
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 **entire worksheet by the end of today – 11:59 PM EDT 10/12/23 (Thu)**. You may hand in a written copy or upload to Canvas.

In today's lab, we will carry out our first electronic structure calculations (specifically Hartree-Fock theory), and we'll be using CPU resources to do so. To carry out this tutorial, we will need to become familiar with new electronic structure software – QChem (<https://www.q-chem.com>), and we will use Expanse from 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 angled brackets without additional spaces such as <Your Username>, you should replace everything (including the brackets) with the specified prompt. Optional arguments are provided in square brackets – e.g. [out <outfile>] means that you can specify an output file with the name provided for <outfile>.
- 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.*
- 5) Text in **red** indicates that you will need to make an action. Text in **blue** indicates optional notes or hints for if you get stuck.

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 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
`ssh <ACCESS username>@login.expense.sdsc.edu`
- 2) Copy the input files for Lab 4 into your home directory and go into the new folder
`cp -pr /expense/lustre/projects/itm101/dbkchu/lab4/ ~/`

This command copies a directory called "lab4" into your home directory. This directory should contain separate directories for the required ethylene and optional benzene subsections of the lab.

A. Ethylene: Geometry optimization.

Today, we will first be using Hartree-Fock (HF) theory to conduct a geometry optimization of ethylene. We'll be using the cc-pVTZ basis set (discussed in *Topic 5*).

- 1) First, change directories into the one which contains the files for the ethylene calculation:
`cd ~/lab4/ethylene`

Take a look at the files included in this folder:

```
ls
```

You should see seven files: two for the geometry optimization (**ethylene_optimize**), two for the normal mode calculation (**ethylene_frequencies**), two for the ionization potential calculation (**ethylene_ionize**), and one Python script (**get_opt_xyz.py**). For each type of calculation, there are two files. The **.qcin** file is the input file to QChem and contains the initial atom positions, the electronic structure method (in this case, Hartree-Fock), the type of calculation (e.g., geometry optimization), etc.

Open the **ethylene_optimize.qcin** file. A few of the settings are left blank, and you will need to fill them in.

```
nano ethylene_optimize.qcin
```

The JOBTYP should be specified as "Opt" for a geometry optimization. The EXCHANGE should be set to "HF" for a Hartree-Fock calculation:

ethylene_optimize.qcin (changes indicated as <> in red):

```
$rem
  JOBTYP <jobtype>      # Specify the type: optimization, frequencies, single point, etc.
  EXCHANGE <method>     # Specify the electronic structure method
  BASIS cc-pVTZ         # Specify the basis set, will be covered in Topic 5
  GUI=2
$end

$smolecule              # MMFF94 optimized guess structure (nuclear coordinates in Å)
  0 1                    # The charge(0) and spin multiplicity(singlet) of the molecule
  H    1.16103    0.06607    1.02388
  C    0.65790   -0.00446    0.06398
  H    1.33519   -0.08301   -0.78143
  C   -0.65783    0.00446   -0.06385
  H   -1.33549    0.08301    0.78127
  H   -1.16080   -0.06606   -1.02386
$end
```

- 2) After modifying the **.qcin** file, you are ready to submit the job.

```
sbatch ethylene_optimize.q
```

Monitor the progress of the job. When it finishes, you may move on to step 3:

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

This calculation should take under two minutes once it starts. The jobscript (i.e., **.q**) file does not require any changes, but here it is:

ethylene_optimize.q (no changes)

```
#!/bin/bash
```

```
#SBATCH -J ethylene-optimize # The job name, which appears with the squeue command
```

```
#SBATCH -o ethylene_optimize.%j.%N.out # The queue output file
```

```
#SBATCH -p shared
```

```
#SBATCH -A itm101
```

```
#SBATCH -N 1
```

```
#SBATCH -n 1
```

```
#SBATCH -t 00:10:00
```

```
module load cpu/0.15.4 gcc/10.2.0 mvapich2/2.3.6 # Loads modules we need to run QChem
```

```
module load qchem/6.0.2 # Gets the QChem program ready
```

```
qchem ethylene_optimize.qcin ethylene_optimize.out # Run QChem, save output to .out file
```

- 3) After the calculation finishes, there should be three new files in the directory. The **ethylene_optimize.o** file contains the output of the jobscript. The **ethylene_optimize.fchk** file is a checkpoint file that enables restarts and contains the final wavefunction. The **ethylene_optimize.out** file, includes the final energy, the optimized structure, and a record of how the job proceeded. The following commands will help you to extract this information from the output file.

a. To extract the final energy from the calculation in units of Hartrees, use:

```
cat ethylene_optimize.out | grep 'Final energy is'
```

b. To obtain the nuclear coordinates and energy at each step of the optimization cycle:

```
cat ethylene_optimize.out | grep -A 10 Standard Nuclear'
```

c. To obtain more information about the self-consistent field (SCF) convergence during each step of the optimization:

```
cat ethylene_optimize.out | grep -A 15 'A restricted SCF'
```

Linux hint: The first command **cat** is providing the full text of the file and piping it (the **|**) to a **grep** command, which is pattern matching the text in quotes. The extra flag **-A** tells the command to also print a specified number of lines after the pattern match.

Worksheet prompt: What is the final energy (in Hartrees) of the optimized ethylene molecule from this HF/cc-pVTZ calculation?

- 4) We will use the optimized structure of ethylene to calculate its Hessian and ionization potential. We need to extract the optimized geometry and save it as an **.xyz** file (here we name it as **opt_ethylene.xyz**). You may look in **ethylene_optimize.out** to find the optimized geometry (**Window 1**), but we have provided a Python script which will extract this.

```
python get_opt_xyz.py
```

Note: The script will also create a file called **ethylene_molecule** which will be used to create new input files starting from the optimized structure by providing the molecule block

format that QChem expects. We'll use the xyz file to visualize in Avogadro.

```

*****
**  OPTIMIZATION CONVERGED  **
*****
-----
Standard Nuclear Orientation (Angstroms)
-----
I      Atom      X      Y      Z
-----
1      H
2      C
3      H
4      C
5      H
6      H
-----
Coordinates will appear here!
-----

```

Window 1. Screenshot illustrating where to find the optimized geometry in `ethylene_optimize.out`.

- 5) Copy this optimized structure back to your local machine so we can measure the bond lengths and angles using Avogadro. You can copy the file to your local machine two ways:

- Copy the files using **scp**. This is recommended if you are more comfortable with the command line. **Open a new terminal window** on your local machine and type the following commands, which will copy the `.xyz` file to your current working directory. This will ask you for your ACCESS password. It may also ask you if you want to add the host key to your list of known hosts. If so, you should respond "yes":
`scp <ACCESS username>@login.expense.sdsc.edu:~/lab4/ethylene/opt_ethylene.xyz ./`
- Copy the coordinates using your clipboard. This method may be tedious but requires less use of the command line. Output the contents of the file on ACCESS with:
`cat ~/lab4/ethylene/opt_ethylene.xyz`
 Then highlight the contents shown in the terminal and copy them with **Ctrl+C** or **Command-c** on a Mac. Paste the results on your local machine.

Once on your local machine, open the file with Avogadro and use the measurement tool to find the geometric properties listed below.

- 8) Optimize an ethylene molecule with MMFF94 in Avogadro (see instructions in *Lab 1*).

Worksheet prompt: The experimental results for the geometric properties of ethylene are provided in the table below (as reported in <https://doi.org/10.1063/1.1744018>). What are these values for the optimized ethylene molecule geometry of MMFF94 and HF/cc-pVTZ?

	C=C bond length (Å)	C-H bond length (Å)	H-C-H angle (°)
HF			
MMFF94			
Exp.	1.330	1.074	116

Worksheet prompt: Which method (MMFF94 or HF/cc-pVTZ) is in better agreement compared to experiments for the C=C bond? For the C-H bond? Why might this be?

B. Ethylene: Hessians and ionization potential calculation.

- 1) We will use the optimized structure of ethylene (obtained in *Section A*, `opt_ethylene.xyz`) to calculate its Hessian and ionization potential.
- 2) Modify the `ethylene_frequencies.qcin` file to set the job type and include the optimized structure you obtained. The JOBTYP should be specified as "**Freq**" for the calculation of the Hessian.

`nano ethylene_frequencies.qcin`

`ethylene_frequencies.qcin` (changes indicated as \diamond in red):

```
$rem
  JOBTYP  $\diamond$ jobtype      # Specify the type: optimization, frequencies, single point, etc.
  EXCHANGE HF              # Specify the electronic structure method
  BASIS cc-pVTZ            # Specify the basis set, see Topic 5
  GUI=2
$end
```

\diamond molecule \diamond

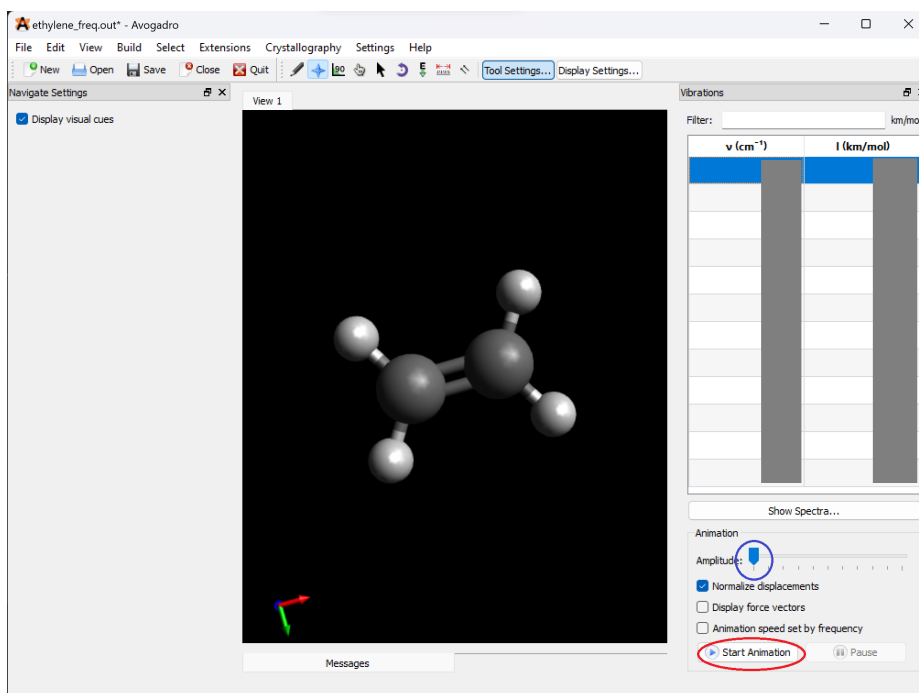
You may include the molecule block by using the command:

`cat ethylene_molecule >> ethylene_frequencies.qcin`

Linux hint: `>` is a redirect operator which redirects the output of the previous command.

Using `>>` tells it not to overwrite the file contents but to append the output instead.

- 3) Submit the Hessian calculation:
`sbatch ethylene_frequencies.q`
- 4) Monitor the progress of the job. It should take < 1 min.
`squeue -u <ACCESS username>`
- 5) When the Hessian calculation is finished, copy the results back to your local machine to visualize the normal modes (i.e., eigenvectors of the Hessian). Since the file is large, we would recommend that you copy the file back to your local machine using `scp` (it is a useful tool!) rather than copy-and-paste. Copy the files using `scp` (**open a new terminal window**):
`scp <ACCESS username>@login.expense.sdsc.edu:~/lab4/ethylene/ethylene_freq.out ./`
- 6) Open `ethylene_freq.out` with Avogadro. Avogadro should automatically open the "Vibrations" toolbar when it detects the output file format. Use this to visualize each normal mode and identify which are C-H bend/stretch and C=C



Window 2. Avogadro interface for visualizing normal modes (vibrations). The button to animate the normal modes is circled in red. The recommended amplitude for visualization is circled in blue. The frequencies for each normal mode have been omitted but note that the first one is selected (highlighted in blue).

Worksheet prompt: What is the wavenumber (in cm^{-1}) of the bending normal mode (with nonzero intensity)? The symmetric stretch? The asymmetric stretch?

	bending (cm^{-1})	symmetric stretch (cm^{-1})	asymmetric stretch (cm^{-1})
HF			
Exp.	~900	~3000	~3100

Worksheet prompt: How do your results compare to the experimental values?

7) Now we're ready to compute ionization potentials of ethylene. Modify the **ethylene_ionize.qcin** file to set the job type, charge, and spin multiplicity. Again, use the optimized ethylene structure by HF/cc-pVTZ.

cat ethylene_molecule >> ethylene_ionize.qcin

Note: The **JOBTYPE** should be set to "**SP**" (single point) for this vertical ionization potential calculation. The **CHARGE** should be specified as "**1**" to calculate the wavefunction of a ethylene molecule with one fewer electron. You will also need to adjust the **SPIN MULTIPLICITY** to "**2**" for a doublet.

ethylene_ionize.qcin (changes indicated as <> in red):

\$rem

JOBTYPE <jobtype>

EXCHANGE HF

BASIS cc-pVTZ

GUI=2

Specify the type: optimization, frequencies, single point, etc.

Specify the electronic structure method

Specify the basis set, see Topic 5

\$end

<molecule>

should start with \$molecule

second line is <CHARGE> <SPIN MULTIPLICITY>

- 8) Submit the vertical ionization potential calculation.

`sbatch ethylene_ionize.q`

Monitor the progress of the job. **Timing note:** It should take << 1 min.

`squeue -u <ACCESS username>`

- 9) Compare the energy of the ionized ethylene molecule to the energy of the neutral ethylene molecule. Use the difference between the two values to calculate the vertical ionization potential of ethylene:

To obtain final energy of the neutral ethylene molecule in Hartrees use:

`cat ethylene_optimize.out | grep 'Final energy is'`

To obtain the final energy of the ethylene cation in Hartrees use:

`cat ethylene_ionize.out | grep 'Total energy'`

Note: In the first command, the `grep` command gets the last energy after a geometry optimization, whereas for the ionized molecule, we only carried out a single point energy calculation. So, the way the code prints the energy for the single point is different than the way we obtain the last energy of a geometry optimization.

Hint: Useful conversion factors are **1 Ha = 27.211 eV = 627.509 kcal/mol**.

Worksheet prompt: What is the vertical IP from HF/cc-pVTZ total energy differences? Report your answer in eV.

- 10) Compare the IP calculated in *step 7* to the one that would be obtained from Koopmans' theorem (i.e., $IP = -\epsilon_{HOMO}$). The code will print these orbital energies in units of Ha. Obtain the eigenvalues of the occupied molecular orbitals in the optimized ethylene molecule with:

`cat ethylene_optimize.out | grep -A 2 'Occupied' | tail -2`

Hint: Remember that ethylene has 16 electrons in 8 orbitals, and this command should show the eigenvalues of these 8 orbitals.

Worksheet prompt: What is the vertical IP of ethylene according to Koopmans' theorem? Report your answer in eV.

Worksheet prompt: How do the calculated IP values from total energy differences and Koopmans' theorem compare to the experimental value of 10.5 eV? (reported on [NIST](#))

Worksheet prompt: Given what you have observed about HF frequencies and IP values, what missing ingredient from HF do you believe could improve agreement with experiment? Explain your reasoning in the context of both frequencies and ionization potentials.

Extra credit prompt (1.5pts): Use Avogadro to visualize the calculated spectra, overlay it with the experimentally reported spectra, report the scale factor & Gaussian width used, and label the peaks as C-H bend/stretch/etc.

Download the experimental IR of ethylene as a JCAMP-DX (.jdx) file from [NIST](#) or from the Lab 4 module on Canvas.

- In Avogadro, click "Show Spectra..."
- In the new window that opens click "Advanced >>"
- Set the Scheme to Publication
- Click "Import..." and select the .jdx file

- e) Optionally, change the color of the imported spectra via the “Set Color...” button
- f) Click “Infrared Spectra Settings”
- g) Adjust the Scale Factor and Gaussian Width until you get something that aligns reasonably with the experimental spectra. (**Note:** the suggested value for HF was ~0.9)
- h) Check the “Label peaks” box.
- i) “Export Image”, label the peaks with the type of vibrational mode, and submit the image

C. Benzene: Frequencies and ionization (**Extra Credit**)

This is an **Extra Credit** section and is optional!

- 1) First, change directories into the one which contains the files for the benzene calculation:

```
cd ~/lab4/benzene
```

- 2) Set up the HF frequency (Freq) calculation:

```
nano benzene_frequencies.qcin
```

benzene_frequencies.qcin (changes indicated as <> in red):

```
$rem
  JOBTYP<e> <jobtype>      # Specify the type: optimization, frequencies, single point, etc.
  EXCHANG<e> <method>      # Specify the electronic structure method
  BASIS cc-pVTZ
  GUI=2
$end

$molecule                  # HF/cc-pVTZ optimized structure
  0 1                        # The charge and spin multiplicity of the molecule
  C 0.5625564505 -1.2630382788 -0.0000000000
  C -0.8125449922 -1.1187073105 -0.0000000000
  C 1.3751014434 -0.1443309749 0.0000000000
  C 0.8125449922 1.1187073105 -0.0000000000
  C -0.5625564505 1.2630382788 -0.0000000000
  C -1.3751014434 0.1443309749 -0.0000000000
  H 0.9992934054 -2.2435900151 0.0000000000
  H -1.4433589500 -1.9872086629 -0.0000000000
  H 2.4426526484 -0.2563813992 0.0000000000
  H 1.4433589500 1.9872086629 0.0000000000
  H -0.9992934054 2.2435900151 0.0000000000
  H -2.4426526484 0.2563813992 -0.0000000000
$end
```

- 3) Run the calculation (**Timing note:** may take up to 20 minutes to run!)

```
sbatch benzene_frequencies.q
```
- 4) While you are waiting, run the ionization calculation (this should run in about 30 seconds)

```
sbatch benzene_ionize.q
```
- 5) Once the single point calculation finishes, get the energy via

```
cat benzene_ionize.out | grep 'Total energy'
```
- 6) Compute the ionization potential by comparing to the neutral-charge state energy computed during the frequency calculation (**Note:** this can be done before the frequency calculation finishes)

```
cat benzene_frequencies.out | grep 'Total energy'
```


- 7) Get the Koopmans' theorem IP via the orbital energies

```
cat benzene_frequencies.out | grep -A 7 'Occupied' | tail -3
```

Extra credit prompt (0.5pts): Compare the calculated values of the IP to the experimental IP of 9.24 eV. Does HF give greater or lesser error for the IP of benzene than the IP of ethylene? Why might this be?

Extra credit prompt (1pts): When the frequency calculation finishes, transfer the **.out** file to your local machine and use Avogadro to visualize the spectra as in the Extra credit part of *Section B*. Again, download the experimental spectra (from [NIST](#) or on Canvas), determine the values of the scaling factor (start with 0.9 that was suggested in a journal article discussed in class) and Gaussian width which best align the experimental and HF computed spectra. Include the overlaid image of the two spectra in your worksheet submission.

Extra credit prompt (0.5pts): Comment on the fit of the 0.9 scaling factor for the spectra in benzene and in ethylene. Is there a difference? Why might this be?