experimentally. But for example like water, we similarly have the equation of state; we have the diffusion coefficient of the water.

Now, there is oxygen which is negatively charged and two hydrogen which are positively charged by large law from quantum chemical calculation for charge then size, however, there are many things we are not taking into account, while we are taking the two water molecules interacting and there are pair wise additive that means 3 particle, 4 particle interactions are not taking into account, that has to be absorbed in my interaction range.

This technique of finding good interaction potential is called force field is a highly demanding and highly respected area of research. So this is the force field once I get forceful I can do the simulation but a lot of statistical mechanics and so back and put and back and put them, finally we settled on a force field and that they can be computer simulation.

(Refer Slide Time: 35:19)

Take-home Message

- Computer simulations help to study complex systems with many-body interactions
- Produces phase space trajectories (Molecular dynamics)
- Produces configuration space trajectories (Molecular dynamics and Monte Carlo)
- Important books:
- (1) Statistical Mechanics for Chemistry and Materials science B. Bagchi
- (2) Understanding Molecular Simulation, From Algorithms to Applications D. Frenkel, B. Smit

23

(3) Computer Simulation of Liquids – M. P. Allen, D. J. Tildesley

So take-home message of this lecture is that, the study of the systems with many-body interactions it really takes the statistical mechanics, huge rejection so to an extend the modern success of statistical mechanics of the subject is hugely due to the computer simulations and there are many books, my book is there, but in computer simulation of the Frenkel and Smit and then the number 3 Allen and Tildesley is a very respected book.

And I strongly recommended a lot of materials on the internet and Google and you can see many simulations, actually the trajectories, the configurations evolving to computer simulations. I strongly recommend you that you see some of these things.