4 Post-Hartree Fock Methods: MPn and Configuration Interaction

In the limit of a complete basis, the Hartree-Fock (HF) energy in the complete basis set limit (E_{CBS}^{HF}) yields an upper boundary to the physical energy of the system:

$$E^{total} = E_{CBS}^{HF} + E^{corr}, \tag{81}$$

where the difference between the total and the Hartree-Fock energy is given by effects due to *electron correlation*, E^{corr}. The neglect of electron correlation in Hartree-Fock is a consequence of the restriction to a single Slater determinant. In weakly correlated systems, this may be a reasonable approximation; in the majority of the cases, however, this approximation breaks down more or less drastically. In such cases, it is inevitable to include electron correlation; Hartree-Fock results may, in the worst case, be almost (or even totally) meaningless (recall that Hartree-Fock predicts a non-bonding F_2), since the the contribution of correlation is significant in these systems. Inclusion of correlation can be based on *e.g. Post-Hartree-Fock methods*, such as *Configuration Interaction* (CI) or *Møller Plesset Perturbation Theory* of n^{th} order (MPn). In this set of exercises, you will compare the performance of HF, MPn and CI in describing a bond dissociation energy, and you will assess the influence of electron correlation on structural parameters (bond lengths and angles).

4.1 Bond Dissociation in CH_3F: HF vs. MPn

An accurate description of thermochemical properties, such as reaction enthalpies, will usually call for a more accurate energy than the Hartree-Fock energy. Møller-Plesset perturbation theory is often precise enough to describe such processes, and will therefore be the method of choice especially for larger systems, where other, more elaborate methods become computationally untractable.

Møller-Plesset Perturbation Theory

According to Rayleigh-Schrödinger perturbation theory, an instantaneous perturbation of a system described by a Hamiltonian $\hat{\mathrm{H}}_0$ is described by the perturbed Hamiltonian

$$\hat{\mathrm{H}} = \hat{\mathrm{H}}_0 + \lambda V. \tag{82}$$

For a sufficiently weak perturbation, the eigenstates and eigenvalues may then be expanded in a power series:

$$\langle\Psi| = \langle\Psi^{(0)}| + \lambda \langle\Psi^{(1)}| + \lambda^2 \langle\Psi^{(2)}| + \dots \tag{83}$$

$$E_\Psi = E_\Psi^{(0)} + \lambda E_\Psi^{(1)} + \lambda^2 E_\Psi^{(2)} + \dots \tag{84}$$

The perturbing operator $\hat{\mathrm{V}}$ which is introduced in Møller-Plesset perturbation theory is the difference between the true ground state Hamiltonian and the Hartree-Fock Hamiltonian; hence, the perturbation may be written in terms of excited Slater determinants

(which also imposes orthonormality). By carrying out the expansion, collecting terms of the same order and applying the Slater-Condon rules (*cf.* chapter 5.4 of the course script), one concludes that only doubly-excited Slater determinants can contribute to the second- and third-order term. Only the fourth-order energy will include up to quadruply excited determinants. By truncating the expansion at second order, one arrives at the MP2 expression for the energy, where electron correlation is now included as a perturbation to the uncorrelated Hartree-Fock wavefunction:

$$E_0^{(2)} = \frac{1}{4} \sum_a \sum_b \sum_r \sum_s \frac{|\langle ab| \, |rs\rangle|^2}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s}, \tag{85}$$

where a, b denote occupied, r, s denote virtual orbitals and ϵ are the respective energy eigenvalues. The total energy is given by the perturbative contribution and the Hartree-Fock energy:

$$E_0^{total} = E_0^{HF} + \sum_{n=1} E_0^{(n)}. \tag{86}$$

(We note *en passant* that the first-order Møller Plesset energy is nothing but the expression for the total Hartree Fock energy, such that the zero-order term is given by a sum over orbital eigenvalues.) The MP approach is often accurate enough; however, it is not variational, *i.e.* the resulting MPn energy may be lower than the true ground-state energy.

Homolytic Cleavage of the C-F Bond

In this exercise, you will investigate the influence of correlation on the BDE in a radical process by comparing Hartree-Fock to MPn results. Consider the reaction enthalpy of the homolytic cleavage of the H_3C-F bond in monofluoro methane:

$$CH_3F \quad \rightarrow \quad H_3C\cdot \quad + \quad F\cdot. \tag{87}$$

As in the exercise on hydrogen dissociation, you may calculate the bond dissociation energy as the difference of the energy of the fragments:

$$E^{BDE} = E^{CH_3F} - E^{H_3C\cdot} - E^{F\cdot}, \tag{88}$$

where E^{BDE} was experimentally determined to be 109.2 kcal mol^{-1} (*J. Phys. Chem. A*, **2000**, 104, *436*).

Practical Application of Post-Hartree-Fock Methods

Bond dissociation energies of large systems often require the use of large basis sets of triple-ζ quality; you may either use Pople's 6-311+G* basis, or the larger aug-cc-pVTZ by Dunning and coworkers (the choice is up to you; either will take comparably long). As Post-HF methods require orbitals that diagonalise the Fock operator (*i.e.* the Hartree-Fock orbitals), the HF orbitals will need to be calculated at the start of every run.

The Post-HF treatment itself may take some considerable time. It is therefore not advisable to run electronic structure calculations with computationally expensive methods starting from a random guess for the geometry (since this guess may be considerably off the equilibrium). It is usually more effective to pre-optimise the geometry and wavefunction at the Hartree-Fock level, and only then carry out the more elaborate calculations on the basis of the converged Hartree-Fock result. Therefore, one must make the converged geometry and wavefunction available to the electronic structure code. In Gaussian, this information is written to the *Checkpoint* file that will be used in the following calculations. First, create a Hartree-Fock input file for CH_3F:

```
%NProcShared=2
%Mem=1GB
%Chk=CH3F_HF
#P UHF/[...] Opt Symmetry=None

CH3F - HF geometry optimisation

0 1
F       -0.650  -0.000   0.000
C        0.749   0.000   0.000
H        1.112   0.000   1.028
H        1.112  -0.890  -0.514
H        1.112   0.890  -0.514
```

You will need to replace `[...]` by the basis set you chose to use. The new line corresponding to `%Chk=` instructs Gaussian to create a Checkpoint file that will be named `CH3F_HF.chk` in this example. This file will contain all the relevant information for Gaussian to read in binary form. Create two other input files for the radical fragments $CH_3\cdot$ and F$\cdot$, and make sure that you can keep track of the naming of the files. *Do not forget to adjust the name of the Checkpoint file for the new inputs, as the old file will get overwritten otherwise.* Consider that the multiplicity changes when moving from a closed-shell species to a radical. Submit all three jobs to Gaussian. Quickly investigate the output structures in Molden, as in the previous exercise. If the geometries look reasonable, you may continue using Post-HF methods.

You may now create the MPn inputs. As the HF wavefunction and geometry should be read from the Checkpoint file, the syntax for the input will be slightly modified:

```
%NProcShared=2
%Mem=1GB
%OldChk=CH3F_HF
%Chk=CH3F_MP2
#P MP2/[...] Opt Symmetry=None Guess=Read Geom=Checkpoint

CH3F - MP2 geometry optimisation from HF checkpoint

0 1
```

Note that you will only have to give charge and multiplicity, but no coordinates. Instead, `Geom=Checkpoint` makes Gaussian read the optimised (converged) coordinates from the *old* Checkpoint file, the name of which is specified by `%OldChk=`. `Guess=Read` similarly instructs Gaussian to avoid the initial guess for the wavefunction optimisation; it will instead be reading the converged HF wavefunction from the old Checkpoint file. Note that we need to specify a name for the new Checkpoint file in `%Chk=` in order to avoid overwriting the old Checkpoint file when submitting the calculation.
For MP3 and MP4, one often expects the geometry not to change too much anymore, and a reasonable estimate for the energy may already be given by a single-point calculation on the optimised MP2 geometry. Create inputs for MP3 and MP4, where you restart from the MP2 Checkpoint file rather than the HF Checkpoint file. As a reminder, the keyword for a single-point wavefunction optimisation is `SP`.
(In general, it is sometimes necessary to optimise the geometry at a lower Post-Hartree-Fock level, followed by a higher level single point calculation. This is due to a lack of available analytical gradients for certain higher order methods (such as MP4), which causes the computational time for a geometry optimisation to become exorbitantly high.)

a) Complete the following table. Please ensure that before running the Post-HF run, you have appropriately renamed the new MPn Checkpoint file in the input and that for MP3 and MP4, you make sure to start from the MP2 and not the HF geometry. Note that although you will find the statement `SCF Done` in all the outputs, the relevant keyword for the energy in the outputfile will not be `SCF Done` anymore, but will be given by `EUMP2, EUMP3, UMP4(SDTQ)` for MP2, MP3 and MP4 respectively.

Method:	HF	MP2	MP3	MP4	Exp.
$E^{F\cdot}$					n/a
$E^{H_3C\cdot}$					n/a
E^{CH_3F}					n/a
E^{BDE}					109.2 kcal mol^{-1}

b) Is the homolytic cleavage of the H_3C-F bond likely, based on the BDE that you calculated? (Think of radical processes in general.)

c) Why is it a reasonable choice to use Hartree-Fock geometries and wavefunctions as a starting point for optimisations at the Post-Hartree-Fock level? What is the advantage and how is this approach justified?

d) What is the trend in energies when moving from MP2 over MP3 to MP4? Does this series converge?

e) Why can't you extract the MPn energies by grepping `SCF Done`, as you did for the HF calculations?

f) Comment on the accuracy of the HF approach for this system. If the difference between HF and MPn is large, how is the use of MPn still justified? (Hint: Think in terms of the scale of the absolute energy of the system, rather than the reaction.)

g) Why did we choose not to carry out a geometry optimisation for the methods where no analytical gradient is available (such as MP4)?

4.2 Geometry Optimisation of HNO_3: HF vs. MP2 vs. QCISD

The influence of electron correlation on structural parameters such as bond lengths and angles is not negligible. Hartree-Fock calculations may give bond lengths that deviate considerably from the experimental value. Comparably cheap methods such as MPn may often remedy sufficiently for this problem, but some smaller deviations may remain. The most rigorous approach in such cases would be *e.g.* a full configuration interaction (full CI/FCI), but this is computationally untractable for any larger molecule using a reasonable basis. One may therefore restrict the method to singly and doubly excited Slater determinants, assuming that the main effects of correlation are captured well enough in such a truncation. However, such truncations introduce size-consistency problems. Quadratic configuration interaction corrects for any size-consistency errors (*cf.* chapter 2.4 in the course script) in the truncated CISD approach (configuration interaction with single and double excitations); the resulting approach is denoted QCISD, and the results obtained from QCISD are in principle of better quality than those obtained from many-body perturbation approaches such as MP2. QCISD, however, requires more computational resources, and its performance is still outrun by methods such as coupled-cluster approaches (for instance, CCSD(T)).

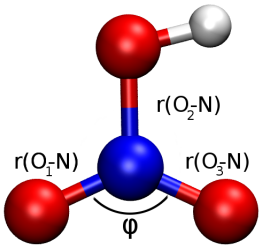
a) Assess and comment on the performance of HF, MP2 and QCISD with respect to the prediction of bond lengths and bond angles in HNO_3 (*Phys. Chem. Chem. Phys.*, **2014** 16, *19437*). For the Hartree-Fock level, you may use the following input:

```
%NProcShared=2
%Mem=1GB
%Chk=HNO3_HF
#P UHF/6-31+G* Opt Guess=Mix Symmetry=None

HNO3 - HF geometry optimisation

0 1
O            1.326   -0.373   -0.127
N            0.015   -0.015    0.415
O           -0.329    1.292   -0.145
O           -0.944   -0.973   -0.135
H           -1.811   -0.738    0.223
```

For the Post-HF methods, proceed as in the previous exercise: Make Gaussian read the converged HF wavefunction (`Guess=Read`) and Geometry (`Geom=Checkpoint`) from the HF Checkpoint file, and do not forget to name the new Checkpoint file differently for every run. Analyse the optimised structures in Molden and compare the following structural parameters:



Method:	HF	MP2	QCISD	Exp.
E [kcal mol^{-1}]				n/a
ϕ_{O-N-O} [°]				130.2
$\mathbf{r}(O_1 - N)$ [Å]				1.198
$\mathbf{r}(O_2 - N)$ [Å]				1.410
$\mathbf{r}(O_3 - N)$ [Å]				1.213

b) Bonus: Even QCISD breaks down for the radical derived from nitric acid, $NO_3\cdot$. Why could that be? (Think of the common basis of all Post-HF methods.)