NICOLA MARZARI: As you have heard that this is actually the last class for the course, it will be a split lecture. I'll start for half an hour. Then Professor Ceder will continue and wrap up the course, and then at the end, I will also spend 15 minutes doing course evaluations. So we'll just leave you here with the forms and filling up.

On my side here, I think I'll end up discussing a topic that's actually very dear to me. I think it's been a very intriguing and very fruitful application of modeling and, in particular, modeling based on quantum mechanical energy schemes to an interesting set of scientific problems. We have said it over and over again. A set of simulations allow you to do sort of computational experiments in areas where maybe things are happening too fast to follow them with real experiments, or maybe they happen on time scales-- sorry, on length scales that are too small to really follow. Or as in this case, they take place under certain thermodynamic conditions that are very difficult or even impossible to reproduce with an experimental apparatus.

Let's look, actually, at the first case, and this is research which a lot of work has been done at University College in London by Dario Alfe and Mike Gillan. So this is actually a cut out of the Earth interior, and in case you're not familiar with this, the way it works is that we are actually sitting on a large liquid region. So you see, on top here, we have sort of the Earth crust. And what is very important for us is actually this region, dark yellow and light yellow, in which we have the so-called core of the planet that is mainly constituted by iron.

Iron actually is a very stable nucleus. So instead of the early formation of stars, it's one of the elements that sort of, once created, sort of was more stable than the other. And what happens in our planet is that we have a sort of increasing pressure going inside, and the inner core, so-called, again, sort of mostly iron is actually solid.

And we know very well what is the pressure here at the boundary between the outer core and the inner core. This part is liquid. Pressure is increasing here. So this part is liquid. This part is solid. It's mostly iron. And for this sort of case, we'll assume that is 100% iron, although there is another problem of really how much impurities they are diluted.

And we know very well what is the pressure there, because we can sort of look at seismic waves and how seismic waves get deflected by that discontinuities. And the pressure is roughly 330 gigapascal, but we actually don't know what is the phase diagram of iron at that pressure. And you see, if we knew what the phase diagram of iron at that pressure was, we could pinpoint exactly the temperature here, because these points here are sort of points of coexistence at 330 gigapascal between the liquid and the solid phase at that boundary.

So if you want, by knowing the pressure, we would be able via the phase diagram of iron to actually pinpoint what is the temperature inside the Earth. And before sort of the set of experimental and theoretical work that I'll show, the estimates, actually, for this were really sort of ranging in a very wide spectrum of possibilities from 3,000 Kelvin to 8,000 Kelvin. And this is where sort of our initial simulations become very useful, because really, once you believe your energy model, there is no reason to sort of think that it would fail, either because the pressure is too large or because the temperature is too high.

In all of this current experimental capabilities to study the high pressure system, the best instrument is what is called a diamond anvil cell. That is basically made by 2 diamonds. So the more perfect they are, the better that they squeeze whatever you're putting inside. And you know diamond is transparent to a lot of wavelengths, so they are very useful to probe them with spectroscopic techniques.

But basically, now, the best you can do-- and you still sort of break up hundreds of diamonds in the process-- is that you can get to 300 gigapascal if you are doing this at room temperature. But as soon as you want to study things at a higher temperature, really 200 gigapascal is the limit that you can reach. Well, we want to figure out what's happening here.

There are other techniques-- experimental techniques to study matter under conditions of very high pressure and very high temperatures done with what is called a gunshot cannon in which you really sort of shoot pellets and try to establish what's going on in the instant in which they hit their target. And so they find themselves at very high condition of pressure, but it's very difficult to get accurate data out of this.

And so we'll see here actually sort of how ab initio simulation and, in particular, thermodynamic integration has been used to figure out what is going on in this region-- in this region here. And as usual, you need to make sure that you believe your energy model. So these are a couple of validation curves. This is for an HCP iron. That is one of the high pressure phases, and the comparison between the sort of experimental results and simulation where the simulation are the sort of solid lines. And you see it agrees very well in describing the equation of state, pressure versus volume in this case.

And then in particular, this is a calculation of the phonon modes of the vibrational frequency. This is something that actually you could very easily do with the quantum mechanical tools that you have done as a sort of post-processing of your laboratory three, but again, it works. It works. It works very well.

So once you believe iron, now the issue is trying to figure out what is its phase diagram at, say, which temperature you have coexistence between the solid and the liquid. And you have actually done this in your laboratory for a sort of explicit method to find out the transition temperature by just doing a simulation of a liquid and a solid phase, and figuring out at which temperature the boundary between the two phases doesn't really grow into each other. So the solid doesn't keep growing, or the liquid doesn't keep growing.

And you see an example of this in sort of what follows. The caveat that in doing this kind of simulation is that they are still sort of very expensive to do ab initio. You usually need at least 400 or 500 atoms to do such kind of simulation and sort of to kill the final size effects. And this is, even now, borderline for what we can do even for the simplest elements.

There is another approach that I wanted to introduce first, and this is the one in which we actually sort of determine explicitly the Gibbs free energy. So in this case, what you want to find out is thermodynamic stability at a given pressure and at a given temperature. So the Gibbs free energy is your correct thermodynamic potential. And so your goal, if you want, is being able to calculate the Gibbs free energy of your solid system as a function of temperature and the Gibbs free energy of your liquid system.

And of course, the point where these two free energy cross is really your coexistence point. Above that critical temperature, the liquid phase is more stable. And below that critical temperature, the solid-- that solid phase is more stable.

And sort of one possible way of calculating, actually, explicitly this free energy or, in particular, calculating the free energy difference with respect to a known phase is to use thermodynamic integration. That is a technique that sort of has been discussed in some of the previous lecture by Professor Ceder, so I've just put up a reminder here. And the nomenclature is that we are using, actually, a Hamiltonian or an energy model that depends on a parameter lambda. And the parameter lambda is something that, say, brings us from the solid to the liquid, because ultimately what we want to sort of figure out is, what is the difference in Gibbs free energy, say, between the liquid and the solid, or maybe between two solids, between one reference phase of the solid and another reference phase of the solid?

And again, the idea behind thermodynamic integration is fairly simple once you skip all the math here. Again, one can write out explicitly the partition, or the partition function is written here. And then it just requires a little bit of math, sort of figuring out that the derivative of the, in this case, Gibbs free energy-- well, sorry, this is actually a constant volume. So it would be a sort of [? isentropic ?] simulation. No, sorry. It would be constant volume, constant temperature simulation. So [INAUDIBLE] energy.

So the derivative of the free energy with respect to this parameter lambda-- you see once you sort make the appropriate substitution-- that is, you put the logarithm of the partition function, and you work out explicitly what is that logarithm of the partition function in there, you have 1 over q times the q over the lambda, and 1 over q is the partition function here. And this is the derivative of the partition function with respect lambda.

You basically get an expression that is nothing else than the average in your ensemble normalized by the partition function of the u over the lambda. So in order to calculate what is the derivative of the free energy with respect to lambda, you need to calculate what is the thermodynamic average in your ensemble of the derivative of the energy with respect to lambda. And so this is a quantity that you can always calculate, and this is just an average over your ensemble.

So let's look at sort of how this is done in a specific case. So again, I just returned here the partition function, the logarithm-- the free energy, sorry, as the logarithm of the partition function. u will be our internal energy, and this is what you calculate with your quantum mechanical model. And for simplicity, I've actually broken this in three parts, and that is I've written out what is the energy at the equilibrium and then the energy for any possible configuration of the ions can be written as the energy at equilibrium plus what is called the harmonic term.

So when you expand the energy of a system around equilibrium, the first derivatives are 0 by definition. And so here you have the quadratic terms in the displacement of the position around that equilibrium position. And then you have all the nonlinear term that will call sort of anharmonic.

And it's actually, say, in the case of a solid, this makes a lot of sense, because it's sort of fairly easy to calculate explicitly the integral of the partition function for this term in the overall energy function. And then this sort of is reserved for an explicit integration than using molecular dynamics. And because basically you have a logarithm of an exponential, if you sort of break out of this thing in the sum of different terms, all the terms that contribute separate free energy pieces to your overall system.

And so we could calculate, let's say, for a solid system what is the additional vibrational-- in this case, free energy-- that comes on top of the equilibrium ground state energy, and coming just from the harmonic term. And I haven't worked out the algebra explicitly, but it's actually fairly easy to do these integrals by writing the harmonic term in terms of the interatomic force constants.

So if you sort of figure out with electronic structure codes what are the normal modes of your system-- that is, what are the modes that give you an expression for the harmonic term that is diagonal in the displacement-- and you find out the frequencies, well, you can actually substitute that expression in here, work out the integral, and basically what you obtain is that your harmonic vibrational free energy is just given by a sum over the logarithm of all your vibrational frequency.

So this omega here would be your phonon frequency if you go back to what we had seen before for the case of iron when we have calculated the phonon spectrum. Here it is. So here, what we have calculated is all the possible frequency for iron, depending on the wave vector. So in principle, we have a continuum of frequency, and we just need to basically integrate these curves or sum over our course representation of them to get the vibrational free energy-- at least the harmonic term. And this is it.

And so this is sort of the first step. This can give me, say, what is the change in the vibrational free energy in going, say, from one phase to the other. And I just calculate the phonon frequency in sort of two different phases. What is more intriguing is the calculation of all the nonlinear terms that can be very important at high temperature.

And this is sort of the slightly more difficult task, because especially in a sort of ab initio framework, it requires very extensive integration of your thermodynamic ensemble. So you would do something like molecular dynamic simulations or Monte Carlo simulation to sample as much as possible a set of representative microstates, but it tends to be very expensive. So there is actually sort of a very useful, if you want trick that is often used in order to calculate this anharmonic term, that remember the anharmonic term is nothing else than the overall vibration of free energy minus the harmonic term.

And the trick are consistently used in introducing a reference classical potential so that a lot of your expensive sampling of the phase space can be done with classical simulations and a very expansive ab initio simulation are only used to figure out what is the sort of difference in thermodynamical terms between your quantum potential and your classical potential. So again, sort of just a mathematical term. I mean, we can sort of think at our overall energy. So this would be the overall, apart from this sort of equilibrium term.

So what we are trying to calculate is the difference between the whole term minus the harmonic term, but we actually introduce a classical potential that sort of hopefully reproduces as closely as possible the energies, the interactions between all our atoms. And so we can actually break down this term in two parts-- one that has to do with the difference between the harmonic crystal and the reference system, and one that has to do with the difference between sort of our overall energy term here, or the overall free energy term, including non-linear fx minus their reference-- minus the reference system.

And now, this here is a calculation that involves only sort of classical potential calculation of your phonon frequency. So this can be done very extensively over longer classical simulation, and in a way, it captures all the complexity of the phase space. While this, that is very expensive, because you need to calculate the internal energy with sort of the quantum mechanical model-- is sort of you are trying to sample the difference between two quantities that are as similar as possible. And in the ideal case in which your classical potential is able actually to reproduce your quantum potential, this goes towards 0. So it's something that is very simple to integrate.

So by introducing basically a reference potential, you are shifting a lot of the thermodynamic integration to the classical simulation. And you are just using the quantum mechanical simulations to sort of sample the difference. And so this can be done, actually, with much shorter and much smaller simulations. And I want to go actually into the details of what you need to do for the case of iron that is what is a good potential, but people have figured out that something as simple as basically a sort of Lennard-Jones just like kind of potential on top of anharmonic term is actually good enough to provide a reference potential.

So basically with all this formalism, what one is able to do at the end is calculate the free energy differences in which, if you want the fundamental components are calculating the phonon frequency of your system-- in this case, specifically for the case of a solid-- that gives you the first step in the vibrational free energy. And if nothing else, then this case, a density function or perturbational theory calculation.

You want the phonon spectrum, and then you need to do a number of molecular dynamic simulations on very large and sort of very extensive for classical potential. And then if you want, you just correct your classical simulations with the difference between the classical and the quantum potential. And once you have all of data, you are basically able to calculate all sort of the free energies that you need. And I guess this is an example. I'll show in a moment a more telling example in which, again, sort of we compare actually sort of what our prediction are for the pressure, volume, curve coming from the ab initio simulation and experimental results.

But more importantly, if you want, we are actually able to figure out when the free energy of the solid and the free energy of the liquid cross. And we can do this as a function of pressure. And so this is, at this stage, our best prediction for the phase diagram of iron as a function of pressure. In particular, the black line here is sort of the set of simulation that I've described. And there starts to be a reasonable agreement between a number of Shockley-Gunn experiments and a number of simulation that puts roughly the temperature at 330 gigapascal, just sort of above 6,000 Kelvin.

And so this was-- when it came out, the scientific literature, it was sort of hailed as taking the temperature of the core of the Earth. And at this stage, it's, again, sort of our most accurate prediction of what is going on inside. Just sort of to make the point that this is not the only way, remember that one can also do the simulation of the coexistence between the solid and the liquid exactly as you have done in your laboratory, in your lab number four.

This is actually a paper, again, sort of *Science* 2000, really sort of state of the art research. And in these particular simulations, they used classical potential. So again, not very far from what you have done with the only caveat that the classical potential had been extensively optimized to describe solid and liquid iron. And the way that optimization was done, it used what is called a force matching method, in which basically you do, again, extensive but small ab initio molecular dynamic simulation.

And you sort of tune your classical potential in order to have this more or less the mean square error with respect to the ab initio simulation. And that often, because you have basically so many configurations, and so you sample so many possible different environments for your atoms, it can somehow give you a reasonably robust classical potentials, again, in sort of very simple cases. This case is just an elemental system, and sort of in a range of temperature and pressure.

So this was sort of an example that I wanted to discuss a little bit more in detail. I think all this sort of research area of studying the properties of matter at extreme condition has been extremely fruitful from the point of view of ab initio simulations. And sort of there are other cases in which actually the phase diagram of, say, methane under high pressure has sort of led to a number of interesting discoveries. There are all the planets-- Jupiter and beyond often are made by a mixture of methane, water, or often, say, in the case of Jupiter, a large amount of hydrogen.

And the phase diagram of this system, again, at very high pressure is largely unknown. One of the most interesting sort of prediction that was made already in the '30s by Wigner was that hydrogen, under high pressure, should really start to behave as all the other alkali metals in the first group. If you think hydrogen is in there, and there's a sort of molecular gas at ambient condition, but everything below it-- it's really a simple metal-- lithium, sodium, potassium, and so on and so forth.

And at ambient condition, hydrogen, if you want, would belong much more to group seven with the halogens that really form dimer and gases, like a chlorine or bromine, than on group one. But if you keep compressing, compressing, there should be a point in which it becomes metallic. And we haven't really reached the state. There are indications from the shotgun experiments that it could reach the metallic state at 350, 400 gigapascal, but this is extremely important, because, say, for a planet like Jupiter, there is a huge difference if we are going to have in the inside of the planet, say, a molecular solid, or a metallic solid, or a metallic fluid. If we have a hydrogen that all of a sudden inside is a metallic fluid, well, we can have there right away, like for the case of Earth, a significant source of magnetic field if we start to have a sort of circulation in that fluid.

So all of this is very important, and I think one of the most intriguing prediction that actually sort of had originally come just as a hypothesis by Martin Ross in 1981 was that, in some of this planet, there could be an interesting decomposition of methane. So basically, what could be going on is that inside, say, Neptune, as you increase the pressure, your methane molecule under the effective pressure starts decomposing. And this actually has been seen in quantum mechanical simulations. And so what you have is that-- on one side is that the carbon atoms-- and you have a breaking of the carbon-hydrogen bond.

And so you have formation of pure carbon, of diamonds, and you have formation of higher hydrocarbons. And actually, there has always been a sort of experimental observation of an anomalous presence of high order hydrocarbons on the surface of this planet. And so this simulation for the first time provided a reason why there should be higher order hydrocarbons. They are basically created when methane collapses under pressure, and then somehow from deep inside the planet, these sort of convective currents bring these hydrocarbons up.

But the other prediction that, I think, even if not confirmed at the time is the most appealing, is that at the same time you have nucleation of diamond, and so there is this beautiful picture of basically a rain of diamonds taking place inside the planet and basically converging to the inside. And again, this was just a prediction, but I think in 2002 or 2003-- maybe it's mentioned here. No, 1999, so just two or three years after the previous paper, finally there were sort of diamond anvil cell experiments made by Raymond Jeanloz at Berkeley. And all of a sudden, methane under pressure sort of gave rise both to higher hydrocarbons, but also to the clear signature of diamonds.

So they saw basically the nucleation of pure diamonds just under pressure, and I guess this brought in general, both in the geophysical and in the planetary community, a lot of resonance for the importance of this ab initio simulation in the field. And the last example that I wanted to show in terms of sort of geophysical or planetary importance is actually the one of water that, as a very intriguing phase that was sort of suggested for the first time, again, in 1999-- and what I'm showing here is just a network of solid oxygen.

So here I'm just showing the oxygen for a high pressure phase. In this simulation, we were looking at 20 gigapascal of pressure. And what happens is that, if you increase the temperature enough, actually water becomes superprotonic-- that is a very exotic phase. So the covalent bonds between hydrogen and oxygen breaks apart, and the system is still a solid. Again, high pressure even if there is high temperature, but the hydrogens start moving around like a liquid.

So you have a coexistence of phases between a solid phase of one sublattice and the other sublattice that has gone liquid. There are a number of sort of materials that, even at ordinary conditions, do this. There are some salts silver iodide that does this, and there are interesting fuel cell materials that do this, because, again, in this case, say, when you have a superprotonic, in this case, phase-- that is, when you have that the protons behave as a liquid-- well this system becomes a very intriguing ionic conductor.

But again, this was a prediction from 1999. And actually, just a few weeks ago came the experimental confirmation from Lawrence Livermore, and lo and behold, the superprotonic phase of water has been found. It has been found that 47 gigapascal instead of 20 gigapascal as usual, I mean, because of all the errors involved in density function theory, and in this specific case, also, the the errors involved with the sort of quantum mechanical description of the vibrational excitation that, remember, is a Bose-Einstein gas.

And I'll conclude with the last slide that, again, it's not any more geophysical or planetary, but is, again, another very intriguing example of what was discovered in the high pressure regime. And this is, again, some work from the late '90s that has to do, actually, with the behavior of alkali, like sodium or lithium under pressure.

These are metals that we really sort of describe and consider-- simple metals, where the valence electron behave as almost nearly free electron gas. Think at lithium for a moment. You have a core with two electron in the 1s orbital, and then you have a single electron on the 2s orbital that becomes very delocalized.

And now this is what was discovered that, I think, it was very appealing. And at some level, it was also confirmed recently by the Carnegie group that does a lot of these high pressure experiments, but basically, when you start compressing your system, what you have is that the 2s electrons start overlapping more and more. If you want that very high pressure, even the core of your lithium atoms start overlapping.

And what's happening is that Pauli principle becomes exceedingly relevant. That is, you start to have 2s electrons that, by pressure, would be forced to be almost in the same quantum state. They start to overlap very significantly.

And what happens is that, in order to escape this, the electrons actually self localize in the interstitial position. So the 2s electrons, from sort of being around the lithium atom, sort of moves around and jumps into the interstitial position. It localizes, and it actually pairs with another electron.

So you have that your unit cell becomes double or sort of becomes more complex. And actually, your system is a phase transition from the metallic case of the sort of ordinary simple metals to an insulating case in which, actually, really a pair of electrons localize in the interstitials. And again, these are actually sort of cutting edge research that one nowadays can do on a simple computer, because studying something like the equation of state of a simple metal in a variety of [INAUDIBLE] lattices and a variety of crystallographic incarnation is actually very simple to do. It's nothing different from what you have done for, say, silicon or the semiconductors in your first laboratory.

With this, I'll conclude my part, and Professor Ceder will continue from here. And again, as a summary, just wanted to give you an example of our whole line of research that has come out from modeling and very accurate modeling, and really going in directions where experiments would be-- or still are exceedingly difficult, or even impossible to do. And with this, I'll switch on to the next lecture. Let me--

GERBRAND CEDER: So let me hammer in again the sort of overview, the kind of guideline by which we set up the course. I think by now hopefully you'll see the structure. We covered energy methods. We covered simulation methods, and then the last third of the course was really a kind of showing you how these are combined together, and that's often still a very hard part, how you combine all these things to make impact on something.

And this is-- being at MIT, impact is the one thing we worry a lot about. And if you see the kind of things that-- to integrate modeling more with materials research, there are really these four poles that you need. There's obviously method development, and we told you a lot about that. We told you what goes into DFT, what goes into molecular dynamics, Monte Carlo, there's an important factor of dissemination, which is getting better and better done, I think, in modeling.

There's now a lot of stuff where you really just can fairly easily get access to codes. You get quantum codes, good MD codes-- sometimes for money, but in many cases, for free. And then there's, of course, the education part, which we try to address here. This is actually a picture from last year, or two years ago. I actually think I have one from longer ago where, see, Professor Marzari still has a beard, which is how you know how old that picture is.

But anyway, but I think, then, the fourth pole is the one you shouldn't forget. You think that for modeling you only need modeling, but I think what you need more sometimes than modeling is the basic signs of your field. I think, without understanding the field in which you work-- the theory, the material science of it-- there's really very little impact you can do with modeling, because you can be a good modeler, but if you don't know what to model, you're never going to have impact, unless you choose to remain on the pure theory and method development side. But I would argue that, even then, you need to understand. You have to have field specific knowledge, because otherwise you never know what is going to be important to develop.

So I sort of hope you don't forget that. So this being the end of the course, I'll give you a few more maybe sort of loaded perspective-- personal perspectives on the field, and what's good, and what are the potential directions about it. This is what people want from you. They want properties. They want behavior.

And this is where you essentially start, and typically you see people draw these kind of slides of the multi scale perspective. You go from atoms to microstructure to continuum. This is essentially-- has almost never worked. This is great for raising money.

It's a great sort of cartoon view grab of what you do. It's rarely that you actually ever do this. There are very few properties where you make it across all the length scales. And so what you really do is you kind of connect at some level. You sort of connect these through your brain. In many cases, there are properties we determine just by doing stuff at the electronic level.

And we don't barter coarse-graining that explicitly through the length scales. And what you're really doing is you're using your basic materials theory knowledge-- for example, that maybe when you work on color of materials. I mean, I had this one-- jewelers once came to me and asked if I could predict color for gemstones and for thin layer-- thin surface layers that they could use to coat. So in many cases, those are fairly local properties, especially in oxides.

So there you sort of decide yourself, gee, there's kind of one length scale that's important-- this one-- and I'm going to directly calculate my property. So this is you. This is, I think, why your field specific knowledge is, in some sense, more important than all these multi scale techniques.

This is my pet topic that-- it may seem obvious that you can-- an obvious goal to try to simulate all the way from electrons to properties by coarse-graining the length scale. But you have so much more impact if you just substitute that idea by knowledge. I think, if all the money had been spent on trying to integrate length scale had been spent on actually trying to integrate materials modeling with material science by connecting your field specific knowledge, by trying to figure out what to compute, which is often the hard part, we'd actually done much better.

So I'm going to give you an example, which I really like. It's not my work. It's done by people at Northwestern. It's on intergranular embrittlement of iron. And this is sort of one of these things that you really think you need a multi scale sort of big kind of modeling approach on. Essentially, the observation is that there's quite a few impurities in steel that embrittle steel, and there's some that actually enhance the internal cohesion.

Sort of the typical examples are phosphorous, which is really bad. Phosphorus is a really bad embrittler of high strength steel, whereas boron tends to enhance intergranular cohesion. So when you sort of hear about this for the first time, you think, well, how am I going to study this? Maybe I'll set up a big simulation with a grain boundary in, and I'll put phosphorus there, or I'll put born there, and I'll try to pull it apart.

And that'd sort of be the brute force, maybe multiscale approach. You could embed that in a continuum theory, but there's a sort of very simpler-- a very basic theory that tells you there's a much simpler way to understand this. And that's the Rice-Wang theory, which essentially states that the embrittling tendency of a solute depends on the difference in segregation energy between the segregation energy for that element to a grain boundary and to a free surface. And you can sort of intuitively-- this is actually mathematically justified. It's actually a quite elaborate theory that you can catch in two sentences, but you can sort of see why this is true from a driving force perspective.

If you have an impurity that has a large segregation energy to a surface rather than a grain boundary, then its presence is going to enhance surfaces, because then it can have that benefit from that large segregation energy. So those are going to be embrittlers, because they will tend to promote decohesion. Whereas the ones that have a very large segregation energy at the grain boundary, rather than the free surface, will tend to promote cohesion. And so if you believe Rice-Wang theory, then this problem is very simple. All you have to do is calculate segregation energies, and that you can do.

See, now you're back to the atomistic level. You could set up a supercell with a grain boundary and really calculate the energy there. Do the same on the surface, and you really don't have to do any dynamics. And that's what these folks did. This is our work. This is work that was published in *Science* about 10 years ago and already by Art Freeman's group and Greg Olsen, who sort of runs a company that does atomistic modeling and other modeling for steel development. And essentially, you see that this theory works out very well. On one side, you have the strong embrittlers, phosphorus and sulfur. And they essentially have much higher-- this is the difference in segregation energy between the surface and the grain boundary.

These very much want to segregate to the surface. And that's why they tend to create internal surfaces, whereas carbon, to a lesser extent, but boron to a large extent promotes decohesion. So you can get a fairly easily good insight of what are embrittlers without having to coarse-grain across all the length scales.

AUDIENCE: Why is the separation different?

GERBRAND   I'm sorry?
CEDER:

AUDIENCE: Why is the separation different in both boron and [INAUDIBLE].

GERBRAND   Yeah, why it's different. That's actually what that paper is about, why it's different, essentially. And I don't
CEDER: remember anymore, but I know it had to do with the bonding. This is why these pictures are here. There's a very-- I think it probably has to do with the-- well, first of all, boron is smaller. There's definitely a size effect, boron and carbon. These two are much smaller than these two, but there's another effect, I think, which has to do with their-- these, I think, tend to prefer-- in some sense, have much more bonding potency.

So in a grain boundary, they have a lot-- they have higher coordination. So they gain a lot more from covalent bonding than these two. So these two don't gain as much from being coordinated by iron atoms. I think that's what I seem to remember, the gist of it. But that's exactly what that paper is about.

To sort of end, I wanted to mention one thing that we actually haven't talked about that I think that is becoming a fairly big issue in computational modeling. There's sort of two ways that I've mentioned to go from all this stuff you can calculate at the atomistic and electronic level to the macroscopic level. One is to just sort of think your way through it, what's important, and I'll just calculate that.

And then one is to sort of do this sort of much more explicit coarse-graining or simply handing off information. With diffusion, you can just hand off information. You can calculate activation barriers. You can put that in a kinetic Monte Carlo to get diffusivities. And then you can hand that off to macroscopic diffusion theory and do simulations.

There's really a third one that's emerging, and that's sort of essentially giving up on trying to understand the relation between macroscopic and microscopic, but rather try to determine it statistically. And this has been done in chemistry for a long time. It's really only now, I think, reaching material science. Chemists have used things like QSAR, which are a quantum structure-- and I forget what the A is-- relations or something.

But essentially, their idea is that maybe there are certain macroscopic properties that are so hard to link to the electronic structure-- that is, you just try to correlate them. And I'll give you an idea-- an example. One where this is done extensively is toxicity of molecules.

How do you predict toxicity of molecules? This is a very difficult question. I mean, how do chemists do it? Well, gee, they got a lot of experience, essentially. All these groups probably react with this part of your body, but how do you even coarse-grain towards toxicity. It's essentially, what is toxicity?

It's the reactivity would have an enormous large potential of molecules, what your molecule can do with it. So people have tried to do that with QSARs. Essentially what they say is, for a lot of molecules for which I know that they're either toxic or not toxic, I'm going to calculate with quantum mechanics, as much as I can, bond lengths, electron densities, electron negativities. I mean, people literally have parameterized charged density surfaces for molecules.

And then they do a correlation study. They do something like principal component analysis or linear regression, and they see which factors at the atomic scale correlate most to the output. And there are powerful data mining techniques that do exactly this. This is where the field of science really intersects with a lot of other fields.

When you apply for a credit card, the first thing that people do on the information you provide-- that the companies do is they data mine you, and they predict essentially your risk of default. And that's all done with correlation studies. They have enormous large databases of people who have been in default, and they know that, well, if you're 23 years old you're probably more likely to default than when you're 47. And so all that comes out of the data.

And people have essentially tried to do the same thing with the relation between macroscopic behavior and electronic structure. And in chemistry, it's very well advanced, I would say. You can buy QSAR software. You do a search on Google. You'll find it everywhere.

In sort of solid condensed matter, it's really sort of in its infancy, but you're starting to see it. And one of the reasons I think it's in its infancy, it's harder-- the link between those properties and electronic structure is harder. There are intermediate scales which can mess up the relation. This would never be successful probably for mechanical properties because we know that things that-- two materials that look almost identical on the microscopic scale can have very different mechanical behavior depending on the impurity level, depending on the microstructure.

So you kind of already know ahead of time you're lost if you try to correlate just to atomistic properties, but it may work for other cases. The other thing that makes this possible is that essentially we get enormous amounts of computing power, and we can do very high throughput data generation at the atomistic level. And I was going to show you just one example.

So this is the idea that you essentially correlate these things, but let me skip through that-- that you use some kind of learning method like neural networks, or Bayesian statistics, or some kind of correlation method to correlate output to calculable input. So let me show you one example. This is something I had gotten into myself, and still I mean since a few years-- sort of a very old problem, and it was essentially trying to predict crystal structure of materials. It's a really old premise, essentially an unsolved problem, and it's an intriguing one, because you can actually show quite dramatically that the energetics of quantum mechanics, the way we do it in our density functional theory-- is actually pretty much most of the time accurate enough to predict the stable crystal structure.

So what do I mean with predict? What it means is that the true crystal structure will have the lowest energy, but the problem is finding it. So what I'm saying is, if I gave you a list of 20 and I say the winner is among here, you're done, because you calculate the energy of all 20.

And the lowest one will be the real one, but you can't make that list of 20. That's the problem. It's a search problem. It's an optimization problem, rather than a sort of energy problem.

So the idea we have is that, rather than trying to calculate the energy of all crystal structures, can you guess them by calculating a few and correlating their relation with a whole lot of other ones? So it's the same idea. Let's say you have a really big vector. This is like the energy of all possible crystal structures, whatever it is. In the end, we have to use finite numbers, you'll see.

But can I just calculate a few that goes fast-- that's the one in blue-- and then predict all the ones in red without calculating them so that I get to sort of a complete vector? And the way you want to do this is by correlation methods. The philosophical question is really, if you calculate a few pieces of energetics of materials, do you have enough information in there to say a lot more about the material?

And you can do that explicitly. Some people would say I'm going to calculate a few pieces of energetics-- fit potentials to it, say. And then with those potentials, calculate a large amount of other things. That's the explicit approach. The question you can ask yourself, can you do this implicitly?

And that's what we did by building correlations. Essentially, we build a sort of a smart algorithm-- call it a neural net or whatever-- that essentially builds this relation. So you train it on large amounts of data sets for which you have the full vector. Essentially, you're asking yourself, is there a correlation between the pink data variables and the blue ones?

And it turns out that there is. Gee, OK. Well, this is the picture I forgot. You can actually, with a few calculations, predict very well the energy of other things. And this is sort of the summary. This is out of a very large collection of binary metals where we did the test, and this is essentially out of how many calculations do you need to do to get all the ground structures in the system with a given accuracy? Because of course, this is a statistical method, so you're never going to get 100% accuracy, unless you calculate everything.

But what you see is that with something like 15 calculations, you get 85% to 90% accuracy with-- this is with the data mining technique. So this is an example of you build a predictive method without a theory, purely based on statistics. And you see more and more of that done. There are people who built correlation methods between very basic electronic structure input data and finite temperature data-- for example, like melting point, which I think, as you've seen, is a non-trivial thing to actually calculate, because melting point you sort of need the intersection of the liquid and the bulk free energy. So you could argue, how is that melting point already hidden in the energetics, and can I extract that correlation?

So I'm going to skip through a bunch of slides because I want to give you enough time to do the evaluations. These were actually slides on examples of materials design. Well, maybe I should-- a few more minutes-- say a few things about it. One was band gap engineering. The reason I put it in there is because bandgap is the one thing people always think with DFT, well, I can't do anything useful. The bandgap is always like a factor of 2, sometimes more off.

But you can correct for that. People have made empirical pseudopotentials that correct for it. There are actually now methods out there that will give you the bandgap more accurately. But the whole idea is that you still get a fairly systematic ordering of the bandgaps, and you look at-- can look at bunches of semiconductors.

And what I want to show you here-- this was actually a Monte Carlo optimization of the bandgap of super lattices. These were aluminum gallium arsenide super lattices, and essentially the configuration space over which they searched was, how do you arrange the aluminum and the gallium on the cation sublattice? And so essentially this was a scheme-- a Monte Carlo scheme to sort of go to the largest possible bandgap in that system.

So it's a fairly well known example of materials designed purely by computation, and you find that, after about 10,000 Monte Carlo moves, you seem to find pretty much what the maximal bandgap is in that system. And then you can try to make that will then be-- a thermoelectric is one where there's been several ab initio predictions for very high thermoelectric coefficient. Then I'll show you one was this lanthanum antimony, essentially.

So my last two slides are on the future of modeling. I can't imagine a better feel to be in. And I think there's a few reasons of-- the two things that we essentially benefit from tremendously-- one is computing power. I mean, when I started in this field in the early '90s, I mean, I can't even thinking back anymore how slow and expensive computers were. And that's a good 10 years ago.

So we ride this curve with an anomalous benefit. In my department talks, I usually show numbers. I think, in the last 20 years, you've had an improvement of sort of performance versus price of something like 10 to 50 million. That's how much more computing power you get for your buck now than actually about 15, 20 years ago.

That's enormous. There are no other techniques in material science that do that, period. So you ride this enormous curve. And for whatever people say, this is likely to continue at least for quite a while.

I think another one is that we benefit tremendously from all the developments in condensed matter theory, because this allows us to compute. Those allow us-- the benefits in condensed matter theory allow us to compute things much more accurately. And the third leg is the basic materials theory development, which is probably the sort of weaker leg of the three-legged stool. But so this is one of the reasons this field is still very good.

You should keep into account scaling, which we've only sort of cursorily said something about. But here's some examples of why scaling can kill you. I put out some hypothetical examples of linear scaling examples, cube scaling examples, and worse. If you take an example of linear scaling, I think molecular dynamics with real space potentials is probably pretty well up to linear scaling these days, because you can partition space. As soon as you start partitioning space with finite range potential, you can get basically linear scaling.

So I figured out, what can you do in 40 years, assuming we get about a factor of 10 to the sixth. That's what you would get from Morse's law. 40 years is about when you retire. So I mean, I'm long gone by then, but you guys are going to retire.

So where will you be? Well, let's say we can do 10 to the 8 atoms. So what is that, 100 million? The record stands more at a billion now, but nobody kind of performs at the record all the time. You have to essentially take what you can do on an almost daily basis.

So if you do 100 million atoms with potentials, you'll be able to do 10 to the 14 atoms. What does that buy you? I'm not totally sure.

You have to ask yourself, what can you do with 10 to 14 atoms that you can't do with 10 to the 8? Because a lot of things don't depend on the number of atoms. They depend on the length scale. And so of course, the length scale is the cube root of the number of atoms. So there, you've only gained about-- well, you've gained about a factor of 100 in length scale. But if you go to something that scales like the cube. And I know this is a little overdoing it-- LDA scales in large problems is much more like n squared log n. But let's say on smaller problems it's just like the cube of n.

So if you can do 1,000 atoms now, when you retire, you'll be able to do 100,000 atoms. So where is that going to take us? I think it's going to allow us to do some things that we have a hard time with-- biological species, for example, where even this is really cutting it close. You want to do proteins and DNA. Yeah, people do dabble in it now, and you can do it, but it's fighting against the frontier.

You've got to do solvent, and so people now make more ad hoc models for including the solvent. I think, when you do 100,000, you are going to do biology much more. Are you going to do mechanical behavior? I don't think so. Not in the direct ab initio scale. This is probably by no means enough to sort of span the length scale from atoms to microstructure.

Of course, if you believe that we're all going to be doing [? LDA ?] [INAUDIBLE], now you're up to-- what is this, a billion atoms? So starting to count. Of course, when you have a billion atoms, you're going to have to figure out how to look at them and how to extract data from them, which is non-trivial at these large simulations.

The people who sort of broke these frontiers-- there was a group at Louisiana State that, for the longest time, drove this forward. And I think they may have done the first billion atom simulation. A lot of the research is in data transfer, data analysis, visualization, because think about it. When you have a billion atoms, how do you decide what's going on? You just look at it, or like-- so that's actually a big chunk of that research.

And then I took one that's kind of excessive, but let's say you do [INAUDIBLE] configuration interactions, which is as close to-- essentially as close to the exact solution maybe that you could get for quantum mechanics, except maybe quantum Monte Carlo now. But that scaling is depending on who you believe is somewhere from n to the fifth to n to the seventh.

It's such small numbers that it's not even really sure you can quite calibrate the scaling. But so if you can do 10 atoms now, you can only do 100 by the time you retire. So these things are useful for calibration, but they're not going to allow us to, say, do DNA and do it more accurately.

So the only reason I'm saying is that you will still not be able to rely on Morse's law for all your problems. You still have to figure out ways to, in a sort of smart way, use computations. I always find it funny the proposals I read of we can't quite compute this, but next year we can. Well, most stuff you can't compute today you won't be able to compute next year either.

And so you'll still have to find shortcuts. If you count on being able to sort of put materials in a box, and simulate the hell out of them, and sort of figure out what comes up, I think it's unlikely that that will happen in any reasonable time for a lot of properties. There are properties you can do it for. But for most things, I don't think you can.

So it was a pleasure teaching you. Professor Marzari and I had a lot of fun. We always have a lot of fun with this course. We're going to leave you with the course evaluations, which have to be done. And can I count on you? We're not allowed to touch these, so one of you has to return this to department headquarters, or to Kathy Farrel. Can I count on one of you just to pick them up and like-- OK, thanks, everyone. OK, thank you.