Configuration Interaction and Semi-empirical methods

In this exercise, you will use a method that is at a higher level than Hartree-Fock, the so-called configuration interaction (CI) method, as well as methods that are at a lower level and have been fitted to experimental data and high level ab-initio methods, so-called semi-empirical methods.

A. Dissociation curve of H_2

Calculate the energy curve of H_2 using the **QCISD(T)** method and several different basis sets. This calculation will be similar to the one performed in lab 3, part D. However, instead of using the 'scan' function (as in lab 3) we will use the 'paras' (short for 'parameter-scan') function of ORCA. In both 'paras' and 'scan' one degree of freedom is chosen (bond length, angle or dihedral) and dragged from an initial value to a final value in discrete steps. While all other degrees of freedom are kept fixed ('paras') or relaxed ('scan'), respectively. The 'paras' function is more suitable for the problem at hand, i.e., calculating the energy curve of H_2 using QCISD(T), since no gradient computation is required when using the 'paras' function. Below is an example of an ORCA input file for a parameter-scan:

```
!UHF 6-31G

%paras
R = 0.40,2.00,17
end

*xyz 0 1
 H      0.0   0.0   0.0
 H      {R}   0.0   0.0
*
```

This input will instruct ORCA to change the x-coordinate of the second H-atom (labeled by {R}) from 0.40 Å to 2.0 Å in 17 steps and compute the energy of the H_2 molecule in each step, using UHF/6-31G.

Now create an ORCA input file requesting the QCISD(T) method and also the **6-311G** Pople basis set. Perform a parameter scan, from $R_{HH} = 0.4$ Å to 2.0 Å in 33 steps. Submit the calculation. The values of the energy obtained for the various values of the parameter (here, bond length) are listed in a table in the output file. Search for the section titled **The Calculated Surface using the 'Actual Energy'** and **The Calculated Surface using the SCF energy** close to the end of the output file. The **'Actual Energy'** is the QCISD(T) energy you are interested in. The **SCF energy** is the single Slater-determinant approximation on which the post-HF calculation is based on.

Repeat the calculation for for the following basis sets: **6-311G****, **6-311++G****, **6-311++G(2d,2p)**, **6-311++G(2df,2pd)** and **6-311++G(3df,3pd)**.

In 1965 Kolos and Wolniewicz did highly accurate calculations of the binding curve of hydrogen. Their results can be taken as the exact values and it is interesting to compare them with your results. You can find their results on the website of the course next to this handout.

> **Q1:** *Show the binding curve calculated using QCISD(T) for the various basis sets together with the results of Kolos and Wolniewicz. Do the same for the Hartree-Fock results. How do they compare?*

To examine the effect of perturbative triples, you can use the energy curve for QCISD (i.e. no perturbative triples included) which is found in the section titled **The Calculated Surface using the MDCI energy minus triple correction**. You can also search for the occurrence of **RHF TRIPLES CORRECTION**

in the output file, where two energy values are given, one for the single plus double excitations, which are evaluated self-consistently in the energy minimization, **QCISD**, and one including the perturbative correction from triple excitations, **QCISD(T)**.

> **Q2:** *How do the QCISD and QCISD(T) values compare? What would you expect a full CI calculation to give? Explain the trend.*

B. Bond angle and singlet-triplet splitting in CH_2

Calculate the energy and geometry of the singlet and triplet states of the CH_2 molecule using QCISD(T) and the 6-311++G** basis set. To have a good initial guess for the molecular geometry, use the output file of the 6-311++G** HF calculation of the triplet you did in the previous assignment. Select geometry optimization in QCISD(T) using the 6-311++G** basis set. ORCA is unfortunately not able, to evaluate the energy gradients needed for the geometry optimization analytically, when using a QCI method. However, it can evaluate the gradients numerically, which can take quite some time for larger molecules, but is still acceptable for methylene. To instruct ORCA to use numerical gradients, add the keyword **NUMGRAD** in the simple-input line.

This calculation will take a few minutes. When it has finished, record the following values from the output file (search for the last occurrence of the mentioned keywords): The H-C-H angle, the QCISD(T) energy (**E(QCISD(T))**), the QCISD energy (**E(QCISD)**) and the HF energy (**E(0)**).

Repeat this calculation for the singlet state of methylene.

> **Q3:** *How does the geometry and singlet-triplet splitting obtained using QCISD(T) compare with the Hartree-Fock results (from the previous assignment) and with the experimental values of the bond angle ($134°$ (triplet) and $102°$ (singlet)) and a singlet-triplet splitting of 9.05 kcal/mol?*

> **Q4:** *Does the QCISD energy differ from the QCISD(T) energy? Why is that?*

> **Q5:** *Is the singe Slater determinant ('Hartree-Fock') energy from these calculations, E(0), different from the one you calculated in the previous assignment? If so, is it higher or lower? Explain your observation.*

C. The C-C bond length in hydrocarbons

Errors in calculated bond distances using the Hartree-Fock approximation are greater for double and triple bonds than for single bonds. For example, while Hartree-Fock carbon-carbon single bond lengths are quite close to experimental values, the corresponding double and triple bonds are typically too short. This can be rationalized by recognizing that corrections to the Hartree-Fock approximation in one way or another 'mix' in excited state determinants. Bond distances in excited states will tend to be longer than those in the ground state, meaning that any mixing of excited states with the ground state will lead to bond lengthening.

Furthermore, as excited states will generally be more accessible (lower in energy) for unsaturated systems as compared to saturated systems, it is reasonable to expect that changes from Hartree-Fock results will be greater for double and triple bonds.

Optimize the geometry of ethane (C_2H_6), ethylene (C_2H_4) and acetylene (C_2H_2) using Hartree-Fock and the 3-21G basis set and record the C$-$C bond lengths.

> **Q6:** *Compare the calculated C$-$C, C$=$C and C$\equiv$C bond lengths with the experimental values ($1.531\,Å$, $1.339\,Å$ and $1.203\,Å$, respectively). What is the error in each case and what is the trend?*

Display the HOMO and LUMO for ethylene (see lab 1). The first excited state of ethylene might be viewed as resulting from excitation of an electron from the HOMO to the LUMO.

Q7: *Is the HOMO classified as bonding, non-bonding or anti-bonding for the carbon-carbon bond? How about the LUMO? Would you expect the carbon-carbon bond in the first excited state of ethylene to be longer, shorter or unchanged from that in the ground state? Elaborate.*

Q8: *What effect, if any, on the carbon-carbon bond length in ethylene would be expected when excited state determinants are inlcuded?*

Calculate the equilibrium geometries using the QCISD approximation and compare the bond length you obtain with the Hartree-Fock results. As QCI calculations with ORCA require numerical energy gradients, the calculations will take long if the initial geometry is very different from the equilibrium structure. To save time, use the final structure of the HF calculations. Remember to add **NUMGRAD** in the input file.

Q9: *How does the addition of the excited state determinants change the calculated bond length?*

D. Semi-empirical calculations of nitrobenzene

Now we would like to optimize the geometry of nitrobenzene using three different (but of same methodology) semi-empirical methods and compare the results to a Hartree-Fock calculation. Do a equilibrium structure search for nitrobenzene and use the **MNDO** method (instead of e.g., HF). You do not need to specify a basis set for these calculations, since semi-empirical methods employ a minimal basis set by construction. Repeat this for the other semi-epmirical methods **AM1** and **PM3**. Also, for Hartree-Fock and the **3-21G** basis set.

Q10: *How do the optimized structures of the semi-empirical calculations compare with the Hartree-Fock results?*

The semi-empirical methods avoid evaluating the most tedious integrals involved in HF calculations and use instead functions that are fitted to experimental data or high-level wavefunction theory calculations. They are called semi-empirical because they involve information from empirical (expermental / high-level calculations) data. The accuarcy of these methods can be quite good for the types of molecules that are used in the fitting, but can fail drastically if they are applied to molecules that are quite different from those that were used in the fitting.

Q11: *Use the web or other resources to find out what types of molecules the three semi-empirical methods you have used, **MNDO**, **AM1** and **PM3** are made for. What are their strengths and weaknesses? Does the performance of these methods in describing nitrobenzene agree with what you would expect given this information?*