

**NICOLA
MARZARI:**

Welcome, everyone. Lecture 8-- hopefully this will be a somehow simpler class to follow than last class. That was sort of heavily theoretical. What we are going to do today is really, after two or three slides of recap of what we have seen in the past lecture, we'll go into really sort of practical application of density functional theory. We'll first discuss how you set up calculations. That is, what are the parameters, and what are the objects that you manipulate when you do an electron [INAUDIBLE] calculation?

And then we start seeing a number of examples to give you the feeling for the properties that we can calculate using density function theory for the accuracy and for the pitfalls you need to be careful with. And in the next class on Thursday, we'll finish up with more case studies.

So let me actually spend three slides to recap what we have seen in the last lecture. Starting somehow from this, that, I would say, is the most fundamental one. And it's really rewriting quantum mechanics. So this is really the Schrodinger equation from the '60s, density functional theory. And this is what it says-- that we don't really need to solve a many-body differential equation to find that the ground state charge density, and the ground state energy, and the ground state properties for a system.

But we can actually solve it using a variational principle on the charge density, and the functional that has to be minimized is written here. That is, for any given charged density and prime, we have, at least in principle, a well defined functional, and we need to change to vary the charge density and prime. And the minimum value that is functional will take is actually the ground state energy. Somehow, sort of the weakness of all of this is that everything is well defined in principle, but it doesn't work in practice.

To remind you what were the sort of conceptual steps, what Hohenberg and Kohn proved from their first theorem is that, given any charge density and prime-- more or less any charge density and prime, an external potential v prime is well defined, for which that charge density is going to be the ground state charge density. This was sort of the inverse part of the first Hohenberg and Kohn theorem. And then, at least in principle, the solution to the Schrodinger equation corresponding to v prime is well defined.

And we call that c prime. And so one can construct a functional f that is just the expectation value of p -- ψ prime over the many-body kinetic energy operator plus the many-body electron-electron interaction. And so this here is the well defined, at least in theory, density functional term that is here. And the other term is just the integral of the external potential times the charge density.

So everything is well defined in principle, and so in this, it's a perfect reformulation of the Schrodinger equation. The real problem is that we don't have an exact representation of this universal density functional. And so what Walter Kohn and Lu Sham did was they tried to figure out an approximation to get a universal functional. And so in the total energy functional, these were-- this was their choice for the approximation. That is what they said-- is, well, let's try to figure out what are the most relevant physical terms in this functional, and then we'll sort of leave-- sweep under the rug all the many-body complexity of this problem. And we'll call data an exchange correlation function.

It is really where the fine detail of your solutions are. And so what they said, in this universal functional, we can extract terms like the Hartree electrostatic energy, written here, that is a very simple and well defined functional of the charge density. The sort of second term was the tricky one, because they wanted to figure out what was a significant contribution to the overall quantum kinetic energy.

And as we have seen, there isn't really a good way to extract kinetic energy that's basically a second derivative of the wave function-- of curvature of the wave function from a charged density. The sort of most remarkable case is that of a plane wave. A plane wave of any wavelength gives you a constant charge density, but gives you a very different second derivative, very different kinetic energy.

And so what Kohn and Sham did-- they introduced a related system of non-interacting electrons that, for a given charge density n , they would have the same-- sorry, they would have the same ground state energy of our original problem. But because these are non-interacting electrons, one can define exactly what is the quantum kinetic energy of this set of non-interacting electrons.

So if you want, this is just a definition, but it turns out to be a definition of the second piece that makes our third piece here very, very small. So by extracting, selecting the green and the red term, we were left with an exchange correlational term that, if you want, still contain all the complexity of this problem that it's well defined in principle, but we don't know how to solve. But for this, then, they applied that the same idea of Thomas and Fermi. That is, they said, well, maybe we could try to approximate this exchange correlation term with a local density approximation. That is, we try to construct this energy term by integrating an infinitesimal volume by infinitesimal volume, and each infinitesimal volume will contribute to this many-body problem and exchange correlation energy density that is the exchange correlation energy density of the interacting homogeneous electron gas at that density.

So when, in 1980, Ceperley and Alder first did very complex calculation-- quantum Monte Carlo calculation of the interacting, but homogeneous electron gas as a function of the density, this functional, at least in principle, was, for the first time, parameterized over a whole set of density. Often we call that the Perdew-Zunger parameterization, and so we had finally a working expression.

Once we have a working expression, then the problem of finding the ground state becomes our electronic structure computational problem. That is how to find the minimum of this functional. Note the very significant difference with respect to Hartree-Fock. In Hartree-Fock, we had a variational principle, and that led us to our expression for the energy for the Hartree-Fock equation, in which there was a well defined exchange term. It was written somehow complicated-- complex, but well defined.

But because it was basically coming from a variational principle, we had the possibility of making Hartree-Fock better and better by extending our class of search function from Slater determinant to more flexible classes. We could just take, say, combination of Slater determinant. So Hartree-Fock, in principle, can be made better and better in a systematic way. And the computational cost that you will pay is horrendous, but it gives you an avenue.

Density functional theory doesn't give you an avenue. It sort of monolithically states that there is going to be, in principle, a well defined universal functional, and in the Kohn and Sham sort of the composition that is going to be in particular one single given exchange correlation functional, but it doesn't give us any sort of systematic route to find better and better functional.

And that's why for many years there wasn't really much applied work using density functional theory. Sort of in the early '70s, people started studying very simple system, like the surface of a metal sort of represented as a step in a free electron gas solution. And then it's only sort of the beginning of the '80s, when we had the parameterization of LDA from the calculation of Ceperely and Alder, that people could sort of put together an overall working algorithm. And I've shown you in the last class sort of the phase diagram of silicon, the first time that we see somehow that this theory is going to give us a very accurate practical result and will be able to give us the lattice parameters, the bulk models of silicon, just by plotting the energy as a function of the lattice parameter, or it will give us even the phase transition to high pressure phases of silicon.

What I had written in the previous slide was the total energy in this density functional paradigm, and we had, as a computational goal now-- and you see how we address this-- the task of minimizing that total energy. And there are two main routes that we can take. Once the exchange correlation potential is written in an explicit form, this is a well defined, even if nonlinear, functional of this independent single particle orbitals-- the Kohn and Sham orbitals of which the charge density is just the sum of the square moduli.

And so we can look at this as a nonlinear minimization problem. We need to find the orbitals that minimize this expression, or as it often happens, we can write the associated Euler-Lagrange equation. We have a variational principle so we can take the functional-- the functional differential of that, and that gives us a set of differential equation that is what we call the Kohn and Sham equation.

And again, they are written here-- sort of not very different. If you look at it, from the Hartree-Fock equations, there are really the same terms. There is a quantum kinetic energy. And now these are single particle orbitals. There is a Hartree term. There is an external potential term, and the only difference is in the way the exchange or exchange correlation term is calculated.

In Hartree-Fock, it was an explicit integral of the orbital. In density functional theory, it's going to be some kind of complex function of the charge density. And we are going to try and find out what are the eigenstates of this Hamiltonian, of which the most important part that you need to remember is that this Hamiltonian is what we call self-consistent. That is, it's an operator that actually depends on its own solution, because you see, what we have is that the charge density is given by the sum of the square moduli. And the charge density goes into the expression of the Hartree operator, and goes into the expression of the exchange correlation operator.

So the Hamiltonian, acting on this single particle orbitals, depend on the charge density. And the charge density is a function of the orbital themselves. And so the problem has become slightly different from your usual problem. The Schrodinger equation gives you an operator for which you need to find the eigenvectors. Here, you have an operator for which you find eigenvectors, but then these eigenvectors give you a charge density that, put back in here, gives you a different operator.

And so exactly like in Hartree-Fock, you have found your ground state solution only once you have become self-consistent. You have Hamiltonian whose eigenstates give you a charge density that gives you the same Hamiltonian you had calculated the eigenstates for. And so all the computational approaches to solve the density functional problem, as those that solve the Hartree-Fock problem are iterative approaches. We can't just find a solution to an Hamiltonian, but we really need to make that problem self-consistent.

So this is where we concluded in last class. And so now we'll actually go into practical feature of density function theory. So starting from reminding you that all the quality of your calculation depend ultimately on the quality of your functional. And for many years, the only functional that was used was really the local density approximation functional.

In the late '80s and early '90s, people started to develop what are called generalized gradient approximation. That is, they constructed functionals of the charge density that didn't only depend on the charge density itself, but also on its gradient. Not in the sense of a Taylor expansion, because a Taylor expansion wouldn't actually satisfy a number of symmetry properties that we know that the exact exchange correlation functional would do.

And so in this sense, all these new approximation are called generalized gradient approximation. And there is a little menagerie of acronyms and symbols that really are sort of build up upon the names of the few people that have really done a lot of work in this field. And so you'll see a lot John Perdew in this, represented by a P, or Axel Becke in Canada by a B, or Kieron Burke in Rutgers by another B.

So these are some of the most popular functionals. I would say that, by now, sort of almost every one is standardized on PBE as being the most reasonable and the most accurate advanced approximation beyond local density. And so this is what we'll use. There are classes of more complex functionals. The quantum chemistry has done-- community has done a lot of work in looking at hybrid functionals, functional in which a certain percentage of the exchange correlation functional comes from the density function approximation. It could be LDA, or in most cases, could be something like PB.

And some amount of Hartree-Fock exchange is mixed in. So they really take Kohn and Sham equation, in which they exchange correlation term as both a component that is density functional theory-like and a component that is Hartree-Fock-like. And some of them are, like these two mention here, can work very well, especially for molecules. Somehow, again, Hartree-Fock comes from atoms and molecules, and it tends to work better in that limit.

Density functional theory is built on an approximation like the LDA that ultimately comes from the homogeneous electron gas. So it tends to work better for solids, and as usual, the most difficult cases in which you have a combination of the two. If you want to study a molecule on a solid surface, then none of these two approaches work really exceedingly well.

There is a lot of work in developing more complex functionals that tend to work better than any of these that I have mentioned here. They tend to be exceedingly complex. And so again, there are sort of meta-GGA functional-- there is a lot of more complex functional that starts to depend not only on the density and on the gradients, but then maybe on the Laplacian. They might depend on the orbital themselves. So you find terms like exact exchange functional and so on.

There is a lot of current work, and there isn't a unique solution. And also they tend to be so much more computationally expensive that, at this stage, I would say they are still limited to a development effort more than an application effort. But you see, once we have a functional-- and this is even done sort of with standard LDA or TGA, we can actually describe with remarkable accuracy a number of property for very different materials.

This is a table that Chris Pickard in Cambridge had given me, and you see we have a combination of metals, semiconductors, oxides, alloys. And what we have here, say, is a list of what is the experimental lattice parameter, and what is the theoretical prediction. And you see we are sort of in the range in which the errors are around 1%.

And I would say, for most materials, we are really in this ballpark. The error on a lattice parameter of a material that doesn't have exotic electronic properties is not a high-Tc superconductor, is not as strongly correlated electronic material can be expected to be in the range of an error that is between 1% and 2%.

So very, very good. You see predictive power. When we calculate with density functional theory the lattice parameter of this, we are really not trying to fit any potential at all. When we calculate the lattice parameter of silver, we are just saying there is going to be an array of Coulombic potentials attracting the electrons with the atomic number of silver.

And we are looking at how the total energy of this system varies with lattice parameters. And that's why, if you want, electronic structure approaches tend to be extremely powerful, because they are not fitted. We are just trying to find the electronic structure solution.

So how does it work in practice? And what I'm going to describe here is what's become known by now as the total-energy pseudopotential approach. This is really the approach that has been developed to study solid systems. Again, you have to keep in mind these two communities-- one of sort of solid-state extended studies of matter, and one of atoms and molecules. And really density functional theory comes from these solid state approaches. And for reasons that we'll see in a moment that are really related to what we use as a basis set to describe electrons in a solid, what become apparent very early on is that describing accurately the core electron in an atom or in a solid would have been exceedingly complex.

If you think at it, whenever you have an atom that forms a molecule or a solid, you have a lot of electrons that are going to be basically unaffected by the chemical environment. If you take an iron atom, and you put it in iron as a metal, you put as iron in a transition metal oxide, or you put as iron as a center, say, in a heme grouping-- hemoglobin or myoglobin-- what happens is that the valence electrons or iron will redistribute themselves. But really, the core electron of irons are so tightly bound to the nucleus by order of magnitude in energy with respect to the sort of typical energy of valence electrons that they are basically unaffected.

And so in some ways, we really don't want to carry all the computational expense of describing core electrons when we know that their contribution is going to be a rigid contribution that doesn't change depending on the chemical environment. The core electrons are certainly important. They are there. They are bound to the nucleus, and they screen the nucleus charge.

In iron, the 1s electrons-- the two 1s electrons will be so tightly bound to the nucleus that the nucleus doesn't look to the 2s, or 2p, or 3s, 3p electrons as having 26 protons. But it really looks like having 24 protons, because those two 1s electrons screen completely. And so the 2s and 2p really will also spin almost completely the nucleus from the point of view of the 3s and the 3p.

So we want to find out a way of somehow taking into account the presence of that core electrons, but we don't want to carry that on in our calculation, because it's very expensive. Not to mention that the spatial variation of core electrons is going to be extremely sharp. I'll show you an example in a moment.

And so we need a lot of computational information to describe all the sharp wiggles that core electrons will do around the nucleus. And this problem has been solved that again from the late '70s to the early '80s by what are called pseudopotential approaches-- in particular, something that is called-- and you'll see this term often-- norm conserving pseudopotentials that, in many ways, are some of the most complex part of all these electron structure approaches. And I'll just give you a sort generic flavor of how they approach this problem.

Once we have sort of removed the core electrons from our problems, we still need to find out what are the Kohn and Sham orbitals that minimize our density functional problem. And as always-- and again, we'll see this in the next slide-- we need to find an appropriate computational representation of these orbitals. And that is in particular will expand those orbitals on a basis. That is, we'll represent any possible orbital in a problem as a linear combination of simple function. This is what is called a basis set.

So in our one dimensional analysis problem, when you have got an arbitrary function, you can expand-- you can do, say, Fourier analysis. You can expand it in a series of sines and cosines. So the same general problem applies here. We have arbitrary functions in three dimensions that are our orbitals, and we want to find an appropriate basis set that describe them. So this basis set needs to be accurate-- that this, needs to be flexible enough to describe all the possible wiggles of our orbitals, but also need to be computationally convenient. And we'll go in that in a moment.

So once we have sort of put together these elements, we have sort of our external potential being represented by this set of nuclei, and somehow, via the pseudopotential approach, also the core electrons. And once we have decided what is our basic set, we really need to go and solve self-consistently either the Kohn and Sham equation that you have seen before, or we want to minimize the nonlinear energy functional.

And so there are sort of several steps. Usually, what we'll do is we'll start from a trial solution. Remember that, since the Hamiltonian depends on the charge density, we need to have a guess to our initial charge density to even construct our operator, because our operator depends on the ground state charge density. So since that we don't have the ground state charge density when we start our problem, we need to start with a trial solution.

It could be a trial charge density. Could be trial orbitals. But once we have that, all our operator is well defined. So we can calculate all the different terms-- the quantum kinetic energy, the Hartree energy, the exchange correlation terms. And we can try to solve that Hamiltonian from that Hamiltonian.

We'll find new orbitals, and with those new orbitals, we'll calculate the ground state charge density, and we'll obtain a new Hamiltonian that then will sort of keep iterating, finding new orbitals, new charge density, new Hamiltonian, until it reaches a fixed point, until we reach self-consistency. Or we can just take the approach of minimizing the total energy functional to self-consistency. I've written in a more compact way the problem here. Well, it'll come in a moment.

So let's describe first our first computational approximation in this problem-- that is, the introduction of pseudopotentials. And as I said, what we want to do is we want to get rid of the electrons in the inner shells, in the core of the atoms, because they really don't have any contribution to the valence chemical bonds.

They are just there. They're important because they screen the nucleus in a very specific quantum mechanical term, but they really don't change much when the valence coordination changes. We call that, actually, a frozen core approximation. That is, we take an atom-- again, we take an iron atom. We solve the density functional problem for the iron atom. That is, we find a density functional orbital for the 1s, 2s, 2p, 3s, 3p, and so on orbitals.

For an atom, it's reasonable to do that [INAUDIBLE] solution, even for the core electrons, because you have spherical symmetry. And the problem is still sort of fairly simple. And at this point, we say, well, from now on, we are not going to describe the full iron atom, but we are going always to deal with a pseudo iron atom in which the nucleus and the 1s, 2s, and 2p electrons have been frozen.

And so what really the valence electron need to see is a pseudo nucleus that is not given just by the bare Coulombic potential, but by the bare Coulombic potential screened by this 1s, and 2s, and 2p electron frozen in their atomic configuration. And there is a theorem that we won't dwell into by-- from Barth and Gelatt from the '80s-- that says that this freezing of the core electrons is actually a very good approximation. It doesn't really matter. It's been verified over and over again.

It is the last thing we have to worry about. To be precise, the only thing that we have to worry about in density functional theory, apart from making sure that our technical approximations-- computation approximation are all accurate is the exchange correlation functional. That is truly the only source of error.

And so here, if you want, I've represented the idea of a pseudo potential. This would have been the standard solution for an atom. Actually, here I've chosen something simpler than iron. I've chosen aluminum.

And so we would have had a Kohn and Sham set of equations for aluminum in which, again, this term here contains the Hartree potential, the exchange correlation potential, and external potential that would be just the bare nuclear potential. We solve this problem. What we find?

Well, we find this hierarchy of states. You see, we find two electrons in the 1s state. We find two electrons in the 2s, two electrons in the 2p. But really, the electrons that do all the chemistry are the 3s and 3p electrons. And you see the enormous difference in energy scales.

So the binding energy of the 1s electrons to the nucleus is 1,500 electron volts. There is no way, unless you are sort of throwing X-rays of very high energy to those electrons, that you are going to affect or perturb in any significant way these 1s electrons. Remember, the energy of a hydrogen bond is 0.29 electrons-- is a fraction of an electronvolt. So this is four orders of magnitude larger.

So what we want is we want to say this set of electrons are so tightly bound that we don't need to consider how they change during the formation of a chemical bond. And so what we want to find is a new potential that we call our pseudopotential, such that the eigenstates in the presence of this pseudopotential give us solutions that really reproduce exactly the valence electron solution of the original atom.

So you see, what we want is we want to go from $13/r$, the Coulombic potential that is contained in here for the bare aluminum nucleus, to a new potential that somehow contains both the bare potential of the nucleus, the screening from the core electrons, and it is constructed according to a well-defined prescription so that the eigenstates of this new set of Kohn and Sham equation are in order. The lowest eigenstates are actually the valence eigenstates of our regional problem.

We get the same eigenvalues, and we get, in a way that I'll specify in a moment, the same eigenfunctions. And so here, it's how we would actually look at this problem. And this is a figure courtesy of Chris [INAUDIBLE].

So you see, what we usually ever in an atom is a Coulombic potential, is this thin red line here. This is the potential as a function of the radial distance. So it diverges. It goes to minus infinity Coulombically. For aluminum, it would go as $13/r$.

What are the solutions, say, for the 1s state in this-- sorry, what would be the solution for a valence electron in this potential? Well, there are going to be all the sort of core electrons, but then a valence electron would look something like this. It has-- you see this enormous number of oscillations. The reason why those oscillations are there is that eigenfunction of the Kohn and Sham equation need to be orthogonal to each other.

This is another one of these sort of fundamental quantum mechanical rules. It's basically the Pauli principles. You can't have two electrons in the same quantum states. And in particular, electrons corresponding to different quantum states need to be orthogonal to each other. So the integral of the ψ^* of the 3s electron in aluminum times the ψ of the 1s electron in aluminum needs to be 0.

And so what it means is that the higher you go, the more wiggles you have. And these wiggles-- that is, these changes of signs is what allows you orthogonality. So if you have a 1s electron that looks like this, a 1s electron will have actually sort of exponential to the minus r decaying wave function.

The wave function say-- already the 2s wave function needs to be orthogonal to this electron, and so it needs to change sign to allow orthogonality when you take the product of these two orbitals. So orthogonality creates a lot of wiggles, and these wiggles also make the charge density coming from a valence electron more spread towards the outside of the nucleus.

There is, if you want, another quantum mechanical derived effect in which valence electrons are moved outwards because of this orthogonality constraint. And so you see this would be, in light blue, the exact wave function for a valence electron. And what we want to find is a pseudopotential that substitutes for the original bare Coulombic potential-- and it's written here in sort of-- with a thick red line-- that is constructed so that its ground state wave function-- that is here in a thick blue-- has the same energy eigenvalue of the valence electron. And in many ways, it's identical to the original wave function.

And the way we make it identical is actually we only require it to be identical to our original solution outside the core, because outside of the core of the atom is where chemistry takes place. Chemistry will be here when atoms bind with other atoms. Inside here, this is the region where core electrons sit, and that's a region that will never overlap with other atoms, again, because there are sort of already these very strongly bound electrons. And to bring two atoms to have their core regions to overlap requires enormous pressure, hundreds of gigapascals. So we never really have to worry.

So what we are trying to do here is we want to construct a pseudopotential in thick red such that its ground state wave functions in thick blue have the same eigenvalues of my original eigenfunctions in thin blue lines. And in many ways, they represent the same physics and the same chemistry, because they are sort of identical outside the core. So the thick red line is how we describe an atom. The bare Coulombic potential-- but we also built all the screening action from the core electrons, and you see the central feature of the pseudopotential is that, if you want, it becomes repulsive around the center of the atom because we really need to keep the valence electron to be mostly delocalized in the valence region.

How to build this pseudopotential is actually a complex and somehow dark art. By now, it's been perfected, and that-- there are basically tables of pseudopotential. So it's a problem you almost never have to worry anymore. An electronic structure code will have a set of pseudopotential.

And in the course of the years, there have been two flavors that have been developed. The first one that was actually, I would say, invented at Bell Labs at the end of the '70s by Hamann, Schluter, and Chiang is what are called norm conserving pseudopotential, of which I'll describe this plot in a moment. There was a sort of improvement of this norm conserving pseudopotential that are called ultrasoft pseudopotential, developed by David Vanderbilt at Rutgers University.

They tend to be much less expensive to be-- to use in practical calculation, basically because they are smoother and smoother, and so they can be described with sort of a smaller basis set, and they tend to be more accurate over a broader range of coordination and energies. But so the general idea of the pseudopotential from a different way-- in the previous slide, I showed you how it looks like.

But the genetic idea of a pseudopotential is that we want to create something that contains both the nucleus and these core electrons. And that basically will act on the valence electron in the same way as the true problem of core electron explicitly and nucleus. And so what we say is that we want a pseudo potential to scatter an incoming wave, an incoming electron in the same way as our real set of core electrons in an atom would do that.

And in order to do that very accurately, for this pseudopotential, what we need to do is actually to have the pseudopotential not just be a single local form, as in the previous slide. It's shown you the sort of red thick line that was just a unique form of the potential as a function of r , but we actually want our pseudopotential to be different, depending on the angular momentum component of the electron coming.

So if you have a valence electron, and you have a valence wave function, you can actually say that that valence wave function will have a 20% angular symmetry of the S type, 30% of the P type, 30% of the D type, and so on and so forth. And what you do is you act differently on the S component of the wave function, on the P component of the wave function, on the D component, and so on and so forth.

And so this would be the different radial parts of the pseudopotential, depending on which components they act on. They all have this overall $1/r$ or $1/z$ asymptotic trend. But as we get more close to the nucleus, they start differing. You see, from the core region inwards, they are all different. And in particular, they are all, or at least in part, repulsive to again represent this frozen set of electrons.

And this is how the corresponding wave functions would look like. And I've made-- again, this is for an indium atom, a comparison between, say, the p all electron wave function and the pseudo wave function. And you see they are identical in the valence region, but in the core region, the true all electron valence wave function is a lot of oscillation, while the pseudo wave function is much smoother.

And this smoothness means that we can describe this pseudo wave function in sines and cosines with a much smaller set of waves at different wavelengths than we would have to do if we were to-- if we had the need to describe all this high frequency oscillation. So at the end, pseudopotential are just there to sort of give us a chance to avoid describing all these high frequency oscillation of the true valence wave function, and also to give us the possibility of avoiding the explicit description of the core electrons that really don't change their shape in going from the atom to the solid.

So now with this, we have an exchange correlation functional. We have this new approximation to the external potential given by a sort of array of pseudopotential centered on each nucleus. And so we are back to this sort of self-consistent problem of having to solve these Kohn and Sham differential equations. And again, the way we want to attack this problem on a computer is by making this problem into a numerical matrix algebra problem.

And the way we go about this is by choosing a busy set and expanding our orbitals in a basis. That is, we say by saying that any sort of generating function can be thought of as a linear combination of a set of simple function. Simple function, for which we can do analytical operation-- that is, we are able, say, to calculate the matrix elements, say, of the kinetic energy between this simple function, or we are able to calculate analytically the second derivative of this basis function.

Well, by making this ansatz, by saying that a generic wave function is a linear combination of this, our problem transforms from a differential analysis problem to a problem of finding these coefficients. And again, sort of the basis set that we'll choose in all sort of our practical computational lab will really be a basis set of sines and cosines, or what we call plane waves. That is especially adapted and especially useful for the case of solids.

And I've given you an example here. Suppose that what we want to describe is a Gaussian charge density, something that is a little bit similar to the charge density in an atom. Well, this localized charge density can be actually expressed fairly accurately as a linear combination of very few sines or cosines with different wavelengths. You see, I go from this wavelength to smaller and smaller wavelength.

And if I choose my coefficients such that I have constructive interference here in the center and destructive interference outside, I actually get just by something nine terms here, something that basically describes perfectly my charge density. So this will be our general sort of approach to the problem. We choose a basis set. That is, we choose a set of elementary functions that are particularly suited to describe our problem, and for which we can do all the analysis that we want. That is, we are able to calculate second derivatives.

We are able to calculate matrix elements analytically. And then our computational problem becomes really just a problem of finding the coefficients. And this is one of those approximations for which you need to test the accuracy. That is, if you use a very small basis set, if you use only 10 plane waves with different wavelengths, your calculation will be very inexpensive, because you have only 10 coefficients to work with. But it probably won't be very accurate. If you use one million plane waves, your calculation will be extremely accurate, but it will be also very expensive.

So this is an approximation, but an approximation-- and this is fundamental-- for which you can control the accuracy. That is, you just need to make sure that your basis set is good enough. And that is something that you can always check and test. You can't test, apart from comparison with experiment, if your exchange correlation functional is good enough. That's why we put so much importance, and you should never forget that really all the accuracy or all the errors that ultimately come in your calculation come from the exchange correlation function.

All the other computational numerical approximation using pseudopotentials, using a finite basis set are all approximations that can be tested. And we always assume that a well done electronic structure calculation will have those under control. You are using enough basis sets. You are using accurate pseudopotential, and so on.

As I said, in solids, for a reason that you'll see in a moment, really the universal basis of choice is that of plane waves-- basically the $e^{i\mathbf{k}\cdot\mathbf{r}}$ complex exponential, and you'll see why. But in a sort of-- atoms and molecules-- their almost universal basis choice is that a combination of Gaussians. And the reason why Gaussians have been chosen, and actually why the most famous quantum chemistry code is called Gaussian is that it's very easy to do the analytical integrals that are present in the exchange term for Hartree-Fock.

So if you want an exchange term for Hartree-Fock, the term means the choice of Gaussian as a particularly convenient basis set. And again, the product of two Gaussians is a Gaussian, and that's why this is ultimately a good choice. This is what has developed. At the end, you can still use Gaussians to study solids, or you can use plane waves to study molecules. And there is even a sort of richer phenomenology. But generally speaking, these two approaches, that of localized functions in chemistry and of delocalized function in extended systems, are the basic categories.

We'll start from the point of view of solids, and then you'll see why we actually describe-- we use plane waves to describe these orbitals. And when you deal with solid, you need to be aware of one of the fundamental symmetries that the eigenstates of our solution in our periodic potential, as what happens in solids, satisfy. That is, in particular, the Hamiltonian of a solid is periodically invariant. That's the characteristic of a solid. It's a regular periodic array of external pseudopotential that [INAUDIBLE].

So in the language of quantum mechanics, we say that the Hamiltonian commutes with the translational operator. Translation operator by a vector-- that is, a direct lattice vector. So if you take out cubic system and you displace, you translate your set of potentials by 1 cubic lattice vector, you have the same operator. That is, trivially to say that, if you have a regular array of blue dots that is infinite in space, and you displace them, and the displaced one \mathbf{r} crosses, and they have been displaced by this lattice vector \mathbf{r} here, and both the red dots and the blue dots are infinite, well, by displacing the blue dot into the red dots, you have the same identical problem.

When two operator commuting quantum mechanics, what we know is that we can find eigenstates that are common eigenstates. That is, they are eigenstates both of the first operator, and then on the second operator. I won't dwell on this in case you are not familiar with parts of this quantum mechanical problem. But what I want to highlight is that the net result, the fact that we can choose eigenvectors that are common eigenvectors of the Hamiltonians and of the translation operator, tells us that a genetic eigenvector for aperiodic potential needs to have a very well defined symmetry and a very well defined symmetric form. And that's sort of summarized in the Bloch theorem.

And that tells us that an eigenvector will be given by a term-- a function that has the same periodicity of the crystal times a plane wave-- again, a complex exponential that modulates these periodic parts and that can have any wavelength. So in other words, what the Bloch theorem here is telling us is that the solutions, the eigenstates of aperiodic Hamiltonian can have any possible periodicity.

You don't have to think at the electronic eigenstates of aperiodic Hamiltonian as having the same periodicity of your lattice. But the shape that they have is actually this. That is the product of two function-- one that has any arbitrary wavelength, and another term-- the so-called periodic part of our orbitals-- that does have the same periodicity of the lattice.

This can be proven in a variety of ways, and somehow I summarized here why ultimately this is the shape that the eigenfunction of the Hamiltonian needs to have, basically because if we have a translation and the Hamiltonian is invariant, the charge density of your solid must also be invariant. And this solution satisfies that, because if you look at the wave function-- not at r , but r plus a direct lattice vector-- this term will be unchanged by definition because this term has the periodicity. This term will become $e^{i\mathbf{k} \cdot \mathbf{r}}$ plus a direct lattice vector.

But when you look at the charge density-- that is, the square modulus-- this term disappears. So only this remains in the charge density. And you also want two translations being equivalent to the sum of the other two. And again, if you translate that by two consecutive lattice vectors, this term doesn't change. And the exponential of a sum of two lattice vectors is going to be just the product of two exponentials for each lattice vector.

But the fundamental concept here is this-- that the wave function are modulated by this wavelength and we actually classify them. We do these two quantum numbers. As we classify orbitals in the periodic table, we give them quantum numbers. We say this is going to be a 3s electron with spin-up.

We are saying what the quantum number n, l, m and spin are for that electron. In a solid, these are our quantum numbers. We have a band index. That is, if you wanted solid equivalent of your energy levels, and it's actually a discrete integer index that somehow classifies the different eigenvectors, but those eigenvectors can have also any wavelength.

And so the wavelength of the block orbital is another quantum number that characterize our system. So when you actually look at a molecule, you think that in terms of energy levels, and you draw an energy diagram in which you will have a 1s, 2s, 3s energy level. When you think of the possible energy levels from four electrons in a solid, you'll draw a more complex band diagram in which you will sort of classify levels, not only in terms of discrete integers, but also in terms of the wavelength.

And so this is what is usually band energy diagram, of which I have an example here. So this is-- would be the solid equivalent of the energy levels. So for a molecule, you would have a discrete set of energy level. For a solid, you don't have only a set of discrete quantum numbers n , but you have also a continuous set of wavelengths. So you have bands that show you how your energy of that eigenvector changes depending on the wavelength of that eigenvector.

And what I've plotted here is actually the band energy diagram for a free electron in an FCC lattice. What I'm saying is that, suppose that I'm considering an electron that doesn't feel any potential. What would be its band energy? Well, its band energy is actually going to be a parabola.

So it's something that we would represent as this sort of energy, as being really proportional to its wavelength. That is, again, because a free electron is represented by a plain wave $e^{i(kr - \omega t)}$. And so the second derivative of a plain wave-- that is, the quantum kinetic energy-- is just the k squared term. And if an electron is free, there is no potential.

So we can think of the energy as a function of wavelength for a free electron to be perfectly parabolic. But now what we can think for a moment is also how to represent this parabola. In principle, this parabola is a single parabola infinite as a function of wavelength. But we can also, for a moment, represent the data in an FCC or in any arbitrary periodic lattice.

So what we could be saying is that suppose that now there is an imaginary geometric structure that is, say, in real space Bravais lattice with FCC periodicity, and that, in reciprocal space, will correspond to a BCC reciprocal lattice. And what I want to do is I want to take this parabola and fold it. So instead of having a single parabolic branch that extends to infinity, every time I hit in Brillouin zone in the reciprocal space a boundary, I fold this parabola back.

So what I've shown here is really the band energy diagram that is a parabola, supposing for a moment that my electron leaves in the geometry of an FCC lattice. And this is just to give you the feeling that sometimes something that looks very complex is just a representation of something very simple in a specific geometry. And now you see in a moment what happens if we study the electron in a system that truly has the FCC Bravais lattice in real space. And so it is a BCC Bravais lattice in reciprocal space.

And that is silicon. And you see that the band energy dispersions for electrons in silicon looked, actually very, very similar to free electrons. You see now for a moment something that you thought it was always extremely complex, like the band diagram for silicon, turns out to be just a parabola, just a free electron slightly modified. And you can clearly see all the same pieces, but obviously now these electrons, these valence electrons in silicon feel the periodic potential.

And so the overall solution is modified. And for those of you familiar with solids, there is a band gap between valence electrons and conduction electrons. And here, we would be at a gamma point. This would be the lowest valence band, and here we would have the heavy and the light whole bands for the top valence-- for the top valence bands.

But what this is telling us is that electrons in a solid really look like free electrons with a little bit of a perturbation that is enough to transform this parabola into something different. And again, because they are almost free electrons, and because they have the periodicity-- the periodic part of the Bloch orbitals need to have the periodicity of the crystal lattice that we are studying-- it will actually be extremely appropriate to use as a basis set to describe at each k point in the Brillouin. The periodic part of the Bloch orbitals will be extremely convenient to use plane waves.

And this is the reason, ultimately, why I written here, sort of again reminding some basics of crystallography. I've taken, say, the zinc blende structure. So again, an FCC lattice here, where these three here in green would be the primitive lattice vectors of the FCC lattice-- so what I am calling here a_1 , a_2 and a_3 .

And these three lattices-- three vector lattices in $1/2, 1/2, 0, 0, 1/2, 1/2$, and $1/2, 0, 1/2$ really represents my Bravais lattice by repeating blue atoms at any linear combination of three-- of these three vectors. I span all the infinite blue sublattice of the cations in a zinc blende. And the zinc blende is just a parallel compound of silicon. It's just that the second atom in the unit cell is different.

And again, by applying translations that are linear combination of these green vectors, I span all the blue lattice. And silicon would be identical to this. Just the red and the blue would be identical. So having defined direct vector lattices, the a_1 , a_2 , and a_3 , what I can define is a set of dual vectors that we call reciprocal lattice vectors g_1 , g_2 , and g_3 that are such that the scalar product between g_1 and g_2 is either 1 or 0. And this is the [INAUDIBLE] δ times 2π .

This is a definition. This is what we know is the definition of the reciprocal lattice vector, but what is very important for us is that these primitive reciprocal lattice vectors are the fundamental descriptor of all the plane waves that I want to use in describing a solid and the reason for that is that a plane wave written like this exponential of iGr , where the G here, where this vector is just a linear combination with integer number of my primitive lattice vectors-- so it's going to be l with l integer G_1 times G_2 to times n G_3 . So well defined by the triplet of numbers l, m, n .

Well, these plane waves, e to the iGr , for any linear combination of g_1 , g_2 , and g_3 is a function, defined in real space, that is a periodicity compatible with the periodicity of my direct lattice. So suppose that we were in one dimension for a moment, and I would have a system that has this periodicity. And remember, the periodic part of the Bloch orbitals will have the same periodicity in real space.

So maybe the periodic part of that will look something like this, with the same periodicity. Well, then what I want to do is really expand this periodic function as a linear combination with appropriate coefficient of plane waves with different wavelengths, but I need to choose these plane waves to all have the same periodicity of the green lattice.

So my plane waves will only look like this, and this will be the sort of fundamental harmonic if you want. And then another plane wave could have this periodicity, and so on and so forth, going to higher and higher wavelength. But the compact mathematical representation of this is in here. That is, given a direct lattice defined by the three vectors a_1 , a_2 , and a_3 , I can define a reciprocal lattice that is represented by the three vector G_1 , G_2 , and G_3 . And any linear combination with integer numbers of this G_1 , G_2 , and G_3 will give me a vector G such that the complex exponential e to the iGr is actually a periodicity that is compatible with my direct lattice.

That is, the moment I create a direct lattice, there are an infinite number of wavelengths that are not compatible with the periodicity that I've thrown into my system. Those disappear. I don't want to have anything to deal with. What I'm left is with a numerable infinity of plane waves that are all compatible with the direct lattice that I'm choosing from.

And you see now there is a natural way of both choosing plane waves-- they will be here-- and also a natural way to choose the most important one, because when we choose our basis set for our calculation, we'll start from the plane waves of longer wavelength. That is, that the blue wavelength that I've plotted here-- and will decrease the wavelength. That will include more and more plane wave that sort of have finer and finer wavelength, finer and finer resolution.

And that can be just naturally done by including plane waves with G vectors that have a larger square modulus. That is the larger this G vector here, the finer the resolution, the higher the wavelength of this plane wave. So there is truly a natural way to choose basis set now, because, again, what we have is a reciprocal space. And so we'll have a Brillouin zone in red.

We'll have, if I'm in two dimension, the two G_1 and G_2 vectors. And so the linear combination G's that I've described will be given by the infinite, but discrete set of G vectors represented here. And so only the G vectors denoted here with a cross give rise to a plane wave that has the compatible periodicity with my periodic boundary conditions.

And so in order to choose a basis set, what I do, I just draw a circle, or in three dimensions, a sphere, and I decide to include all the vector that sit inside that sphere. And by making that sphere larger and larger, I'm going to include G vectors that have larger and larger modules that is finer and finer wavelength.

And that is what is called your cutoff in your basis set. So one of the fundamental parameters of your calculation will be choosing your cutoff for your plane wave basis set-- that is, choosing the radius of that sphere that includes plane waves with wavelength compatible to the periodic boundary conditions up to a certain resolution. You make this cutoff sphere larger and larger. You systematically include finer and finer resolutions, so you become better and better at describing your problem.

And you'll always need to check in your first calculation that your basis set is good enough. That is, by increasing your cutoff radius, you will not see any significant physical change in the quantities that you are calculating. You don't see a physical change in the energy as a function of volume. You don't see a change in the forces acting on atoms, and so on and so forth.

This will take forever. Again, this is not the only choice of basis set, but as you can see, it's really the appropriate one for periodic systems. And it has sort of a set of nice advantages, of which I would say the most important is the fact that it's systematic in the sense that you can continuously improve the resolution in your problem by including more plane waves.

There is a negative side to this that says that it doesn't have any information on how valence electrons should look like close to a nucleus or close to a core. It requires a very large number of basis elements. So plane wave calculation tend to have a lot of elements in that. There are a number of other important practical consequences. In particular, as much as Gaussians, they allow for very easy evaluation of some of the analytical terms that we need.

So it's very easy to take the gradient of a plane wave or the Laplacian of a plane wave, because basically the derivative of a complex exponential, if you have $e^{i\mathbf{G}\cdot\mathbf{r}}$, the derivative with respect to \mathbf{r} of this is just the function itself times the vector \mathbf{G} . This would be the gradient, and the Laplacian-- that is, the second derivative-- is just G^2 times the wave function.

And so all these terms are easy to calculate, and there is a sort of more subtle conclusion of this, that if you start to have a calculation in which atoms move, like a molecular dynamic calculation in which you need to calculate things like forces, you need to calculate the derivative of the energy with respect to the position of an atom. But now the energy in itself is an expression that involves linear combination of your basis set.

Well, this basis set does not depend on the position of the atom. So there is no term in the force in the derivative of the energy with respect to the position that comes from this. If, on the other hand, you are using, say, a Gaussian that was centered on an atom, a Gaussian centered on an atom would have in there the position of the atom. And so in order to take the derivative, the force, you would also need to take a derivative of your basis set. And these are what are called Pulay terms. And that just add another layer of complexity to your calculation.

And again, this is, I think, ultimately, is one of the reasons why in the solid-state community problems like molecular dynamics and forced relaxation were developed earlier, basically because it's much simpler to calculate forces. And I want to give you an example on how certain parts of your problem become very easy in your solution. In particular, remember sort of in the construction of the self-consistent operator from the charge density we needed to construct the Hartree operator from the charge density integral of n over \mathbf{r} minus \mathbf{r}' .

Well, that's basically a solution of what is called the Poisson equation. That is, given a charge density, the Hartree potential is really coming from the solution of this differential equation. That is, the Laplacian, the second derivative of the Hartree potential is equal to minus, in atomic units, $4\pi n$ the n electronic charge density taken as positive here in particular, just to get the signs right.

This is a differential equation. We need to solve it in the course of our self-consistent problem, because remember, we have sort of diagonalized Hamiltonian. We have gotten some orbitals. From those orbitals, we have calculated a charge density. But now, from the charge density, we need to calculate the electrostatic potential. And that's how it works.

Well, this differential equation is actually trivial to solve if, for a moment, you think at a plane wave solution. That is, you think that your potential-- that is, a function of \mathbf{r} is now being written as a linear combination where the coefficients are called v of g of plane waves. So I'm taking my potential. It's going to have the same periodicity of the reciprocal lattice, and I'm writing it out as a linear combination of waves. And I'm doing the same for the charge density here.

So the charge density in itself is going to be given by a linear combination with coefficient G of plane waves. So this is my expansion in plane waves of this real space functions. And then the algebra is trivial, because to take the derivative of this red term here, what I obtain is nothing else than the sum over g . Second derivative will give me $\mathbf{r} - \mathbf{g}^2 v$ Hartree of $G e^{i\mathbf{G}\cdot\mathbf{r}}$. And that is-- needs to be equal to, well, minus 4π sum over G n of $G e^{i\mathbf{G}\cdot\mathbf{r}}$.

So I have done nothing else than inserting the explicit expansion in plane waves of my potential of my charge density in the differential equation. And now, well, sort of mathematically I can actually sort of multiply the left and the right hand term for something like $e^{-i\mathbf{G} \cdot \mathbf{r}}$. And then I can integrate in the \mathbf{r} .

And this is just the mathematical operation that on an orthonormal basis set, as this is, tells me that, in order for this equality to be satisfied, what I really need to have is that each coefficient corresponding to the same \mathbf{G} vector on the left hand side and on the right hand side is separately equal. So the solution to the Poisson-Boltzmann equation is trivial, because what I need to have is that $\mathbf{G}^2 \phi_{\mathbf{G}} = 4\pi n_{\mathbf{G}}$. That is, each coefficient needs to be identical $\phi_{\mathbf{G}} = 4\pi n_{\mathbf{G}} / \mathbf{G}^2$.

So if I have a charge density in real space, I can expand it in plane waves. And this coefficient and nothing else than a Fourier transform, so a computer is very good, given a periodic function in real space, to give me these coefficients. And once I have this coefficient $n_{\mathbf{G}}$ of \mathbf{G} , I will know instantly what are the coefficients of the potential that is the solution of this Poisson equation, because those $\phi_{\mathbf{G}}$ coefficients are really just my charge density coefficient multiplied by 4π and divided by \mathbf{G}^2 .

So you see with plane waves, a lot of the analytical work in a transaction problem becomes trivial. And here, it's contained one of the trickiest part of electronic structure calculation. And the reason why electronic structure calculation of large system becomes more and more computationally ill-defined, and more and more difficult to do as the size of the system becomes larger, because as the size of the system becomes larger, your cell in real space will become larger. And your Brillouin zone in reciprocal space will become smaller and smaller.

As the three a_1, a_2, a_3 vectors becomes larger, the three dual reciprocal space vector-- $G_1, G_2, \text{ and } G_3$ -- becomes smaller. So the smallest of the \mathbf{G} vector will become smaller and smaller as the supercell grows. And so then you see that there is an instability, because this component of the Hartree potential is given by this $n_{\mathbf{G}}$ coefficient divided by \mathbf{G}^2 .

As your system becomes larger, \mathbf{G}^2 becomes smaller. And so as smaller numerical instability in your charge density gives you a large instability in your potential. And so you need to be more and more careful. You need to become more and more careful in your electrostatic solution as the problem becomes larger and larger.

This will take forever to say. I think that I need to stop here for today, and what we'll see in the next class on Thursday will sort of wrap up the last few technical details that you need to go into the computational lab on Tuesday, and we'll start discussing case studies for our density functional problem.

As a reminder-- and I've emailed and hope all of you have received my email-- in order to do the computational lab on Tuesday, it is essential that each and every one of you has an independent personal computer account on the computer class, because this calculation will become expensive. And we need to run them not on the main a central node. That will collapse if more than two or three of you run on it. But we need to spool it via queuing system on the nodes of the cluster that lie beneath the master nose.

This is all for today. Enjoy the snow, and see you on Thursday morning.