

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseware continue to offer high-quality, educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseware at ocw.mit.edu.

**TROY VAN
VOORHIS:**

All right. Well, good morning, everyone. Hope you had some at least rest over the long weekend. Looking forward to getting back to talking about quantum. So last time, which was a whole week ago-- long time ago-- we started talking about electronic structure. We talked about the Born Oppenheimer approximation and how that gives us an electronic Schrodinger equation, how we look for-- what we're looking for are the eigenvalues and eigenvectors of that Schrodinger equation, how we can get lots of interesting information out of that.

And then we broke it down so that we saw that there were going to be two main knobs that we were going to have to turn when we were making our approximations here. We were going out to look at our atomic orbital basis sets and how we compute the energy. And we spent the last half of class last time talking about choosing basis sets. We came up with some understanding of these strange names that are written at the bottom here.

And then today, we're going to talk about some of these ways of computing the energy. And then I'm going to flip over, and we're going to talk about computational chemistry tools and how we sort of assemble that into doing a calculation.

But again, the idea with basis sets was that we were sort of questing after this mythical complete basis set limit where the basis set was so big that it was infinitely big, that making it bigger didn't change your answer at all. And we never actually get there. But if we get close enough, we decide that's good.

All right. So now we're going to talk today about computing the energy. And thankfully, most of the methods for computing the energy start from something that I believe you are familiar with already from 5.61. You guys talked about the Hartree-Fock approximation? Yeah? Yeah. OK. So-- or maybe we talked about-- did we talk about singled did we talk about determinants? Yes. All right. So there we go.

So then we talked about a determinant. The idea of a determinant is that you have a set of orbitals, they're independent, and then you compute the average energy. And so, in one form

or another, I would write the energy of a determinant something like this. I would say that there's a piece here which is the energy of each electron by itself. So these are one-electron energies. So for every electron in my system, if I have n electrons, I add up the energies of each of those electrons.

But then, because my Hamiltonian has electron-electron repulsion terms, I end up with some additional contributions that can't be assigned to just one electron or another electron. And those terms are here in this sum. They're pairwise so for every μ and ν summing over independent pairs, I have a coulomb, term j , and an exchange, term k . And what these things show me is that there is an average repulsion here between the electrons. The electrons are independent. They can't avoid each other. But they still interact. There's still the electron-electron repulsion between electrons here.

So that's an energy expression that, in one form or another, should be somewhat familiar, I think, from things you guys have done before. The idea of Hartree-Fock is relatively simple, which is we have an energy, then. This is true for any Slater determinant, any set of orbitals. So I have some orbitals here. And the idea of Hartree-Fock is to choose the orbitals to minimize the energy. This independent particle model energy, I want to minimize that. I'm going to choose the orbitals that minimize that.

And the way I'm going to do that is I'm going to write my orbitals-- ψ_μ -- as a sum over some index, i , some coefficients $c_{i\mu}$, times my basis functions, ϕ_i . And I'll just emphasize that what I have here are functions from my AO basis, the things that I talked about last time, like the actual 1s-like, 2s-like, 2p-like, 3p-like orbitals. And then on the other side, what I have here are my molecular-like orbitals. The molecular orbitals are linear combinations of my atomic orbital basis functions. So that's the general idea of Hartree-Fock there, is that we're just going to choose these orbitals to minimize the energy.

And because the orbitals are defined by these coefficients, I can think of this energy as a function of the coefficients, c . So I can say, all right, that means that I have an independent particle model energy which depends on these coefficients, c . And if I want to minimize that energy, I then need to look for the places where the derivatives are equal to zero. So I take the derivative of that energy, set it equal to 0.

Now as you might suspect, if I look at this expression, I have an orbital here, here, here, here, here, here, here, here, here, here. I have a lot of orbitals appearing in there. And then I'm

going to have to use the chain rule, because every one of those orbitals depends on every one of the coefficients. So I'm going to have a chain rule expression.

You might expect that there's going to be a lot of algebra involved in taking this derivative, and there is. This is where I get to use my favorite abbreviation, which is ASA. ASA stands for After Some Algebra. It means I'm not going to work through the algebra. There's some algebra. There's several pages of lecture notes where algebra is gone through. After you do the algebra, you end up with the left-hand equation reducing to an equation that looks like this.

So there is a matrix, which I will call h , and then when I dot that matrix into a vector, c , the coefficient vector, I get some energy, e , times the coefficient vector again. So that takes a lot of algebra to get to that, but you can get an equation that looks like this. And that's encouraging, because this is familiar to me. This is an eigenvalue equation.

And that's nice because I remember eigenvalue equations from-- I mean, the Schrodinger equation is an eigenvalue equation. So it's nice that when I do this minimization, I get back an eigenvalue equation. That's nice. The one complication is that the Hamiltonian here depends on the coefficient, c . So it's some effective Hamiltonian. I'm intentionally collecting terms so that it looks like a Schrodinger equation, but it's a little bit funny because the Hamiltonian here depends on the coefficients that solve the equation. So this is sort of an eigenvalue equation.

And so what we're seeing here is that H depends on the wave function somehow. It seems like, depending on what the wave function is, I have different Hamiltonians. And to understand why this is the case, I want us to think about the potential, v , that an electron feels in an atom or in a molecule. And there's generally going to be two pieces here. And they're actually-- I can color-code them with the terms that exist in the independent particle expression. So there is the one-electron terms, the terms in the energy that just come from one electron. And those are going to be things like, well, I might have a nucleus of charge plus z . And I might have my electron out here, the one that I'm interested in. And it's going to have some attraction to that nucleus. So that just involves one electron. It involves a nucleus, but just one electron. So there's some nuclear potential that each electron is going to feel.

But then there's the term that correspond to these average repulsion terms. So there are other electrons in my system. And each of those electrons interacts with the electron that I'm looking at. So there's going to be some electron-electron repulsion term here. And that electron-electron repulsion depends on where the other electrons are. And those other electrons have

some wave function. They're kind of spread out here, all over the place. And depending on how spread out those other electrons are, I might have more or less electron repulsion. So if they're very, very concentrated in the region that I'm looking at, the electron repulsion is going to be very large. If they're very spread out, and very diffuse, they're not going to have very much electron repulsion. But overall, the electron-electron repulsion depends on the wave functions that I've chosen.

So the potential here is called a mean field, or average field. And it's because it's the potential that comes from the average repulsion. It's the effective potential due to the averaging, the smearing out, over all the other electrons. So what we do have, then, is that the Hamiltonian does indeed depend on the wave functions, because the wave functions determine that average potential. But then if I know the Hamiltonian, I can solve for the wave functions, because the wave functions are the eigenfunctions of the Hamiltonian. And so I have this sort of chicken and egg situation, that the Hamiltonian defines the wave functions, and the wave functions define the Hamiltonian. And the solution to this is to do what are called self-consistent field iterations.

And you can more or less think about these as the simple process of, guess some wave functions, build the Hamiltonian, get some new wave functions, build a new Hamiltonian, get some new wave functions, build a new Hamiltonian, and keep going until everything becomes consistent. And the idea, then, is that the wave function we're looking for in Hartree-Fock are the eigenfunctions of H of ψ $H\psi$, which we just defined to be equal to H Hartree-Fock. So there's some Hartree-Fock Hamiltonian. And the Hartree-Fock wave functions are the eigenfunctions of that Hamiltonian. Does that make sense? Any questions about that?

All right. So Hartree-Fock is the starting point. So we have that as an approximation. It's a useful starting point. In many cases, Hartree-Fock is not good enough. And so the real thing that we're trying to figure out is, how do we make something that uses what we've learned from Hartree-Fock about orbitals, and energies, and average repulsion-- how do we make something that's like Hartree-Fock, but better?

So one way to do this is something that you've spent a lot of time on this semester, which is perturbation theory. You have something that's pretty good as a starting point, and you want to make it better, perturbation theory is the natural thing to turn to do that.

And so the way we apply perturbation theory in Hartree-Fock is we say, well, we've got this

Hartree-Fock Hamiltonian, which we can solve, and then we've got the actual Hamiltonian, H , which we can't solve. And so what we can do is then we say, aha, well that means that if I just take the difference between the Hamiltonian I can solve and the Hamiltonian I can't solve, and I treat that as a perturbation, then I can do a well-defined perturbation theory where I know the solutions to H_0 , and then I treat that additional term as a perturbation.

And so then if I do that, I end up with a perturbation expansion where I can go to first, second, third, fourth, fifth order, and that additional term that's the difference between the two. And if I look at the first two terms added together, they give me the Hartree-Fock energy. So if I'm the one making up a perturbation expansion, I will always choose it so that the first order correction is 0. So that if otherwise I have chosen, I will just re-engineer my perturbation so that's true. So up through first order, I just get what I started with. I get Hartree-Fock.

But then I have these additional terms. And so I can go to second order. And I will get a method that I call-- it's called MP2. So Hartree-Fock is named after Hartree and Fock. MP stands for Moller and Plesset, the names of the two people who wrote the paper in 1938, about this particular thing. And the 2 is second order. And then I could go on, and I could do MP3, MP4, dot dot dot dot dot. So in general, this is MP, and some number, n , telling me the order I go to in perturbation theory.

So you could do this. In principle, this gets better and better answers. However, in practice, about 20 years ago, people discovered something that was rather disturbing. So if I take a look at the energy minus the [INAUDIBLE]. So let's say I have a case where I know the exact energy. And I look at the energy of these different approximate methods for, say, one particular simple thing, like a neon atom. So I take a neon atom, and I look at the energies of these MP_n approximations, and I look at them as a function of n , so 1, 2, 3, 4, 5, 6, 7, 8.

And I'll say that for neon-- neon is something for which Hartree-Fock is already pretty good. So the difference between Hartree-Fock and the exact answer is not too bad. So Hartree-Fock might be, let's say, there on this particular scale. And then if you do MP2, MP2 is better. So my MP2 is maybe there. MP3 is also a little bit better than that, but it overshoots a little bit. MP4 corrects back in the opposite direction. And then we've got 5, 6, 7, 8. And you can-- if I connect the dots, you can see what's starting to happen here.

So that's a series. So mathematically, I've got something that there is a value for everything in n . That's a series. And the name for a series like this is one that is not converging. So there's

an answer which is the infinite order answer, but these partial sums are not converging to the infinite answer.

And very often this turns out to be the case with perturbation theory. There can be perturbation theories that are very useful at low orders for describing qualitatively how things shift. But if you go out to infinite order, they are series that, in principle, you can prove, give the correct, exact answer, because we've derived it as such. But just because a series in principle sums to something doesn't mean that any finite number of terms will give a convergent answer, or give smaller errors.

And people find that for perturbation theory for electronic structure theory, this is-- not all of the time, but very often the case, that the perturbation expansion doesn't converge. And so this is one of those places where you make a nice empirical compromise, and you say, well, gee, every term-- it doesn't necessarily get better after MP2. MP2 usually improves on Hartree-Fock, but higher order might be better or worse. And certainly, higher order is more expensive. So MP2 is a good place to get off the elevator, because it's just going to get more expensive, and not necessarily more accurate. So this is a good place to stop. And so you'll see a lot of calculations out there that are variations on MP2, because it's the least expensive, and it's somewhat more accurate.

OK so now we're going to take a pause, and we're going to talk-- and this is where I need audience participation. So we're going to discuss two different things that you might want to know. We've now got two methods, Hartree-Fock and MP2. And there are two things that you might want to know. One is, how accurate will one of these calculations be? If I do it, will it give me the right answer? And the other one is, how long will it take? In other words, will it finish by Friday when the P set is due? Will it finish by the end of the semester? Will it finish by the end of my time at MIT? What time scale are we looking at?

So we'll start off talking about properties. So people have done many, many, many calculations using many different electronic structure methods, so that now there's sort of an empirical rule of thumb for just about any property you want to know about. There's a rule of thumb for how accurate one of these methods should be.

So what's a property that you might want to know? This is where the audience participation comes in. I will only give you information about properties that you say are interesting.

AUDIENCE: Time and space complexity.

TROY VAN VOORHIS: Time and space complexity. All right. Could you be more specific. What type of time and space complexity?

AUDIENCE: How well-- I mean, for a size of molecule, how long it takes--

TROY VAN VOORHIS: Ah, OK.

AUDIENCE: --and also how much space it takes.

TROY VAN VOORHIS: Yes, so that will be the second thing that we will talk about. Those are the two variables we'll talk about, in terms of how long and how fast. I'm interested here, where I'm saying, like, molecular properties-- things you might want to know about a molecule, or a reaction, or something like that.

AUDIENCE: Like the energy of a molecule in transition state.

TROY VAN VOORHIS: OK, so that would be sort of activation energy, right/ OK, what's something else? We only want to know about activation?

Yeah.

AUDIENCE: Charge distribution.

TROY VAN VOORHIS: Right. OK, so I'll give one variable that probes that, which is a dipole moment, for example. You might want to know a dipole moment.

AUDIENCE: What the MOs look like [INAUDIBLE]

TROY VAN VOORHIS: OK, yes. So that is something we want to know about, but it's a qualitative one. So what I'm going to give is I'm going to give you things where I can say, oh, for a dipole moment, it's within this many debye, or this much percent. So MOs are qualitative, and we can get that, but I can't say, well, this MO is 75% right.

AUDIENCE: Average distance of an electron from an atom.

TROY VAN VOORHIS: OK, so that would be like size.

AUDIENCE: Yeah.

TROY VAN VOORHIS: Yeah, yeah.

VOORHIS:

AUDIENCE: [INAUDIBLE].

TROY VAN VOORHIS: What was that?

VOORHIS:

AUDIENCE: Color.

TROY VAN VOORHIS: Yeah. I mean, so yeah, that's about an absorption-- like an electronic absorption spectrum.

VOORHIS:

Other things people might want to know?

AUDIENCE: Bond length.

TROY VAN VOORHIS: Bond length. There you go.

VOORHIS:

Anything else? We've got a good list. OK, we'll use that as our list.

OK, so I will start off with bond lengths, because that's the one where Hartree-Fock does the best of these things. So it makes you feel like it's encouraging.

So bond lengths-- Hartree-Fock usually predicts bond lengths to be a little bit too short. So the molecules are a little bit too compact. But it's not too bad. They're usually too short by about 1%, which is not too bad. It gets 99% percent of the bond length right. MP2 does better, so that it doesn't actually have a systematic error. It typically gets bond lengths correct, plus or minus one picometer. So a picometer is 10^{-12} meters. So 0.01 angstroms, since a bond length is usually about an angstrom.

So we'll go from that to size. So for the size of an atom or a molecule, this has to do with the Van der Waal's radius, basically. And actually, Hartree-Fock does a fairly good job of describing Van der Waal's radii. I would say that it more or less goes with the bond length prescription, so that the size is a little bit too small. So it might be minus 2% to 3% to small. And then there's very few people that actually do sizes with MP2, for various reasons. But I

would say that it's-- I'll put "accurate" there. Because basically the size of an atom or a molecule is fuzzy. So I can talk about the radius, and it's a little bit fuzzy. The errors in MP2 would be smaller than whatever fuzziness I could have in my definition of size.

For activation energies, Hartree-Fock is pretty poor, so the activation barriers here are too high, by 30% to 50%. So the barrier heights are very high in Hartree-Fock. MP2 has barrier heights that are still too high, but only by about 10% on average.

Dipole moments in Hartree-Fock-- there are no systematic studies on that. I will just say that they are bad. They are bad enough that nobody does a systematic study. But for MP2, they're quite a bit better, and typically accurate say, 0.1 debye. And then for excitation absorption energy-- so this is going to be the difference between the lowest electronic state and the next highest electronic state-- so you can get these things through various means. The typical absorption-- and also, I'll say that doing that involves an extension beyond what I've talked about so far, because I've been focused on the ground state. But if you use the logical extensions of these things, the absorption energies in Hartree-Fock tend to be too big, and too big by, say, half an eV or so. And then for MP2, they're more accurate, but not that much. So plus or minus, say, 0.3 eV. Electronic inside states tend to be fairly difficult to get.

But we see that MP2 is doing better on all accounts, than Hartree-Fock. So it's improving. That's good. But then, all right, what is this going to cost me? How long is this going to take? Well, there's two different ways that we can measure computational complexity. One is by the amount of storage that's required. It's the amount of information you have to store in order to do the calculation. The other way is by how long it's going to take, how many operations the computer has to do to solve the problem. And so for Hartree-Fock, the main memory piece that we need is actually the Hamiltonian. So we need that Hamiltonian matrix. We've got to store that in order to compute its eigenvalues. And it is an end-- it's a number of basis functions by number of basis functions.

And so then now we have to say, all right-- well, first of all, we have to figure out how much RAM. So when I'm figuring out how much RAM is required-- does anybody know how much RAM they have on their laptop or desktop computer? Nobody knows. 16 gigs. All right, that's what's on my mine, too. 16 gigabytes, which is 16 times 10 to the 9th bytes, roughly.

So now we have to say, all right, for storing n basis times n basis numbers, how much does that take? Well, let's say that we have a atoms. So we've got-- the number of atoms we have

is a. Then the number of basis functions we have-- last time, we figured out that a DZP basis, which is a decent-size basis for carbon had around 15. It was 14 basis functions per carbon. I'm just going round to 15 because it's easier math. So that means that the number of basis functions is 15, roughly, times the number of atoms.

So there's 15a times 15a things that I need to store to get this matrix. And that means I need about 225 a squared numbers. So these are how many numbers I have to store. And then I have to know that each number-- I'm usually storing these in double precision. So each number requires eight bytes in double precision. So that means that I have something like 1,700 a squared bytes to store that object.

So now that says, all right, if I've only got 16 times 10 to the 9 bytes of storage space on my computer, and I need 1,700 a squared bytes for a molecule with a atoms, that just gives me a natural limit on the number of atoms. If you back it out, that implies that a is less than or about equal to 3,000, which is a big number. So your molecule could have up to about 3,000 atoms in it, and this calculation would run. And that rough calculation turns out to be about right.

Now let's take a look at the CPU time required. So the CPU time, we're going to measure this in hours, because that's my human unit of time. So one hour is 3,600 seconds. And then what I need to know is well, what actually takes the computer time? What takes the computer time is doing mathematical operations-- so add, subtract, multiply, divide, if, things like that. And usually, we measure the speed of a CPU in terms of the number of operations you can do per second. And does anybody know the order of magnitude of operations per second the CPU on your computer can do? It's about a billion-- some number of billion operations per second, depending on how recent a model you have, and whether you play video games or not, you will do more or less than that.

So that means that our computer, in one hour, can do about one-- and I'm going to round up, and say you've got a really good one. So we're going to say that your thing can do 10 billion operations per second. If you've got multiple cores, they can each go independently. So 1 times 10 to the 10 operations per second. And that means that you can do about 3 times 10 to 13 operations in an hour. I'm just going to set the hour as my patience limit. I don't want to wait longer than an hour.

And so, then, for Hartree-Fock, the rate-determining step is computing the eigenvalues of the Hamiltonian. And that requires n cubed operations, where n is the dimension of the matrix. So

if our matrix n basis on a side, it requires n^3 basis functions cubed operations. And so then, again, using our translation of that into atoms, that's 15 times the number of atoms cubed. Which, if I did my math right back in my office, that's about 3,000 a cubed. And then if I back that out-- again, if I compare that to the number of operations, that gives me a limit on the number of atoms that I can handle in an hour, and that turns out to be about 1,000. So similar sizes. So I run out of storage about the same time as I run out of patience. If I was willing to be a little bit more patient, I might be able to do a few more. But order of magnitude, we're at 1,000.

So I'll summarize. What we found here is that the storage requirements were something like a squared of order a squared. CPU time was of order a cubed. The maximum feasible number of basis functions that I could do here was something like-- so if I have 1,000 atoms, that was my limit. Something like 15,000 would be the maximum feasible number of basis functions. And maximum feasible number of atoms was about 1,000.

So we won't go through the same exercise for MP2 in such detail. I will just tell you that everything is worse for MP2. Everything takes longer, it takes more storage, everything's worse. So it requires order a cubed storage. It requires order a to the fifth CPU time. The maximum feasible number of basis functions is therefore smaller, about 2,000. And in MP2, you actually require more basis functions per atom to get an accurate answer, so that the number of feasible atoms is more like 50 rather than 1,000. So that's the cumulative effect of larger basis sets and worse scaling with number of atoms that makes MP2 deal with much smaller systems.

So questions about that. Yes.

AUDIENCE: So what are some examples of [INAUDIBLE]?

TROY VAN VOORHIS: So C60 is 60 atoms. That's an easy one. There are a number of small dyes, for example, that we worked with in our group, where you might have seen absorption spectra, or emission spectra, or HOMOs, or LUMOs, or hole transport properties, or electron transport properties, that are around 50 atoms in size. And then the main thing that people get interested in that are on the 1,000-atom regimes are enzymes and peptides. Those are the sort of-- when you start saying, I want to do 1,000 atoms, it's usually because it's really some polymer or heteropolymer.

Another case where you also are sometimes interested in things where you do simulations

with 1,000 atoms, if you're interested in a surface. Because if you have a chunk of gold, for example, 1,000 atoms is just 10 gold atoms, by 10 gold atoms, by 10 gold atoms, which is just a little chunk-- a very, very small chunk of gold, but still a chunk.

Other questions. OK, so I'm going to spend seven minutes talking about density functional theory. And then we're going to go over and do some examples.

So the idea of density functional theory is that it'd be really nice if you had some magical way to do a Hartree-Fock calculation, but have a give you exactly the right answer. That basically would be the dream because we saw, well, for Hartree-Fock, we can do big systems, it's cheap, it has good scaling, all these kinds of things. It just doesn't give us very good answers. The results are pretty poor.

So what if we had something that scaled like it had the computational cost of Hartree-Fock, but was, say, the accuracy of MP2, or even better than that? And so in density functional theory, what we use is the idea of looking at the electron density ρ of r as the fundamental variable. So you can actually work out, for a determinant, what the electron density is. And it's just the sum of the squares of the orbital densities-- or the sum of the orbital densities. So you square each orbital, then you add that up, and that gives you the total electron density. It's basically just the probability of finding an electron at a point, r .

OK, so that's a nice observable. The thing that's kind of amazing is that there is a theorem, a mathematical theorem that you can prove in about two pages using just sort of proof by contradiction. You can prove that there exists a functional, which is always given then in e sub v of ρ , such that when you're given the ground state density, ρ_0 , if you plug that into this magical functional, it will give you the ground state energy. So if you found the ground state density lying around in the gutter, and you picked it up, and you put it into this functional, it would give you the ground state energy.

And e_0 is not just the approximate ground state energy. It is the exact ground state energy, the exact thing. Further, for any density, ρ' , that is not the ground state density, if you plug that density into eV , you get a higher energy.

So what you could do is you could say, well, let me just search. Let me try density, see what energy it gives. Then I'll try another density. If it gives a lower energy, that's a better density. And I'll keep going, and going, and going, and going, until I find that ground state density. So the idea is that then you would say, aha, if I solve that equation-- so I search over densities,

and that's not going to be a very hard search, because the density is just a three-dimensional object. It's not like the wave function that depends on a bunch of coordinates. It's just three dimensional.

So I solve that equation, that will give me ρ_0 . And then I can take that ρ_0 , plug it back in there, and I will get the ground state energy. That gives me a closed set of conditions that lets me find the ground state energy, and then report back the energy of that density.

AUDIENCE: Where do we get this [? EV ?] from?

TROY VAN VOORHIS: That's a great question. The proof is a proof by contradiction. It's not constructive. It proves that it exists, but gives you no way of constructing it. So kind of the frustrating thing is like, oh, cookies exist, but we don't know how to make them. But the idea that such a thing exists gives you hope to say, well, maybe we can construct-- if we can't find the exact one, maybe we can find one that's very, very good. And that one that's very good, we would just use. We would pretend it was exact, minimize the energy, find the density, and then report back the ground state for that approximating.

And that's actually what you do in density functional theory. You have approximate functionals. And they're all-- just like with basis sets-- named after the authors of the people who wrote the papers. Almost all of them are. So there's the Local Spin Density Approximation, LSDA. There's the Perdew-Burke-Ernzerhof functional. There's the Becke-Lee-Yang-Parr functional, the Perdew-Burke-Ernzerhof Zero functional, the Becke 3 Lee-Yang-Parr functional, and then on and on and on.

Now these things have been sort of-- there's a mishmash of different exact conditions that were used to derive these functionals. And then, empirically, people have gone and shown that PBE is better than LSDA. It's about as good as BLYP. BLYP is not as good as PBE0, but it was just about as good as B3LYP.

So we have, then, sort of an idea that, OK, if we go over towards the right-hand side of this, we're going to get better results out of DFT. But DFT, because it's based off of a Slater determinant, has exactly the same computational scaling, and storage, and everything, as Hartree-Fock. It's the same kind of expense. But if I go back to this accuracy thing, and I say, well, what would I get from that best density functional B3LYP, well what I find is that activation energies are still a little bit difficult. Now I underpredict activation energies. But my dipole

moments are good. My sizes are again accurate. My absorption energies are just as accurate as MP2, and my bond lengths are just as accurate as MP2.

There are other things, but basically B3LYP is as accurate or more accurate than MP2 for virtually every property you would want, but has Hartree-Fock-like cost. And so that makes density functional theory the workhorse of computational chemistry these days. Because it's inexpensive, you can do lots of calculations with it, but it's accurate enough for grunt work.

And then there's a blank column there that, if I had time, I would talk about. There's a whole other category of things where you say, well, perturbation theory doesn't work, but are there more constructive ways that I could use wave functions to approximate the correlation energy? And the answer is yes, you can get good energies and good answers out. But then it comes at the cost of the calculation getting much more expensive. So there are methods that improve all these properties on the wave function side, that just require more. So if you want to know more about those, they're in the notes.

And so now I think I will switch over and show you how we do some calculations with this. So now I will switch.

There's two different tools that you have available to you for free on Athena that you can use to do electronic structure calculations. One is Gaussian, and the other one is Q-Chem. There may be even some other ones, but those are the two that I've collected notes about how to use. So I'm going to show you how we use Q-Chem. I'm not going to do it on Athena. I'm going to do it on my laptop, but it's basically the same.

So there's this GUI called IQmol. And when you open it up, it gives you a window that looks like this. And there's a few things here. So there's, like, an open file, you can create something, you can open a file, you can save something, you can move something around after you've built it. And then these are the build tools over here. So that just turns it into build mode. If I click on this, I can choose any item in the periodic table. Right now, it's on carbon. This lets me select various fragments, and also what the attach point of the fragment's going to be in some cases. And I don't want to do it, so I'll select it for now, but then I'm going to go back to this one.

This is the Add Hydrogens button. So if you have something that-- you just wrote your Lewis structure, it had no hydrogens in it, you click that button, it'll put all the hydrogens where they should be, or where it thinks they should be.

This is a Minimize Energy thing, which basically if your structure looks really crazy, it'll kind of make it look less crazy. And then this is the Select button, which lets you pick certain atoms. This lets you delete certain atoms, take a picture, do a movie. And then that changes it over to fullscreen mode. And that's the life preserver for Help.

But we'll do building a molecule. So does somebody want to tell me a molecule that has less than 15 atoms in it that they would like me to draw here, that I can actually draw?

AUDIENCE: Diethyl ether.

TROY VAN VOORHIS: Diethyl ether. Yes, I can do that. OK. All right, so put a carbon, carbon, then I got to switch to an oxygen, oxygen, carbon, carbon. And then I will click the Add Hydrogen button.

[LAUGHTER]

There we go. I've got my diethyl ether now. And while my geometry here is not perfect, it doesn't look totally crazy. So I'll just go with that. And then the thing we're going to be most interested in is this little tab here, called Calculation. So calculation here has a little button that says Q-Chem Setup. So I'll do Q-Chem Setup.

And now it's got a bunch of things that I can specify. There are a few things that are most important. So the first thing is, what kind of calculation I want it to do. So I could have it just compute the energy of the molecule for the configuration I specified. That would be one useful thing. I could compute forces. In other words, the forces on the atoms, where the atoms want to move. I could optimize the geometry, which would then relax the thing down, and figure out what the best geometry is. I could do various things that scan across the potential energy surface. I can compute vibrational frequencies. I can compute NMR chemical shifts. I can do molecular dynamics. I can compute properties-- so lots of things.

I'm just going to optimize the geometry. That's going to be what I'm going to choose for now. And then I get to choose a method. So I can do Hartree-Fock. I can do B3LYP. I can go down-- so all the things above this dotted line here are different density functions. And then, down here, we've got MP2, MP2 bracket v-- I'm not actually sure what that is. I've got various versions of MP2 that try to make it faster. And then I've got a couple of cluster methods that we didn't talk about yet, but that are in the notes. And then various methods for getting electronic excited states, down here at the bottom.

So for now, let's just do B3LYP, and then I get to choose a basis. Because this is on my laptop, and we're in class, I'm going to choose a small basis. So I'm going to leave it at 321g, which is a small basis. And then I can also specify the charge, which is like the total charge of the molecule, and the spin multiplicity. So I want it to be a neutral molecule, spin 1-- or spin multiplicity 1, which is spin 0. Because it's 2s plus 1 is the spin multiplicity.

And then there's various things about convergence control that you probably won't need to worry about. And then there's advanced things that you also probably won't need to worry about.

So now I'm happy with that. I can submit my job. And I will call this diethyl ether, hit OK. And now it doesn't appear to be doing anything. In the background, it is doing something. I think this will finish in just a few seconds, but if you get concerned that the job might not be running or something might have happened, you can go over to the Job Monitor. Methane-- that was the job that I ran this morning to make sure that it works. Diethyl ether, we just submitted it at 10:50 and 21 seconds. Ah, there we go. Now it's been running for four seconds. Let's see how long it takes. OK, it's taking a lot longer than I think. All right.

AUDIENCE: They close [? in 15 ?] [INAUDIBLE].

TROY VAN VOORHIS: That's all right. We still we've got a few minutes. So what will happen here in a minute, when we--

AUDIENCE: [INAUDIBLE].

TROY VAN VOORHIS: That's all right. It's going along.

So when it's done, what will happen is, over here, it'll let us know. And over here will appear a thing where we can start looking at the results. It might have gone faster if I had minimized the energy to begin with, because then it would do less geometry steps. It's going along. I don't know. We might run out of class time before it finishes. It's 10:51. We've got four minutes. Let's see.

Come on, laptop. I've got the MacBook Pro. Maybe I should have upgraded the processor, this kind of thing.

AUDIENCE: How long do you usually let your group calculations go for?

**TROY VAN
VOORHIS:**

Well, in my research group, we have 1,000 cores that we-- maybe around 1,200 cores, even, I think, actually. Oh, there we go. It finished. But we have around 1,200 cores for around 10 people. So that means that you can let a job run for a very, very, very long time, on multiple cores, and it doesn't get in anyone's way. So it's not unusual for us to put something in there that lasts a week.

So now it's finished. We want to copy the results from the server. This is an arcane way of saying things. Because it's designed for supercomputers, where you are sitting at a terminal in your office, and then it sends all the results over to the supercomputer, it runs the calculation, and then you have to copy the results back to your little computer. I was actually running it on my computer. It's just the computer is both-- I'll just put it-- New Folder. OK.

There we go. Now it has a little star next to it, which lets me know that it's finished. So let's see, Info-- so I can show the dipole moment. So there we go, that little blue arrow-- which, magnitude and direction is the dipole moment of diethyl ether. It does have a dipole moment, with its positive end here, pointing in that direction. That seems about right. Let's see. I can look at the files. These are only if you're really, really interested in details. You see these are very verbose. They have every bit of information you could possibly want in them. So if there's ever something that you can't get out of the GUI, if you just kind of go-- you can search. And I can look for converge. Oh, look, optimization converged, and then it tells me everything after that.

I can look at the different atoms in the molecule, and it will highlight them for me. And then I can look at bonds, and it will tell me-- I don't if you guys can read, but down here, it has the bond length in the corner. So that C-H bond is 1.093. All these C-H bonds are going to be 1.09, plus or minus a bit. The carbon-oxygen bond there is 1.46, rounding. That one is also 1.46.

And then this here-- so as it did the geometry optimization, it was adjusting the geometry trying to find the minimum. And these are the different total energies that it had as it went through the optimization. You'll see that eventually it sort of slows down. The energy is going down every time, and eventually it sort of slows down to where it's not changing any more. It says, all right, that's the minimum.

You also notice that these are in atomic units. So that's minus 232 Hartrees. So that's like 4,000 kcals. No, that's more than 4,000 kcals. Anyway, it's tens of thousands of kcals. The

reason it's such a low number is this is the energy it would take to rip every electron off of the molecule, including all of the 1s electrons. So the energies are typically going to be very, very large negative numbers. Don't worry about that. It's energy differences that matter.

So you notice, even my bad geometry was already 232.1. The correct answer is 232.3. So the total energy change wasn't that much.

And then the last cool thing that I want to show you is that we can plot orbitals. So I'm going to count some orbitals. There we go. So I just had to calculate the HOMO and the LUMO. I didn't go through all of it with you. But if you go there, it gives you some things-- I should actually go back and look, show you. So you get to choose a range of orbitals. The iso value is basically where it draws the surface. So it's inverse of what you expect. A larger iso value is-- it's an isosurface, so if you make it larger, the place where the iso value is larger is closer in. So it makes a smaller bubble. And a smaller iso value makes a bigger bubble. And you can change the number of points that it samples to make the surface, change the colors, change the opacity of it in terms of how you want it to look.

But I already did this. So there is the HOMO. So the HOMO is something in the pi manifold. It's pi star here. And then I can take out the HOMO and show the LUMO. And the LUMO is, in this particular basis set, not very useful, because I don't have a lot of basis functions. So my basis functions, just there's not enough of them to really describe the LUMO, or the LUMO plus 1. If I want to get a LUMO, we'd need a bigger basis set than this.

So that's the kind of thing that you can do with Q-Chem or with Gaussian. Gaussian has a GUI called GaussView. You can use either one. But they both work pretty well, and are pretty intuitive.

And so I think, on the next problem set, you will have some homework problems that involve running some calculations with these things. And I hope you enjoy it. Well, I don't know if homework is ever really enjoyable. But I hope it's marginally enjoyable anyway.