GERBRAND CEDER: --temperature finite temperature methods of which you did molecular dynamics earlier with Professor Marzari. So what we're going to do is sort of continue that for a bit, showing you Monte Carlo simulations and how you can essentially use it. And after that is kind of a series of fairly random topics. In some cases, clearly we don't even know what they are. Even when it says case study, that really means that the instructors don't have a clue yet what they're going to be teaching about.

Somebody's listening in I think. OK. So I want to have a bit of a discussion on how you calculate certain properties. This is a good way to see the difference between molecular dynamic simulation and Monte Carlo. So in a molecular dynamic simulation, one way you calculate certain properties is by tracking them over time. So if you can calculate a certain property for a given microscopic state, for a given configuration, say the energy-- let's take some simple ones-- or the volume, you calculate their macroscopic averages, which are their thermodynamic qualities, really as an integral over time.

So there's sort of two issues with that, that obviously in that average you will only include states that you can reach with the molecular dynamic simulation. And so since you simulate over a finite time, if there are excitations, say, which can determine the average energy in reality that occur over a longer timescale, you will not sample them in that average. The second thing is that if all you care about are the average properties but not the dynamics of how you go between them, then sometimes you can do simpler techniques than molecular dynamics, or things that are less computing intensive. You can literally just go and sample those averages.

And so that's basically what we're going to talk about this lecture and the next lecture, talk about how you do that sampling. So if you only care about the averages but not the dynamics itself. You'll see that for some models of materials you'll actually need to do that, because you won't know the dynamics. Remember that in molecular dynamics you basically assume that the atoms move with Newton's equation of motion. So you know the dynamic.

There will be certain models for which you have no idea about the dynamics. I'll give you an example, the study an Ising model, which is a spin model of magnetic moments sort of being oriented in space. In many case you don't know the spin dynamics. It's a fairly complicated problem. So if all you want to know is the average magnetic moment, there's no way of doing dynamics on the spin. You'll actually have to sample them.

So I want to give you one quick example of a time scale problem, because I'm going to use this example throughout the lecture to illustrate how we use Monte Carlo. Here's a binary mixture, sort of green and blue atoms. You may want to know the average energy of that system. So in any given configuration, let's say you have an energy method to calculate the energy of that system.

So at high temperature the atoms will simply sort of hop around and give you some average energy. So I want to give you an idea of what you need-- do you mind passing around-- Thank you. Thank you, [INAUDIBLE]-- of what kind of timescale you need to get that average right. Thanks. And you saw, from Professor Marzari's last lecture, the relation between the diffusion constant and the root mean square displacement. Essentially, the root mean square displacement scales linear with the diffusion constant, where d is the dimensionality of the system-- so in this case 2-- and so D is the diffusion constant.

OK. To get an idea of the kind of hopping rate you need, you can assume for a second that the atoms do a random walk. If you do a random walk, then the root mean square displacement you take is the number of jumps you take-- N times the jump distance squared. You may remember that for a random walk.

So that means that now we know we can relate the jump rate, the N dt to the diffusion constant. And I'll call the N dt, which is the jump rate, gamma. So if you know how long your simulation can last, you sort of know what jump you need to see atoms jump in that simulation. So you can relate that to-- excuse me-- the diffusivity you need. And let me sort of do that on the next page.

OK. We know that the jump rate is some vibrational frequency times a success factor. When atoms have to jump over a barrier, if Ea is the activation barrier, we essentially assume that they attempt with some vibrational frequency, which is, you could say, the frequency with which they sort of try to go up the hill times the success rate. And the success rate is that Boltzmann factor, the exponential minus Ea over kT.

If you assume that the vibrational frequencies of the order of 10 to the 13 Hertz, which typically we are, anywhere from 10 to 12, 10 to the 14, let's say you want to get a jump rate of 10 to the 10 hertz. Why would we want 10 to the 10 12? So 10 to the 10, that's one jump or 100 picoseconds. Is that right? Yeah. 0.1 jump for 0.1 nanoseconds.

So if you did a 10-nanosecond simulation, you'd have about 100 jumps per atom. Which would be reasonable for some equilibration, but in many cases still not enough. OK. This is sort of a reasonable jump rate to expect.

That means that your success rate, your Boltzmann factor, has to be greater than 10 to the minus 3. So 1 out of 1,000 times the atom has to jump across the barrier. So you can deduce from that what the activation energy should be. Essentially, the activation energy should be about less than 7 kT. And then you'll see jumps of the order of once per 100 picoseconds.

And just to put these numbers in your head, because they're sort of important to remember, at 300 k, 7 kT is 180 millielectronvolt-- that's a low barrier. If you simulate at 1,000 degrees, you can do much higher barriers. Now, 7 kT is about 600 millielectronvolt. Now you're starting to be in the range of typical to low barriers in solids.

Of course, in liquids you'll have much, much faster transport. So in liquids you'll have no problem. If you do the liquid at 1,000 degrees, you'll have a tremendous amount of hopping. Solid barriers are anywhere typically from 0.5 to 2 electronvolt. So to put-- it's kind of important, I think, that you kind of get an idea of these scales that you need.

And so, like I said in the beginning, if the average is all you care about, it may be better to simply do a sampling technique. And that's essentially what Monte Carlo is. Monte Carlo is often mistaken for another form of dynamics, but it is really nothing more than a sampling method. So now remember what the issue is. So if you want to get a sample-- this is in any statistical methods-- there are two things you have to decide, which population you sample from-- how do you set constraints on the population of states that you sample from-- and with what probability do you sample them, or should you sample them unbiased for example. So there are two issues that always come up with sampling, the probability or the rate with which you sample a particular state and the group of states, the constraints on the states, over which you sample.

So before I get into that, I want to briefly review the statistical mechanics and thermodynamics altogether in 15 minutes. And I know a lot of you have seen that. Some of you have not seen that so much. I sort of want to try to put us all on the same page.

You know, I find when I see Monte Carlo simulation that people do, the biggest mistake people always make is at the outset it has nothing to do with the technical implications of Monte Carlo, it's defining the boundary conditions, which is essentially defining the constraints on the states you sample. Those constraints come from the thermodynamic boundary conditions you impose on your system. I have sometimes really funny stories. I've seen discussions at meetings where people argue strongly back and forth about problems, and all it is a misdefining of their thermodynamic boundary conditions.

About two years ago, I was at a meeting where people had enormous discussions about where hydrogen sits on palladium. Palladium is a pretty good hydrogen absorber, and people would argue back and forth. It sits on the surface-- and somebody says, no, it's at subsurface, layer subsurface. All that is-- it actually can sit both, it's just a matter of thermodynamics. If your hydrogen chemical potential, your external field, is high enough, it sits everywhere. Period. And so thinking about your thermodynamic boundary conditions before you make any conclusions is quite important.

So for those of you-- let me just remind you of the relation between the thermodynamic boundary conditions and the statistical ensemble. The statistical ensemble is essentially the group of states over which you sample. So we're always going to make the relation between a microscopic ensemble-- so this is the group of states for which you sample. You could think of it, it's the boundary conditions on the states over which you sample-- should you sample any configuration, should you constrain them-- and the macroscopic thermodynamics. And the relation here is made between statistical mechanics.

In the end, it's very easy. Any fixed thermodynamic variable-- I'll come to that in the second, what they are-- sets a constraint on how you can sample. And more particularly, the constraints come from what are called the extensive thermodynamic variables. So remember, extensive thermodynamic variables are things at scale with size, things like volume, number of atoms of a given type. And intensive variables are the conjugates. Those are the ones that don't scale that size-- temperature, pressure, chemical potential.

You'll see in a second that the probability with which you sample depends on the relevant Hamiltonian, and the macroscopic pound counterpart of that is a free energy function. So a Hamiltonian in microscopic space will correspond to a free energy function in macroscopic space. OK. So here's all I'm going to say about thermal is sort of one slide that I think represents pretty much all you need to know for doing sampling.

In thermodynamics variables appear in groups, what we call conjugate pairs. And one way you can see how they belong, it's the couples that appear together in the first law. First law tells you, essentially, that the change of internal energy of a system is the net energy flow in or out of that system. And you can write that from different contributions. TdS you could think of as the heat flow, pdV is the mechanical or volume work done, mu dN as the chemical work. So if u is the chemical potential, N is the number of atoms of a given type.

So remember, these are conjugate pairs. S is an extensive, T is an intensive, V is an extensive, p is an intensive, N is an extensive, mu is an intensive. Essentially, the rule is very simple. To define your boundary conditions, you always need to specify one out of each pair. So you need to say whether you're at constant entropy or constant temperature, whether you're at constant pressure or constant volume, whether you're at constant chemical potential or constant number of particles.

You could not leave one unspecified. If you do that, you're actually always making an implicit assumption. OK. If you have a closed box and nothing can go in and out, you're obviously under constant number of particles.

A really classic mistake is simulating on the constant number of particles, when you're really, in reality, under constant chemical field. Surfaces are a really good example of that. If you study a surface, you will often study it in simulation under constant number of particles. It's sort of the obvious thing to do. But real surfaces are actually equilibrated with their environment. So they're under constant chemical potential field of the environment or under constant chemical potential from their bulk.

And in some cases that doesn't matter. But in other cases that actually will. So the number of atoms on a surface can change.

In statistical mechanics, we tend to not use this formulation of the first law, but we tend interchange the role of the internal energy and the entropy literally by moving this to the left-hand side and dividing by T-- so riding the differential of the entropy instead. And you'll see in a second that that's a much more useful thermodynamic notation to correlate to statistical mechanics. So this is called the entropy formulation. And in the entropy formulation, U and 1 over T are conjugate variables. U is the extensive, 1 over T is the intensive, V is the the extensive, p over T is the extensive. I hope you can see this is just a rewrite of the standard first law.

OK. So how do the thermodynamic boundary conditions relate to the ensemble's over which you average? The ensembles are determined by the extensive variables you keep constant. So if you think about, if you go back for a second the previous slide, the simplest is that you keep the energy, the volume, the number of particles, constant. That's all the extensive variables.

All the extensive variables, that's the ensemble at constant energy, volume, number of particles. That's what's called the microcanonical ensemble. A realization of that would be if you did a system in a box, Newtonian dynamics. If it's in a closed box, the number of particles can't change, the volume is fixed. And if you do Newtonian dynamics, the energy is fixed. So the basic form of molecular dynamics, you could say, goes through a microcanonical ensemble. If you start thermostating your energy to fix the temperature, then you go out of the microcanonical ensemble.

So if you release the constraint of fixed energy, then you have to control the average energy. And you do that with the temperature. Then you end up in what's called a canonical ensemble.

If you release the constraint of fixed number of particles, again, then you have to take into account the field that sets your average number of particles. So then you have an ensemble with T, V, and mu. That's called a grand canonical ensemble. You can keep on making other ensembles by switching in intensive variables for extensive, it's just that we run out of names. These are three classic names-- microcanonical, canonical, and grand canonical. After that, we sort of have no names for them anymore, but they do exist.

OK. So that defines the states from which you should sample. The question is, now, how do you sample? Well, first I'm going to tell you how you should sample, and then I'll tell you how you really sample in practice. You should sample with the correct probability. And the probability of a state in an ensemble is proportional to the exponential of minus beta, the Hamiltonian-- again, remember, beta is 1 over kT-- and then normalized by the sum of that over all the states, which is called the partition function, PrZ.

And the free energy is essentially the log of that partition function. Now, what should that Hamiltonian be? You've probably seen this kind of probability function a lot already when you substitute in the energy. In more general terms, that should be the relevant microscopic Hamiltonian, and it should include all the things that can fluctuate in your system.

Typically, the way I sort of remember what goes into Hamiltonian, you think of your thermodynamic boundary conditions, you take the relevant Legendre transform of your entropy, and it's essentially those pieces that go in your Hamiltonian. For example, if you work at fixed N, V, and T, the relevant free energy is the Helmholtz free energy, which is S minus E over T. This is the part that goes into your Hamiltonian. It's whatever is the Legendre transform part of the entropy.

And that's what you see. In a canonical ensemble, you weigh states with the exponential of minus beta the energy. That's the one you're used to.

But let's say you go to a grand canonical ensemble. So now you're at fixed chemical potential so the number of states can fluctuate. So now the relevant thermodynamic potential, the relevant Legendre transform of the entropy has a minus E over T and a plus mu over T in it. So that's essentially the part that appears in your Hamiltonian. So now you weigh with minus beta times the energy minus mu N. Just sort of think of that as a Legendre transform Hamiltonian almost.

And if you were to simulate that constant pressure , you would have a P over T times V term in your Hamiltonian. If you sort think about it, it makes sense that you need these terms in your Hamiltonian. Because if they don't appear in your Hamiltonian, you have no way of controlling the number of particles. OK. It's sort of going backwards.

So this is here sort of the summary. The macroscopic boundary conditions are set by the constant extensive thermodynamic variables. That in turn defines the relevant ensemble and the probability which you sample over an ensemble. And when you do that all correctly, you can get the averages of properties, such as the energy and the volume with those probabilities. OK.

Just one brief aside-- there's an implicit assumption you make in here, which is common to statistical mechanics. We've essentially said that following a system in time will give the same averages then sampling over its ensemble. The assumption in there is that when you follow a system in time, it will reach essentially all the states in its ensemble. Or, maybe simpler said, it will reach all the states it can-- it will reach all the states that are within the boundary conditions.

So if you put-- think about this. What you're really saying is if you put a system in a box, fixed number of particle, say, at constant temperature, not constant energy, you're essentially saying that that system will, if you wait long enough, reach any possible configuration of atoms. Because that's the one that will be allowed in your ensemble. And that's called the ergodicity principle. The ergodicity principle says that if you wait long enough a system will reach, finally, all its states . Because if you know that it's going to reach all its states and you know the probability of them, then you can just sum over it. And that's what you do in statistical mechanics, rather than doing the full dynamics.

This is where we usually have to give you the caveat-- the lawyer's caveat-- that not all systems are ergodic. Most practical ones are. The examples that are often given as non-ergodic system are a bit fake. And one that I like to bring up is the harmonic oscillator. If you think of a harmonic oscillator, think of something, a mass, that is in contact with a rigid object to a spring.

If you think that the two coordinates-- what are the two coordinates of this system? Well, there's a position coordinate, where this mass is, and there's a velocity coordinate. Well, because it's a harmonic oscillator, the two are actually correlated. So if you think of the phase pace of the system as-- here's the spatial coordinate, here's the velocity coordinate. The system actually goes through an ellipse. It actually never goes to states like this and this and this.

So people say the harmonic oscillator is not ergodic. I'd say it's not ergodic because you took a stupid phase space. It's because you essentially said that the system has to coordinates. So in a two-dimensional space, it's not ergodic. In a one-dimensional space, it is ergodic.

This system has only one coordinate. That's its normal mode. So if you write it in the normal mode, it is ergodic. In one dimension, it is ergodic. So along that trajectory, if you say that's your phase space, the system is ergodic. So it's a bit of a matter of definition. In the obvious coordinate space, this one is not ergodic.

And then there are systems that for practical reasons are essentially non-ergodic. They just never get there going through all their phase spaces. Glasses are really difficult to do thermodynamics on, to integrate phase space over, because you really don't know over what region to integrate.

If I give you atoms-- I say, give me the free energy of a glass. Really, when you think about it, you don't know what ensemble to integrate over. Because if you say, well, what is a glass? Well, it's sort of random. So can I just average over all the possible positions in the box? Then I'm actually getting the free energy of a liquid. The glass actually never samples all the possible positions. So you could argue that it's a liquid that's not ergodic. And if you sort of think of it, it's a liquid that's frozen [? in. ?]

OK. So finally we're going to get to the Monte Carlo simulation. Before that, I want to introduce one toy model. And you'll see the relation with the mixing problem. And that's the Ising model, which is essentially a spin model.

So we have lots of sites. At every site we can have an up spin or a down spin. So there's only a binary variable at every site. We'll mark that with a plus 1 or minus 1-- say plus 1 for up and minus 1 for down. And so the simplest Hamiltonian or energy function you can write for this system is one that counts the nearest neighbor bonds and associates some interaction with them.

So I and J are only nearest neighbors in this model. So let's say if J is positive, then, since there's a minus sign here, you will want to make this positive so you'll get a ferromagnet. That means if you have a plus 1 somewhere, you'll want a plus 1 next to it to get flow energy. If J is negative, you will want the spin product sigma I sigma J to be negative. That means if you you have a plus spin, you'll want a negative spin next to it. So you have antiferromagnet.

So it's the simplest possible magnetic model you can make. And that was Ising's PhD thesis. And his advisors really hated it. His advisor was so mad at him for that PhD thesis that Ising left science and he became a businessman. So he never knew he became famous later on.

But what you're going to see is that the Ising model is sort of a conduit. It's a model I will use for any two-state system. So any time you have a model where there's sort of two possibilities on the lattice side, we'll use this model. Because I call it a magnetic spin. I mean, you could call it low fat, high fat cheese, whatever-- any time you have two states.

So we're going to use it for binary alloys, for example. If we have a mixture of A and B on a lattice, you could say the plus 1 is A, the minus 1 is B. So that's why it is such a useful model. OK.

So how would you sample a model like that? Well, let me first tell you a little bit about where the ideas of sampling and Monte Carlo sampling came from. It's typically credited to the people at Los Alamos, Monte Carlo samples, who built the ENIAC computer, which is essentially, I think, the first computer or one of the first computers that was built. One of these things with vacuum tubes, and you could calculate for 10 seconds and then one of the tubes would break. And then you'd replace the tube and calculate another 10 seconds.

But I believe idea of Monte Carlo sampling really goes back to Fermi. People before that sort of had sampling ideas already. They sort of knew how to use sampling for things that were difficult to integrate. And the first example that I thought is known is by the Comte de Buffon-- which I think is a beautiful name., buffon is like-- who thought of this as a method for integrating functions.

And the idea is very simple, that if you have this function that you can't integrate, or maybe it's numerically defined, a really simple way to integrate is to draw a box and throw darts in that box. Randomly take points in that box, and you'll hit here once in a while, and here, and here, and here. And if you think about it, if you track the number of times that you're below the curve, that's telling you something about that integral under the curve. Because the number of times below the curve is the integral divided by the total area.

So the area of the square is the total area, and the integral is a fraction of that. And so that fraction is the probability that you hit below the curve. So by just knowing how many times you're below the curve, you have a reasonable estimate of the integral. And if you do this with a lot of points, you'll get a good approximation to the integral.

This is how kids in school learn to integrate. My daughter's in third grade, and the way they learn to integrate is by putting grids on this and counting how many grid points are below the curve. Which you could say is a sampling method-- it's just not a random sampling methods. You use sample with a fixed grid rather than with a random grid.

So I was going to give you an assignment. This is one of these assignments for when you're sitting in a boring lecture-- which is not this one, but another boring lecture-- or you're sitting on an airplane and they hold you for two hours on the tarmac with nowhere to go and no food. You should try to determine pi. And here's a way of determining pi. You know that pi essentially relates to the ratio of the area of a circle divided by the area of the square that's circumscribed.

Because this area here, if you just take, say, a quadrant of this, is 1 over 4 pi r squared. And the area of the square is r squared. So the ratio of those two is 1/4 times pi. So if you throw darts randomly at this thing, if you count the fraction of the times you're within the circle divided by the total number of points, you'll get an estimate for pi.

So you should try this, like throw darts. You can't bring darts anymore in an airplane unfortunately, so you'll have to kind of like hold your pen randomly. And you should see once how quickly or slowly you actually approach pi. You get the first few digits pretty quickly actually-- not so bad. But then, after that, of course, it goes really slowly. Because, essentially, to get the third digit, you're going to have to almost do 1,000 points. But when you're bored, it's not a bad thing to do.

So let's talk more about sampling. There are essentially many ways of sampling, but I'm going to give you two extremes. One is called simple sampling. One is called importance sampling. Later, in one of the later lectures, I'll sort of show you that you could kind of do anything in between as well.

So what is simple sampling? It's kind of what these previous methods were. You would randomly pick points in your phase space. So if you randomly pick points, that means now states of atoms or whatever the states of your system are, you have to weigh them with the correct probability. But you could do that.

So you could randomly pick states. Let's say the state is new here. The property in the state is A. So the way you get the average of A is by weighing that A with the probability of that state. And we know how to calculate the probability. Probability we know is proportional to that exponential of minus beta of the Hamiltonian.

What I don't know is the partition function that I have to normalize with. But what I can do is after I sampled a bunch of states I can literally define an approximation to my partition function as the sum of that exponential over all the states that I've sampled [INAUDIBLE]. And I need that, because I need to normalize my probabilities to 1.

Simple sampling, it's called simple because, as I said, economists use it. That was a joke, but, you know. It works for small spaces-- small phase spaces. It does not work at all for materials for a reason I'll tell you in a second. And you think that's obvious. Did you think that was obvious that simple sampling does not work for materials?

Who thought that was obvious? Nobody. So it's not obvious. Good. Because otherwise, it was a trick question. I'll show you in a second that it's not obvious.

But why does it not work? You essentially, with simple sampling, pick states in proportion to their degeneracy. So if you think of your spin model, let's say you're spinning at a low temperature where this thing wants to be a ferromagnet with all the spins aligned. How would you pick spins? Well, if you pick them randomly, you would never end up with a ferromagnetic configuration. Because you picked the spin here up, here down, here up, here down. So you would sample a lot of states in the phase space but never the relevant one.

OK. And typically you pick states with higher entropy. And you know that, because the entropy is essentially the log of the multiplicity. The entropy at a given energy is the log of the number of states of that energy. So remember, if you do simple sampling, you're going to pick states proportional to omega.

So you're always going to pick states with high energy, because we know that d log dE is 1 over kT is a positive number. So that means the number of states with a given energy is an increasing function of energy. So that means that if you simple sample you're almost never going to pick down here, because there are almost no states. You're always going to pick up here, because there are a lot of states there.

See the parallel with your spin model is-- see, these are the ferromagnetic states. You have low energy, everything's nicely lined up. These are all the random states, the paramagnetic states. In reality, your system lives somewhere in between. At some temperature it has some average energy, and then it has a fluctuation of energy around it. So your real system kind of samples these states, and you're always hanging out here-- you're always sampling the wrong thing.

If you had said that this was obvious that it wouldn't work, I would have thrown this at you. Here's a paper from a famous physicist that I respect very well, but he was so wrong. This was somebody who's been great in density functional theory. He has done beautiful work in it and at some point tried to do a phase diagram. And realize, of course, I blanked out the name out of respect. But of course you could find out.

They were really good at doing quantum mechanics and getting energies of states. And they wanted to get free energies now at finite temperature. This was on the aluminum lithium system. I mean this is 1988. This was like if there was any heydays ever of aluminum lithium, this was it. Everybody was working on aluminum lithium, because everybody thought that was the alloy of the future.

Because lithium is the lightest metal. It's the lightest solid in the periodic table. So any atom you can replace by lithium will make your alloys lighter. So that was a big dream. You could make really light airplanes and whatever. And [INAUDIBLE] built anomalous capacity for making aluminum lithium alloys, et cetera, and it never panned out. Turned out you couldn't weld them very well, and they have really bad fatigue resistance.

So you really don't want to make an airplane out of them. Ironically, they make the doors of airplanes out of them. I never know how to figure that one out. And the seat frames, there's aluminum. But the seat is supported on is out of aluminum lithium. But the doors always scares me, when I sit by the door and I know this is aluminum lithium, which has poor fatigue resistance and very poor welding characteristics. But anyway, everybody worked on aluminum lithium.

And so these guys were good at doing quantum mechanics. So how did they get the phase diagram? That they actually took a box of atoms and randomly made up configurations and then calculated the energy along with quantum mechanics. So this was essentially simple sampling.

And then they summed that and got the partition function. If you have the partition functioning, you have the free energy. And this was the phase diagram that came out to the right. It sort of violates all kinds of phase rules, and it's completely wrong. And so it's not that obvious that simple sampling doesn't work, because even the best people have done it.

I like this example. This comes out of a great book. This is the book by Franklin Smith. And I'll show you a list of references at the end. This is a great book on Monte Carlo if you want to delve deeper into it.

They have this great analogy of simple sample versus importance sampling. The question is, how do you measure the average depth of the Nile River? And simple sampling is essentially, well, you know the Nile is in Africa, so you just walk around Africa with a dipstick and you sort of measure how deep the Nile is. So of course, most of the time you're not in the Nile, so you get a lot of zeros.

So what's a smarter way to do it? Well, a smarter ways to first find the Nile. That's putting a bias on your sample. You really want to get to the Nile. And then once you're in the Nile, you want to stay in the Nile.

So that means you still want to have some amount of randomness to walk around in the Nile, otherwise you're not getting the average depth of the Nile. But you want to be biased, so you stay in it. And that's the idea of importance sampling.

Remember, what is importance sampling? You want to sample the important states. Those are the ones with low energy. If you want to set up a thermodynamically-relevant ensemble, obviously you know that your system spends most of its time in the states with low energy and then fluctuates around it. So you want to get to the low energy and stay around it. So you want a sample with bias and then correct your probabilities for that bias. That's the whole idea.

OK. So here's the idea of importance sampling. Remember, in a random sample, you pick randomly. And to average a property, you weigh the property in each state with the correct probability. Now, if you think about it, maybe the best sample we could make-- couldn't we sample right away with that probability? So could we sample with a probability that's directly proportional to the real probability the state should have in the ensemble?

So if you do that, if you could sample these states with the propr probability, then their average is just determined by summing them, by a trivial average. So does everybody see the difference? In simple sampling, you randomly pick things. In importance sampling, you right away try to shoot for getting them with the proper probability so the low energy state's a lot more than the high energy state. Keep that in mind.

There will be properties where you need the high energy states. And then you're going to have to adjust your Monte Carlo sampling algorithm to go there. Because in some cases maybe some properties are determined by extreme in your distribution. Maybe when it gets to a certain energy your system does something weird. If you want to see that, then you should bias your system towards there. But if we're interested in just thermodynamic behavior, we want to sample around the energy minimum.

So there are various ways that you can construct the probability-weighted sample. A common way is what's called a metropolis algorithm. And this is the one we'll use pretty much exclusively in the [INAUDIBLE]. And the idea is that you construct what's called a Markov chain, which is really just a sequence of states, where, over a long time, if you do it long enough, you visit each state with the proper probability.

So I'm going to define two quantities, and then I'm going to tell you what rules they should obey. So obviously you have to start with some state. Well, in a metropolis algorithm you can pick that state randomly. So I call that state i. Then you have to pick a new state called j from i. And the rate at which we pick that state I call omega 0 i to j.

But you don't always accept that state. You accept that with some probability p i going to j. So the total transfer rate, how many times you pick j from i, and put it in the Markov chain is determined by the product of these two. It's times you actually pick it as a possible transfer state times the success rate in getting there. That's p i j.

The two conditions that a metropolis algorithm has to satisfy are typically these two. The first one isn't actually a real condition, it's just sort of a practical one to work with. Usually people look for what's called equal a priori probabilities. What that means is that the rate at which you pick j froim i should be the same as the rate at which you pick i from j.

That's actually not absolutely necessary. But if you don't satisfy this, you have to correct later on. And we'll do that two lectures from now-- we'll give you an example of that. But it's easiest to work with equal a priori probabilities.

In most this is a condition that's amazingly easily satisfied. You'll see that. Next lecture, we'll do some examples of complicated Monte Carlo moves, and it's hard to violate this one. You have to work on it. But it does happen.

The more important one is what's called detailed balance. Detailed balance is that the rate at which you actually transfer from i to jay and the rate at which you transfer from j to i has to be inversely proportional to the probability that you're in the initial state. So essentially the rate at which you go from i to j over the rate at which you go from j to i has to equal p j over p i, or these are the probabilities that you are in that state.

It's actually fairly easy to convince yourself that if that's true you're going to end up with a correct probability distribution. I mean, think about it. If you look at this carefully, what is this really saying? That if you were doing a lot of samplings in parallel, this gives you a steady state distribution of states. Because if you look at the left-hand side of the green box, that's the number of times you're going out if i into j. Because it's the probability that you're in i times the rate at which you transfer to j.

And the right-hand side is the probability at the rate at which you go from j to i. Which is, first of all, you have to be in j-- otherwise you can't go from j to i-- times the rate W j i, the rate at which you go from j to i. So when these two are equal, you sort of have a net zero flow of probability density. So you have a steady state distribution. If this holds for all pair of states i j, your probability distribution isn't changing anymore. So you end up with a steady state distribution.

You can be fairly creative about how you make metropolis algorithms. But here's a very simple one, which you'll often see a lot in a Monte Carlo codes. So you first have to design some pick rate, some rate at which you choose new states for i. So that's omega 0. So you have to have some mechanism for picking potential moves. So omega 0 i to j.

And then you have to have an acceptance criteria. And a typical acceptance criteria is that you accept the probability from i to j is 1. That means you definitely accept when the energy goes down-- so when the new state has lower energy than i, than the previous state. And if the new state has higher energy than the initial state, then you accept with a Boltzmann factor. So then you accept with a certain probability.

You can easily show that this satisfies detailed balance. Because if you-- let's see if I can squeeze this in here. Detailed balances that essentially p i p i to j has to equal p j p j to I. Well p i is the exponential of minus beta i over the partition function. p i to J, let's assume that i is the state with low energy is 1. That has to equal p j, which is the exponential.

Sorry for my poor writing here. This pen is displaced from where I see it. This is exponential minus beta E j over z, times p i j, which is the exponential. I'm going to write it below here. The exponential of minus beta Ei minus Ej. This Ej cancels with this one, and so you get that the exponential is minus beta Ei equals the exponential minus beta Ei. So this set of transfer rates satisfies detailed balance in an almost trivial way.

So how would you practically implement a metropolis algorithm for sampling phase space? You would start with some configuration you would choose a perturbation of the system. So that's essentially determine your mechanism for picking omega 0 i j. You compute the energy for that perturbation. If the change is negative, you accept the perturbation. If it's positive, you accept it with some probability.

And then you go on and you choose the next perturbation. And so you essentially cycle this. So you build up a sequence of states. This is your sample.

This must be a Pentium 286. It's checking the equations. That's-- OK.

So what you will get then is some property as a function of sample size or sampling, which is often called Monte Carlo time. Which is very confusing, because this is not a dynamical trajectory. You could think of it as a trajectory through phase space, but there's nothing here that says that it's a dynamical trajectory.

I'm going to later confuse you even more and talk about what's called kinetic Monte Carlo, which is a way of mimicking dynamics with Monte Carlo. But standard Monte Carlo simply has nothing to do with kinetics. It's an efficient way to sample phase space. So you'll go through some property, and then you just essentially track the average.

Typically what people do is they often, just for practical consideration, cut off the first part and remove it. Because, remember, you started with a random state. You may be very far away from the low energy state. And so by cutting that off, the first part, you often get a much faster relaxation of the average. Because the initial state may have really very high energy and thereby some weird property.

And so you only start sampling after the system has relaxed towards the minimum. So you often talk about, in Monte Carlo, people using a certain amount of equilibration passes and after a certain amount of sampling passes. And you only average over the sampling pass.

So let's go through this for the very simple magnetic model. How will you do it in a magnetic model, in a simple Ising model? You could randomly initialize the spin. You could start with that. You could also start with a ferromagnetic configuration. The closer you start to your equilibrium state, the faster this is going to go.

An obvious perturbation is you pick a spin and you flip it over. So if it was up, you make it down. If it was down, you make it up. Calculate the energy for perturbation. If it goes down, you accept it. If it doesn't, you accept it with a certain Boltzman probability and you go back.

And as you can see, you can do this very fast. If you think of computer code now, how you write this, these are trivial operations. You have to calculate a random number to randomly pick a spin in the lattice, or two if it's a two dimensional. You have to compute the energy, but your Hamiltonian is just like j sigma i sigma j. That goes really fast. It's like multiplying a few numbers. And then you have to compute that exponential also.

So you can cycle this through literally millions of times per second on a modern computer. Even on a very basic computer, if you did just this two-dimensional Ising model, if you coded it well, you could do probably 50 million of these per second. So that's how fast you can sample through your ensemble.

So here's what comes out. Here's real output. So this is the 2D square lattice model I showed you. The temperature is in units of j over k-- so the interaction constant or the Boltzmann constant. So kT over j equals 2, that's actually below the period temperature of this model. So this is a ferromagnetic model. And so if you heat it up, it would go paramagnetic at some point.

This is still below that temperature, but it's not far. The transition temperature would be about 2.23 or something like that, 2.24 in this model. So you're below the paramagnetic transition temperature, but not much.

So I started it up randomly I think. And so the energy is sort of high. It kind of fluctuates for a while and then starts going down. And here it seems to have relaxed fairly well, and then it just kind of fluctuates around the average. And you average this and you'd get the energy.

If you want to know the magnetization, well, you would keep track of the magnetization, which is sort of the difference between how much up spin do I have and how much down spin do I have. You track that, and this is that quantity as a function of temperature, and see it's clearly ferromagnetic. You have preferred up spin.

It slowly goes down with temperature. And then this is the transition temperature. It's the paramagnetic transition temperature, beautiful second-order transition.

So just to give an idea, doing this with Monte Carlo is a one-page code without the bells and whistles. And it runs before you blink. I mean, it's so fast that you would sample this. OK.

How do you detect phase transitions? This is sort of a critical issue in any simulation method to be honest. Whether you do Monte Carlo or molecular dynamics, it's one thing to look at the atoms, and it's another thing to figure out exactly what they're doing. You may have seen that if you try to do things like melting with molecular dynamics, it's not trivial to exactly find where the atoms melt. Because what do you do? You just stare at them.

But it seems kind of a fairly random decision about-- let's say I gave you this lattice model, and I asked you, where does it become disordered? Where is the paramagnetic transition temperature? You could just stare at the spins. But I'll tell you, when you're close to that transition, just below and just above, it looks pretty similar. And one is still the ferromagnetic phase. The other one's the antiferromagnetic phase.

So you need a more systematic way of looking at phase transitions. And the beauty is you can do that through the thermodynamic quantities you sampled. If you sample the energy, you typically use the fact that the energy is discontinuous at first-order transitions. It's not so at second-order transitions.

If you work at constant chemical potential-- and I'll come to that later. It's often a useful thing to do in Monte Carlo. it tends to equilibrate faster-- you will have concentration discontinuities at first-order transition.

Another thing to usually track is the heat capacity of the system. The heat capacity is the fluctuation of the energy. If you go back-- can I go back? The heat capacity is essentially a measure of how much you fluctuate around the average. It's your sigma squared.

So the heat capacity is defined as the fluctuation of the energy-- so E squared average minus E average squared. That tends to show singularities at second-order transitions. But it's even useful often to spot first-order transitions, even though it's a little harder to. I'll show you a lot of sort of case studies and examples on more complicated systems later on.

But for example, here's the heat capacity of that 2D model as a function of normalized temperature. So notice how the heat capacity the transition is here of course, about 2.24 or so. [INAUDIBLE] has a beautiful singularity. This is actually a logarithmic divergence to infinity. So you can't miss this transition. Even if you tried, you wouldn't miss it.

There's a great Java program on the web that does two-dimensional lattice files if you want to play with it. And you can literally type in temperatures and interaction strength, and it just simulates on the screen for you and plots like thermodynamic quantities. Pretty cool. OK.

So in the last bit of time I want to go through some variants of this lattice model, how we use it for other fields. And then, in the next lecture, I'll pick up on that. Like I said before, you can use this for anything that has essentially two states on a fixed topology.

So like I said, you could use it for A and B atoms on a lattice. It's very often used for surface absorption as well. If you say you have a surface-- let's say this is a triangular lattice, so that could be the surface sites on a 1 1 1 FCC metal, like platinum or palladium. If stuff can absorb on the lattice side-- now you can really see that my pen is miscalibrated. I'm drawing on the latest points.

So if you want to know if stuff absorbs, well, you see that becomes a binary problem. At each site you want to know, is there an absorbent there or not? So you can define a binary variable and whether it's plus or minus 1 doesn't matter. People often write things in not spin variables but occupation variables, which are maybe a little more natural. They're mathematically less elegant, but you could say the occupation variable is 1 when the side is occupied and is 0 when the side is not.

The advantage then is that you only get interaction when two things are occupied. So one occupied side with one occupied side gives you no contribution to the Hamiltonian, because one of the p i's is 0. Whereas if you use spins, then this is a plus 1, and this is minus 1. So it gives you a contribution to the Hamiltonian. Mathematically it all works out, but it's a little more confusing.

So you can even write that out for a binary alloy. You can sum over the probability that you have A B bonds on i j and multiply with the A B interaction. You can get the probability that it's A A and multiply it with the AA interaction, and you get the probability that it's B B and multiply with the B B interaction. OK.

I think the coolest thing about Monte Carlo is that you have anomalous degrees of freedom about the moves you do. These are essentially the perturbations you attempt to get to the next state in the Markov chain. And the one thing you should never do is let yourself be guided there or constrained by physical principles.

People often do moves that they think are physical. But remember, it's not the point to do a physical move. The point is to efficiently sample phase space. And so on the spin model, well, you think, a spin going from up to dowr or vice versa is a physical move. Spins flip up and down all the time. But in some cases, for example, it's much more efficient to flip over whole patches of spin.

When you're near a second-order transition in a material, the fluctuations in the system are very long range. So it's sort of like the wind blowing to a wheat field or something. There are whole patches of spin that move together. So you'll get much better convergence there if you flip whole patches of spin. So rather than attempting one at a time, you could say I'm going to attempt 10 at a time close to each other.

So you can always pick your dynamic. So I've given you an example. Here's an example of two types of dynamics. Let's say you have A and B atoms on a lattice.

So if you want to equilibrate that, you could pick the obvious dynamics. You could interchange A and B. So if you have a lot of us and you have A here and B here, you could interchange them. That could be a possible move.

And it's sort of an obvious move, because it sort of mimics diffusion. Atoms interchange and equilibrate like that. That's called Glauber dynamics on the lattice model. Turns out that that's useful, but a much faster move to equilibrate is actually exchanging the identity of the atoms. So rather than interchanging A and B at different positions, you take an atom, and if it's A, you try to turn it into B. And if it's B, you try to turn it into A. That's called Kawasaki dynamics.

Now, you see that if you do Kawasaki dynamics, you will need some kind of chemical potential. Because in Glauber dynamics, you constrain-- the composition is fixed. You just interchange the positions of the A and B, but you're not changing the number of A or B. You conserve the concentration.

In Kawasaki dynamics, you don't change the number of particles, but you do change the concentration of the A and B species. Because whenever I flip A to B, I'm changing the concentration. So you will need some kind of chemical potential term in your Hamiltonian, and that chemical potential term is the difference in chemical potentials of A and B.

So that difference will drive the concentration. So it's essentially multiplied with-- this is essentially the concentration. It's the sum of the probability that you have A on each side. That's essentially the concentration. OK.

You'll find that Kawasaki dynamics equilibrates 100 times faster easily in many cases. The reason is that, if you think about it, if you started with a situation where you had too much A in one area and too much B in another, if you do Glauber dynamics, you can really only equilibrate that system by slowly diffusing these. A has to go this way, B has to go that way.

In Kawasaki dynamics, the A here feels right away that there's too much of it, and it starts turning to B. The B here starts turning to A, and you're done. So it tends to be a much more efficient way.

This is sort of a generic thing. Open systems tend to equilibrate faster than closed systems. So if you can at all make your system open and end up with the same end result, it'll go much faster. And it's essentially because you give open systems more degrees of freedom. You can change the chemistry of them.

Yes.

AUDIENCE: I mean, principally [INAUDIBLE] complexity of the system. First of all, you're getting [INAUDIBLE].

GERBRAND CEDER: Not in the thermodynamic limit. That's sort of an idea of statistical mechanics is that you get the same averages.

AUDIENCE: [INAUDIBLE].

GERBRAND CEDER: You can. But but you still-- so that'll definitely make it go faster if you do random exchanges, but you still lose efficiency. I'll tell you why. First of all, you have to pick a dissimilar pair. So if you think about it, half of the time you'll pick a similar pair. So you will pick A here and A here, and so you have to give up on that move, because it's not going to do anything. So that's one thing. That's not a major thing.

But the fact that you can fluctuate the concentration of your system will in many cases still allow you to equilibrate faster. OK. And usually this is not-- it's a good thing you pointed it out, because I haven't talked about it. The choice of the ensemble should not matter for thermodynamic properties. So any time you're doing something with periodic boundary conditions and your objective is to study the bulk of something, you're OK.

When you go to finite systems-- you're studying, I don't know, the thermodynamics of nanoparticles-- then the ensembles do matter. The problem that you run into there is that it's not obvious what the right ensemble is. Like if you study a nanoparticle, now you have to-- there is essentially no thermodynamic limit of a nanoparticle. It's a finite system by definition. And you just have to sort of figure out, can it exchange things within the environment or not. But in general, the choice of the ensemble doesn't matter.

Let me sort of skip this. I've given you examples on a lattice model right now. But you can, of course, do off-lattice Monte Carlo. For example, this the example of a liquid. Let's say a liquid is a bunch of atoms in a box. And again you could study the electrodynamics and track them. You could also get their thermodynamic averages just by sampling all the possible positions of the atoms.

So which perturbations would you pick? Again, anything that's consistent with the degrees of freedom of your ensemble is fine. So here an obvious way to do this would be pick an atom and you then pick a random displacement vector. So you say, I'm going to pick a random displacement vector, and that's my perturbation. I calculate the energy, it goes down, I accept. If it doesn't, I don't accept it. OK.

The question is, do you truly pick this displacement vector random? But what about its magnitude? You probably pick its orientation random, but what about its magnitude? Do you pick it out of the whole box length? Do you pick it out of a very small length?

So typically you'll do it somehow, like you pick a delta x-coordinate out of range-- call it minus a over 2 to plus a over 2, and then the same for a delta y and z-- and then you randomly pick a coordinate in that range for x, y, and z. But the question is, what do you take A? If you take A really big, you're not going to accept a lot of your perturbations. Because if you take A of the order of the box length, then many times if this is a dense liquid your atoms is going to end up very much on top of something else.

The density of liquids is what? In many cases, 10% less than a solid. At best they're still dense. So most space is filled by atoms.

So if you take A very big, you're going to sort of equilibrate in principle the composition easily-- I mean the density-- but you're not going to accept a lot of your moves. So you say, well then, the solution is I pick A small. But if you pick A too small, then you're probably going-- you have a high acceptance rate. So if you just displace A-- if A is like 100 of an Angstrom, then you're atom is just moving a very little bit. So most time, you're going to accept it.

But what's the problem now is you have high degree of correlation in your average. Essentially all the states you put in your Markov chain look all the same. If you barely this place your atom, they all look the same. So it's going to take a really long time to get good averages. Actually, you could almost think of Monte Carlo with very small perturbations on the system, it's almost like molecular dynamics. It's like the atoms sort of move slowly with a kind of random force, but they still move slowly along trajectories and are real trajectories, but they don't wildly go through phase space.

The nice thing about models like this is that you can actually, during the simulation, adjust A. And this is often what's done. Often people will adjust it so that they get a reasonable acceptance rate. If you set A too big, then you get low acceptance rate. If you set A too small, you get a high acceptance rate people. Typically shoot for anywhere from 0.3 to 0.5 acceptance rate. So you want to be somewhere where the order of a 1/3 to 1/2 of the moves you try are accepted. Because if they don't get accepted, you're wasting all your effort calculating energies.

And this depends a little bit on the cost of your energy function. If you're doing this with quantum mechanics, then you really want to optimize your Monte Carlo moves, because any energy evaluation costs you a lot of time. If you're doing this with maybe some simple hard sphere model or something, you can afford to have a lot of moves-- calculate a lot of energies on which you don't execute.

So the last thing I want to say is, how do you pick random numbers? How do you implement probabilities? Because it's sort of not obvious in a computer. Computers are not probabilistic. And that's done by random number generation. Sort of maybe obvious, but I want to show it anyway.

So remember, what you want to do is you want to implement a move with a given probability with this Boltzmann factor. Remember, if delta E is positive, that Boltzmann factor is between 0 and 1. So let's say this is the value.

Well, what you do is you pick random numbers between 0 and 1. And what's the probability that the random number is less than this exponential? Well, that probability is that exponential of minus beta E i. We'll call this p i. If you take a random number and the probability that that random number is below p i is p i.

So you pick the random number if it's below that exponential. You execute. If it's above that exponential, you don't execute. So you execute now with the correct probability. So random numbers give you a way of implementing probabilities.

There's a few caveats. A lot of cheap random number generators are not at all random. I used to not believe this until I tried it. I always thought that was kind of like people write books and they just talk about it because it's kind of cool to say how non-random. They actually are very non-random.

You will often find that if you plot the density of random numbers you generate, that there are discrete holes where you have very little coverage. And it's because of the sort of fractional algorithms by which they work. In many cases, that really doesn't matter. For a lot of practical things, it doesn't matter.

I'll tell you where I've seen it matter. It matters when you're in this regime here, really close to 0. Because essentially, a random number generator, the way it generates random numbers between 0 and 1-- it doesn't do it between 0 and 1. It generates between 0-- or is it-- between 0 or 1. No, between 1 and some integer. No, sorry, between 0 and some maximum integer, like 2 to the 16 or something like that. And then it divides by 2 to 16. And that's how you get it between 0 and 1.

But that means your set is discretized. So often what's important is the spacing between 0 and the smallest random number. Because if you work with very high energy barriers, very high energy excitations, they fall in that first gap. And so you have a completely discretized probability there. And there are problems.

It's the kind of thing I thought I'd never run into, and I've actually run into several times within that [INAUDIBLE] research. If your system is stuck and it doesn't have a lot of low energy excitations left, it tends to execute high energy excitations with a way too higher probability. Because think about it-- if the biggest number is 2 to the 16th, your random number hits 0 with a probability of 1 over 2 to the 16th. That's way too high. It should be 1 over infinity.

You should hit 0 with a rate of 1 over infinity if you had a perfect continue of random numbers, but you hit 0 with a rate 1 over 2 to the 16. That means if you hit that random number you execute any move, even if it's 100 electronvolt above the ground state. Because that exponential is not 0. It's small, but it's not 0, So you will execute. And you'd be surprised how often this is a problem, especially if you have a wild phase space in which you can go.

So how do you solve this? We solve it in a trivial way. We add a small epsilon to a random number. Which is so, trivial but you have to know it I. I've never seen this in books actually. But see you should not hit 0. Because when you hit 0, you execute everything.

OK. OK. So I'm going to stop here. And let's see, today's Thursday. So yeah, I'll see you on Tuesday for the rest of Monte Carlo.