You will need to complete a worksheet with questions about key results from today's lab. While completing this lab, record your response to each **worksheet prompt** as you encounter them. We ask you to submit **the worksheet by 11:59 PM EDT 10/17/23**.

In this week's lab, we will continue to carry out Hartree-Fock electronic structure calculations, and we'll be using CPU resources. To carry out this tutorial, we will continue to develop familiarity with the electronic structure code QChem (https://www.q-chem.com), and we will use Expanse at ACCESS.

Notes:

1) You should type the commands because copying and pasting directly from the PDF may lead to some errant characters. It's also just good practice to get used to typing the commands we need to use in this class.
2) To help you follow along the tutorial, we will show all commands and input files in courier font like this. However, we will show the commands we need you to type in **`green and bold courier font like this`**. Optional commands will not be colored green.
3) If you see brackets without additional spaces such as `<ACCESS username>`, you should replace everything (including the brackets) with the specified prompt.
4) *Comments inside input files to explain them in the document will be prefaced with a "#" but will not appear in your local copy of the file.*

Today, we will be SSHing into **Expanse** to run the calculations.

Quick login reminder:

Login to Expanse directly with:

```
ssh <ACCESS username>@login.expanse.sdsc.edu
```

Quick overview of running jobs on ACCESS:

You can submit a job with

```
sbatch <jobscript>
```

You can check the status of your jobs with

```
squeue -u <ACCESS username>
```

You can delete a job with

```
scancel <job #>
```

To watch the output file of a run as it proceeds:

```
tail -f <filename>
```

Note: In QChem, the total energy and eigenvalues are reported in units of Hartrees:
1 Ha = 27.211 eV = 627.509 kcal/mol.

Getting started:

1) Login to Expanse (recommended)
   ```
   ssh <ACCESS username>@login.expanse.sdsc.edu
   ```
2) Copy the input files for Lab 5 into your home directory and go into the new folder
   ```
   cp -pr /expanse/lustre/projects/itm101/hkulik/lab5/ ~/
   cd lab5
   ```
   This command copies a directory called "lab5" into your home directory. This directory should contain separate directories for the ethylene, ammonia, and tetracene subsections of the lab.

A. Ethylene basis set extrapolation for energies.

In the last lab, we geometry optimized an ethylene molecule at the HF/cc-pVTZ level of theory. Now, we want to compare its energy and properties with different basis sets: cc-pVDZ, cc-pVTZ, and cc-pVQZ for basis set extrapolation to the complete basis set (CBS) limit.

2. Make sure you are in Expanse and in the lab subdirectory for ethylene basis set extrapolation:
   ```
   cd ~/lab5/ethylene
   ```
3. We will use the coordinates from your geometry optimization (**`opt_ethylene.xyz`**) that you ran in *Lab 4*. Copy it over with the following command:
   ```
   cp ~/lab4/ethylene/opt_ethylene.xyz ./
   ```
   Note: Let us know if you do not have access to the file from the last lab.

4. I've provided a queue script called **`ethylenecbs.q`**. This script generates an input file for each basis set and incorporates the **`.xyz`** coordinates from the *Lab 4* geometry optimization. To take a look at it using nano (**you do not need to change anything yet**):
   ```
   nano ethylenecbs.q
   ```

The commands this script will run are shown in the annotated window below.

```
#!/bin/bash
#SBATCH -J ethylene-cbs
#SBATCH -o ethylene_cbs.%j.%N.out
#SBATCH -p shared
#SBATCH -A itm101
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -t 02:00:00

#----load modules----
module load cpu/0.15.4 gcc/10.2.0 mvapich2/2.3.6
module load qchem/6.0.2

xyzfile="opt_ethylene.xyz" # your optimized ethylene in lab4
#----loop through basis sets----
for basis in cc-pVDZ cc-pVTZ cc-pVQZ; do
echo "Starting $basis ..."
#----write QChem input file for each basis set----
cat > c2h4basis.$basis.in << EOF
\$rem
    JOBTYPE SP
    EXCHANGE HF
    BASIS $basis
    GUI=2
\$end
\$molecule
0 1
EOF
tail -6 $xyzfile >> c2h4basis.$basis.in
cat >> c2h4basis.$basis.in << EOF
\$end
EOF
#----run QChem calc for each basis set----
qchem c2h4basis.$basis.in > c2h4basis.$basis.out
echo "Finished on `date`!"
done
```

Window 1. SLURM queue script and QChem input file generation for neutral ethylene molecule total energy calculations with correlation consistent basis sets.

5. Submit the calculation while inside the **`~/lab5/ethylene`** directory by doing:

   `sbatch ethylenecbs.q`

This calculation will generate several output files with different basis sets (i.e., **`c2h4basis.cc-pV<#>Z.out`** where <#> = D, T, or Q).

Timing note: *These basis sets will finish in under a minute.*

5. Obtain the total energies from each single point energy calculation for each different basis set. You can do this a few ways:
   (a) You can **`nano`** each basis set file and find where it says **"Total energy"**.
   (b) You can use grep to extract every instance of the phrase **"Total energy"**:

   `grep 'Total energy' *out`

   (c) The most streamlined way to do this is to use the Python script we provided you that will replace the letters with the cardinal numbers and sort the energies by increasing number of zetas (e.g. the cardinal number) in the basis set. Please specify after the command the prefix you'd like to match (This makes it easier once we've run later steps of this part of the lab to repeat the command). For example, the prefix here should be "h2obasis" since your output files are named as "c2h4basis.<basis set>.out":

The Python script requires a package named **`scipy.`** To fulfill the requirement, you would want to first load **`anaconda`** by

```
module load anaconda3/2021.05
```

Then you can install **`scipy`** by

```
pip install scipy
```

Now you are all set to run the final script:

```
python energy-grab.py c2h4basis
```

Script note: If you'd like to see what Python wizardry we used to do that, try **`nano energy-grab.py`**.

6. Extrapolate the HF energy of ethylene to the complete basis limit using a fit to:

$$E(n) = E(\infty) + Ae^{-\alpha n}$$

where the constant is the complete basis set limit energy, n is the cardinal number, and A and α are fitting constants.

Running the **`energy-grab.py`** will print out the CBS energy (in Ha) and α as well.

Worksheet prompt: What is the difference between HF/cc-pVTZ and the complete basis set result from the exponential fit? Report your answers in kcal/mol.

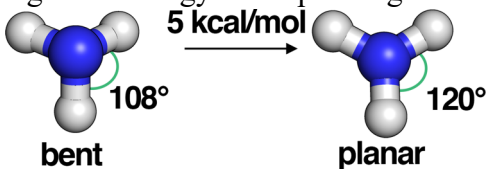
7. Compare the CBS limit from the formula you just used to what you would get using n -1 = 2 and n = 3 in this two point formula, then repeat for n -1 = 3 and n = 4

$$E(\infty) = E(n-1) + \frac{E(n) - E(n-1)}{1 - (1 - 1/n)^3}$$

Worksheet prompt: Which choice of n -1 and n in the two-point formula gives the closest agreement with the exponential decay fitting formula?

B. Basis set error and structure: planar or pyramidal NH_3?

In the next part of this week's lab, we'll look at the degree of error cancellation in how very small basis sets can alter the structures and relative stability of molecular conformers. We'll take a look at the ammonia (NH_3) molecule. Due to its lone pair, the NH_3 should have a pyramidal shape, with a trigonal planar structure that is higher in energy corresponding to a local maximum for inversion:



At room temperature, ammonia is believed to invert rapidly because the experimental free energy barrier is modest (ca. 5 kcal/mol). As discussed in class, not all basis sets get this relative ordering or preference correct. We will take a look at which basis sets do best now!

1. Go into the subdirectory where you will set up these calculations:

```
cd ~/lab5/ammonia
```

2. Much like in the previous subsections, we have provided a queue script for you that will generate the input files and run the calculations:

```
#!/bin/bash
#SBATCH -J ammonia-opt
#SBATCH -o ammonia-opt.%j.%N.out
#SBATCH -p shared
#SBATCH -A itm101
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -t 02:00:00

module load cpu/0.15.4 gcc/10.2.0 mvapich2/2.3.6
module load qchem/6.0.2

# ---loop through basis---
for basis in BASIS1 BASIS2  6-31Gs 6-311++Gss; do
# ---loop through geometry type---
for geom in bent planar; do
# ---special trestment with "*"----
if [ "$basis" == "6-31Gs" ]; then
mybasis="6-31g*"
elif [ "$basis" == "6-311++Gss" ]; then
mybasis="6-311++g**"
else
mybasis=$basis
fi
# ---write input file---
cat > nh3$geom.basis$basis.in << EOF
\$rem
    JOBTYPE Opt
    EXCHANGE HF
    BASIS $mybasis
    GUI=2
\$end
\$molecule
    0 1
EOF
if [ "$geom" == "bent" ]; then
cat >> nh3$geom.basis$basis.in << EOF
N -0.84642 0.53821 0.00662
H 0.17090 0.59569 -0.00347
H -1.15802 0.91774 -0.88626
H -1.15802 1.19918 0.71682
\$end
EOF
fi
if [ "$geom" == "planar" ]; then
cat >> nh3$geom.basis$basis.in << EOF
N 0.07038 1.05529 0.11299
H 1.01271 0.74052 -0.04029
H -0.39994 1.59587 -0.59209
H -0.40162 0.82953 0.97137
\$end
EOF
fi
# ---run the calculation---
qchem nh3$geom.basis$basis.in > nh3$geom.basis$basis.out
echo "Finished NH3 $geom basis $basis on `date` !"
done
done
```

Window 2. Script with some additional details about how to make QChem input files for NH_3 optimization and run them. You only need to change the line that has "**BASIS1**" and **BASIS2**" and replace them with the basis sets that we would like to use.

Script note:

- It has two nested **for loops**, the first of which iterates over basis sets.
- The second (inner) for loop distinguishes the initial planar and bent structures and an if statement adds the CartesianC coordinates to the input files.

- Finally, "**$**" in BASH is used to represent variables, and the name of the basis and geometry are both used in the name of the input file. However, QChem also uses "**$**" in its namelists, so we need to escape those with a "****" to not confuse them with BASH variables.
- We're going to compare to the basis set 6-31G*. Due to difficulties with asterisks in file names in BASH, we have already added a special trick to the script to work around this issue. It is called 6-31Gs in your 'for' loop. Do not change this part of the script.
- See more details annotated in the window above.

3. We're going to compare the minimal basis set STO-3G to the split-valence double-zeta 6-31G and the same basis set with polarization functions, 6-31G*, and a triple-zeta basis set that includes the polarization and diffuse function 6-311++G(d,p). The last two basis definitions are already included in the for loop, see above). Open the script to prepare to edit the for loop to make these comparisons
   ```
   nano ammoniaopt.q
   ```

4. Add STO-3G and 6-31G to the outer basis set **for loop,** replacing **BASIS1** and **BASIS2** in the current file and save the file.

5. Submit the job
   ```
   sbatch ammoniaopt.q
   ```
   You can watch your job in the queue using the commands on the first page of this handout.
   You can also do the following to watch as each individual calculation completes:
   ```
   tail -f *.out
   ```

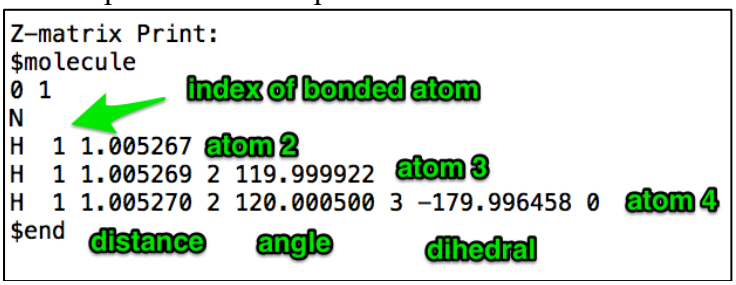
Timing note: *All calculations should take less than 3 minutes in total.*

6. Collect the final optimized energies of the NH_3 bent and planar structures:
   You'll probably want to do this for each basis set one at a time, e.g., for 6-31G* (6-31Gs) is shown:
   ```
   grep 'Final energy is' *6-31Gs.out
   ```
   Or you can also open each file with **nano**.
   Or you can get all of the energies together:
   ```
   grep 'Final energy is' *.out
   ```
   and then sort/collect them in your preferred (e.g., spreadsheet) program for analysis.

Worksheet prompt: Fill in the table on the worksheet with the total energy you obtained for each structure (bent/planar) and basis set (STO-3G/6-31G/6-31G*/6-311++G(d,p)) combination. Then calculate the relative energy of the bent structure over the planar structure in kcal/mol.
Worksheet prompt: Of the three basis sets (STO-3G/6-31G/6-31G*), which one **gives the most incorrect** relative energy of the planar vs bent structures with respect to 6-311++G(d,p)?
Worksheet prompt: Of the three basis sets (STO-3G/6-31G/6-31G*), which one gives the **most accurate** relative energy of the planar vs bent structures with respect to 6-311++G(d,p)?

Now, we're going to see if the basis set choice impacts structures as much as it does energies. Collect the structures of the optimized NH_3 molecules from their Z-matrix representation. The Z-matrix prints once at the end of the optimization and is an internal representation that consists first of a single atom, then of two bonded atoms and their bond lengths, the next row consists of bond and

angle, and finally the last line contains information about bonds, angles, and dihedrals. For more info, Z-matrix internal coordinate systems were briefly discussed in *Topic 2*.
Here is an example of what a planar Z-matrix printout should look like:



Window 3. Example Z-matrix of the NH_3 molecule printed at the end of a QChem geometry optimization.

What we want to focus on is the preferred N-H bond length, which will be shown in the first float column for each of the N-H bonds in Å, then we want to know the H-N-H angle, which should be close to 120° for the planar structure and 108° for the bent structure. This angle will be reported in the second float column.

7. To obtain the Z-matrix for both structures with the 6-31G* basis set try the following command:

   ```
   grep -A6 'Z-matrix' *6-31Gs.out
   ```

   You can also collect all of these Z-matrix results at once and place them in a spreadsheet by using a wildcard:

   ```
   grep -A6 'Z-matrix' *.out
   ```

   Here, **"-A6"** helps to grep the six lines after the keyword **"Z-matrix"** appears.

Worksheet prompt: Fill in the table on the worksheet with N-H distance and H-N-H angle you obtained at each structure (bent/planar) and basis set (STO-3G/6-31G/6-31G*/6-311++G(d,p)) combination.
Worksheet prompt: Of the three basis sets (STO-3G/6-31G/6-31G*), which gives the **best** agreement for the planar ammonia N-H bond length with respect to 6-311++G(d,p)?)
Worksheet prompt: Of the three basis sets (STO-3G/6-31G/6-31G*), which gives the **worst** agreement for the bent H-N-H angle with respect to 6-311++G(d,p)?)

C. Orthogonal basis set contributions: the tetracene molecule.

We can't always extrapolate to the complete basis set limit, as we did in part A. For larger molecules, an *additivity concept* has been proposed (see also *Topic 5* discussion):

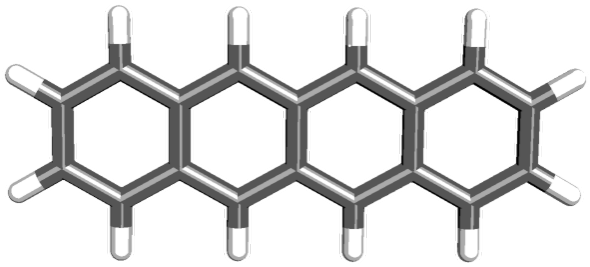
$$E[\mathrm{HF/6\text{-}311{+}{+}G(d,p)}] \approx E[\mathrm{HF/6\text{-}31G}]$$

Polarization functions $+E[\mathrm{HF/6\text{-}31G(d,p)}] - E[\mathrm{HF/6\text{-}31G}]$

Larger valence $+E[\mathrm{HF/6\text{-}311G}] - E[\mathrm{HF/6\text{-}31G}]$

Diffuse functions $+E[\mathrm{HF/6\text{-}31{+}{+}G}] - E[\mathrm{HF/6\text{-}31G}]$

We're going to compute the relative contributions of different basis set features for the tetracene molecule, which is larger (30 atoms) than ethylene (6 atoms):



Tetracene is an interesting molecule we'll revisit in the course that has relevance to umolecular organic semiconductors. We've provided you with a MMFF94-optimized tetracene molecule coordinate file and a script that will run input files you will edit/make to collect the single points for the needed basis sets (HF/6-31G, HF/6-31G**, HF/6-311G, HF/6-31++G, and HF/6-311++G(d,p)).

1. If you are not already in the lab5 subdirectory for this step:
   ```
   cd ~/lab5/tetracene
   ```

2. We've provided you a template submission script, but you need to specify the basis set that are required to compute by yourself. See below for example:

```
#!/bin/bash
#SBATCH -J tetracene-loopbasis
#SBATCH -o tetracene-loopbasis.%j.%N.out
#SBATCH -p shared
#SBATCH -A itm101
#SBATCH -N 1
#SBATCH -n 1
#SBATCH -t 02:00:00

module load cpu/0.15.4 gcc/10.2.0 mvapich2/2.3.6
module load qchem/6.0.2

xyzfile="tetracene_opt.xyz"
for basis in <list-of-basis>; do # FILL IN the basis set!
echo "Starting $basis ..."
if [ "$basis" == "6-31Gss" ]; then
mybasis="6-31g**"
elif [ "$basis" == "6-311++Gss" ]; then
mybasis="6-311++g**"
else
mybasis=$basis
fi
cat > tetracene.$basis.in << EOF
\$rem
    JOBTYPE SP
    EXCHANGE HF
    BASIS $mybasis
    GUI=2
\$end
\$molecule
0 1
EOF
tail -30 $xyzfile >> tetracene.$basis.in
cat >> tetracene.$basis.in << EOF
\$end
EOF
qchem tetracene.$basis.in > tetracene.$basis.out
echo "Finished on `date`!"
done
```

Window 4. Tetracene script for iterating over basis sets with the line for you to update the basis sets indicated.

3. You can edit the file by
   `nano tetracene_loopbasis.q`

4. Run the single point energy calculations:
   `sbatch tetracene_loopbasis.q`

5. Collect the total energy for each single point energy run with:
   `grep 'Total energy' *out`

6. Now, we'll estimate the difference between HF/6-311++G(d,p) and HF/6-31G energies using the additivity concept. First compute the difference between HF/6-311G and HF/6-31G, then HF/6-31G** and HF/6-31G, and finally compute the difference between HF/6-31++G and HF/6-31G. **Collect the wall time of each calculation (the first reported number) by:**
   `grep 'Total job time' *out`

Worksheet prompt: Fill in the table on the worksheet with the total energy you obtained for each basis set.

Worksheet prompt: Fill in the table on the worksheet with the relative energy you obtained for pairs of basis sets that measure the contribution of triple-zeta/double-zeta split valence, polarization, and diffuse functions.

Worksheet prompt: Rank the contributions from largest to smallest: triple-zeta split valence (i.e., vs. double-zeta split valence), polarization, or diffuse functions according to their contributions to the total energy difference.

Worksheet prompt: Estimate the energy (in Ha) of the tetracene molecule at the HF/6-311++G(d,p) level using the additivity concept.

Worksheet prompt: What % of the energy difference between HF/6-311++G(d,p) and HF/6-31G did the additivity concept recover?

Worksheet prompt: What % of the total wall time did the sum of the four needed individual calculations that you carried out take vs. the large calculation we provided you?

Worksheet prompt: If we gave you a much bigger molecule to study that had >50 heavy atoms, how do you expect your answer to the last question to change?