

# An Introduction to Basis Sets:

---

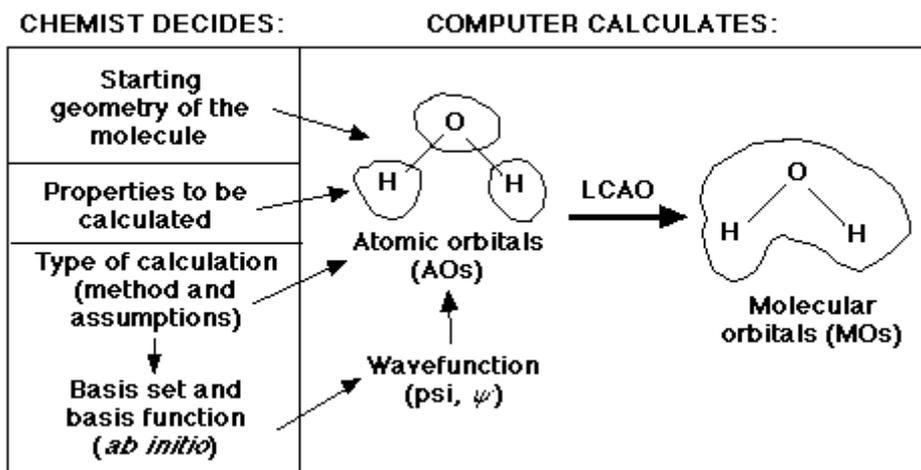
The "Holy Grail" of computational chemistry is the **calculation of the molecular orbitals (MOs)** for a given molecule. **IF** we can calculate the MOs for a molecule, **THEN** we can know lots of things about the molecule, including its:

- energy
- electron density
- electrostatic potential
- transition state (if any)
- frequency

However, calculating MOs is not that easy. The computer, however, does this for us. But, you, the chemist, must tell the computer some information first. **Input files** for all of the major computational chemistry packages contain these three parts:

1. Geometry which includes
  - Bond lengths
  - Bond Angles
  - Dihedrals
2. Kind of Calculations:
  - Single point energy
  - Frequency
  - Transition state
  - Electron density
  - Electrostatic potential
3. Starting set of mathematics and approximations
  - Calculation method (molecular mechanics, semiempirical, or ab initio)
  - Type(s) of approximation (Hartree-Fock, Moller-Plesset, etc.)
  - Basis Set Approximation

The graphic below "captures" the essence of what is the responsibility of the chemist and what is the responsibility of the computer:



In this lab we will be looking at the third ingredient, basis sets.

A molecular-orbital theory calculation is a mathematical expression of an electron in a molecule. Although there are many types of molecular-orbital functions, in this lab we will only look at the Slater Type Orbitals (STOs) and the Gaussian Type Orbitals (GTOs).

Although there is not a major difference in these two methods when calculating small molecules, major discrepancies arise for larger molecules of 30 or more atoms. STOs require more calculating, which takes tremendous amounts of time, however their calculations have been found to be more accurate than GTOs. On the other hand, GTOs, although less accurate, are much faster to calculate than STOs. This forced scientists to compromise time or accuracy.

Eventually, scientists realized that by adding several GTOs, they were able to mimic the STOs accuracy. In fact, as the number of GTOs used increased, the better they were able to model the STO equation.

When using GTOs to model STOs, the new equations are given a new name. They are identified as STO-kG equations where k is a constant that represents the number of GTOs used. For instance, two common equations are the STO-3G and the STO-6G in which 3 and 6 GTOs are used respectively.

## Objective for the "Comparing STOs and GTOs" Lab Activity:

In this lab you will be required to create a spreadsheet that will allow you to compare a set of STOs and GTOs. This will be done for a hydrogen atom which has only one electron. You will make the assumption that the Slater Type Orbital answers are the most accurate and then attempt to model your Gaussian Type Orbitals to fit the STOs.

## Objectives for the "Comparing STOs and GTOs" Lab Activity:

1. To gain an understanding of the relationship between the size of the molecule and the CPU time.
2. Define a formula that calculates the number of basis functions required for molecules with an (x) number of carbons and a (y) number of hydrogens.

# Background Reading for Basis Sets:

## Key Points

The objective of [computational chemistry](#) is to be able to mathematically express the properties of molecules. One of the most important properties is that of the [molecular orbitals](#). Scientists use something called a [basis set](#) to approximate these orbitals.

There are two general categories of basis sets:

### Minimal Basis Sets

A basis set that describes only the most basic aspects of the orbitals

### Extended Basis Sets

A basis set that describes the orbitals in great detail

In the most general sense, a basis set is a table of numbers which mathematically estimate where the electrons can be found. An American physicist named [J. C. Slater](#) developed algorithms by fitting [linear least-squares](#) to data which could easily be calculated. Suddenly, basis sets became much easier to express. Slater's equation is known as the Slater Type Orbital or (STO). The equation is given below.

The general expression for a basis function is given as:

$$\text{Basis Function} = N * e^{(-\alpha * r)}$$

where:

N = Normalization constant

alpha = Orbital exponent

r = Radius in angstroms

This expression given as a Slater Type Orbital equation is:

$$STO = \frac{\xi^3}{\pi^{0.5}} e^{(-\xi r)}$$

At this point, the STO equation looks like a great solution to finding the basis set of a molecule. Although the STO equation is a wonderful approximation for the molecular orbitals, the problem lies in the calculations. You see, calculating the STO of a molecule requires enormous amounts of work. This uses way too much time, even for super computers. Instead, a scientist by the name of [S. F. Boys](#) developed a method of using a combination of Gaussian Type Orbitals in order to express the STO equation. The following is the Gaussian Type Orbital (GTO) equation.

$$GTO = \frac{2\chi}{\pi^{0.75}} e^{-\chi r^2}$$

Notice that the difference between the STO and GTO is in the "r." The GTO squares the "r" so that the product of the gaussian "primitives" (original gaussian equations) is another gaussian. By doing this, we have an equation we can work with and so the equation is much easier. However, the price we pay is loss of accuracy. To compensate for this loss, we find that the more gaussian equations we combine, the more accurate our equation. In Basis Set Lab, you will be looking at a STO-3G basis set:

- The "STO" tells you that you are attempting to represent a STO equation for your molecule.
- The 3G tells you that in order to do this you are combining 3 gaussian primitives.

In this case, STO-NG, the number "N" before the "G" is the number of gaussian primitives that are used to simulate the STO equation. Remember, the bigger your N, the more accurate your results.

All basis set equations in the form STO-NG are considered to be "minimal" basis sets. (Remember our definition of minimal.) The "extended" basis sets, then, are the ones that consider the higher orbitals of the molecule.

The way to think about extended basis sets is to think of cleaning the kitchen floor. First you must sweep it to get all the big pieces of dirt. Then you go back and mop it to get the little bits of grime. Then you go over again and wax it to make sure everything is cleaned off and the floor is shiny. That is what you are doing to the molecules. Using the minimal basis set is like sweeping the floor. You are only describing the big properties. Using the extended basis sets is like mopping and waxing. You are fine tuning your description.

One way to mop up the little grime in your molecule descriptions is to use a method called "Split-Valence Basis Sets." The theory behind this method is that it would be extremely tedious to attempt calculating the expression for every single atomic orbital. Instead, by combining an

expression for an orbital larger and smaller than the one we are looking at, we can approximate the orbital we are interested in. This method entails combining two or more STO's in order to describe an orbital. The STO's differ only by the value of their exponent ( $\zeta$ ) and can be manipulated by a constant "d" so that you can size the orbital to what you are looking for.

Now we need to wax the molecule. This is done by taking into account the Polarization effect. The polarization effect is the resulting effects when two atoms are brought close together. Taking into account the polarization effect means accounting for the d and f-shells by adding STOs of higher orbital angular momentum. The asterisk is the symbol used when taking into account polarization. One asterisk symbolizes accounting for the d-shell and two asterisks symbolizes the f-shell. If an asterisk is in parentheses symbolizes that polarization functions are added only to second-row elements. Standard polarization basis sets are 3-21G\* and 6-31G\*.

It is important that you understand how to read each basis set. For example, let us consider the 6-311G\* basis set. The 6 represents the 6 gaussian primitives used to calculate the s-shell, the 3 represents the number of GTOs for one of the sp-shells and each 1 represents the number of GTOs for the other two sp-shells. The \* represents the consideration of the d-shell. Other standard basis sets are STO-3G, 3-21G, 3-21G\*, 4-31G, 6-31G, 6-31G\*, 6-311G, and 6-311G\*.

When dealing with anions, the use of diffuse functions is recommended. Indeed, this is the case when the electron density of a molecule is far removed from its nuclear center. Diffuse functions are also useful for systems in an excited state, for systems with low ionization potential, and systems with some significant negative charge attached. The presence of diffuse functions is symbolized by the addition of a plus sign, +, to the basis set designator: 6-31G+ or 6-31+G. Again, a second +, such as 6-31++G, implies diffuse functions added to hydrogens. The use of doubly-diffuse basis sets is especially useful if you are working with hydrides.

When choosing a basis set, choose the best basis set for your time limits. Remember, the better the basis set, the more time it will take to calculate!

Now, let's look at a typical basis set printout.

Here is a dump for H and C, showing data for the STO-2G basis set. You can see that for the hydrogen, 2 gaussian primitives are used to construct

the s orbital. The first value of 1.309 is the orbital exponent, and the 0.430 is the contraction coefficient. With carbon, you see the addition value, so the first is the orbital exponent, the second is the s-part of the sp-hybrid, and the third part is the p-part of the sp-hybrid.

---

STO-2G

BASIS="STO-2G"

```
H  0
S  2  1.00
    1.30975638    0.43012850
    0.23313597    0.67891353
****
C  0
S  2  1.00
    27.38503303    0.43012850
    4.87452205    0.67891353
SP 2  1.00
    1.13674819    0.04947177    0.51154071
    0.28830936    0.96378241    0.61281990
```

\*\*\*\*\*

---

Here is a dump of the 6-31G\* for H and C. For H, there is only one valence electron, and it is represented by two orbitals, one constructed of 3 primitives and the other with 1 primitive. With C, the s-orbital is a core orbital, and is represented by 6 gaussian primitives. The sp-orbital, on the other hand, is a valence orbital, and is represented by two orbitals, one with 3 gaussians and the other with 1. Since this is a "\*", or polarized basis set, notice that there is some "d" character to the carbon atom.

---

6-31G\*

BASIS="6-31G\*"

```
H  0
S  3  1.00
    18.73113700    0.03349460
    2.82539370    0.23472695
    0.64012170    0.81375733
S  1  1.00
    0.16127780    1.00000000
****
```

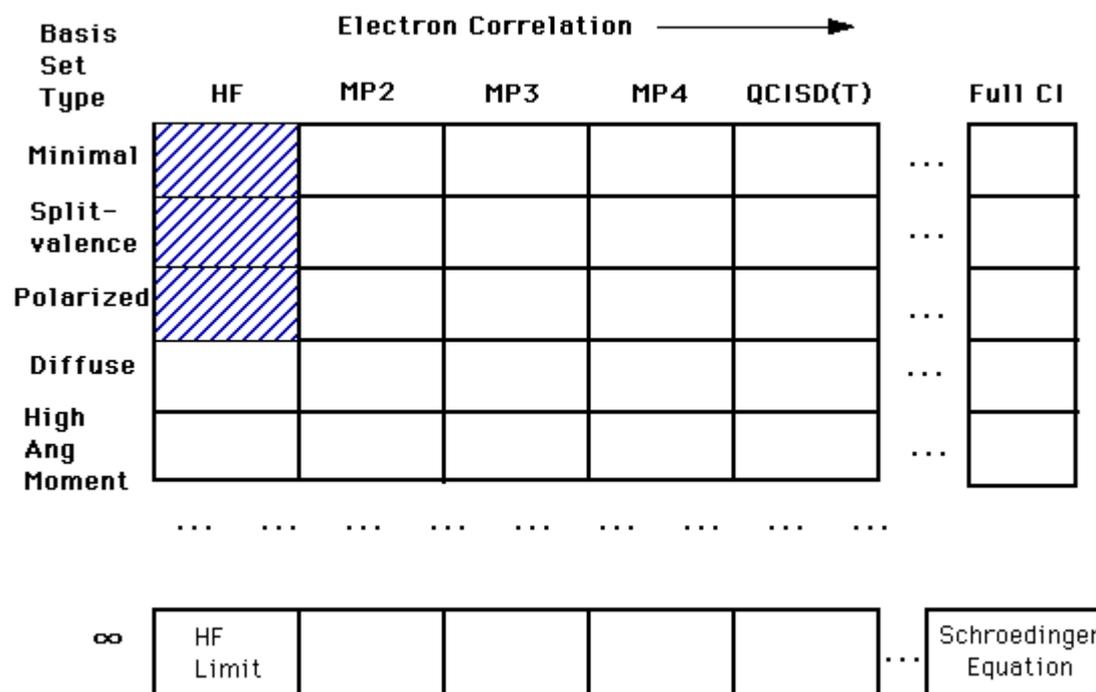
```

C 0
S 6 1.00
  3047.52490000 0.00183470
  457.36951000 0.01403730
  103.94869000 0.06884260
  29.21015500 0.23218440
  9.28666300 0.46794130
  3.16392700 0.36231200
SP 3 1.00
  7.86827240 -0.11933240 0.06899910
  1.88128850 -0.16085420 0.31642400
  0.54424930 1.14345640 0.74430830
SP 1 1.00
  0.16871440 1.00000000 1.00000000
D 1 1.00
  0.80000000 1.00000000
****

```

---

The relationship between basis sets and accuracy is represented in the diagram below. Our ultimate goal is to calculate an answer to the Schroödinger's Equation (right bottom corner). However, we are still a long way from being able to complete this calculation. Right now we are in the top left corner of the chart. In that first box, we are treating each electron independently of the others. As you move across to the right, you find calculations that account for the interactions of electrons. As you move down the column you find more complex and more accurate basis set calculations. You will only be expected to understand the shaded regions.



There are other trade-offs for using each type of basis set. The more complex basis sets are more accurate but, they use up a great deal of computing time. Whenever you run a computational chemistry calculation you will be using time on a computer. Normally, the computer that will be running the calculation will be one that is shared with many other people doing other calculations. Thus, it is important that you act responsibly when choosing which basis set to use. You should pick one that is efficient for your use. This means that you should consider how much time it will take to run the molecule and use the basis set that will run the fastest without compromising your desired level of accuracy.

# "Comparing STOs and GTOs" Lab Procedure

---

## Procedure:

- For this lab, you will need access to a spreadsheet.

## Part I:

Below is the equation for a Gaussian Type Orbital where alpha ( $\alpha$ ) is 0.4166 and r is the radius from the nucleus.

$$GTO = \frac{2\alpha}{\pi^{0.75}} e^{-\alpha r^2}$$

## Data Entry:

1. At the top of the spreadsheet, label the first cell "alpha." In the next cell define the constant.
2. Leaving a few rows blank, identify one cell "Radius." This is where we will give the equation various values of r.
3. Beneath Radius, you are going to create a list of values from -3 to +3. Divide the values into 0.01 increments. In order to do this, you are going to use a "Fill" command. First, type a -3 in the first cell beneath the "Radius" label. Next, highlight the -3 and all the cells down to at least the 600th line. In the menu bar, look under Edit to find the "Fill" command. You want to choose the "Series" option. This will then prompt you for the increment by which you want your column to increase. The default is set to 1, but you will type 0.01. Click on OK. If your values do not reach +3, then repeat these steps until you reach this value.
4. Next to the Radius column, title the next cell "Gaussian" or "GTO."
5. Beneath this label you will need to enter your GTO function as given above. Remember to identify alpha and the radius by their cell names and that  $\pi = \text{PI}()$ ,  $e^x = \text{EXP}(x)$ .

## Hints:

- When referring to a cell, make sure you use, \$, before the column or row if it remains constant.

- The first value should be 0.00869.

	A	B
1	alpha	0.4166
2		
3		
4		
5		
6		
7		
8	Radius	GTO
9	-3	0.00869673
10	-2.99	0.00891648
11	-2.98	0.00914102
12	-2.97	0.00937044
13	-2.96	0.00960481
14	-2.95	0.00984423
15	-2.94	0.01008877
16	-2.93	0.01033853
17	-2.92	0.01059358
18	-2.91	0.01085403

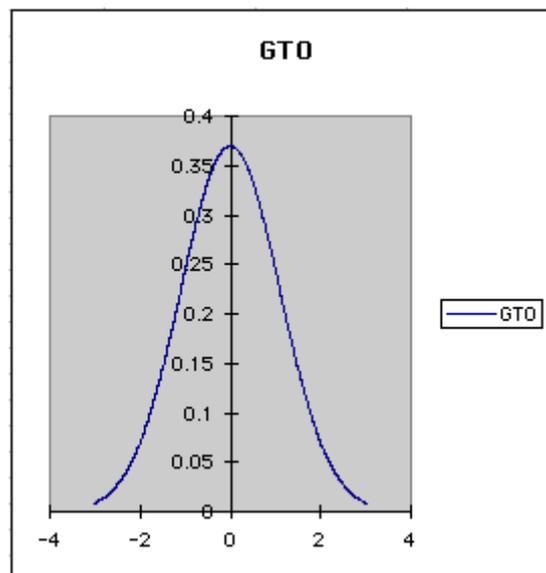
6. Use a "Fill" command, (this time use "Down" instead of "Series") in order to get values for all the radii.

### Graphing:

1. Highlight the lists "Radius" and "GTO" including labels.
2. In the toolbar, click on the button with a wand and bar graph.
3. Select the area on the spreadsheet where you want the graph to appear.
4. Follow the steps to create a line graph.

### Hints:

- In Excel, make sure you choose the XY scatter plot and the second type of line graph.



## Part II:

Now we will compare this approximation to the STO. Below is the STO equation. The "squiggle" is a Greek character standing for an orbital coefficient, similar to alpha in gaussian curves. For

hydrogen, the "squiggle" (better known as "xi",  $\xi$ ) is equal to 1.24.

$$STO = \frac{\xi^3}{\pi^{0.5}} e^{-\xi r}$$

1. Under alpha, type "xi." In the next cell, type the value.
2. Now, label the cell next to your GTO heading "Slater" or "STO." In the cell beneath the heading, enter the Slater equation. Fill down your column.

	A	B	C
1	alpha	0.4166	
2	xi	1.24	
3			
4			
5			
6			
7			
8	Radius	GTO	STO
9	-3	0.00869673	36.6628631
10	-2.99	0.00891648	36.2110506
11	-2.98	0.00914102	35.764806
12	-2.97	0.00937044	35.3240607
13	-2.96	0.00960481	34.8887468
14	-2.95	0.00984423	34.4587976
15	-2.94	0.01008877	34.0341468
16	-2.93	0.01033853	33.6147291
17	-2.92	0.01059358	33.2004801
18	-2.91	0.01085403	32.7913361
19	-2.9	0.01111995	32.3872341

3. With the "Radius," "GTO," and "STO" columns highlighted, create a new graph to compare the GTO and STO calculations.
  - Notice the major discrepancy for small radii values. Also, notice how the approximations become similar for larger radii.

## Part III:

Now we would like to use several gaussians to model the hydrogen orbital.

To get all of the numerical constants required, we will have to use a basis set as a reference. These are stored in a large database, from which you can "order" a specific basis set.

1. Go to the *Gaussian Basis Set Order Format* <http://www.emsl.pnl.gov:2080/forms/basisform.html>. Your browser should look like

this:

## EMSL Gaussian Basis Set Order Form

[Basis Sets:](#)

STO-2G	
STO-3G	
STO-6G	
STO-3G*	
3-21G	
3-21++G	
3-21G*	
3-21++G*	
3-21GSP	

[Elements:](#)

**Output Options**

[Optimize General Contractions](#)  [ECPs](#)

[Show Supported Elements](#)

[Email:](#)

- Order your constants:
  - Select the STO-3G basis set. By selecting the STO-3G we are requesting to receive constants for 3 gaussians.
  - Next type an h for hydrogen in the "Elements" blank.
  - Don't put in an email!!
  - Click Submit.

The result should be two columns of three numbers each. This is a basis set for hydrogen. The left list is alpha values, and the right list is coefficients which will be used to sum up the three gaussians.

- Copy these constants so you can enter them into your spreadsheet.
- At the top of your spreadsheet, next to the other constants, label each constant and define the value in the next cell.
- Next to the STO column, put three new columns, labelled "g1," "g2," and "g3".
- Under each new label, enter the gaussian equation using the values of alpha from the basis set and the radius as needed.
- Fill down the column, and repeat the procedure for the other two columns.
- Change the first number under Radius to be 0. When you do this, the whole list should change to go from 0 to 6 instead of -3 to 3. Also, the values in your gaussian lists should change, since we are computing them for different radii now. Check your values against those of your classmates to make sure you are all getting correct answers.

\*\*\*Since we looked at the equations on both sides of the nucleus before, you should have observed that it is exactly the same on both

sides. This is important! We are only interested in the positive side of the graph now, since we know that both sides will be exactly the same (and since negative and positive do not have any true physical meaning here!).\*\*\*

9. Compare the answers from each new gaussian equation to those of the slater equation.

Notice that no single one of these three gaussians is an approximation on its own; they must be combined. The way of combining them is called "Linear Combination of Atomic Orbitals" (LCAO). The other list of numbers in the basis set are coefficients for adding up the three gaussians. They are used in the following equation.

$$LCAO = d_1g_1 + d_2g_2 + \dots$$

1. Create another column next to the three gaussians, labelled "LCAO".
2. Put the equation, referring to the appropriate constants, in the first cell.
3. Fill the column.

	A	B	C	D	E	F	G
1	alpha	0.4166		g1	a		
2	xi	1.24		g2	b		
3				g3	c		
4							
5							
6							
7							
8	Radius	GTO	STO	g1	g2	g3	LCAO
9		0	0.36957241	0.88848664			

4. Create a graph of all three new gaussians and the LCAO.
5. To make this contrast clearer, we are going to examine several gaussian approximations and compare them to the STO.
6. Label three new columns for the two equations needed for a STO-2G model, and the STO-2G model itself.
7. Do the same for a STO-6G
8. Now go back to the order form for basis sets and find the basis sets, both alpha and d values, for STO-2G and STO-6G.
9. Put these values into your spreadsheet as before.
10. Set up each column with the necessary equations.
11. In two new columns, use the LCAO equation for both of the 2G and 6G approximations.
12. Graph the 1 gaussian (the first one created), 2 G, 3 G, 6G, and STO all on a single graph.
 

\*\*DO NOT put all of the gaussians for each approximation on there, ONLY THE SUMMATIONS.\*\*

# Basis Sets, Functions, and CPU Time

## Lab Procedure:

Procedure:

Group Part:

- In order to conduct this lab, you will need a spreadsheet and a Waltz Interface.
- The molecules we are using in this lab are alkanes. They have the formula  $C_nH_{2n+2}$ . Alkanes are made up of  $n$  carbons linked together in a line with hydrogens jutting off every carbon. All of the bonds in alkanes are single bonds. The entire class will be collecting data to use for analysis. Split up the molecules so that each group will run a geometry optimization for two molecules. The more molecules the class runs, the better your results. When dividing up the molecules, skip every third alkane in the sequence. (ex.  $CH_4$ ,  $C_2H_6$ ,  $C_4H_{10}$ )
- You should set your spreadsheet up as follows:

	A	B	C	D	E	F	G	H	I
1	Molecule	# of C	# of H	STO-3G		3-21G		6-31G	
2				# of Basis Func	CPU Time	# of Basis Func	CPU Time	# of Basis Func	CPU Time
3									
4									
5									
6									
7									
8									
9									
10									

- You will use Waltz to run each molecule THREE times: once with an STO-3G, once with a 3-21G, and once with a 6-21G.
- Once your run has completed, you will need to extract two pieces of information from the output file,

1. The number of basis functions used.

```

      INTERNUCLEAR DISTANCES      (ATOMIC UNITS)      ANGSTROMS
      *****
INPUT UNIT OF LENGTH              =  ANGS
TOTAL NUMBER OF SHELLS            =  5
TOTAL NUMBER OF BASIS FUNCTIONS   =  9
NUMBER OF ELECTRONS               =  6
CHARGE OF MOLECULE                =  0
STATE MULTIPLICITY                =  1
NUMBER OF OCCUPIED ORBITALS       =  3
TOTAL NUMBER OF ATOMS             =  1
NUMBER OF SUBGROUPS              =  1
    
```

2. The CPU time. (Remember that the CPU time is different from the Wall time. The CPU time is the time it took to complete the run on the computer. The Wall time is the time it took for you to get your answer back after you clicked the execute button.)

TIME IS DISCO : 2.46 SEC  
Disco Done

### Class Part:

3. Once the class has collected all the data, compile it all into one spreadsheet.
4. Graph the number of basis functions versus the number of carbons in each molecule. Do this for each of the 3-21G and 6-21G basis sets. You should have two graphs. What do you observe about the shape of each graph?
5. Try to write two equations (one for each basis set) which give the number of basis functions (bf) for a certain number of carbons (c) plus a certain number of hydrogens (h). (Clue: Use [systems of equations](#))

$$y \text{ bf} = (a * c) + (b * h)$$

What are the coefficients (a) and (b) in this equation (both are integers)?

6. Let's test your equation. Calculate the number of basis sets needed to calculate the following molecules.
7. Now graph the CPU time, in seconds, versus the number of basis functions calculated. What relationship do you see? Describe what happens to the CPU time required as the molecule gets extremely large.
8. Try to predict what the CPU time will be for the molecules which haven't been calculated yet. Each group should take one molecule that had been skipped earlier and one that is larger than any that have already been tried. How accurate do you think your predictions are? Do you think that the prediction for a molecule between two that we calculated is any more or less accurate than the prediction for a molecule not in between two that we calculated? If so, why? Discuss the difference between these predictions with your group. Record your predictions.
9. Now, each group should run the calculations for their two molecules. Record the data and share the results with the class.

## Questions for "Comparing STOs and GTOs" Lab Activity:

### Part I:

1. What do you notice about the graph on either side of the nucleus?
2. Are the two sides similar?
3. If so, what does that show?

### Part II:

1. What do you notice about the shape of the three gaussians?
2. How similar are they?
3. How similar is each one to the simple gaussian we first created?
4. How do they relate to the LCAO? Discuss these questions and compare results with you lab parter and classmates?

### Part III:

Remembering that the STO is the most accurate approximation of the waveform:

1. How accurate is each of the gaussian approximations?
2. Which approximation is the most accurate?
3. Which one takes the most time to compute?
4. How does accuracy relate to number of gaussians used?
5. Which is more important, the time required or the accuracy?
6. How should we decide which approximation to use for a specific project?

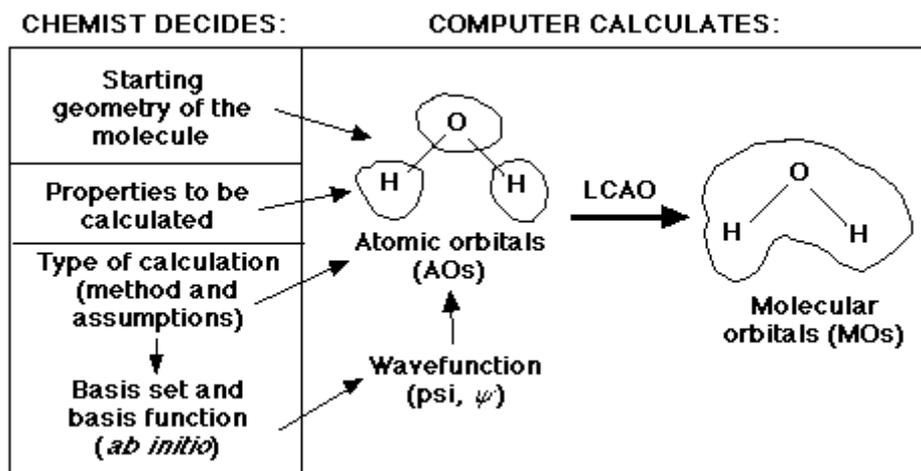
## Questions for Basis Sets, Functions, and CPU Time Lab Activity:

### Class Part:

1. How accurate were your predictions about the number of basis functions? Were all of the predictions of basis functions equally accurate or inaccurate?
2. How accurate were your predictions about CPU time? Were all of the predictions of CPU time equally accurate or inaccurate?
3. Did the patterns of accuracy/inaccuracy fit what you expected? How so? Why or Why not?
4. Which predictions of CPU time were most accurate? Why?
5. Which predictions of CPU time were least accurate? Why?
6. How could you have made these predictions more accurate without doing calculations on a molecule for which  $n$  is larger than what we have already done?
7. Predicting values for data points which are inside a set of data that you already have is called **interpolation**. Predicting values for data points which are outside the data you have is called **extrapolation**. From your experiment, which one do you think gives more accurate predictions in general? Discuss why there is a difference in accuracy between the two types of predictions.

# Background Reading for Basis Sets:

As a quick review, remember that scientists are mainly interested in the properties of molecules. One characteristic of a molecule that explains a great deal about these properties is their molecular orbitals. Recall the following diagram regarding what the scientist must know in order to calculate the molecular orbitals.



One of the three major decisions for the scientist is which [basis set](#) to use. There are two general categories of basis sets:

## Minimal basis sets

a basis set that describes only the most basic aspects of the orbitals

## Extended basis sets

a basis set with a much more detailed description

Basis sets were first developed by [J.C. Slater](#). Slater fit [linear least-squares](#) to data that could be easily calculated. The general expression for a basis function is given as:

$$\text{Basis Function} = N * e^{(-\alpha * r)}$$

where:

N = normalization constant

alpha = orbital exponent

r = radius in angstroms

This expression given as a Slater Type Orbital (STO) equation is:

$$STO = \frac{\xi^3}{\pi^{0.5}} e^{-\xi r}$$

Now it is important to remember that STO is a very tedious calculation. [S. F. Boys](#) came up with an alternative when he developed the Gaussian Type Orbital (GTO) equation:

$$GTO = \frac{2 \chi}{\pi^{0.75}} e^{-\chi r^2}$$

Notice that the difference between the STO and GTO is in the "r." The GTO squares the "r" so that the product of the gaussian "primitives" (original gaussian equations) is another gaussian. By doing this, we have an equation we can work with and so the equation is much easier. However, the price we pay is loss of accuracy. To compensate for this loss, we find that the more gaussian equations we combine, the more accurate our equation.

All basis set equations in the form STO-NG (where N represents the number of GTOs combined to approximate the STO) are considered to be "minimal" basis sets. (Remember our definition of minimal.) The "extended" basis sets, then, are the ones that consider the higher orbitals of the molecule and account for size and shape of molecular charge distributions.

There are several types of extended basis sets:

- [Double-Zeta, Triple-Zeta, Quadruple-Zeta](#)
- [Split-Valence](#)
- [Polarized Sets](#)
- [Diffuse Sets](#)

### Double-Zeta, Triple-Zeta, Quadruple-Zeta

Previously with the minimal basis sets, we approximated all orbitals to be of the same shape. However, we know this is not true. The **double-zeta basis set** is important to us because it allows us to treat each orbital separately when we conduct the [Hartree-Fock calculation](#). This gives us a more accurate representation of each orbital. In order to do this, each atomic orbital is expressed as the sum of two Slater-type orbitals (STOs). The two equations are the same except for the value

of  $\xi$  (zeta). The zeta value accounts for how diffuse (large) the orbital is. The two STOs are then added in some proportion. The constant 'd' determines how much each STO will count towards the final orbital. Thus, the size of the atomic orbital can range anywhere between the

value of either of the two STOs. For example, let's look at the following example of a 2s orbital:

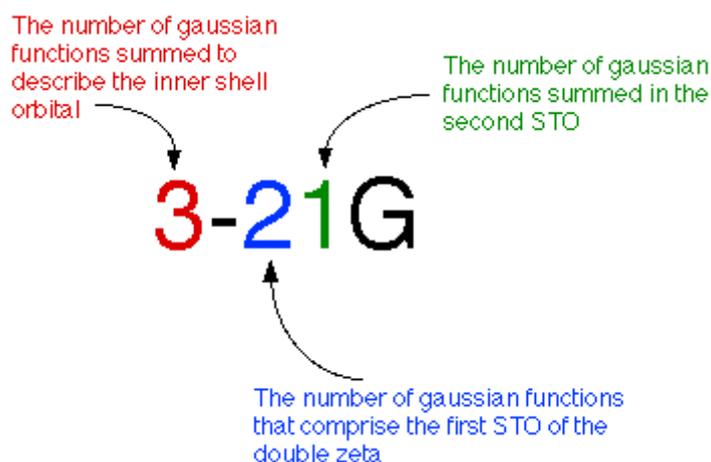
$$\Phi_{2s}(r) = \underbrace{\Phi_{2s}^{STO}(r, \xi_1)}_{\text{Slater Orbital 1}} + \underbrace{d}_{\text{Constant}} \underbrace{\Phi_{2s}^{STO}(r, \xi_2)}_{\text{Slater Orbital 2}}$$

In this case, each STO represents a different sized orbital because the zetas are different. The 'd' accounts for the percentage of the second STO to add in. The linear combination then gives us the atomic orbital. Since each of the two equations are the same, the symmetry remains constant.

The triple and quadruple-zeta basis sets work the same way, except use three and four Slater equations instead of two. The typical trade-off applies here as well, better accuracy...more time/work.

## Split-Valence

Often it takes too much effort to calculate a double-zeta for every orbital. Instead, many scientists simplify matters by calculating a double-zeta only for the [valence orbital](#). Since the inner-shell electrons aren't as vital to the calculation, they are described with a single Slater Orbital. This method is called a **split-valence basis set**. A few examples of common split-valence basis sets are 3-21G, 4-31G, and 6-31G.



An example is given below. It is strongly encouraged that you take part of the class time to walk through this example with your class. It will be a tremendous help to the students understanding of this subject.

Here we are using a 3-21G basis set to calculate a carbon atom. This means we are summing 3 gaussians for the inner shell orbital, two gaussians for the first STO of the valence orbital and 1 gaussian for the second STO. This is the output file from the [gaussian Basis Set Order Form](#) for carbon given a 3-21G basis set.

```

CARBON (6S,3P)->[3S,2P]
S 3
1 172.256000 0.617669000E-01
2 25.9109000 0.358794000
3 5.53335000 0.700713000
L 2
1 3.66498000 -0.395897000 0.23640000
2 0.770545000 1.21584000 0.860619000
L 1
1 0.195857000 1.00000000 1.00000000

```

Here is another common method of displaying data. Notice the numbers are labeled so it is easy to match this data with the corresponding data in the output file.

Atom	$\alpha_{1si}$	$d_{1si}$	$\alpha_{2si} = \alpha_{2pi}$	$d_{2si}$	$d_{2pi}$	$\alpha'_{2si} = \alpha'_{2pi}$
C	172.256000	0.617669000E-01	3.66498000	-0.395897000	0.23640000	0.195857000
	25.9109000	0.358794000	0.770545000	1.21584000	0.860619000	
	5.53335000	0.7007130000				

Once you have retrieved a basis set output file, you can use these numbers to calculate your equations. For a carbon, you will need three equations: 1s orbital, 2s orbital, and 2p orbital.

This equation combines the 3 GTO orbitals that define the 1s orbital.

$$\begin{aligned}
\Phi_{1s}(r) &= \sum_{i=1}^3 d_{1si} \Phi_{1s}^{GF}(r, \alpha_{1si}) \\
&= 0.6176 \Phi_{1s}^{GF}(r, 172.256) + 0.3587 \Phi_{1s}^{GF}(r, 25.910) \\
&\quad + 0.7007 \Phi_{1s}^{GF}(r, 5.533)
\end{aligned}$$

This equation combines the 2 GTO orbitals that make up the first STO of the double-zeta, plus the 1 GTO that represents the second STO for the 2s orbital.

$$\begin{aligned}
\Phi_{2s}(\mathbf{r}) &= \sum_{i=1}^3 d_{2si} \Phi_{2s}^{GF}(\mathbf{r}, \alpha_{2si}) + d'_{2s} \Phi_{2s}^{GF}(\mathbf{r}, \alpha'_{2s}) \\
&= -0.395 \Phi_{2s}^{GF}(\mathbf{r}, 3.664) + 1.215 \Phi_{2s}^{GF}(\mathbf{r}, 0.771) \\
&\quad + 1.000 \Phi_{2s}^{GF}(\mathbf{r}, 0.195)
\end{aligned}$$

This equation combines the 2 GTO orbitals that make up the first STO of the double-zeta, plus 1 GTO that represents the second STO for the **2p** orbital.

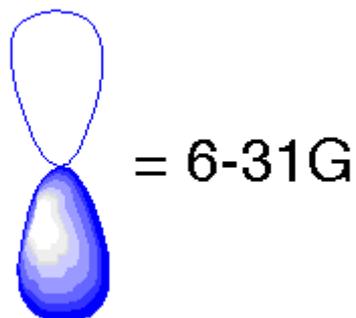
$$\begin{aligned}
\Phi_{2p}(\mathbf{r}) &= \sum_{i=1}^3 d_{2pi} \Phi_{2p}^{GF}(\mathbf{r}, \alpha_{2pi}) + d'_{2p} \Phi_{2p}^{GF}(\mathbf{r}, \alpha'_{2p}) \\
&= 0.236 \Phi_{2p}^{GF}(\mathbf{r}, 3.664) + 0.860 \Phi_{2p}^{GF}(\mathbf{r}, 0.771) \\
&\quad + 1.000 \Phi_{2p}^{GF}(\mathbf{r}, 0.195)
\end{aligned}$$

Now, using these three equations, we can calculate the LCAO for the carbon atom.

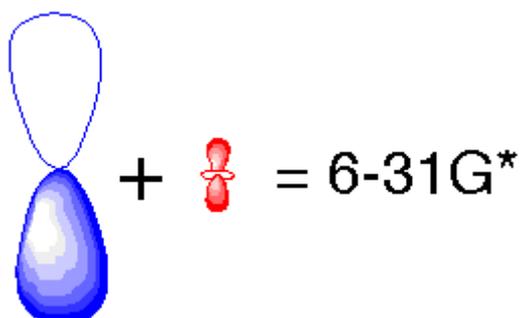
### Polarized Sets

In the previous basis sets we have looked at, we treated [atomic orbitals](#) as existing only as 's', 'p', 'd', 'f' etc. Although those basis sets are good approximations, a better approximation is to acknowledge and account for the fact that sometimes orbitals share qualities of 's' and 'p' orbitals or 'p' and 'd', etc. and not necessarily have characteristics of only one or the other. As atoms are brought close together, their charge distribution causes a polarization effect (the positive charge is drawn to one side while the negative charge is drawn to the other) which distorts the shape of the atomic orbitals. In this case, 's' orbitals begin to have a little of the 'p' flavor and 'p' orbitals begin to have a little of the 'd' flavor. One asterisk (\*) at the end of a basis set denotes that polarization has been taken into account in the 'p' orbitals. Notice in the graphics below the difference between the representation of the 'p' orbital for the 6-31G and the 6-31G\* basis sets. The polarized basis set represents the orbital as more than just 'p', by adding a little 'd'.

Original 'p' orbital

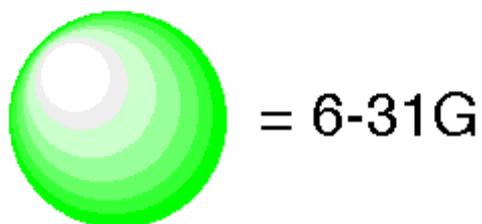


Modified 'p' orbital

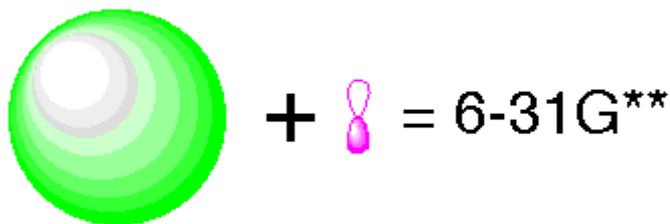


Two asterisks (\*\*) means that polarization has taken into account the 's' orbitals in addition to the 'p' orbitals. Below is another illustration of the difference of the two methods.

Original 's' orbital

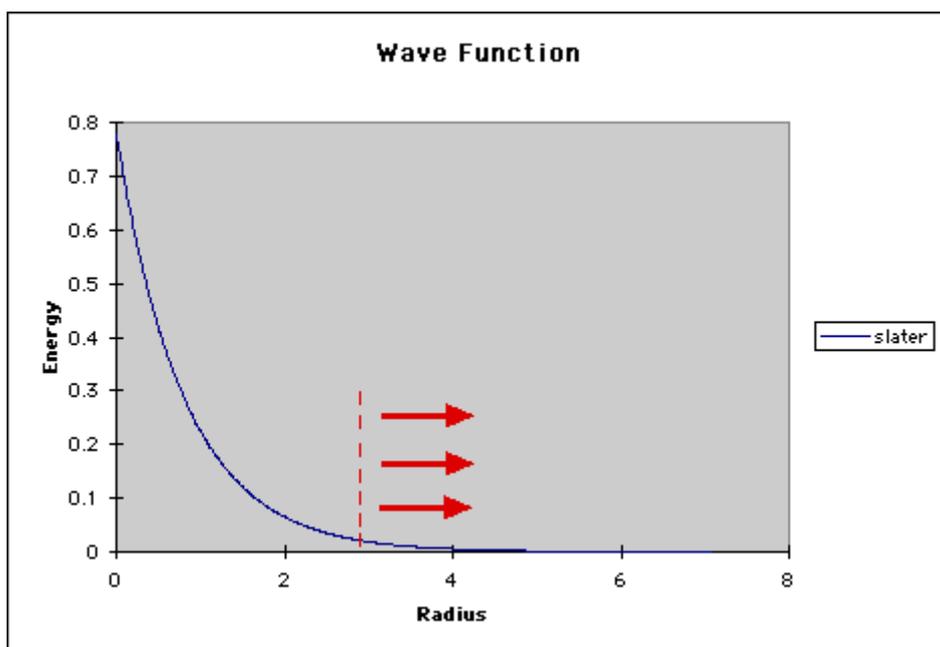


Modified 's' orbital



## Diffuse Sets

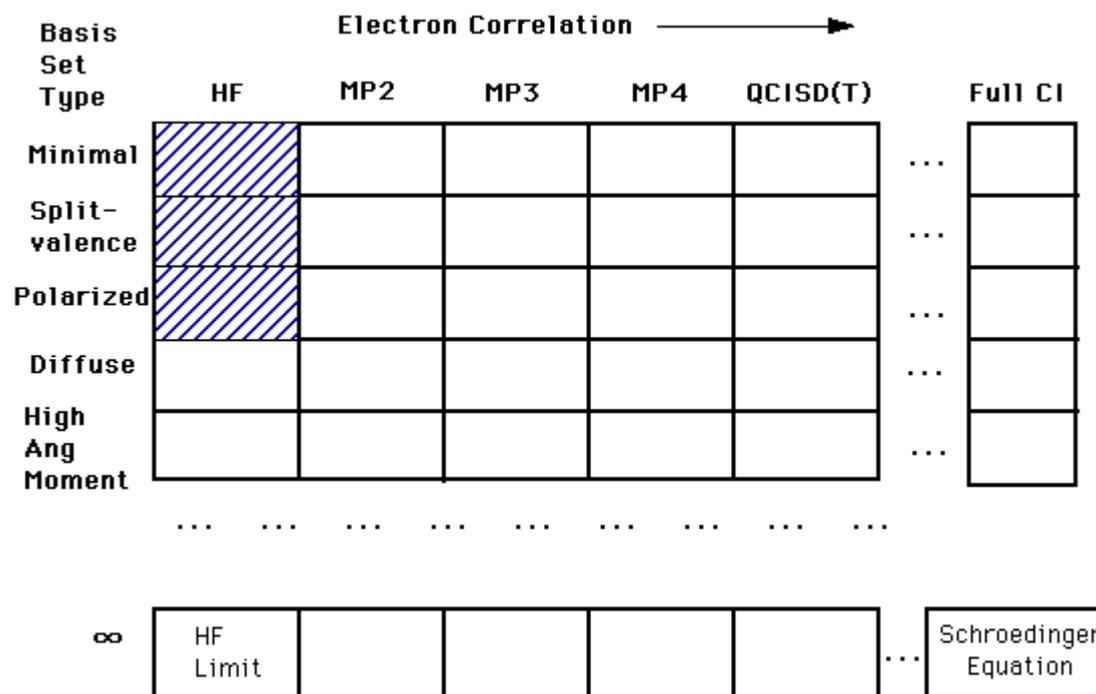
In chemistry, we are mainly concerned with the valence electrons which interact with other molecules. However, many of the basis sets we have talked about previously concentrate on the main energy located in the inner shell electrons. This is the main area under the wave function curve. In the graphic below, this area is that to the left of the red dotted line. Normally the tail (the area to the right of the dotted line), is not really a factor in calculations.



However, when an atom is in an anion or in an excited state, the loosely bound electrons, which are responsible for the energy in the tail of the wave function, become much more important. To compensate for this area, computational scientists use **diffuse** functions. These basis sets utilize very small exponents to clarify the properties of the tail. Diffuse basis sets are represented by the '+' signs. One '+' means that we are accounting for the 'p' orbitals, while '++' signals that we are looking at both 'p' and 's' orbitals, (much like the asterisks in the polarization basis sets).

The tradeoff/relationship between basis sets and accuracy is represented in the diagram below. Our ultimate goal is to calculate an answer to the Schrödinger's Equation (right bottom corner). However, we are still a long way from being able to complete this calculation. Right now we are

in the top left corner of the chart. In that first box, we are treating each electron independently of the others. As you move across to the right, you find calculations that account for the interactions of electrons. As you move down the column you find more complex and more accurate basis set calculations. Students will only be expected to understand the shaded regions.



There are other trade-offs for using each type of basis set. The more complex basis sets are more accurate but, they use up a great deal of computing time. Whenever you run a computational chemistry calculation you will be using time on a computer. Normally, the computer that will be running the calculation will be one that is shared with many other people doing other calculations. Thus, it is important that you act responsibly when choosing which basis set to use. You should pick one that is efficient for your use. This means that you should consider how much time it will take to run the molecule and use the basis set that will run the fastest without compromising your desired level of accuracy.

