

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 to view additional materials from hundreds of MIT courses, visit MIT OpenCourseware at [ocw.mit.edu](http://ocw.mit.edu).

**TROY VAN**

All right, well, good morning, everyone. I'm not Bob Field, but I'm Troy. Nice to see everyone.

**VOORHIS:**

So I'm here. We're going to spend the next couple of lectures talking about electronic structure theory. And at this point, I'll give you the big reveal, which is that at this point, you have already encountered virtually every problem that you can solve by hand in electronic structure theory.

So there's a handful of other things you could do if you really, like, knew your hypergeometric functions and some things like this, but basically, every electronic structure problem that you can solve by hand you've already encountered. And so if chemistry relied on us being able to solve things by hand, we wouldn't get very far. We'd be pretty limited in what we could do.

So we end up using the fact that we have computers. And computers are far more good at doing tedious repetitive things that human beings are. They don't complain at all. And so we can use computers to do fairly neat things.

So I'll show you just a couple of quick pictures here of things that we can do. So if you open up any research paper these days, even if it's just synthesis-- you know, we made this catalyst. We did this thing, whatever. There's always some part of it where they said, well, we did a calculation, and this is what the structure looks like. Or this is what the orbitals look like or something like this. So this was actually a calculation. We are computational chemists, I guess.

I'll use this screen over here. Is this the one that you're-- or you're filming this one. This one?

All right. so I'll film this one-- go to this one. Breaking the fourth wall there.

So this is the picture of the HOMO, the reactive orbital water-splitting catalyst that we computed. So you can see that there's 50, 60 atoms here, probably several hundred electrons. We did not do this by hand. We didn't draw the surfaces by hand. The computer did all of this for us. And you can also do things like look at chemical reactions and run dynamics.

So we can do-- there we go. So you can run dynamics of molecules. So this is a particular molecule in solution. And we can use the electronic energy in order to govern the dynamics of the molecule. And the computer can actually predict for us whether this molecule is going to

react, what it's going to do, how it's going to behave in one solution, say, versus another or at one particular temperature versus another, those kinds of things. And all of these, again, are based off of the idea that we can use computers to solve problems in electronic structure, or at least approximately solve problems we couldn't do by hand. And so what the goal over the next couple lectures is for us to learn enough of the basics so that you can do some calculations like this yourselves and then see what you can do with that.

So now I will switch over. You'll find where the other differences, which is that I don't like the feeling of chalk on my fingers, so I use my iPad as a chalkboard. So the rest of this will all be on the iPad. There we go.

So for those of you who want to-- you may not have gotten in this time. But I've literally posted what I'm going to be writing on here, these blank notes, online. You can download them, print them off, if you want.

If you have an iPad or a computer, and you like to take notes on that, you can download them to your iPad or computer and take notes that way. But I find it's good because there are some things that take a long time for me to write at the board and that it would take a long time for you to write in your notes. And we can just have them written down in advance.

So the starting point for all of these things that you can do on a computer for electronic structure theory, we start with the Born-Oppenheimer approximations. So the Born-Oppenheimer approximation was the idea that because the nuclei are very heavy, you can clamp their positions down to some particular values. So in this particular case here, in this equation, I have it such that big  $R$  here are the nuclear positions. So there's going to be more than one nucleus generally.

So there's going to be  $R_1$ ,  $R_2$ ,  $R_3$ . All those crammed together, I'm just going to denote big  $R$ . And that's where the nuclei are. And so in Born-Oppenheimer approximation, we clamp the nuclei down. And then what we're left with is the electrons.

The electrons whiz around. They move in the field dictated by those nuclei. So we have a Schrodinger equation that describes the motion of the electrons in the presence of the nuclei. And so here I'm using the lower case letters, lowercase  $r$ . So they'll be, again, many electrons, so  $r_1$ ,  $r_2$ ,  $r_3$ , so on and so forth. And it's the electrons that are really doing all the dirty work in this equation.

They're the ones that are moving around. I've clamped the nuclei down. They're just parameters in this equation. And so what I have is I have an electronic wave function that describes the distribution of electrons. I have an electronic Hamiltonian that governs the motion of the electrons.

Then I have an electronic energy after I've solved this Schrodinger equation. And the nice thing about this is that for an arbitrary number of electrons and an arbitrary number of nuclei, I can in one line, in about 10 seconds, write down the Hamiltonian.

So the Hamiltonian is pretty easy to write down. I've even written it right here. So I did use the cheat of using atomic units. So the reason there's no  $\hbar$  or mass of the electrons or any of that appearing in here is because I chose atomic units.

But then the ingredients of my Hamiltonian are all pretty well-defined. So I'll have some kinetic energy. And that kinetic energy will just be for the electrons because the nuclei are fixed. I'll have an electron-electron repulsion term.

So of course, the electrons repel each other. I'll have an electron nuclear attraction term because the nuclei are fixed in positions, and the electrons feel a potential due to that. And then I have a nuclear repulsion term. And in this case, because the nuclei are fixed, that nuclear repulsion is just a number because I know where all my nuclei are. They have some particular repulsion energy. That's just some number that I add onto my Hamiltonian there.

But if I put all those things together, I get the Hamiltonian. So I can write down the Hamiltonian without too much difficulty. But of course, the interesting thing here is this Hamiltonian depends on these nuclear positions.

So I have to know what-- depending on where the nuclei are, I'm parametrically changing this Hamiltonian. That changes  $H$ , and it also changes the eigenvalue, the electronic eigenvalue  $E$ . And so there's an obvious question of, well, what is this electronic eigenvalue? It turns out to be a very important thing.

We will typically call this thing a potential energy surface. And it turns out to govern a whole host of chemical phenomena. So the first thing that it could generate that we can look at-- we've already actually seen potential energy surfaces.

So on the left-hand side here, I have the picture from the case of  $H_2^+$ . So here we had the energy of the sigma orbital. And as a function of  $R$ , it forms some bonds. So as we made  $R$

shorter, the sigma orbital formed a bond.

Then we had also this sigma star orbital as a function of  $R$ . When we brought the atoms together, in that one there was no bond, no swarming energy. It just went straight up. So we've got two of these things, and I should label my axes.

So what I'm plotting here is the energy as a function of the bond distance  $R$ . So this was our first example of a potential energy surface, the electronic energy as a function of  $R$ . And we see that it taught us about the bond, the bond strength, so forth.

There's one other thing that this picture reminds me of, and that's that for the same problem here, I actually had two different potential energy surfaces. I have the sigma potential energy surface and sigma star. And that's because I left something out in my Schrodinger equation up here.

There's, of course, an index. I have different eigenfunctions of the Schrodinger equation. So I have an electronic Hamiltonian. It'll have many different eigenfunctions. So I have an index  $n$  for those eigenfunctions. And each of those eigenfunctions will have a different energy. And each of those energies is a different potential energy surface.

So the sigma state was the lowest solution. The sigma star state was the next lowest solution. They had different potential surfaces.

But I will often not refer to potential energy surfaces plural but instead to potential energy surface singular because most chemistry occurs on the lowest potential energy surface. So it involves just the lowest potential energy surface. And the reason for that is because electronic energies tend to be very big in chemical terms.

So the difference between sigma and sigma star is several eV, typically. And chemical reactions on that same scale are usually tenths or hundreds of an eV. So the chemical reaction energy might be somewhere way down here, way way, way below the amount of energy you would need to get all the way up to sigma star. And so the sigma star, potential energy surface is just irrelevant. It's there, but it doesn't participate in the reactions at all.

So in most cases, knowing about the lowest potential energy surface gets you most of chemistry. But then we can also, sort of, schematically talk about how these potential energy surfaces will look for a more complicated example. So let's say I had the water molecule, HOH.

So this is a more general case. We have at least two coordinates here,  $R_1$  and  $R_2$ . There's also going to be an angle, but I can't really plot three-dimensional hypersurfaces, even with my computer. So I'll stick with two.

There's a third coordinate there, which is  $\theta$ , which I'm not going to play around with. But we'll play around with just  $R_1$  and  $R_2$ . So I can change individually the OH bond lengths in this molecule. So then this is going to give me the energy as a function of  $R_1$  and  $R_2$ . So it'll be a surface.

So it's a function of two variables. That's a surface. And that surface might look like what I have plotted there, where it would have various peaks and valleys on it. The very, very lowest valley, the very lowest minimum, just like we found for  $H_2$ , that lowest minimum was the equilibrium bond-- bonding configuration.

So for this two-dimensional potential surface, that lowest minimum's going to be the equilibrium configuration. But there's also the possibility of other minima on this potential energy surface. Those would be metastable intermediates, things that you could get trapped in, things that would live for a while, but eventually go down towards the lowest minimum.

Then there'd also be reaction barriers. So these are the things that we learned about governing how fast things convert from these metastable states into the most stable states. So that's what you would get if you had just two coordinates. And so then you can sort of try to generalize in your head, well, what if we had 3, 4, 5, 6, 127, 368, lots and lots of coordinates?

In some many dimensional space, it would still have the same qualities. There'd be minima and barriers and so forth. And we would get those out of the Schrodinger equation, the electronic Schrodinger equation. And so that is the electronic structure problem.

How do we accurately solve for the electronic eigenvalue and the electronic wave function for an arbitrary molecule? And the key word here is accurately. I didn't just say how can we solve for electronic energy and the electronic wave function because, in general here, exact solutions are impossible. And I don't mean impossible for 5.61. I mean impossible for humans or computers.

So in general, electronic structure theorists love approximations. We love to make approximations because that's the only way that you can make progress. And so what we're going to learn about today and next Monday are the, sort of, different categories of

approximations that we make and what the pluses and minuses of those are.

So I'll pause there and see if anybody has any questions. All right, so we'll move on then. So we have an out-- so there is actually an outline of how different modern electronic structure methods make approximations. And these follow the same kinds of steps that you would use in a molecular orbital calculation.

So the first thing that you typically had to do in these calculations was to choose a set of basis functions. So you had to choose like 1s function over here or 1s function over there or a 2s function or 2p. You had to choose some set of atomic orbitals that you're going to use as the starting point of your calculation. Then you made linear combinations of those things to get better functions. And so then after you'd chosen that basis, you had to build some matrices.

Then you had to solve the eigenvalues problem for that matrix. Then you had to pick out which orbitals you were actually going to occupy. Those were the next few steps. And then finally, you had to compute the energy.

Now, that's already kind of a detailed outline. If it seemed like we had to understand the nuances of every one of those five steps, we'd be kind of-- well, you can do that. It would be an entire course unto itself. And that's not what we're going to be doing.

The reason that we can avoid going over every single nuance is twofold. So first, we can avoid, more or less, steps two through four because they're done automatically. So once you have chosen your basis that you want to use, the computer knows how to build matrices. It knows how to diagonalize matrices and find eigenvalues. It knows how to pick out which orbitals to occupy.

It knows how to do all of those things without you telling it to do anything. You don't really have to change anything. So the only things you have to worry about are choosing a good basis and then also telling the computer a good way to compute the energy.

The other reason that we can actually do something useful here is because using a computer to solve an electronic structure problem is much like using an NMR to get a spectrum of a compound. I doubt that anybody in this room could build their own functioning NMR. I know generally the principles of how an NMR works. But building from the ground up-- oh, so we do have any takers? Anybody able to build and NMR, huh?

**AUDIENCE:** [INAUDIBLE]

[LAUGHTER]

**TROY VAN VOORHIS:** From the ground up. I'm betting-- no. See? You have to actually, like, make your own superconducting magnet. You've got-- from the ground up, actually building your own NMR would be very, very difficult.

But that doesn't preclude you from knowing how to use an NMR because you know, OK, well, this piece is-- this dial is roughly doing this, so I change this over here. I shim this. That's how I make the NMR work.

It's the same thing with electronic structure code. You need to understand a little bit of how they work in order to use them effectively. But you don't have to understand every single line of code that went into building it, which is good because there's several hundred thousand lines of codes in many of these electronic structure packages.

So we're going to try to understand the principles that go into this so that we know how to be good users of computational tools. And we can actually, then, organize. So roughly speaking, we'll be focusing on step one today and step five next Monday.

But we can organize these approximations on a sort of two-dimensional plot, which just sort of summarizes the whole idea of what we're going to be getting at. So the idea here is that on one axis we have the step one thing, which is the basis, the atomic orbital basis. Then we can organize those basis, those choices of basis, from, roughly speaking, on the left hand, we have bad choices. And on the right hand, we have good choices.

Now, you might say, well, why would we even include choices if we know they're bad? Because the bad ones also happen to be the fast basis. And the good ones tend to be slow. And so it's a time trade-off.

So you have a lot of time, you might choose a really good basis. If you have not a lot of time, you might choose not a very good one. And then we can do the same thing with the energy.

We have different choices of how we compute the energy. And we can roughly arrange those from bad choices to good ones. But again, these bad choices are fast, and the good ones are slow. And so the general thing that we want is we want to get up here into the upper right-hand corner, where the exact answer is.

Somewhere up there, with a very good energy and a very good basis, we're going to get the exact or nearly the exact answer. So that's where we want to get. But moving from the lower left to the upper right. Every time we move in that direction we're making the calculation slower and slower and slower and slower and slower. Until eventually we just lose patience. And so then we cut it off and say, all right, this is as far as I'm willing to go.

What's the best answer I can get? What's the most best-cost benefit analysis I can do? And so in terms of being good users of electronic computational chemistry, it's about knowing, OK, well, if I've only got three hours for this calculation to run or two days or however long I'm willing to let the computer run on this, what's the best combination of a basis and an energy to get me a decent result in that kind of time?

So questions about that before I move on? OK, so we'll focus on choosing an atomic orbital basis. So we already have chosen an atomic orbital basis before. So when we did H<sub>2</sub>-plus, which is, sort of, a standard problem that you can actually solve by hand if you choose the right basis, when we did that calculation, we chose to have an atomic orbital on A, the 1s function on A, and the atomic orbital on B, 1s on B. And those two things together formed our basis.

So we wrote our molecular orbitals as linear combinations of those two functions. We got sigma and sigma star out. And we were able to work out the energies.

Now, this type of basis is a useful basis. And it has a name, and it's called a minimal basis. And it's minimal because you could not possibly do less. So if you want to form a bond between two atoms, you need at least one function on each atom.

If you had only one function, period, you couldn't really form a bond. So it's as low as you can go, can't go lower than this. Now, you could think about the possibility of adding additional functions.

So I could add in, say, the 2s function on A, the 2s function on B, dot, dot, dot, dot, dot. I could come up with various things. Now, you might ask, why would I want to do this? And there's a very good reason for doing this, which is one of the principles that we need to learn in constructing basis functions, choosing our basis, which is that adding basis functions always improves the calculation.

So even though I think that adding 2s A and 2s B, why would those be important? I can't really

say. Adding basis functions always make things better. And the reason for that is because we were doing a variation calculation. We're trying to approximate the lowest energy of the system. And just choosing the 1s A and 1s B functions doesn't give you the lowest energy. It may be close, but it's not exact.

So then by adding the 2s functions, the result could get better by adding a little-- by making C3 and C4 not quite 0 here, maybe the answer gets a little bit better. Maybe it doesn't, but it can't get worse because the calculation could always choose C3 and C4 0. So adding basis functions always makes things either better, or at least not worse than they were. And so what you'll find is that when we're talking about AO basis, choosing an AO basis, we're going to be choosing bases that are much bigger than your chemical intuition would suggest.

So for H<sub>2</sub>, I think, oh, 1s A, 1s B, that should more or less describe things. And if I was doing things by hand, that's what I would do. But if the computer's doing the work, well, no skin off my nose if this computer wastes some time doing some 2s or 3s or 4s intervals. I'll let the computer do that, as long as it gives me a better answer.

So the bases we have will be much bigger because of this. But the other thing I want to note is that when I talk about these 1s functions, you probably all have in your head-- you know what 1s functions look like. They look like  $e^{-a r}$ . So they just look like exponential decay functions.

Turns out that for practical reasons, these are inconvenient things to use on a computer. And that is because integrals involving exponentials are not analytic in three dimensions primarily because of the cusp that occurs at  $r = 0$ . So if you multiply two of these things times each other and try to do an integral, you have a cusp over here and a cusp over here. And the integral is just not something that can be worked out. And so when the going gets tough, the tough get empirical. And so instead of using these exponential functions, what we use in practice are Gaussians.

So Gaussians are things instead of looking at like  $e^{-a r}$ , they look like  $e^{-\alpha r^2}$ . So if I was to plot a Gaussian here on the same axes and choose the  $\alpha$  value appropriately, I could get a Gaussian that might look like that. So it would be similar to that actual 1s function. But particularly near the origin, where it doesn't have a cusp, and particularly in the tails, where  $e^{-\alpha r^2}$  decays very quickly, they don't actually look that similar.

So whole reason to do this is not that we have some physical reason to think that Gaussians describe atoms is better than hydrogenic functions. They don't. It's just that we can do the integrals. There's easy integrals here. And if we really do want a hydrogen-like function, we can get that by just including more than one Gaussian.

So one Gaussian doesn't look very much like an exponential. But if I take two Gaussians and choose their exponents and their coefficients appropriately, I can make a linear combination of two Gaussians. So this was the one Gaussian result. With two Gaussians I could make something that might look like that. And with three Gaussians I might be able do something that would look like-- and then four Gaussians and five Gaussians and six Gaussians.

But it's clear, then, that by using a large number of Gaussians, I can get whatever I want just by brute force. And so those are the two things about building basis, choosing basis, AO bases. And at this point, you should start to feel a little bit intimidated because I've said that for atoms you're going to need more basis functions than you thought just to try to get a variation a lower energy. And those basis functions are likely to be constructed out of Gaussians, of which you'll need a bunch of Gaussians to even really approximate one hydrogen-like orbital.

So you're going to have lots and lots and lots of Gaussians on every atom. And you're going to choose the exponents of every single one of those Gaussians yourself. And that would be just horrible to have to do. And the thing that comes to our rescue is a thing that's known as a basis set.

So a basis set is constructed in the following way. A graduate student, probably long before you were bored, spent years of their life going through and figuring out, OK, for carbon, what is a good combination of Gaussian exponents? OK, they're 113.6 74.2, 11.3, and 1.6. They wrote that down.

They said, OK, now, for nitrogen, what are a good set of Gaussian exponents? And then they went through, and they did this for every element, or at least many, many, many elements in the periodic table. Then they wrote a paper that said here is-- I did it.

So I'm going to say here is the Troy basis set. And the Troy basis means when you say you're doing the Troy basis set means you're using those exponents that I wrote down for carbon or for nitrogen or for oxygen or for fluorine or for hydrogen. And it's all predefined. It's all laid out. And if I was really diligent, it's laid out for every element in the periodic table.

If I was less diligent, maybe it's only the first two rows or something like that. But the result is that all you have to say is I want the Troy basis set because Troy makes really good basic sets. And so I'm going to use that. And then the computer can go and just say, OK, well, there's some file that has all those exponents. And the computer looks up the numbers and says, for carbon, you need this; for oxygen, this; for nitrogen, this. And so the key idea here is that these are predefined sets or AO basis functions.

So somebody already defined these. You don't have to make a choice. Other than choosing the set, there's not other knobs that you have to turn. So you can obviously see the benefit of this, which is you don't have to put down hundreds and hundreds of numbers in order to get the calculation to run. So that's a big win, which hopefully you're remember for the next 25 minutes.

Because the downside, then, is that you just noticed I called my basis set the Troy basis set, which would give you absolutely no idea of what was in that basis set, just that I made the basis set. And then somebody else might call it the Cambridge basis set because they did it in Cambridge. And somebody else might number their basis sets. You know, this is basis set number 17. And none of those things actually tell you what's actually in the basis set and how they designed it.

They just give it a name. But then that name is-- you have to know the name in order to be able to specify the basis set. And you have to know what-- you have to sort of memorize, oh, well, this basis that does this. This basis set does this. Or at least remember where to look up what those basis sets do.

So for the next 20 minutes or so, we're going to talk about some of those things. And you'll be annoyed at the fact that these basis sets have these funky, weird names and design principles, I think. But just remember, I don't have to put in 100 numbers. This is the price you pay for not having to put in 100 numbers.

So the first thing that we'll talk about here-- so I'll say that basis sets are typically grouped by row. Of course, the basis sets are going to be different. So hydrogen's going to need a different number of basis functions than carbon. Argon's going to need a different number from carbon or hydrogen just because they have different numbers of valence functions and different numbers of core functions. And I'm also going to introduce a shorthand.

So first of all, I'll note that we already have-- so actually, I'll just introduce the shorthand. So, say, for nitrogen and for the basis set we've talked about so far, the minimal basis set, the smallest basis set I can come up with for nitrogen would need to have the 1s function, the 2s function, and the 2p function. I need at least those functions to just have places to put all of my nitrogen electrons.

Now, I'll get writer's cramp if I have to write all of these things out. So I'm going to develop a shorthand, which is when I have 1s and 2s, I'm just going to denote that as 2s. So I'm not indicating that 2s is the last function. It's indicating that there are two of them. So I have two s-type functions, one that's sort of core like and one that's sort of valence like. But there's two of them. And then I have one p-like function. So I'll put 1p So that's just telling me how many of each of these things I have.

So again, this is the number of s-type functions. This is the number of p-type function. And so then we can go, and we can talk about the one basis set we've dealt with so far, which is the minimal basis set.

So for hydrogen and helium, it's just going to be that 1s function on each of those. And then when I go down to the next row, it looks just like nitrogen. They all are going to need in an s core function and an s valence function and a p valence function. So there'll be 2s, 1p-like. And then I go down one more row for sodium through argon, and I'll need another s function.

So it'll be total of three s-type functions. And then I need two sets of p, a core p and a valence p. So there's 2p. So that's the minimal basis. That's the smallest basis that I could conceive of for any atom. And then we're going to have various ways of trying to make these things more elaborate, more broke. And the most common way to do this is to note that, well, when I bring atoms together, it's the valence functions that actually either contract or expand in order to describe the electrons moving around.

The core functions don't really change very much. And so in order to give the valence functions flexibility to change, it makes sense to add more valence-like functions. So if the valence was s, it makes sense to add another s function. Or if the valence was p, it makes sense to add another p function. And so we get a name here, this concept, which is minimal basis is what's known as a single zeta basis set because it has just one set of valence functions.

It might seem like would make better sense to call it a single valence basis set. But history

made a different choice. It said that it's zeta. I don't know why they chose the name zeta, but they did. So the zeta just means that there's one set of valence functions. And then you can think about making a double zeta.

So you'd take every valence function, just add another valence function that has a different exponent so it would allow more flexibility. So that'll be a Double Zeta basis set, DZ. So for hydrogen and helium, it's all valence. So we'd just get two s functions.

For things in the same row as lithium and neon, the valence functions are in s and a p. And so when we double those, we're going to add another s-type function and another p-type function. So we would go 3s 2p here because we'd add one s function and one p function. And then for sodium through argon, we would have 4s 3p. And then we could go on and talk about, oh, well, what about a triple zeta?

So triple zeta would be 3s for hydrogen helium. And then we add an s and a p for the next row. So it's 4s 3p. And then we add s and p for this, and we end up with 5s 4p. And so again, the idea here is that for carbon, we might have 1s that's the core. And then we'd have 2s and 2p as the first valence and then 3p. And these would respectively be-- so the single zeta or-- and then you include the next cell. And it becomes-- or sorry, single zeta would be all of this, sorry.

Oh, it's the core. And then you could say, all right, well, but then I include another set. And this gives me a double zeta basis set. And then I could include another set, which would give me a triple zeta kind of basis set.

Then you can go on and on and on. So you could just make a quadruple zeta and quintuple zeta basis set. And then this is where the names get annoying.

So the name of the most common minimal basis set is STO-3G. And that stands for Slater-Type Orbitals 3 Gaussians, just in case you're wondering why it's STO-3G. And then for the double zeta basis set, there are ones that are called 3-21G, 6-31G, another one that is somewhat confusingly just called DZ, Double Zeta, as if there was only one such basis set. And then for a triple zeta, you can do 6-311G triple zeta valence. And there are others.

So there are other abbreviations out there as well. But that's how you would do these kinds of-- how you change the zeta, the valence. So questions about those? Yeah?

**AUDIENCE:** So is there [INAUDIBLE] in the example you're including as your additional zetas, like 3s 3p 4s. But would there be an advantage to including things that weren't atomic orbital-like things?

**TROY VAN VOORHIS:** Yes. So there are occasionally times where people include basic functions that don't look like atomic orbitals. So the most common one is to make bond-centered functions. Like say, I want a bond here, and you could put basis functions there.

There's a couple of reasons that becomes undesirable. One is that, then, you can't use a standard basis set because basis sets have to be sort of anchored to the atoms. That's how we catalog them. If you say, oh, I'm going to put one in the middle of the bond, you say, well, it would depend on both atoms involved in the bond and how long the bond was and things like that. So that's a little bit undesirable.

The other one is that unless you pin that bond function down to a particular position, if you let the center of that bond function move, there's actually numerical instability that happens when two Gaussians come on top of each other. And that can sometimes cause calculations-- the computer to choke. So that in practice, we usually stick to increasing the atomic basis sets.

Oh, but there is also one other case. In physics, they are much more fond of using solids. That's sort of one of the things that chemists tend to think about molecules as the example. Physicists tend to think about solids. And in solids, they're much more fond of using plane waves as their basis set.

So they start off with the free-- instead of starting off with atoms as the reference, they start off with free electrons as their reference. And then say, oh, and then we introduce these potentials which are perturbations to the free electrons. So you get plane waves as your basis set. You make linear combinations of plane waves instead of atomic functions. Yeah?

**AUDIENCE:** [INAUDIBLE] atomic orbitals do we need [INAUDIBLE]?

**TROY VAN VOORHIS:** That's exactly the next-- you've hit on exactly the next thing, which is that none of those include any d functions. I mean, I can go up to like 18 zeta, and I would never have any d functions. And on hydrogen and helium, I can go [INAUDIBLE]. I wouldn't even have any p functions. And certainly there are situations where you would want these higher angular momenta.

One of them is if you're in an electric field. So that polarizes electrons and tends to actually

require slightly higher angular momenta functions in order to describe that polarization. And the other one is for just the directionality of bonding. That if you had, like, an SN2 reaction or something like this, where temporarily something was either actually hypervalent or sort of hypervalent, you would need these higher angular momentum functions to describe the hypervalency. And so those are called polarization functions.

So here polarization functions are things that add higher angular momentum and valence. So you need this for things like polarizability, which is why they're called polarization functions. But you also need it for things like hypervalency. And so the nomenclature for polarization functions is also a bit weird.

But the simplest one is that you just add the letter P, so P standing for Polarization to the basis set. And then the idea is that you're just going to add one function with that P. You when you add the P, that means you're adding one function that has one unit higher angular momentum.

So if P was your highest angular momentum, you're going to add a d function. And if s was your highest angular momentum, you're going to add a p function. So for hydrogen and helium, that's going to add a P to lithium. To argon, it's going to add a D.

If you did this also the transition metal atoms, you would note the transition metal atoms have d functions already. So the polarization would add an f function, just adding one higher angular momentum. So then going back to our table up here, let me use my magic, copy the double zeta basis results here-- copy.

There's my double zeta. Now I'm just going to add the polarization to this, so double zeta plus polarization. Now I add my polarization functions. And so for hydrogen and helium, I'm going to add one p function to get a DZP basis set. Lithium to neon, I'm going to add 1 d function. For sodium to argon, I'm going to add 1 d function. And then I can do the same thing with triple zeta.

So if I make a TZP basis set, then I'll add 1p, 1d, and 1d. And I'll note that you could also do a TZ 2p basis set. And then you're adding-- the 2p just indicates that I'm adding even more polarization functions. So I'm adding two sets of d functions instead of just one or two sets of p functions instead of just one. And then these things have names.

So there's a 6-31G star, which is a very common DZP basis set. Or there's also that DZ basis that I mentioned above has a DZP generalization of it. Or you could have for the triple zeta,

the TZVP basis, or the longest basis set name, 6-311G (d, p).

This is where the annoyance started to come in. You're like this is annoying. But it's less annoying than the alternative. So we deal with it. So questions about that? Yep?

**AUDIENCE:** [INAUDIBLE] 1p-- I'm sorry. That 1p, they're just like, we're going add DZ squared. Or [INAUDIBLE]?

**TROY VAN VOORHIS:** It's one set of d functions.

**AUDIENCE:** Oh, OK.

**TROY VAN VOORHIS:** Yes. Yeah. Yeah? Yep?

**AUDIENCE:** Why can't we have a single [INAUDIBLE]?

**TROY VAN VOORHIS:** Oh, we could. But, yeah, we could do that. So part of this has to do with-- and there are a couple of basis sets that do that. This mainly represents the hierarchy of empirically what people have found is most important. So if you just start with a minimal single zeta basis set, you say, well, what's the next most important thing? Usually almost always going to double zeta is more important than the polarization functions.

So very few people bother making single zeta plus polarization. They'll say, OK, we'll do double zeta. And then after you do double zeta, then maybe you want to go to polarization. Or maybe you want to go to triple zeta, depending on what you're doing. Does that makes sense?

**AUDIENCE:** [INAUDIBLE]

**TROY VAN VOORHIS:** It would be 1s 1p for hydrogen, 2s 1d for carbon. It would just be to add that extra polarization function on top of the existing set.

**AUDIENCE:** And would the polarization be the-- which AO did that come from?

**TROY VAN VOORHIS:** What do you mean?

**AUDIENCE:** So for hydrogen, would it be adding the 2p basis or--

**TROY VAN**

**VOORHIS:**

It would be like adding-- now, so for hydrogen, it would be like adding-- yeah. For hydrogen, it would be like the 2p function. But again, when we go back, we remember that we're not actually using the hydrogenic functions. We're using Gaussians to approximate them. And so all it really means is that it has the same angular distribution as a p function. But the radial part is whatever the person who made the basis set decided to make it.

All right, then I have-- so then when I teach this class, I often have clicker questions. So this would have been the clicker question. So we'll do it by show of hands instead of by clicker, since we're small enough.

So here's a test of whether we understood-- I can't show-- well, anyway. So I'll go back and put things up in just a second, if you need them. But my question is, how many basis functions will there be for C60 in a DZP basis set? I guess should say [INAUDIBLE].

So option A is there'll be 60 basis functions. So that would be how many basis functions per carbon atom? One. All right, or 180, that's three per carbon. 360, that's six per carbon. 840 is-- I can't-- anybody do the math better than me? Is that 14? No. No way. Yeah, it's 14. Yeah, 14-- 14 per carbon. This is-- huh?

**AUDIENCE:**

18.

**TROY VAN**

**VOORHIS:**

18. And then this is 21, then, right? Yeah, so we've got, 1, 3, 6, 14, 18, 21 as the number of basis functions per. And I'll go back and throw this up for those who want the table up. We didn't write all that down. So here's our EZP basis set. Carbon's in this group here.

All right, so then now I'll go back. Everybody feel like they got their answer? OK. So by a show of hands, how many people think the answer is A, 60? Nobody, all right. How about B? C? We go most on C, OK. D? I don't know, E? F?

Everybody says C. C is not the correct answer. And every year people miss this question. So the key thing to get this right-- and we're going to re-poll in a minute here. But the key thing to get this right was actually the question that you asked, which is that you said, when I add a d function, do I add DZ squared, or do I-- and I said, no, you add a set of d functions, which is how many functions in d? 5.

All right, so now we'll go back and look. All right, but don't feel bad. Every year, the majority of people choose that same answer. So now let's redo our math.

**AUDIENCE:** I've never gotten it right.

**TROY VAN** Never gotten it right? OK.

**VOORHIS:**

[INTERPOSING VOICES]

All right, so now we'll go through. OK, how many people think A? How many B? Nobody's going to say C because I already told you that was wrong.

How many people say D? All right, good, you're all right. Yes, the answer is D. In a DZP basis there are 14 basis functions per carbon atom. Now, you can already start to see even for this DZP basis, which isn't the biggest basis we've talked about, for C60, you've got 840 basis functions. That's a lot of basis functions.

So that's why we're really glad that the computer's doing the work and not us because 840 integrals-- well, 840 by 840 matrices are hard to diagonalize by hand. Doing all those integrals is really, really a pain. So then going beyond those basis-- so those two things are the key basis set ideas.

I'll just touch on a couple of other things before we finish up. So the first thing is diffuse functions. So occasionally, if you have anion, so you have an extra electron, that extra electron is not very well described by the valence because negative electrons tend to spread out a great deal. Bob is one of the world experts in Rydberg states, which is where these electrons spread out a really, really significant amount. And so these Gaussian functions that decay quickly don't describe those well. And so you have to add a Gaussian with a small value of alpha. And it's mostly useful for anions.

So you add these in order to describe anions better. And the notation here is either you use the phrase aug or plus to the basis set. So if the basis set has the word aug in front of it, that means it has some diffuse functions on it in order to describe those weakly bound electrons. Or the plus indicates it has some of those functions in it. And again, those are added with the same angular momentum as the valence.

They're not usually polarization and diffuse at the same time. And then the final thing is you can generalize these ideas to transition metals. It's a little bit hazy because a lot of this is predicated on us knowing what the valence of the element is. And for transition metals, it's like,

well, is s in the valence? Yeah, probably. Is p in the valence? Well, maybe.

I don't know. D is definitely in the valence. But depending on what you think the valence is, your definition of some of these things is a little bit fuzzier. Sometimes you add a p function. Sometimes you don't when you go from double zeta to triples for a transition metal.

But the overall idea here is that we're trying to approach this particular limit, which is known as the complete basis set limit, which is the result that you would hypothetically get with an infinite number of atomic orbitals. So you just crank up, include all the atomic orbital in the universe. You would get to an answer that would be the right answer. And we're just trying to asymptotically approach that by making our basis sets bigger and bigger until we run out of steam.

All right, so that's everything about basis sets. And tomorrow-- or not tomorrow, next Monday we will talk about how you compute the energy. All right, Happy Thanksgiving.