EPSRC National Service for Computational Chemistry Q–CHEM Workshop

*Imperial College, 1st – 2nd April 2004*

Dr Andrew Gilbert

University of Nottingham
1. Introduction
Objectives of the Workshop

By the end of this workshop it is hoped that you will be able to do the following:

- Be able to quickly set up Q–CHEM input files for a range of molecules and job types.
- Know how to submit Q–CHEM jobs on columbus.
- Be able to interpret the output obtained from running Q–CHEM.
- Have an overview of the capabilities of the Q–CHEM package.
- Know about some of the features that are unique to Q–CHEM.
- Appreciate the limitations of the software.
- Know what to do when things go wrong!
The focus of the first day of the workshop will be on introducing Q–CHEM and running common job types. The outline for today is:

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30</td>
<td>Introduction to Q–CHEM, running single point</td>
</tr>
<tr>
<td></td>
<td>calculations, the basis set and ECP library</td>
</tr>
<tr>
<td>11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:30</td>
<td>Geometry optimisations and transition state</td>
</tr>
<tr>
<td></td>
<td>searches</td>
</tr>
<tr>
<td>1:00</td>
<td>Lunch</td>
</tr>
<tr>
<td>2:00</td>
<td>Frequency calculations and reaction pathways</td>
</tr>
<tr>
<td>3:30</td>
<td>Coffee break</td>
</tr>
<tr>
<td>4:00</td>
<td>Exercise session</td>
</tr>
<tr>
<td>5:30</td>
<td>End of day one</td>
</tr>
</tbody>
</table>
Tomorrow the emphasis will shift onto more specialised topics including:

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:30</td>
<td>High level correlated wavefunction methods</td>
</tr>
<tr>
<td>11:00</td>
<td>Coffee break</td>
</tr>
<tr>
<td>11:30</td>
<td>Methods for calculating excited states</td>
</tr>
<tr>
<td>1:00</td>
<td>Lunch</td>
</tr>
<tr>
<td>2:00</td>
<td>Other topics including parallel capabilities of Q–CHEM, linear scaling methods, NMR calculations and solvent effects</td>
</tr>
<tr>
<td>3:30</td>
<td>Coffee break</td>
</tr>
<tr>
<td>4:00</td>
<td>Exercise session</td>
</tr>
<tr>
<td>5:30</td>
<td>End of day two</td>
</tr>
</tbody>
</table>
Strengths and Capabilities of Q–CHEM

- Highly efficient integrals package based on the COLD PRISM and J–matrix engine.
- Wide range of local, gradient–corrected and hybrid DFT exchange and correlation functionals.
- Linear scaling algorithms for large systems: CFMM, LinK, CASE
- Wide range of high–level correlated methods including CC based methods (CCSD), perturbation methods (MPn) and the recently developed OD methods.
- Local MP2 methods for fast correlation energy evaluation.
- Extensive range of excited state methods: CIS, XCIS, TDDFT, SF–TDDFT, EOM–CCSD, SF–CCSD and VOOD
- Analytic first and second derivatives for HF, DFT and CIS calculations.
- Analytic first derivatives for MP2, CIS(D) and OD.
- Fully integrated ECP package for calculations on heavy elements.
Strengths and Capabilities of Q–CHEM

- Parallel capabilities including:
  - Hartree-Fock and DFT single-point energies
  - Analytic Hartree-Fock and DFT geometry optimisations
  - Analytic Hartree-Fock and DFT frequency analyses
  - Finite difference geometry optimisations for several methods
  - Typical speed-ups of 13-14 on 16 nodes for single-point SCF calculations.

- Properties package allows the calculation of:
  - Optimised structures
  - Transition structures
  - Vibrational frequencies
  - Reaction pathways
  - Solvent effects
  - NMR chemical shifts
  - Electrostatic potentials
2. Basic Input Layout
Input Sections

- A Q–CHEM input file is made up from $section blocks.
- Each $section block is terminated by a $end statement.
- Every input file must contain at least the $molecule and $rem blocks, other blocks are only required for certain types of calculation.
- All input is case insensitive.
- Here is a list of the possible $section blocks:

<table>
<thead>
<tr>
<th>$section</th>
</tr>
</thead>
<tbody>
<tr>
<td>$molecule</td>
</tr>
<tr>
<td>$isotopes</td>
</tr>
<tr>
<td>$rem</td>
</tr>
<tr>
<td>$multipole_field</td>
</tr>
<tr>
<td>$basis</td>
</tr>
<tr>
<td>$occupied</td>
</tr>
<tr>
<td>$comment</td>
</tr>
<tr>
<td>$nbo</td>
</tr>
<tr>
<td>$opt</td>
</tr>
<tr>
<td>$plots</td>
</tr>
<tr>
<td>$ecp</td>
</tr>
<tr>
<td>$intracule</td>
</tr>
<tr>
<td>$xc_functional</td>
</tr>
<tr>
<td>$external_charges</td>
</tr>
</tbody>
</table>
The \$molecule\ Section

The \$molecule\ section defines the geometry of the system that is being studied. The section consists of up to three parts:

- A line containing the total charge and the multiplicity of the system (separated by a space).
- A series of lines (one per atom) that specify the coordinates of the nuclei.
- A blank line followed by the specification of any parameters (bond lengths etc.) used in describing the coordinates.

The last section only applies to z–matrix inputs, and only if parameters were used in constructing the z–matrix.

All lengths are in angstroms and bond angles in degrees. (Bohr can be used with the input\_bohr rem.)

Coordinates may be entered in either cartesian or z–matrix format.
The following two inputs are equivalent descriptions for the singlet state of a neutral water molecule:

```
$molecule
  0 1
  o 0.000 0.000 0.113
  h -0.758 0.000 -0.452
  h 0.758 0.000 -0.452
$end
```

```
$molecule
  0 1
  o
  h1 o oh
  h2 o oh h1 hoh

oh = 0.945
hoh = 106.6
$end
```

Cartesian input

Z–matrix input
The Z–matrix Input

- The z–matrix is a powerful way of inputting molecular geometry information in terms of quantities more familiar to the chemist, (e.g. bond lengths and bond angles).
- Z–matrices are particularly useful for inputting small, symmetric molecules. They are less useful for molecules without symmetry.

- New atom:
  \[ c_1 \]

- New atom, old atom 1, bond distance:
  \[ h_1 \ c_1 \ c_1 \]

- New atom, old atom 1, bond distance, old atom 2, bond angle:
  \[ o_1 \ c_1 \ c_1 \ h_1 \ h_1 \]

- New atom, old atom 1, bond distance, old atom 2, bond angle, old atom 3, dihedral angle:
  \[ h_2 \ o_1 \ o_1 \ c_1 \ c_1 \ h_1 \ h_2 \]

- Subsequent lines follow the same structure as the fourth line.
Consider the methanol molecule:
\[ \text{CH}_3\text{OH} \]
Consider the methanol molecule:
\[ \text{c1 o1 cl co} \]
Consider the methanol molecule:

\[
\begin{array}{c@{\quad}c@{\quad}c@{\quad}c@{\quad}c@{\quad}c}
\text{c1} & \text{o1} & \text{cl} & \text{co} & \text{h1} & \text{ol} \quad \text{oh} \\
\text{cl} & \text{hoc}
\end{array}
\]
Consider the methanol molecule:

\[
\begin{align*}
\text{c1} & \quad \text{o1} \quad \text{cl} \quad \text{co} \\
\text{h1} & \quad \text{o1} \quad \text{oh} \quad \text{cl} \quad \text{hoc} \\
\text{h2} & \quad \text{cl} \quad \text{ch} \quad \text{o1} \quad \text{hco} \quad \text{h1} \quad \text{180}
\end{align*}
\]
Consider the methanol molecule:

\[
\begin{align*}
&\text{c1} \\
&\text{o1} \quad \text{c1} \quad \text{co} \\
&\text{h1} \quad \text{o1} \quad \text{oh} \quad \text{c1} \quad \text{hoc} \\
&\text{h2} \quad \text{c1} \quad \text{ch} \quad \text{o1} \quad \text{hco} \quad \text{h1} \quad 180 \\
&\text{h3} \quad \text{c1} \quad \text{ch} \quad \text{o1} \quad \text{hco} \quad \text{h1} \quad 60
\end{align*}
\]
Putting It Together

Consider the methanol molecule:

```
c1
o1  c1  co
h1  o1  oh  c1  hoc
h2  c1  ch  o1  hco  h1  180
h3  c1  ch  o1  hco  h1  60
h4  c1  ch  o1  hco  h1  -60
```
Consider the methanol molecule:

\[
\begin{array}{cccc}
  c1 & o1 & cl & co \\
  h1 & o1 & oh & cl & hoc \\
  h2 & cl & ch & o1 & hco & h1 & 180 \\
  h3 & cl & ch & o1 & hco & h1 & 60 \\
  h4 & cl & ch & o1 & hco & h1 & -60 \\
\end{array}
\]

\[
\begin{align*}
  ch &= 1.08 \\
  oh &= 0.95 \\
  co &= 1.43 \\
  hoc &= 150 \\
  hco &= 113
\end{align*}
\]

Z–matrix coordinates are not unique, and some inputs will be easier than others.
Bond angles of $180^\circ$ should not be used in z–matrices. If required, a
dummy atom should be introduced and the angle redefined in terms
of two $90^\circ$ angles.

$\text{C}_2\text{H}_2$
Dummy Atoms – $\text{C}_2\text{H}_2$

Bond angles of 180° should not be used in z–matrices. If required, a dummy atom should be introduced and the angle redefined in terms of two 90° angles.

$$\begin{align*}
\text{c1} & \quad | \\
\text{c2} & \quad \text{c1} \quad \text{cc}
\end{align*}$$
Dummy Atoms – C₂H₂

Bond angles of 180° should not be used in z–matrices. If required, a dummy atom should be introduced and the angle redefined in terms of two 90° angles.

c1

c2  c1  cc

x1  c1  1  c2  90
Dummy Atoms – C$_2$H$_2$

Bond angles of 180° should not be used in z–matrices. If required, a dummy atom should be introduced and the angle redefined in terms of two 90° angles.

c1

<table>
<thead>
<tr>
<th>c2</th>
<th>c1</th>
<th>cc</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>c1</td>
<td>1.</td>
</tr>
<tr>
<td>x2</td>
<td>c2</td>
<td>1.</td>
</tr>
</tbody>
</table>
Dummy Atoms – $\text{C}_2\text{H}_2$

Bond angles of 180° should not be used in z–matrices. If required, a dummy atom should be introduced and the angle redefined in terms of two 90° angles.

```
c1
  c2  c1  cc
x1  c1  1.  c2  90
x2  c2  1.  c1  90  x1  0
h1  c1  ch  x1  90  x2  180
```
Dummy Atoms – $\text{C}_2\text{H}_2$

Bond angles of 180° should not be used in z–matrices. If required, a dummy atom should be introduced and the angle redefined in terms of two 90° angles.

```
c1
cc = 1.21
```

```
c2  c1  cc
x1  c1  1.  c2  90
x2  c2  1.  c1  90  x1  0
h1  c1  ch  x1  90  x2  180
h2  c2  ch  x2  90  x1  180
```

```
ch = 1.07
```
Bond angles of $180^\circ$ should not be used in z–matrices. If required, a dummy atom should be introduced and the angle redefined in terms of two $90^\circ$ angles.

```
c1  
c2  c1  cc  
x1  c1  1.  c2  90  
x2  c2  1.  c1  90  x1  0  
h1  c1  ch  x1  90  x2  180  
h2  c2  ch  x2  90  x1  180  
```

$cc = 1.21$

$ch = 1.07$

Placing a dummy atom along the principal axis of rotation is usually a good idea.
The $\texttt{rem}$ Array

- Global options for Q–CHEM jobs are stored in an array called the $\texttt{rem}$ array (Latin for “thing”).

- Most $\texttt{rem}$ options have suitable default settings.

- At the very least, the basis set and treatment of exchange must be specified.

- A minimal $\texttt{rem}$ block is given by:

  $\texttt{rem}$
  
  exchange hf

  basis 6-31G*

  $\texttt{end}$

- The order of the rem variables does not matter.

- All $\texttt{rem}$ entries are actually integers, although all $\texttt{rem}$ keys and many $\texttt{rem}$ values have macros defined that make the input more human–friendly.

- Unfortunately not all $\texttt{rem}$ entries are quite so human–friendly!
Basic $\text{rem}$ Options

In addition to the basis set and exchange, here are a few other basic $\text{rem}$ variables that you should be aware of, the defaults are shown in brackets:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>jobtype</td>
<td>(sp)</td>
<td>correlation</td>
</tr>
<tr>
<td>max_scf_cycles</td>
<td>(50)</td>
<td>xc_grid</td>
</tr>
<tr>
<td>xc_smart_grid</td>
<td>(true)</td>
<td>mem_total</td>
</tr>
<tr>
<td>mem_static</td>
<td>(32)</td>
<td>print_orbitals</td>
</tr>
<tr>
<td>scf_print</td>
<td>(0)</td>
<td>scf_algorithm</td>
</tr>
<tr>
<td>scf_guess</td>
<td>(sad)</td>
<td>unrestricted</td>
</tr>
<tr>
<td>correlation</td>
<td>(none)</td>
<td></td>
</tr>
<tr>
<td>xc_grid</td>
<td>(0)</td>
<td></td>
</tr>
<tr>
<td>mem_total</td>
<td>(2000)</td>
<td></td>
</tr>
<tr>
<td>print_orbitals</td>
<td>(false)</td>
<td></td>
</tr>
<tr>
<td>scf_algorithm</td>
<td>(diis)</td>
<td></td>
</tr>
<tr>
<td>unrestricted</td>
<td>(varies)</td>
<td></td>
</tr>
</tbody>
</table>

Other more specialised options will be introduced as required. Note that any input on a rem line after the first two tokens is ignored. This can be useful for adding comments to options that are not obvious:

```
xc_grid  1     (Use the SG-1 quadrature grid)
```
Now that we can write a basic input file, let’s run a job!

- Write an input file `jobname.in`.
- Write a submit file `jobname.run` with the line:
  
  ```bash
  $CHEM/runqchem nproc jobname
  ```
- Make the script executable `chmod +x jobname.run`.
- Submit the script to one of the queues (`bqueue -l` will give a list of the available queues and their limits):
  
  ```bash
  bsub -n nproc -q queuename -o batchout jobname.run
  ```
- A file `jobname.out` will appear with the Q–CHEM output.
- Check the `batchout` file for any error messages.
- `man runqchem` will bring up a man page for the script, and contains full details about running Q–CHEM jobs on columbus.
Exercise: The Water Molecule

Using your favourite editor, create a file *water.inp* with the following:

```
$molecule
0 1
 o
h1  o  oh
h2  o  oh  h1  hoh

oh     0.967
hoh    108
$end

$rem
   jobtype       sp
   exchange      hf
   basis         3-21G
$end
```

Run the file using the `runqchem` command to obtain the Q–CHEM output.
Structure Of Basic Output file

Here is an outline of the output file you have just created, and the steps Q–CHEM goes through to create it:

- **Echo the input options.**
- **Orient the molecule into the standard nuclear orientation. This allows the symmetry to be determined.**
  - Check to see that the point group is as you expect it to be.
  - Round–off in the input may result in a descent in symmetry.
- **Determine the shell information for the basis set.**
- **Print internuclear distances (if the number of atoms is < 15).**
- **Apply the standard electronic orientation quadrupole field.**
- **Determine the shell–pair information.**
  - Important step in setting up the internal data structures.
  - Cutoffs are applied to reduce the number of significant shell–pairs.
Structure Of Basic Output file

- Compute the overlap matrix and its eigenvalues:
  - Small eigenvalues ($< 10^{-5}$) indicates near linear–dependence.

- Generate an initial guess for the SCF calculation.
  - The default SAD guess will print a warning.

- Print out information about what sort of SCF calculation will be performed.
  - RHF, ROHF, UHF or Kohn–Sham (DFT)
  - What convergence algorithm and criterion will be used.

- The SCF calculation:
  - Prints energy and convergence metric (DIIS or gradient) on each cycle.
  - Any problems will be reported.
  - The `scf_print` rem can be set to increase the output on each cycle if required.
Structure Of Basic Output file

- Wavefunction analysis:
  - Orbital energies and symmetries.
  - Mulliken analysis (and spin densities).
  - Cartesian multipole moments.

The basic structure above is repeated for iterative jobs such as geometry optimisations, reaction pathways and finite difference calculations.
# Job Types

A list of the different types of job Q–CHEM can run is given below. They are selected using the `jobtype` rem:

<table>
<thead>
<tr>
<th>Type of Calculation</th>
<th>Jobtype Rem Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single point energy</td>
<td><code>single_point</code></td>
</tr>
<tr>
<td>Force calculation</td>
<td><code>force</code></td>
</tr>
<tr>
<td>Equilibrium structure search</td>
<td><code>optimization</code></td>
</tr>
<tr>
<td>Transition structure search</td>
<td><code>ts</code></td>
</tr>
<tr>
<td>Intrinsic reaction pathway</td>
<td><code>rpath</code></td>
</tr>
<tr>
<td>Frequency calculation</td>
<td><code>frequency</code></td>
</tr>
<tr>
<td>NMR chemical shifts</td>
<td><code>nmr</code></td>
</tr>
</tbody>
</table>

These different job types will be addressed in the following sections.
Sometimes it is useful to concatenate two or more jobs in a single input file. For example, a geometry optimisation followed by a frequency calculation using the optimised geometry.

- Separate inputs are delimited by a single line with `@@@`
- The molecular geometry, guess MO coefficients and Hessian information can all be read in from the previous calculation.

```
...previous job...
@@@
$molecule
  read
$end
        $rem
  scf_guess       read
  geom_opt_hessian read
   ....
$end
```
# 3. Self–Consistent Field Calculations

- Initial Guesses
- SCF Algorithms
- RHF and UHF
- DFT Calculations
- Quadrature Grids
- Basis Sets
- ECPs
Introduction

The simplest calculation Q–CHEM can do is a single point (SCF) energy calculation. This involves solving the Roothaan–Hall (restricted HF), Pople–Nesbet (unrestricted HF) or Kohn–Sham (DFT) equations within a given basis set:

\[ FC = \varepsilon SC \]

These equations are non-linear, \( F = F(C) \), and must be solved iteratively:

\[ C \rightarrow P \rightarrow F \rightarrow E & C \rightarrow P \rightarrow F \rightarrow E & C \rightarrow \ldots \]

In direct calculations, the two–electron integrals must be recalculated at each SCF cycle, therefore ensuring the SCF converges quickly is very important.

Two key factors influence the rate of convergence:

- Quality of the starting point (initial guess orbitals).
- The algorithm for obtaining the new Fock matrix, \( F \), from \( P \).
The SAD Guess

- **Q–CHEM** has five options for generating an initial starting point for the SCF. These can be selected using the `scf_guess` rem variable.

- The default guess uses a superposition of atomic densities (SAD) to generate the density matrix for the entire molecule.

- The SAD guess is superior to the other methods for most systems, especially for large systems and calculations using large basis sets. It usually saves several SCF cycles.

- The SAD guess is usually the best guess for slightly charged species where it will get the number of electrons wrong.

- The SAD guess generates an initial density matrix, and cannot be used if orbitals are required before the first SCF cycle:
  - When orbital occupancies are altered (`occupied` section).
  - When direct minimisation is used (`scf_algorithm = gdm`).

- The SAD guess is not idempotent, and therefore at least two SCF cycles must be performed before convergence to ensure the energy is variational (hence the warning).
The Core and GWH Guesses

- The core guess is the simplest, but also the crudest and is therefore not often used. It is requested by
  \[ \text{scf\_guess} = \text{core} \]

- In the core guess the two–electron part of the Hamiltonian is simply neglected, leaving only the kinetic and potential energy terms.

- The core guess works best for small systems with small basis sets.

- In larger systems, more inter–electron effects are ignored. This degrades the guess.

- The Generalised Wolfsberg–Helmholtz guess is a more sophisticated guess based on the core Hamiltonian.
  \[ \text{scf\_guess} = \text{gwh} \]

- GWH suffers the same fate as the core guess when applied to larger systems.

- Both these guesses are simple options when no SAD guess is available (e.g. for a custom basis set), or when initial orbitals are required.
MO coefficients from a previous job can be read in to provide an accurate initial guess (this is the default option in subsequent cycles of a geometry optimisation).

```
scf_guess = read
```

The necessary information must be generated, and saved, by running Q–C HEM with three arguments:

```
prompt> qchem job1.inp job1.out job.save
```

These can then be used in a subsequent calculation:

```
prompt> qchem job2.inp job2.out job.save
```

Alternatively the two jobs can be run in the single file using the `@@@` construct.

The user must ensure that the same basis set is used across the two calculations, no internal checking is performed.

This approach can be useful for generating guesses for open–shell species by first generating the orbitals for the corresponding closed–shell anion.
Basis set projection is useful for custom basis sets for which no SAD guess exists.

A guess density matrix is obtained using a small basis set specified by the `basis2` rem.

The density matrix is projected onto the larger basis set to form an initial guess.

Minimal basis sets (e.g. STO–3G) should be used for the guess, and little is gained using a larger basis, but more time will be spent on generating the guess.
Modifying The Initial Orbitals

- In some cases it may be desirable to alter the guess orbitals:
  - To obtain a state of different symmetry, or orbital occupation
  - To break spatial symmetry
  - To break spin symmetry, for example in unrestricted calculations with an even number of electrons.

- A simple breaking of symmetry is obtained by mixing the HOMO and LUMO orbitals of the initial guess:
  
  \[
  \text{scf\_guess\_mix n}
  \]
  
  (adds \(n \times 10\%\) of the LUMO to the HOMO, default is 0 \%)

- Orbital occupancies can be altered using the \$\text{occupied} section (first line specifies \(\alpha\) orbitals, second line \(\beta\) orbitals):
  
  \[
  \$
  \]
  
  \[
  \begin{array}{cccccc}
  1 & 2 & 3 & 4 & 5 & 6 \\
  1 & 2 & 3 & 4 & 5 & 7 \\
  \end{array}
  \$
  \]

- These approaches cannot be used with the default SAD guess.
Exercises: Lithium

- Construct an input deck for the lithium atom using HF exchange, the 6–311+G basis and the GWH guess. Run the calculation and make a note of the SCF energy.

- Repeat the calculation, only this time alter the orbitals occupancies as follows:

  \[
  \text{occupied} \\
  1 \\
  5 
  \]

- Compare the energies. Are they what you expected?

- Construct another input deck for Li$_2$ using the same options as above for the basis and exchange. Use a bondlength of 5Å.

- Experiment with the unrestricted and scf_guess_mix rems, and determine what combination gives the lowest energy.

- Unrestricted HF theory will be mentioned later.
SCF Algorithms – DIIS

- The second important factor influencing the convergence of the SCF is the algorithm used to generate the new Fock matrix.
- The most widely used algorithm is direct inversion of the iterative subspace (DIIS).
- The DIIS algorithm makes use of the property that the density and Fock matrices should commute:

\[ SPF - FPS = 0 \]

- Prior to convergence, this equation will not be satisfied, giving rise to an error vector:

\[ SP_i F_i - F_i P_i S = e_i \]

- The new Fock matrix is determined by mixing previous Fock matrices. The mixing coefficients, \( c_j \), are determined by least-squares fit of the error vector, \( e_i \):

\[ F_i \leftarrow \sum_{j=i-N+1}^{i} c_j F_j \]
During the cycles of the DIIS algorithm, the system is never on the true potential energy surface because the new Fock matrix is obtained via an extrapolation.

This property allows DIIS to “tunnel though” energy barriers in MO space and quickly locate the vicinity of the global minimum, rather than finding local minima. This is usually a desirable property.

Having found the approximate location of the minimum, however, this same property of DIIS can cause it to perform poorly when close to convergence.

The DIIS history may contain bad information which prevents the projection getting close enough to the true potential energy surface. This can result in oscillations in the energy and non–convergence.

These problems can sometimes be fixed by changing the iterative subspace size, \( N \) (diis_subspace_size).
Geometric Direct Minimisation (GDM)

- Direct minimisation use the fact that the molecular orbital coefficients at each SCF cycle are a linear transformation of those of the previous cycle:

\[ C^{n+1} = U^n C^n \]

- The energy is then minimised with respect to the orbital transformation \( U \):

\[ \frac{\partial E}{\partial U} = 0 \]

- Direct application of this approach (DM) takes you off the energy surface at each step as it doesn’t take account of any constraints on \( U \), in particular:

\[ U^\dagger U = I \]

- Geometric direct minimisation (GDM) correctly takes account of any constraints, and steps along curved geodesics rather than straight lines in MO coefficient space.
Hybrid Algorithms – DIIS_GDM

- The direct minimisation based algorithms (DM/GDM) are more reliable for finding the exact minimum once in the vicinity as they always step in a direction that lowers the energy.

- Away from the minimum, however, DM/GDM can behave erratically and be slow to converge.

- DIIS is good at quickly locating the quadratic energy well, but can be slow to tighten down on the convergence criterion.

- DIIS and GDM work in complementary ways ⇒ combine them for the best of both worlds.

- DIIS_GDM is the recommended fall-back when DIIS fails.

- If the DIIS error fluctuates around $10^{-3}$, DIIS_GDM can often help.

- If DIIS fails to get even this far, then other factors may be the cause:
  - Quality of the initial guess.
  - Quality of the quadrature grid (for DFT calculations).
  - Basis set linear dependence.
  - Inappropriate level of theory (ROHF).
$\text{rem Variable Settings}$

The following gives a list of the main rem variables that affect the SCF:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>max_scf_cycles</td>
<td>$n &gt; 0$</td>
<td>50</td>
</tr>
<tr>
<td>scf_guess</td>
<td>SAD/GWH/Read/Core</td>
<td>SAD</td>
</tr>
<tr>
<td>scf_convergence</td>
<td>$n \Rightarrow 10^{-n}$</td>
<td>5/8</td>
</tr>
<tr>
<td>scf_algorithm</td>
<td>DIIS/GDM/DIIS_GDM</td>
<td>varies</td>
</tr>
<tr>
<td>thresh_diis_switch</td>
<td>$n \Rightarrow 10^{-n}$</td>
<td>2</td>
</tr>
<tr>
<td>diis_subspace_size</td>
<td>$n$</td>
<td>15</td>
</tr>
<tr>
<td>unrestricted</td>
<td>true/false</td>
<td>varies</td>
</tr>
</tbody>
</table>

The default values of scf_algorithm and unrestricted depend on if the system is open– or closed–shell.
Exercise: CoCO⁺

- Construct an input deck for the linear CoCO⁺ cation using Hartree–Fock exchange, the 6–31G(d) basis set and the GWH initial guess. Use the following bond lengths:

  \[
  \begin{align*}
  \text{R(Co–C)} & \quad 3.613 \, \text{Å} \\
  \text{R(C–O)} & \quad 2.141 \, \text{Å}
  \end{align*}
  \]

- Compare the performance of the GDM and DIIS algorithms. (Note you will need to adjust the maximum number of SCF cycles to around 200).

- How much longer does the calculation take if the convergence threshold is tightened to \(10^{-6}\)?
Open–shell systems can be dealt with in one of two ways:

- **Restricted open–shell Hartree–Fock (ROHF):**
  - Open–shell orbitals are added to the closed–shell ones.
  - Resulting wavefunction is projected to ensure it has the correct spin symmetry.
  - Model breaks down when several configurations become (near–)degenerate, e.g. stretched bonds.

- **Unrestricted Hartree–Fock (UHF):**
  - Spatial parts of the \( \alpha \) and \( \beta \) spin orbitals are no longer constrained to be the same.
  - Able to describe the spin–polarisation of paired electrons in the presence of unpaired electrons.
  - Gives qualitatively correct description of bond–dissociation.
  - Usually better around transition states.
  - Resulting wavefunction cannot be easily projected to give a wavefunction that is a spin–eigenvalue.
Spin Contamination

- For restricted calculations, the wavefunction will be an eigenfunction of the $S^2$ operator with the eigenvalue given by:

  \[
  \langle S^2 \rangle = S(S+1) \quad 0.00 \quad 0.75 \quad 2.00 \quad 3.75
  \]

- For unrestricted calculations, spin contamination may occur if the system has multi-determinant character.

- A value for $\langle S^2 \rangle$ for unrestricted calculations is printed at the end of the SCF, if this differs by more than about 10% of the expected value, the wavefunction is spin contaminated and should not be used as a starting point for post-HF methods.

- Radicals and systems with stretched bonds often exhibit spin contamination.
Exercise: The ·CN Radical

- The ·CN radical is a classic case where the unrestricted HF wavefunction exhibits spin contamination.
- The repulsion of the electrons in the bonding region is not properly accounted for as the effects of electron correlation have been ignored.
- Mixing in the quartet state introduces more Pauli repulsion and allows the electrons to reside further away from each other on average.
- Construct an input deck for the ·CN radical using the bond length R(C–N) = 1.162 Å.
- Perform an unrestricted HF calculation using the cc–pvtz basis, and determine the degree spin contamination.
- Consider the spin–density on each atom and predict if the quartet state is likely to suffer from the same degree of spin–contamination.
- Carry out the quartet calculation to confirm (or refute!) your prediction.
DFT Calculations

- Solving the Kohn–Sham equations in a DFT calculation is very similar to solving the Roothaan–Hall (or Pople–Nesbet) equations in HF theory.

- A DFT calculation is requested by setting the following rems:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options (Functional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>exchange</td>
<td>slater, becke, g96, gg99, becke(edf1), pw91, b3pw91, b3lyp, b3lyp5, edf1, gen</td>
</tr>
<tr>
<td>correlation</td>
<td>none, vwn, lyp, pw91, lyp, p86, pz81, wigner</td>
</tr>
</tbody>
</table>

- The following rem variables are also relevant to DFT calculations:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>xc_grid</td>
<td>details next</td>
<td>0 ⇒ SG–0/SG–1</td>
</tr>
<tr>
<td>xc_smart_grid</td>
<td>true/false</td>
<td>true</td>
</tr>
<tr>
<td>incdft</td>
<td>true/false</td>
<td>false</td>
</tr>
</tbody>
</table>
Many of the integrals involved with DFT calculations cannot be evaluated analytically.

These integrals are evaluated by placing 3–D quadrature grids around each atom, and making the approximation:

$$\int F[\rho(r)]dr \approx \sum_i w_i F[\rho(r_i)]$$

The weights, $w_i$, and roots, $r_i$, define the quadrature grid.

The quality of the quadrature grid can have a significant impact on the results of a DFT calculation.

In Q–CHEM an internal check of the grid quality is carried out by integrating the density. If this integral differs by more than 0.01% from the number of electrons in the system, a warning is printed.

Evaluating the density at the grid points is the rate–determining step in a DFT calculation, therefore the grid must be as efficient as possible.
Specifying Quadrature Grids In Q–CHEM

- The 3-D grid is broken up into radial and angular parts which are treated separately.
- The radial integration uses Euler–Maclaurin quadrature.
- The angular integration uses either Lebedev, or Gauss–Legendre quadrature.
- Lebedev grids are more efficient, but are only available for certain “magic” numbers of points (110, 146, 194, 302...).
- Arbitrarily large Gauss–Legendre grids can be generated for high accuracy.
- The grid is specified with 12 digits (and a negative sign if Gauss–Legendre quadrature is desired) using the rem:

\[ \text{xc\_grid} = \pm 000060000194 \]

- radial

- angular

- The default grid in Q–CHEM (SG–0/SG–1) is designed for maximum efficiency, and has been carefully pruned.
Basis Sets Within Q–CHEM

- Q–CHEM has a wide selection of inbuilt basis sets including:
  - Pople basis sets (STO–\(n\)G, 3–21G(\(m, n\)), 6–311G(\(m, n\))
  - Dunning basis sets (SV, DZ, TZ).
  - Ahlrichs basis sets.

- These basis sets can be requested using the standard symbolic representation.

- Additional diffuse functions are available for calculations on excited states and anions. E.g. DZ(2+), 6–31(+,+))G

- BSSE corrections can be performed.

- A mixture of basis sets can be applied to the atoms in a given molecule.

- User–defined basis sets can be input.
User–defined Basis Sets

- In addition to the library of standard basis sets, Q–CHEM allows the user to input an arbitrary customised basis set.

- User–defined basis sets are requested by setting
  \[ \text{basis} = \text{general} \]

- The basis set is specified within a $\text{basis}$ block.

- Angular momentum types $s, p, sp, d, f$ and $g$ are supported for energy evaluations.

- The basis can be represented as cartesian functions, or the more efficient pure form. The \text{purecart} \text{ rem} \text{ must} be set if a general basis set is being used.

- When modifying an existing basis set, it is a good idea to use the \text{print_general_basis} = \text{true} \text{ rem}, which will print out the basis in a suitable input format.

- User defined basis sets will be considered further in one of the exercises this afternoon.
Mixed Basis Sets

- At times it may be desirable to use different basis sets for the same type of atom within a system. For example, a smaller basis set could be used for the solvent atoms in an explicit solvent calculation.

- Mixed basis sets are requested with $basis = mixed$, and by including a $basis$ section.

- A basis set for *every* atom must be included in the $basis$ section. This differs from the general basis where a basis has to be specified only for each atom type.

- The format is as follows:

  ```
  $basis
  atomtype
  basis
  ****
  $end
  $basis
  atomtype
  basis
  ****
  $end
  ```

- Note that the use of different basis sets for a particular element means that the potential energy surface is no longer unique.
The basis set superposition error (BSSE) arises as a direct result of solving the wave equation in a subspace determined by the selected basis set.

This subspace changes when we add or subtract nuclei (with their associated basis functions) to the system, which can potentially cause problems with any type of dissociation calculation.

To ensure this subspace is static, all calculations should be performed with the total set of basis basis functions, irrespective of whether or not the nuclei are there.
The BSSE can be calculated within Q–CHEM by using the mixed basis set option.

A basis set for each atom must be specified in the $basis$ section.

Ghost atoms (symbol Gh) should be specified in the $molecule$ section for the atoms whose nuclei and electrons are not to be included in the calculations.

The ghost atoms should not be included in the $basis$ section.
Exercise: (H₂O)₂ Binding Energy

Using the given coordinates, construct an input deck for the water dimer and calculate the HF/6–31++G(d,p) energy.

Calculate the binding energy of the dimer at the same level of theory making sure to include the effects of BSSE.

Compare your answer with the high–level MP4/aug–cc–pVQZ value of -4.761 kcal/mol (1 Eₜₜ = 627.51 kcal/mol).

How much effect does the BSSE have on your result?
Effective Core Potentials

- Quantum chemical methods are more difficult to apply to heavier elements for two reasons:
  - The number of electrons in heavier elements is large.
  - Relativistic effects become significant.

- Most of the electrons in these elements are core electrons, and are therefore fairly chemically inactive.

- The relativistic effects are larger for core electrons as they move much more rapidly around the heavy nuclei.

- Effective core potentials (ECPs) address both of these issues by modelling the core electrons by an effective potential.

- This speeds the calculation by removing the basis functions associated with the core electrons, and, if the ECP has been developed based on relativistic calculations, allows for the effects of relativity in the core electrons.

- Loosely speaking, ECPs can be thought of as a kind of basis set that models the core region.
Effective Core Potentials

ECPs can be classified by:
- Their treatment of relativistic effects.
- The definition of the core electrons.

Relativistic ECPs tend to yield better results.

Small core ECPs (where electrons in the outermost two shells are treated explicitly) should be favoured over large core ECPs (where only the outermost shell is treated explicitly), despite their additional cost.

Each ECP has a corresponding orbital basis set which is used for atoms for which an ECP is not defined.

Where analytic first derivatives are available in Q–CHEM, they are also supported for ECPs.

Semi–analytic derivatives are available for ECP DFT calculations.

An ECP calculation is requested by setting the `ecp` rem to the chosen ECP, just like requesting a basis. In most cases the `basis` rem should be set to the same value.
ECPs Available Within Q–CHEM

- Q–CHEM has an inbuilt library with the following ECPs:

<table>
<thead>
<tr>
<th>ECP</th>
<th>Core</th>
<th>Relativistic</th>
<th>Basis</th>
</tr>
</thead>
<tbody>
<tr>
<td>HWMB</td>
<td>Large</td>
<td></td>
<td>STO–3G</td>
</tr>
<tr>
<td>LANL2DZ</td>
<td>Large</td>
<td></td>
<td>6–31G</td>
</tr>
<tr>
<td>SBKJC</td>
<td>Large</td>
<td></td>
<td>3–21G</td>
</tr>
<tr>
<td>CRENBL</td>
<td>Large</td>
<td>✓</td>
<td>6–311G(d)</td>
</tr>
<tr>
<td>CRENBS</td>
<td>Small</td>
<td>✓</td>
<td>STO–3G</td>
</tr>
<tr>
<td>SRLC</td>
<td>Large</td>
<td>✓</td>
<td>6–31G</td>
</tr>
<tr>
<td>SRSC</td>
<td>Small</td>
<td>✓</td>
<td>6–311G(d)</td>
</tr>
</tbody>
</table>

- Q–CHEM also supports user–defined ECPs, however, the development of ECPs is more subtle than for basis sets, and will not be discussed here.

- Note that because the core electrons are not treated explicitly, the total energies will be very different than those obtained from an all–electron calculation.
Application: ECPs

- Platin consists of a single platinum atom surrounded by two chlorine atoms, and two ammonia groups.
- The coordination of the platinum is square–planar, giving rise to two possible isomers.

Cis–platin is used in treating various forms of cancer, including the bladder, ovaries and testes. Unfortunately it is toxic to the whole body and not just to cancer cells.
- Trans-platin is not useful as an anti–cancer agent.
Exercise: Cis–Platin

- Construct input decks for both cis and trans isomers of platin. Base your structures on the following geometric parameters:

<table>
<thead>
<tr>
<th></th>
<th>Cis</th>
<th>Trans</th>
</tr>
</thead>
<tbody>
<tr>
<td>R(Pt–Cl)</td>
<td>2.465 Å</td>
<td>2.484 Å</td>
</tr>
<tr>
<td>R(Pt–N)</td>
<td>2.082 Å</td>
<td>2.038 Å</td>
</tr>
<tr>
<td>R(N–H)</td>
<td>1.055 Å</td>
<td>1.052 Å</td>
</tr>
<tr>
<td>( \angle (H–N—Pt) )</td>
<td>116.2°</td>
<td>111.9°</td>
</tr>
</tbody>
</table>

- Determine the energy of these structures using B3LYP and the HWMB ECP for platinum and chlorine, and the corresponding HWMB basis for nitrogen and hydrogen. Note that both structures have \( C_{2v} \) symmetry.

- Determine if the biologically active isomer is thermodynamically favoured. What practical consequences might this have for using the compound medicinally?
4. Geometry Optimisations
The Optimisation Package In Q–CHEM

- Q–CHEM uses the OPTIMIZE package written by Jon Baker.
- The package optimises using highly efficient delocalised internal coordinates.
- A wide range of constraints may be imposed on the optimisation.
- Eigenvector following or GDIIS algorithms exist for transition structure searches.
- The optimiser is highly customisable, although suitable default options exist for most options.
- The package takes full advantage of the wide range of analytic derivatives available within Q–CHEM.
- A geometry optimisation is requested by setting the \$rem variable
  \texttt{jobtype} = \texttt{opt}
Critical Points

Critical points on the molecular potential energy surface play an important role in chemistry. These points are characterised by a vanishing energy gradient:

\[ \frac{\partial E}{\partial R_i} = 0 \]

- Local minima on the PES correspond to equilibrium geometries.
- Saddle points correspond to transition structures.
The Newton–Raphson Step

The principles behind a geometry optimisation are the same as those of finding the critical points of a one-dimensional function. We seek to minimise the energy with respect to variations in the nuclear coordinates:

\[
E(R_0 + h) = E(R_0) + \sum_i h_i \frac{\partial E}{\partial R_i} \bigg|_{R_0} + \frac{1}{2} \sum_{ij} h_i \frac{\partial^2 E}{\partial R_i \partial R_j} h_j \bigg|_{R_0} + \ldots
\]

\[
= E(R_0) + h^t g(R_0) + \frac{1}{2} h^t H(R_0) h + \ldots
\]

Minimising with respect to the unknown step vector \( h \), and ignoring quadratic and higher order terms in \( h \) leads to

\[
h = -H^{-1} g
\]

At the new point, the Hessian must be updated, the gradient calculated and a new step vector, \( h \), computed. This cyclic process is continued until convergence.
Factors Affecting The Optimisation

Several factors influence the convergence of the optimisation:

- **Quality of the initial geometry** –
  As you move away from a stationary point, the assumption that higher order terms in $h$ can be neglected becomes less valid.

- **The optimisation algorithm used** –
  Certain assumptions about the PES can be made to give the optimiser ‘more information’ about the surface.

- **Quality of the (estimated) Hessian and gradient** –
  If the Hessian is of a poor quality, $h$ will be poorly estimated. This may also be the case if the gradients have been calculated numerically.

- **Coordinate system used** –
  If the coordinate system used has near linear–dependence then the optimiser can be slow to optimise with respect to certain coordinates.
The Hessian

At a critical point the energy gradient is zero. To determine the nature of the critical point we must consider the second derivative information, i.e. the Hessian matrix:

$$H_{ij} = \frac{\partial^2 E}{\partial R_i \partial R_j}$$

The eigenvalues of the Hessian matrix characterise the critical point.

- All eigenvalues positive $\Rightarrow$ an equilibrium geometry.
- Exactly one negative eigenvalue $\Rightarrow$ a transition structure.

Calculating the Hessian is very expensive, and therefore the optimiser will usually try to estimate it rather than calculate it explicitly. In some cases, calculating an initial exact Hessian can significantly speed convergence and converge problematic cases. This should be read in as:

```
geom_opt_hessian read
```
Coordinate systems

Q–CHEM has a choice of three different coordinate systems:

- **Cartesian coordinates:**
  The \((x, y, z)\) coordinates of the nuclei, which are usually coupled.

- **Z–Matrix coordinates:**
  Can be useful if carefully chosen, but are less commonly used.

- **Delocalised internal coordinates:**
  Bond angles and lengths that are well–suited for optimisations.

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>geom_opt_coords</td>
<td>0 ⇒ Cartesian coordinates</td>
</tr>
<tr>
<td></td>
<td>1 ⇒ Internal coordinates</td>
</tr>
<tr>
<td></td>
<td>-1 ⇒ Internal coordinates → cartesian</td>
</tr>
<tr>
<td></td>
<td>2 ⇒ Z–matrix coordinates</td>
</tr>
<tr>
<td></td>
<td>-2 ⇒ Z–matrix coordinates → cartesian</td>
</tr>
</tbody>
</table>

- The default is to use delocalised internal coordinates, which are usually the most efficient, and resort to cartesians if these fail.
Convergence Criteria

The convergence of the optimisation is controlled via the following:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>geom_opt_tol_gradient</td>
<td>$n \Rightarrow n \times 10^{-6}$ 300 $\Rightarrow$ 0.0003</td>
</tr>
<tr>
<td>geom_opt_tol_displacement</td>
<td>$n \Rightarrow n \times 10^{-6}$ 1200 $\Rightarrow$ 0.0012</td>
</tr>
<tr>
<td>geom_opt_tol_energy</td>
<td>$n \Rightarrow n \times 10^{-8}$ 100 $\Rightarrow$ 0.000001</td>
</tr>
<tr>
<td>geom_opt_max_cycles</td>
<td>$n$ (50)</td>
</tr>
</tbody>
</table>

The optimisation will converge if $\text{geom\_opt\_tol\_gradient}$ is below the tolerance, and at least one of $\text{geom\_opt\_tol\_displacement}$ and $\text{geom\_opt\_tol\_energy}$ are satisfied.
Exercise: Ethane

Ethane exists in two rotamers, eclipsed and staggered (gauche).

Optimise both structures at the EDF1/VDZ level of theory. Use the following initial geometric parameters:

\[
R(\text{C–C}) = 1.40 \text{ Å} \quad R(\text{C–H}) = 1.00 \text{ Å} \quad \angle (\text{H–C–C}) = 109^\circ
\]

Try changing the optimisation coordinates to cartesian. Does this improve convergence, or make it worse?

Note that the optimisation is carried out in the point group of your initial geometry.
Constrained Optimisations

The **OPTIMIZE** package allows the user to impose a wide variety of constraints on the system, including:

- Interatomic distances
- Bond angles
- Out-of-plane bends
- Dihedral angles
- Coplanar bends
- Perpendicular bends
- Frozen atoms

Constraints are entered using a `$opt` section.

Initial geometries do not have to satisfy the imposed constraints.

Constraint input is case-insensitive, but there should be no space at the start of each line.

Dummy atoms should be ignored when numbering the atoms in the constraint.
Constraints

1. Stretch
2. Bend
3. Dihedral (torsional) angle
4. Coplanar bend
5. Out-of-plane bend
Constrained Optimisations

The format for the constraints within the $\texttt{opt}$ section is a four letter identifier, followed by the atoms involved in the constraint, and finally the constraint value:

**Interatomic distances (Å, $0.0 < \text{value}$)**

\texttt{stre \ atm1 \ atm2 \ value}

**Bond angles (degrees, $0 \leq \text{value} \leq 180$)**

\texttt{bend \ atm1 \ atm2 \ atm3 \ value}

**Out–of–plane bends (degrees, $-180 \leq \text{value} \leq 180$)**

\texttt{outp \ atm1 \ atm2 \ atm3 \ atm4 \ value}

**Dihedral angles (degrees, $-180 \leq \text{value} \leq 180$)**

\texttt{tors \ atm1 \ atm2 \ atm3 \ atm4 \ value}

**Coplanar bend (degrees, $-180 \leq \text{value} \leq 180$)**

\texttt{linc \ atm1 \ atm2 \ atm3 \ atm4 \ value}

**Perpendicular bend (degrees, $-180 \leq \text{value} \leq 180$)**

\texttt{linp \ atm1 \ atm2 \ atm3 \ atm4 \ value}
Exercise: NH₃

- Construct an input deck for NH₃ and optimise you guess structure using S–VWN/DZ.
- Re-optimise your structure with the N–H bond length constrained to 1.2 Å. What bond H–N–H bond angle do you obtain? How many constraints are required?
- Constrain the H–N–H bond angle to the tetrahedral value (109.47°). What bond length is obtained with this constrained angle?
- Which of your two constrained structures is highest in energy?
- The umbrella mode of ammonia is nicely parameterized by an out of plane bend. If you have time, calculate several points on this potential.
Things To Watch Out For

A few points worth bearing in mind when optimising geometries:

- The Hessian the optimiser uses is only estimated. It is updated at each step, but if the signature of the Hessian is incorrect, a saddle point may be located rather than a minimum.

- To be sure you have a minimum, a frequency calculations should be performed at the optimised geometry. All the eigenvalues should be positive (no imaginary frequencies).

- When using DFT to optimise a structure, the quadrature grid will make the PES ragged. This may confuse the the optimiser, causing it to converge to an incorrect geometry, or to take a long time to converge. It is usually a good idea to use a higher–quality grid when optimising.

- Check your wavefunction has the correct symmetry. If the initial geometry is not a particularly good guess, an incorrect state my be the lowest in energy. Alternatively, a state crossing may have occurred along the optimisation path.
The idea of finding a transition structure is similar to optimising the geometry in that a critical point must be located on the PES.

However, transition states are much harder to find than minima:
- The quasi–Newton algorithms that are often used have a much smaller radius of convergence for saddle points.
- Most optimisation routines rely on the positive definiteness of the Hessian, which does not hold for a transition state. Maintaining the correct Hessian signature can be difficult.

It is recommended that you calculate and read in an initial starting Hessian (geom_opt_hessian = read). This can be calculated at a cheaper level of theory if desired.

The same comments that apply to geometry optimisations also apply to transition state searches. Grid effects, the initial starting geometry and checking your final structure by calculating the Hessian are all just as important.
Job Control

A transition search is controlled by the following rems:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>jobtype</td>
<td>ts → Transition state search</td>
</tr>
<tr>
<td>geom_opt_mode</td>
<td>n → Maximise along the $n^{th}$ mode</td>
</tr>
<tr>
<td>geom_opt_tol_gradient</td>
<td>$n \rightarrow n \times 10^{-6}$ (300 $\Rightarrow$ 0.0003)</td>
</tr>
<tr>
<td>geom_opt_tol_displacement</td>
<td>$n \rightarrow n \times 10^{-6}$ (1200 $\Rightarrow$ 0.0012)</td>
</tr>
<tr>
<td>geom_opt_tol_energy</td>
<td>$n \rightarrow n \times 10^{-8}$ (100 $\Rightarrow$ 0.000001)</td>
</tr>
<tr>
<td>geom_opt_max_cycles</td>
<td>n (50)</td>
</tr>
</tbody>
</table>

The convergence of the transition state searches is controlled by the same rems as for an optimisation.

If `geom_opt_mode` is not specified, the search will follow the lowest eigenvalue (either of the read–in Hessian or of the approximate Hessian that is built up).
Exercise: HNC Isomerisation

- HNC undergoes an isomerisation process where the hydrogen migrates between the carbon and nitrogen atoms.
- Optimise the linear structure HNC using EDF1/6–31G(d) theory. Use the bond lengths of your optimised structure to estimate those for an initial geometry for a transition state search.
- Perform a TS search at the same level of theory.
- What is the H–C–N bond angle at the TS? (Hint: it should be between 60–70°).
- Note that you should use the SG–1 quadrature grid rather than the default.
5. Frequencies and Reaction Pathways
Introduction

- During a frequency calculation, we must solve for the motion of the nuclei within the effective potential:

\[
E^\text{F}(\mathbf{x}) = E(0) + \sum_i \frac{\partial E}{\partial x_i} x_i + \frac{1}{2} \sum_{jk} x_i \frac{\partial^2 E}{\partial x_i \partial x_j} x_j + \ldots
\]

- At a stationary point, the gradient term vanishes by definition.

- To leading order, the shape of the potential, the curvature, is determined by the Hessian (second derivative matrix).

- Assuming the nuclei behave classically, they must satisfy Newton’s equations of motion:

\[
F_i = -\frac{\partial E}{\partial x_i} = -\sum_j \frac{\partial^2 E}{\partial x_i \partial x_j} x_j = m_i \ddot{x}_i
\]
Introduction

- Using the mass–weighted coordinates:

\[ q_i = \sqrt{m_i} x_i \quad \bar{H}_{ij} = \frac{H_{ij}}{\sqrt{m_i m_j}} \]

Newton’s equations simplify:

\[ \ddot{q}_i + \sum_j \bar{H}_{ij} q_j = 0 \]

- It follows from the harmonic approximation

\[ q_i(t) = q_i(0) \cos(\omega_i t) \]
\[ \ddot{q}_i = -\omega_i^2 q_i \]
\[ \omega_i^2 q_i = \sum_j \bar{H}_{ij} q_j \]

- This last equation is an eigenvalue problem.
Introduction

- It is clear that in order to calculate the frequencies we must
  - Ensure the system is at a critical point (either local minimum or transition structure, frequencies calculated at other points on the PES are meaningless, (although the Hessian may be useful).
  - Calculate the Hessian (second derivative matrix).
  - Mass–weight the Hessian (allowing for isotopic substitution).
  - Diagonalise the Hessian to obtain the eigenvalues and eigenvectors.

- The most expensive step is calculating the Hessian, which can be done either analytically (by solving the coupled–perturbed SCF equations), or by finite difference methods.

- Rescaling and diagonalising the Hessian takes little time, and therefore the effects of isotopic substitution can be accounted for relatively cheaply.

- If the system is not at a local minimum (i.e. is at a transition structure), the Hessian will not be positive definite, and imaginary frequencies will exist.
Frequency Calculations Within Q–CHEM

A geometry optimisation or transition structure search should be performed prior to a frequency calculation to ensure the system is at a stationary point.

The frequency calculation should then be performed:
- Using the same basis set.
- Using the same level of theory.
- Using the same quadrature grid (for DFT).

The key outputs from a frequency calculation are the eigenvalues (frequencies) and eigenvalues (normal modes).

Several thermochemical data are also reported at STP (ZPVE, rotational/vibrational entropy/enthalpy).

Imaginary frequencies are represented by negative frequencies in the output.

Currently all frequency calculations are performed within the harmonic approximation, no anharmonic corrections are available.
Job Control

- Frequency jobs are controlled with the following rems:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>jobtype</td>
<td>freq ⇒ Compute vibrational frequencies.</td>
</tr>
<tr>
<td>doraman</td>
<td>0 ⇒ Do not compute Raman intensities</td>
</tr>
<tr>
<td></td>
<td>1 ⇒ Do Compute Raman intensities (Requires third derivatives, not computed by default)</td>
</tr>
<tr>
<td>fdiff_step_size</td>
<td>$n \Rightarrow n \times 10^{-5}$ Default: $100 \Rightarrow 0.001\text{Å}$. May have to be increased for shallow potentials.</td>
</tr>
</tbody>
</table>

- The solving of the CPSCF equations can be broken down into several segments, which reduces the amount of memory that is required. The number of segments is determined automatically based on the available memory, but may be overridden using the cpscf_nseg rem.
The current limitations of analytic derivatives in Q–CHEM are:

<table>
<thead>
<tr>
<th>Theory</th>
<th>Analytic Gradients</th>
<th>Analytic Hessian</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>DFT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>MP2</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>(V)OD</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>(V)QCCD</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>CIS (not RO)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>CFMM</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>CCSD</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>

Where analytic derivatives are available, they exist for all angular momentum types up to $f$. Where analytic derivatives are not available, the derivatives are computed by finite differences.
Things To Watch Out For

- Frequencies are all calculated within the harmonic approximation. This is less valid in systems with:
  - X–H stretching modes
  - Soft potential wells (weakly bound systems)
In both these cases the atoms tend to sample more of the PES (i.e. parts that are not harmonic).

- Check the number of imaginary frequencies is what you expected (usually zero).
  - Check exactly the same level of theory has been used for both the optimisation and frequency calculation.
  - Check the true minimum doesn’t have lower symmetry.

- DFT harmonic frequencies are often a good approximation to actual anharmonic frequencies, but watch for grid effects:
  - Weight derivatives are more important for smaller grids.
  - The default grid in Q–CHEM should be used with caution when calculating frequencies.
Isotopic Substitution

- Isotopic substitution provides a useful way of isolating and identifying vibrational bands in a molecule.

- By default, Q–CHEM calculates the vibrational frequencies using the atomic masses of the most abundant isotopes.

- Non–default masses can be specified by using a $isotopes$ block and by setting the rem variable
  \[ \text{isotopes} = \text{true} \]

- The computation of isotopic effects requires only a re–scaling of the mass–weighted Hessian, and therefore several substitutions can be calculated very easily once the full Hessian has been computed.

- The $isotopes$ block can also be used to change the temperature and pressure at which the frequencies and thermochemical data are computed.
An Example

$molecule$

<table>
<thead>
<tr>
<th>Atom</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>0.690</td>
<td>-0.017</td>
<td>0.000</td>
</tr>
<tr>
<td>O</td>
<td>-0.736</td>
<td>0.072</td>
<td>0.000</td>
</tr>
<tr>
<td>H</td>
<td>-1.587</td>
<td>-0.349</td>
<td>0.000</td>
</tr>
<tr>
<td>H</td>
<td>1.174</td>
<td>0.948</td>
<td>0.000</td>
</tr>
<tr>
<td>H</td>
<td>1.080</td>
<td>-0.539</td>
<td>-0.860</td>
</tr>
<tr>
<td>H</td>
<td>1.080</td>
<td>-0.539</td>
<td>0.860</td>
</tr>
</tbody>
</table>

$end$

$isolopes$

<table>
<thead>
<tr>
<th>n_loops</th>
<th>TP_flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

$isolopes$

<table>
<thead>
<tr>
<th>n_atoms</th>
<th>[T P]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>atom1</th>
<th>mass1</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>17.999</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>n_atoms</th>
<th>[T P]</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>400</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>atom1</th>
<th>mass1</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>atom2</th>
<th>mass2</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>atom3</th>
<th>mass3</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>atom4</th>
<th>mass4</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>2.014</td>
</tr>
</tbody>
</table>

$end$
Exercise: Isotopic Labeling

\[
\begin{align*}
\text{H}_3\text{C} & \equiv \text{C} - \text{OH} + \text{HOCH}_3 \leftrightarrow \text{H}_3\text{C} & \equiv \text{C} - \text{OCH}_3 + \text{HOH} \\
\end{align*}
\]

- Ethanoic acid undergoes Fischer esterification with methanol to give methyl ethanoate and water.
- The question of where the oxygen in the \(-\text{OCH}_3\) group of methyl ethanoate comes from, can be resolved using isotopic labeling.
- Use the supplied coordinates for methyl ethanoate. These have been optimised at the B-LYP/3-21G level of theory using the SG-1 quadrature grid.
- Assume the methanol has been isotopically substituted with \(^{18}\text{O}\) (mass = 17.999). What are the two possible fates for the \(^{18}\text{O}\)?
- Calculate the vibrational frequencies of the ester with the appropriate oxygen atom labeled. By looking at the eigenvectors, suggest a frequency shift that could be used to determine if the \(^{18}\text{O}\) ends up on the ester. What do you predict the shift to be?
A reaction coordinate gives a hypothetical path along the potential energy surface connecting three critical points:

1. A minimum corresponding to the reactant(s).
2. A saddle point corresponding to the transition state.
3. Another minimum corresponding to the product(s).
Intrinsic Reaction Coordinates

- Critical points are independent of the coordinate system used, however, the path connecting them is not. Reaction pathways are therefore not uniquely defined.

- Mass–weighted cartesian coordinates results in the intrinsic reaction pathway orignally defined by Fukui. Q–CHEM also allows z–matrix coordinates to be used.

- One of the most common uses of reaction pathway calculations is to determine which minima are connected to the transition point. From this point of view the type of coordinates used is not crucial.

- A reaction coordinate following calculation must begin at a transition structure. The path is then found by following the gradient downhill:

\[ \mathbf{x}_{i+1} = \mathbf{x}_i - s \frac{\mathbf{g}_i}{|\mathbf{g}_i|} \]
The IRC Following Algorithm

- The differential equation associated with gradient following is stiff, which means small step sizes are required to prevent the algorithm from wandering.

- The IRC following code implemented in Q–CHEM is based on a predictor–corrector method that reduces this wandering:

\[ x^* = x_0 + \alpha \, v(x_0) \]
IRC Job Control

- Reaction coordinate calculations require several steps:
  1. Locate the transition state \( \text{jobtype} = \text{ts} \)
  2. Compute the Hessian at this point \( \text{jobtype} = \text{freq} \)
  3. Perform the IRC following \( \text{jobtype} = \text{rpath} \)

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>rpath_coords</td>
<td>( 0 \rightarrow ) Use mass–weighted cartesians</td>
</tr>
<tr>
<td></td>
<td>( 2 \rightarrow ) Use z–matrix coordinates</td>
</tr>
<tr>
<td>rpath_direction</td>
<td>( \pm 1 ) Determines the search direction</td>
</tr>
<tr>
<td>rpath_max_cycles</td>
<td>( n \rightarrow ) Max number of points along the path</td>
</tr>
<tr>
<td>rpath_max_stepsize</td>
<td>( n \rightarrow s_{\text{max}} = n \times 10^{-3} ) Default: ( n = 150 )</td>
</tr>
<tr>
<td>rpath_tol_displacement</td>
<td>( n \rightarrow \Delta_{\text{tol}} = n \times 10^{-6} ) Default: ( n = 5000 )</td>
</tr>
</tbody>
</table>
Exercise: Transition State of $[\text{N}_2\text{H}_6]^+$

- The $[\text{N}_2\text{H}_6]^+$ cation can form two stable species.
- Using the supplied coordinates, perform a frequency calculation using HF/6–31G to confirm that it is a transition state.
- Carry out a IRC calculation at the same level of theory to identify the two species connected to this structure. Note that you will need to decrease the convergence threshold to about $500 \times 10^{-6}$ and increase $rpath\_max\_cycles$ to about 50.
- The higher energy species is meta–stable, and exhibits an unusual type of bond. Can you identify it?
6. Correlated Methods
Introduction

- The correlation energy arises as a direct result of introducing the mean-field approximation.
- The correlation energy typically accounts for around 1% of the total energy of a system.
- Chemical properties tend to be sensitive to the correlation energy, and it is therefore important to model it as accurately as possible.
- The correlated methods in Q–CHEM fall into two main categories:
  - Density based methods (DFT)
  - Wavefunction based methods.
- The use of DFT methods within Q–CHEM is straightforward, and has already been discussed.
- In this section we will focus on wavefunction based correlated methods.
Wavefunction–Based Methods in Q–CHEM

Q–CHEM has the following wavefunction–based correlated methods implemented:

<table>
<thead>
<tr>
<th>Method</th>
<th>MP2</th>
<th>CCD</th>
<th>QCISD</th>
<th>VOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local MP2</td>
<td>CCD(2)</td>
<td>QCISD(T)</td>
<td>VOD(2)</td>
<td></td>
</tr>
<tr>
<td>MP3</td>
<td>CCSD</td>
<td>OD</td>
<td>QCCD</td>
<td></td>
</tr>
<tr>
<td>MP4SDQ</td>
<td>CCSD(T)</td>
<td>OD(T)</td>
<td>VQCCD</td>
<td></td>
</tr>
<tr>
<td>MP4</td>
<td>CCSD(2)</td>
<td>OD(2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

These methods fall into the following four categories:

1. Perturbative treatments of pair correlations: 80%
2. Self–consistent treatments of pair correlations: 95%
3. Non–iterative corrections for higher than double substitutions: 95%
4. Active space methods for highly correlated systems: varies
Møller–Plesset Perturbation Methods

- Møller–Plesset perturbation is based on the assumption that the Hartree–Fock Hamiltonian is a close approximation to the true Hamiltonian:

\[ H = H_0 + \lambda V \]

- The energy and wavefunction are then assumed to be able to be written as a power series in the perturbation parameter \( \lambda \):

\[ \Psi = \Psi_0 + \lambda \Psi^{(1)} + \lambda^2 \Psi^{(2)} + \lambda^3 \Psi^{(3)} + \ldots \]

\[ E = E^{(0)} + \lambda E^{(1)}_{\text{HF}} + \lambda^2 E^{(2)} + \lambda^3 E^{(3)}_{\text{corr}} + \ldots \]

- Terms involving \( \lambda^k \) can be collected together to give a set of equations for \( \Psi^{(n)} \) and \( E^{(n)} \).

- To leading order, the correlation energy is given by \( E^{(2)} \), the MP2 energy.
Perturbation Theory – Comments

- Like any series expansion, the convergence of the MP\(n\) series is not guaranteed, and in some cases the series may even diverge.
- MP\(n\) theory works best when the perturbation is small, i.e. when HF theory is a good approximation to the exact system.
- MP2 theory fails notably for open-shell systems, and unrestricted systems with a large degree of spin contamination.
- The cost of MP2 theory scales as the fifth power of the system size (due to the AO→MO transformation that is required) which makes it applicable to fairly large systems on modern workstations.
- MP\(n\) theory is size consistent, and therefore scales correctly when the system size is increased.
- MP2 provides a systematic improvement to HF energies and properties.
- MP\(n\) theory accounts for long-range dispersion effects, and is preferred to DFT for dispersion forces and hydrogen–bonded species.
The implementation of MP2 theory within Q–CHEM is based on an efficient semi-direct algorithm.

The algorithm is most efficient when \( O V N \) memory, and \( O V N^2/2 \) scratch space is available.

The disk requirements can be reduced to by batching over occupied orbitals. The two-electron integrals must be evaluated more often, and therefore this is slower.

Analytic MP2 gradients are available, and have been implemented using a highly efficient algorithm. No additional disk or memory is required for gradients using this algorithm, making MP2 geometries available for fairly large systems.

The frozen-core approximation can be used, which is important for basis sets with little core flexibility.

Q–CHEM will automatically select the most efficient algorithm based on the available memory and disk.
## MP2 Job Control

The following rem variables are important for controlling MP2 jobs:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation</td>
<td>Should be set to mp2.</td>
</tr>
<tr>
<td>mem_static</td>
<td>Should be 32–160 Mb to allow efficient integral evaluation and storage of the $OVN$ array (32Mb).</td>
</tr>
<tr>
<td>ao2mo_disk</td>
<td>Affects how often the two–electron integrals in the AO basis have to be evaluated. Should be set to around 80% of the available scratch space (default: 2000 Mb).</td>
</tr>
<tr>
<td>n_frozen_core</td>
<td>Significant savings can be achieved by using the frozen–core (FC) approximation.</td>
</tr>
<tr>
<td>n_frozen_virtual</td>
<td>Sets the number of virtual orbitals to freeze.</td>
</tr>
<tr>
<td>cd_algorithm</td>
<td>Determined by the program, either semi_direct or local_occupied.</td>
</tr>
</tbody>
</table>
Exercise: Ne Dimer

- The Ne dimer is weakly bound by dispersion forces which arise from the induced–dipole/induced–dipole interactions of the atoms. This is a correlation effect which is not modelled by HF theory.
- Calculate the equilibrium bond length of Ne$_2$ at the MP2/aug-cc-pVDZ level of theory. Note: you should decrease the gradient tolerance to around $20 \times 10^{-6}$.
- Compare your answer with the experimental value of 3.09 Å.
- What happens if you try to optimise Ne$_2$ using HF/aug-cc-pVDZ?
Local MP2 Methods

- The cost of computing the correlation energy scales unphysically with the system size \( O(N^5) \ldots \).

- Exact “fast methods” usually rely on spatial locality, which does not take effect until the system has more than several hundred atoms (in 3–D).

- Therefore approximate local methods are required. These should:
  - be size–consistent
  - be uniquely defined
  - yield continuous potential energy surfaces

- The triatomics in molecules (TRIM) model is an approximate treatment of double substitutions. Rather than four atoms being involved, one excitation involves a occupied and virtual orbital on the same atom.

- The more drastic diatomics in molecules (DIM) model forces both excitations to occur on the local atom, thus only two atoms are involved.
Local MP2 Methods

- Both these models require a local atom–centred basis, the basis used in Q–CHEM is based on extracted polarised atomic orbitals (EPAOs):
  - The occupied space is spanned by a minimal set of polarised atomic orbitals projected into the total occupied space.
  - The virtual space is spanned by the full atomic orbital basis projected into the virtual space.
- The TRIM model recovers around 99.7% of the MP2 correlation energy for covalent bonding.
- The DIM model recovers around 95% of the MP2 correlation energy.
- For the water dimer with the aug-cc-pVTZ basis, 96% of the MP2 contribution to the binding energy is recovered with the TRIM model.
- To request a Local MP2 calculation, it is sufficient to set correlation = local_mp2.
Exercise: Benzene Dimer

- Benzene can form a flexible T-shaped dimer where one of the hydrogen atoms points into the ring of the other.
- Using the supplied coordinates, compute the total energy of the dimer at the MP2/3-21G and Local_MP2/3-21G levels of theory.
- What fraction of the correlation energy does the TRIM model recover in this system?
Coupled Cluster Based Methods

Coupled cluster methods are the flagship methods of quantum chemistry, allowing the user to approach chemical accuracy ($< 1$ kcal/mol).

They are strictly size–consistent, but due to the projection used in the determining equations, they are not variational.

To the extent that most of chemistry is determined by interactions of pairs of electrons, the inclusion of only single and double substitutions is expected to give good accuracy.

CCSD is only marginally more accurate than MP2 theory for structures and frequencies, but far superior and more reliable for reactive species and open–shell systems.

The CPU cost of CCSD scales as $O(N^6)$, and the storage space scales as $O(N^4)$ making it very expensive.

Where possible, a small trial calculation should be performed to estimate the cost of a calculation.
The CCSD ansatz is given by:

\[
|\Psi_{CCSD}\rangle = \exp(\hat{T}_1 + \hat{T}_2)|\Phi_0\rangle
\]

\[
E_{CCSD} = \langle \Phi_0 | \hat{H} | \Psi_{CCSD} \rangle = \langle \Phi_0 | \hat{H} | (1 + \hat{T}_1 + \frac{1}{2} \hat{T}_1^2 + \hat{T}_2) \Phi_0 \rangle
\]

\[
\hat{T}_1|\Phi_0\rangle = \sum_{i} \sum_{a} t_i^a |\Phi_i^a\rangle
\]

\[
\hat{T}_2|\Phi_0\rangle = \frac{1}{4} \sum_{ij} \sum_{ab} t_{ij}^{ab} |\Phi_{ij}^{ab}\rangle
\]

where \(|\Phi_0\rangle\) is the reference (HF) wavefunction, and \(t_i^a\) and \(t_{ij}^{ab}\) are the singles and double amplitudes, which must be determined along with the energy.
Optimised Orbital Coupled Cluster Doubles (OD)

- The CCSD equations can be greatly simplified if only double excitations are considered giving rise to CCD theory:

\[ E_{CCD} = \langle \Phi_0 | \hat{H} | (1 + \hat{T}_2) \Phi_0 \rangle \]

- The HF wavefunction (\(\Phi_0\)) is a poor starting point for such a treatment as the orbitals have been optimised in the absence of electron correlation.

- The optimised orbital CCD (OO–CCD or just OD) method is based on CCD, with the added constraint that the cluster energy is minimised with respect to variations in the orbitals:

\[ \frac{\partial E_{CCD}}{\partial \theta^a_i} = 0 \]

- The orbitals thus obtained are similar to Brueckner orbitals, and are the best possible in a variational sense for the CCD expansion.
Pros And Cons Of The OD Method

- The OD method has the advantage of formal simplicity.
- OD does not require the HF orbitals, and therefore can work well in cases where the HF theory performs poorly (artifactual symmetry breaking and non-convergence of the SCF).
- The cost of an OD calculation scales the same as CCSD \(O(N^6)\), but is usually a factor of 2 more expensive.
- The implementation within Q–CHEM supports both energies and analytic gradient calculations.
The following rem variables are important for controlling coupled cluster jobs:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation</td>
<td>Set to cc, od, cc(t), od(2)</td>
</tr>
<tr>
<td>cc_convergence</td>
<td>Main convergence criterion used to determine several others in the coupled cluster code. Set to n for n significant digits in the final energy (default: 8).</td>
</tr>
<tr>
<td>cc_maxiter</td>
<td>Maximum number of iterations to optimise the coupled cluster energy (default: 200).</td>
</tr>
<tr>
<td>cc_print</td>
<td>Controls the amount of printed output from the coupled cluster package (default: 1).</td>
</tr>
</tbody>
</table>
Exercise: The ·CN Radical (Again)

- Recall the ·CN radical from yesterday. It was highly spin–contaminated, making the HF wavefunction a bad reference for post–HF methods.
- Calculate the CCSD and OD correlation energies for ·CN using the cc–pVDZ basis. You should use a bond length of 1.239 Å.
- Which method would you expect to perform the best?
- Compare your answers with the multi–reference coupled cluster value of -92.4715 E_h.
- Use the timing of your job to estimate the time it would take to perform the job with the cc–pVTZ basis.
7. Excited States
Overview of Excited States

- Excited state wavefunctions exhibit a far wider range of behaviours and are therefore more difficult to model.

- Many excited states can lie close together, giving rise to multi–determinant character and resulting in a breakdown of the single–reference model.

- Larger basis sets are required to model both the valence and Rydberg states and thus avoid artifactual mixing.

- Diffuse functions should be included, especially when modelling Rydberg states.

- Because of this, a black–box approach is rarely recommended and each system should be treated on its own merits.
Excited State Options Within Q–CHEM

- Q–CHEM has a variety of methodologies applicable to excited state calculations including:
  - Inexpensive uncorrelated single–excitation methods (CIS, XCIS).
  - Time–dependent density functional theory (TDDFT).
  - More expensive correlated wavefunction based treatments (EOM–CCSD, EOM–OD).

- Cheaper methods can often yield a qualitative picture of the system, but do not usually provide quantitatively correct results.

- The more expensive wavefunction based methods have the potential to yield results for excited states that are of chemical accuracy (< 0.1 eV error), but this comes at a high price.

- Q–CHEM does not have support for multi–determinant wavefunction based methods such as CASSCF, CASPT2 and MRCI.
Types of Excited State Calculations Available

- **Vertical absorption spectra**
  Calculations of electronic excitation energies provide insight into absorption spectra, and are available for CIS, RPA, XCIS CIS(D), EOM–CISD and EOM–OD.

- **Excited state optimisations**
  These provide understanding of the relaxation that occurs in the excited states. Analytic first derivatives are available for UCIS and RCIS, allowing for the efficient computation of excited state geometries. Geometries for CIS(D), CCSD and OD are obtainable via finite difference methods.

- **Excited state vibrational analyses**
  Given a structure corresponding to a minimum on the excited state potential energy surface, Q–CHEM can determine the force constants and hence the frequencies. Excited state vibrational analyses use efficient analytic second derivatives, and are only available for UCIS and RCIS.
A crude (uncorrelated) model for excited states can be obtained by taking the HF determinant, and “exciting” an electron from an occupied to a virtual orbital:

\[ \Psi_i^a = \frac{1}{\sqrt{n!}} |\chi_1 \chi_2 \cdots \chi_a \chi_j \cdots \chi_N \rangle \]

This approach neglects the effects of orbital relaxation that occur on excitation, and is not practically useful.

A better approximation is the CIS wavefunction, which is obtained by considering all possible determinants of this type:

\[ \Psi_{CIS} = \sum_{ia} a_i^a \Psi_i^a \]

The configurations, \( \Psi_i^a \), are interacted via the many–electron Hamiltonian in the space of all single excitations:

\[ A_{ia,jb} = \langle \Psi_i^a | \hat{H} | \Psi_j^b \rangle \]
Diagonalising this matrix yields the CIS excited state energies (eigenvalues) and electron promotion amplitudes (eigenvectors).

Computing all single excitations is expensive, and results in large $A$ matrices. Instead, Q–CHEM considers excitations within a window, which is enlarged iteratively until convergence is achieved.

Brillouin’s theorem states that these singly–excited determinants do not mix with the ground state, and therefore the CIS method for excited states is comparable to the Hartree–Fock method for the ground state.

CIS vertical excitation energies for valence states are usually within 1 eV of experiment.

CIS derives some of its accuracy from cancellation of errors. Improving the basis set by adding polarisation functions can increase the excitation energies as they improve the ground state PES more than the excited state one.
CIS State Properties

- Analytic first and second derivatives are available for CIS allowing optimised geometries, forces, transition structures and frequencies to be computed.
- Calculating geometries for excited states is more difficult than for the ground state as state crossing can occur along the optimisation path.
- CIS properties are usually more useful than CIS energies. CIS frequencies are systematically 10\% higher relative to experiment.
- The semi–direct method is used to evaluate CIS frequencies, and the disk and memory requirements are similar to HF.
- The computer time taken to compute CIS frequencies scales as roughly \( O(N^5) \).
Extended CIS (XCIS)

- ROCIS and UCIS are less effective for radicals than CIS is for closed–shell systems.
- The deficiency can be traced to a particular type of double excitation:

```
        -----------  -----------
        |        |    1
        -----------  -----------
        |    1  |    |
        -----------  -----------
        |    1  |    |
```

- The XCIS method explicitly includes double excitations of this type and is better suited for radicals.
- A ROHF wavefunction is required for XCIS to avoid ambiguity in the correspondence between $\alpha$ and $\beta$ spin orbitals.
The rems used to control a CIS job are:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>cis_n_roots</td>
<td>$n \Rightarrow$ Search for $n$ roots. Default: 0</td>
</tr>
<tr>
<td>max_cis_cycles</td>
<td>$n$ Default: 30</td>
</tr>
<tr>
<td>cis_convergence</td>
<td>$n \Rightarrow 10^{-n}$ Default: $n = 6 \Rightarrow 10^{-6}$</td>
</tr>
<tr>
<td>cis_state_deriv</td>
<td>$n \Rightarrow$ Determine state properties for $n^{th}$ state</td>
</tr>
<tr>
<td>xcis</td>
<td>true/false Default: false</td>
</tr>
<tr>
<td>unrestricted</td>
<td>true/false Default: varies</td>
</tr>
</tbody>
</table>
Application: Excited States of Formamide

- Formamide is an important model for the peptide chromophore found along the backbone in proteins.
- It has a $\pi$ system made up of the three $p_z$ orbitals on the N, C, and O atoms.
- These combine to form the doubly occupied $\pi_b$ and $\pi_{nb}$ orbitals, and the virtual $\pi^*$ orbital.
- The UV spectrum is dominated by the intense $\pi_{nb} \rightarrow \pi^*$ transition occurring around 7.4 eV.
Exercise: Formamide Excited States

- Construct an input deck for formamide. Assume the structure has $C_s$ symmetry and use the following geometric parameters:

\[
\begin{align*}
R(\text{C–O}) & \quad 1.47 \text{ Å} & \angle (\text{N–C–O}) & \quad 109^\circ \\
R(\text{N–H}) & \quad 1.00 \text{ Å} & \angle (\text{H–N–C}) & \quad 123^\circ \\
R(\text{C–H}) & \quad 1.06 \text{ Å} & \angle (\text{H–C–N}) & \quad 127^\circ \\
R(\text{C–N}) & \quad 1.37 \text{ Å}
\end{align*}
\]

- Perform a CIS calculation using the 6–311(2+)G basis to determine which is the $\pi \rightarrow \pi^*$ state, (use \texttt{cis\_n\_roots = 8}).

- Optimise the geometry of your chosen state using CIS, and calculate the CIS frequencies to show that the $C_s$ structure is not a minimum.

- Consider the normal modes of the imaginary frequencies and determine how the system wants to distort.

- Why might the excited state want to distort in this manner?
Time–Dependent DFT

- Time–dependent DFT calculates the excitation energies by computing the poles in the polarisation as a function of the frequency of an applied external field.
- Yields more accurate results than CIS, particularly for valence states with significant doubly excited character, for which CIS fails dramatically.
- Does not capture non–dynamical correlation, and so fails with modelling systems with stretched bonds and significant multi–determinant behaviour (which is often exhibited in excited states).
- Valence excitations are typically 0.3–0.5 eV too low
- Rydberg states are not well handled due to the incorrect asymptotic behaviour of the effective potential for most functionals (Q–CHEM does not have support for asymptotically corrected functionals)
- Excitations with a strong charge–transfer character are not well handled. Why?
Spin–Flip TDDFT

- Single–reference methods, such as DFT, are not well suited to problems that are sensitive to the effects of non–dynamical correlation. These include:
  - Stretched bonds
  - Diradicals
  - Systems with degeneracies

- The difficulties arise due to there being two (or more) configurations that are not equally well described. For example, in the case of a stretched bond, both of the following configurations will compete for the singlet state:

\[ \sigma(\uparrow, \downarrow) \quad \text{and} \quad \sigma(\uparrow)\sigma^*(\downarrow) \]

- Spin–flip methods attempt to remedy this deficiency by choosing a reference that provides a more balanced description of the competing states.

- They are most effective when the non–dynamical correlation originates from a single HOMO–LUMO orbital pair.
Spin–Flip TDDFT

- For systems with an even number of electrons, the reference system is a triplet:

  ![Diagram of triplet state]

- The target open–shell singlet is obtained via a $\alpha \rightarrow \beta$ spin–flip.

- The method can also model the closed–shell singlet ground and doubly excited states.

- For systems with an odd number of electrons the reference system is a quartet, and a spin–flip is used to model the target doublet.
Requesting A Spin–Flip Calculation

- Spin–flip TDDFT calculations are controlled by the following rem variables:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>exchange</td>
<td>Any DFT exchange functional</td>
</tr>
<tr>
<td>correlation</td>
<td>Any DFT correlation functional</td>
</tr>
<tr>
<td>cc_spin_flip</td>
<td>Set to true for SF–TDDFT</td>
</tr>
<tr>
<td>cis_n_roots</td>
<td>$n \Rightarrow$ Calculate $n$ excited states</td>
</tr>
</tbody>
</table>

- Remember to set the multiplicity to either 3 for systems with an even number of electrons, or 4 for systems with an odd number.

- The target ground state will usually be printed as the first excited state, and will have a negative excitation energy with respect to the reference state.

- If only the ground state is required, `cis_n_roots` does not need to be set to more than 1.
Application: Ethylene Torsion

- **D\textsubscript{2h}**

- **D\textsubscript{2d}**

- In the $D_{2h}$ geometry the $\pi$ and $\pi^*$ orbitals are well separated and the HF determinant, $\pi(\uparrow, \downarrow)$, is qualitatively correct.

- When rotated into the $D_{2d}$ structure the $\pi$ and $\pi^*$ orbitals become degenerate and the single–determinant model breaks down for the singlet state.

- The spin–flip model gives a qualitatively correct description for all torsional angles.
Exercise: Ethylene Torsion

- Construct input decks for ethylene in the planar and twisted geometries using the following geometric parameters:

  \[
  \begin{align*}
  R(\text{C–C}) & \quad 1.34 \, \text{Å} \\
  R(\text{H–C}) & \quad 1.09 \, \text{Å} \\
  \angle (\text{H–C–C}) & \quad 122.1^\circ
  \end{align*}
  \]

- Attempt to calculate the rotational barrier of the singlet state by performing unrestricted B3LYP/DZ* single–point calculations at the two geometries. The following rem is recommended:

  \[
  \begin{align*}
  \text{xc_grid} & \quad 1 \\
  \text{xc_smart_grid} & \quad \text{false}
  \end{align*}
  \]

- Use the SF–TDDFT method to improve on the above approach.
CIS(D)

- CIS(D) is a perturbative approximation to the CCSD model, and is therefore a size consistent method.
- Second–order perturbation theory is applied to CIS excited states, and therefore CIS(D) may be considered as an excited state analogue to MP2 theory.
- CIS(D) scales as the fifth power of the basis set size, making it applicable to fairly large molecules.
- The primary strength of CIS(D) is for calculating vertical excitation energies in cases where CASPT2 and CCSD methods are not feasible.
- RMP2 performs poorly in cases where the UHF determinant is spin–contaminated. Generally UCIS wavefunctions suffer from greater spin–contamination, and therefore RCIS(D) is to be preferred over UCIS(D).

CIS(D) is a perturbative approximation to the CCSD model, and is therefore a size consistent method.

Second–order perturbation theory is applied to CIS excited states, and therefore CIS(D) may be considered as an excited state analogue to MP2 theory.

CIS(D) scales as the fifth power of the basis set size, making it applicable to fairly large molecules.

The primary strength of CIS(D) is for calculating vertical excitation energies in cases where CASPT2 and CCSD methods are not feasible.

RMP2 performs poorly in cases where the UHF determinant is spin–contaminated. Generally UCIS wavefunctions suffer from greater spin–contamination, and therefore RCIS(D) is to be preferred over UCIS(D).
CIS(D) Job Control

- The following rems are relevant to CIS(D) calculations:
  
  ```
  correlation  cis(d)
  cc_nhighspin  [i,j,k...]
  cc_nlowspin  [i,j,k...]
  cc_state_deriv  n
  ```

- The CIS(D) output included the $\theta$ diagnostic, which is an indication of how the effect of including correlation causes the singly excited configurations to mix.

- A value of $\theta = 45^\circ$ indicates complete breakdown in the CIS reference.
Exercise: Excited states of $\cdot$CH$_3$

- Construct an input deck for the $\cdot$CH$_3$ radical ($D_{3h}$ symmetry, $R$(C–H) = 1.096Å).

- Run a UCIS(D) calculation to determine the vertical excitation energies of the following Rydberg excited states at the UCIS and UCIS(D) levels of theory:

<table>
<thead>
<tr>
<th>State</th>
<th>Transition</th>
<th>Expt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^2A'_1$</td>
<td>$S_{2p} \rightarrow V_{3s}$</td>
<td>5.729 eV</td>
</tr>
<tr>
<td>$^2A''_2$</td>
<td>$S_{2p} \rightarrow V_{3p}$</td>
<td>7.436 eV</td>
</tr>
</tbody>
</table>

- Do the $\theta$ mixing angles indicate that the CIS(D) approach is valid for these states?

- Could we have predicted this would be the case?
8. Other Topics

- Linear Scaling Methods
- Solvation Effects
- NMR Chemical Shifts
- Plotting Data
- Parallel Capabilities

Exercises 1
Exercises 2
In general, quantum chemical models are expensive, and this has limited their application to relatively small systems.

The cheapest methods, HF and DFT, formally scale as approximately $O(N^4)$, this is reduced to $O(N^2)$ using integral cutoffs.

For very large systems, linear scaling methods are required.
The continuous fast multipole method is an extension of Greengard’s fast multipole method used to evaluate the interaction of point charges in a time that scale linearly with system size.

Interactions between parts of the molecule are classified as either near-field, or well-separated.

The near-field interactions are evaluated exactly.

Local charge distributions are collectively expanded using a multipole expansion.

Well-separated interactions are then evaluated by interacting using these multipole expansions.

The success of the method relies on the hierarchical boxing scheme that is used to determine when charge distributions are sufficiently well-separated.

The accuracy of the method is determined by the order to which the multipole expansion is carried out.
CFMM Options

- Q–CHEM automatically determines when the system size is large enough to warrant using the CFMM.

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfmm_order</td>
<td>( n ) ⇒ Use multipoles up to ( n^{th} ) order (default is 15 for energies, 25 for gradients)</td>
</tr>
<tr>
<td>grain</td>
<td>(-1) ⇒ Program decides the best value (1) ⇒ Turns CFMM off ( n \geq 8 ) ⇒ Use ( n ) lowest–level boxes.</td>
</tr>
<tr>
<td>lin_k</td>
<td>true/false Program automatically sets this if the CFMM is used.</td>
</tr>
</tbody>
</table>

- The grain rem should not usually need adjusting, unless the CFMM is not desired.
- The CFMM method is implemented for HF and DFT calculations.
The coulomb attenuated Schrödinger equation (CASE) method replaces the long–range coulomb operator with a short–range operator. This affects the total energies, but due to the flatness of the neglected part of the potential, energy differences are less affected. The rapid decay of the attenuated operator allows for more aggressive cutoff strategies, leading to linear scaling.
CASE Options

- A CASE calculations is controlled via the following rems:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>omega</td>
<td>( n \Rightarrow \omega = n/1000 )</td>
</tr>
<tr>
<td>integral_2e_opr</td>
<td>-2 \Rightarrow \text{Apply the coulomb operator}</td>
</tr>
<tr>
<td></td>
<td>-1 \Rightarrow \text{Apply the CASE operator}</td>
</tr>
</tbody>
</table>

- A value of \( \omega = 0.2 \) has been found to be effective.
Exercise: Alkanes

- Long alkanes chains are often used to investigate the effects of linear scaling algorithms.

- They give an indication of how the time scales with long molecules, but not globular ones.

- Use the given z–matrix inputs for $\text{C}_{10}\text{H}_{22}$ and $\text{C}_{20}\text{H}_{42}$. Determine the time required to calculate the BLYP/6–311G single point energy with and without the CFMM. Note that CFMM is turned on by default.
Chemistry in solution is often vastly different from the chemistry that occurs in the gas phase.

Explicitly incorporating the effects of solvent molecules in a calculation is problematic for at least two reasons:

- The size of system becomes prohibitive, with more time being spent accounting for the solvent–solvent interactions than the (presumably more interesting) solute–solvent ones.
- The (large) configurational space of the solvent molecules must be considered, requiring many calculations to be performed.

Several models have been suggested that account for the effects of the solvent in a more approximate, but significantly less expensive, way. Q–C\textsc{hem} includes two such solvation models:

- Onsager dipole continuum solvent
- Langevin dipoles solvation model
Onsager Dipole Continuum Solvent

- The Onsager model places the molecule in a spherical cavity. A continuum medium, representing the solvent, is placed outside the cavity.

- Two parameters specify the model:
  - The cavity radius, \( a_0 \)
  - The dielectric constant of the solvent, \( \varepsilon \)

- The cavity radius is typically calculated using

\[
a_0 = \sqrt[3]{\frac{3V_m}{4\pi N_A}}
\]

where \( V_m \) is the molar volume (molecular weight/density).

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>solute_radius</td>
<td>( n \Rightarrow a_0 = n/10000 )</td>
</tr>
<tr>
<td>solvent_dielectric</td>
<td>( n \Rightarrow \varepsilon = n/10000 )</td>
</tr>
</tbody>
</table>
The Langevin Dipoles model provides a more accurate solvation model than the Onsager approach for aqueous solutions.

The solute is surrounded by point dipoles lying on a cubic lattice. The strength and orientation of the dipoles is determined self-consistently by the solute and the other Langevin dipoles.

The free energy of hydration $\Delta G_{hydr}$ is averaged over several different grids of dipoles.

Langevin dipole calculations are available for HF and DFT calculations only.

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>chemsol</td>
<td>true $\Rightarrow$ calculate Langevin dipoles</td>
</tr>
<tr>
<td>chemsol_efield</td>
<td>$1 \Rightarrow$ use Exact solute charge distribution</td>
</tr>
<tr>
<td></td>
<td>$0 \Rightarrow$ use Mulliken charge distribution</td>
</tr>
<tr>
<td>chemsol_nn</td>
<td>$n \Rightarrow$ Average $\Delta G_{hydr}$ over $n(5)$ grids</td>
</tr>
</tbody>
</table>
The Langevin Dipoles model provides a more accurate solvation model than the Onsager approach for aqueous solutions. The solute is surrounded by point dipoles lying on a cubic lattice. The strength and orientation of the dipoles is determined self-consistently by the solute and the other Langevin dipoles. The free energy of hydration $\Delta G_{\text{hydr}}$ is averaged over several different grids of dipoles. Langevin dipole calculations are available for HF and DFT calculations only.

**Rem Variable** | **Options**
---|---
chemsol | true ⇒ calculate Langevin dipoles
chemsol_efield | 1 ⇒ use Exact solute charge distribution
 | 0 ⇒ use Mulliken charge distribution
chemsol_nn | $n ⇒ Average \Delta G_{\text{hydr}}$ over $n(5)$ grids
Exercise: Formaldehyde Vibrational Shift

- Using the Langevin dipole model in Q–CHEM predict the change in the C–O vibrational mode of formaldehyde in water using HF/6–311G(d).

- Note that for the solvent calculation you should relax the gradient threshold to $500 \times 10^{-6}$.

- Analytic Hessians are not available when Langevin dipoles are used. You will need to set $i_{deriv} = 1$ to force the second derivatives to be computed by finite differences.
The shielding tensor is a second–order property, which depends on the applied magnetic field and the nuclear magnetic spin momentum.

The implementation within Q–CHEM based on gauge–including atomic orbitals, and requires the perturbed densities $\partial P_{\mu\nu}/\partial B_i$. These are obtained using a two–stage iterative linear scaling CPSCF approach based on the density.

NMR shift calculations are restricted to HF wavefunctions, and for angular momentum not exceeding $d$ functions.

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>jobtype</td>
<td>nmr ⇒ compute chemical shifts</td>
</tr>
<tr>
<td>d_scf_conv_1</td>
<td>$n \Rightarrow 10^{-n}$ convergence criterion for level 1</td>
</tr>
<tr>
<td></td>
<td>(does not include $2e^-$ integrals)</td>
</tr>
<tr>
<td>d_scf_conv_2</td>
<td>$n \Rightarrow 10^{-n}$ convergence criterion for level 2</td>
</tr>
<tr>
<td>d_scf_max_1(2)</td>
<td>$n \Rightarrow$ maximum SCF cycles for level 1 (2)</td>
</tr>
</tbody>
</table>
Example: Butane/Butadiene Chemical Shifts

- The hydrogen nuclei in butane and butadiene experience different magnetic environments resulting in different chemical shifts.

- Using the supplied coordinates, calculate the NMR chemical shifts for butane and butadiene using HF/6–31G(d).

- Note the values you calculate are raw shieldings. You will need to do a calculation on TMS also to determine the relative shifts.

- Using these shifts, predict the proton NMR spectra for these molecules.
Q–CHEM is able to compute various quantities on a grid for plotting purposes. The following properties are available:

- Molecular orbitals
- Electron Densities
- Electrostatic potential
- Transition densities (for CIS calculations)

Plotting data is requested using the `ianlty` rem, with the grid data being specified in the `$plots` section (pg: 242).

Data is written to files `plot.xxx` (where `xxx` depends on the type of data being generated).

Parameters in the `$plots` section are in Å, but output is in a.u.

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>ianlty</code></td>
<td>200 ⇒ read <code>$plots</code> section</td>
</tr>
<tr>
<td><code>igdesp</code></td>
<td>$n$ ⇒ read $n(0)$ points from the file ‘ESPGrid’</td>
</tr>
<tr>
<td></td>
<td>$-1$ ⇒ use grid defined in <code>$plots</code> section</td>
</tr>
</tbody>
</table>
Parallel Capabilities of Q–CHEM

- HF and DFT energy calculations.
- HF and DFT geometry optimisations.
- HF and DFT frequency calculations.
Exercises 1

- HCN Activation Energy
- The Diels–Alder Reaction
- Dipole Bound States
Exercise: HCN Activation Energy

- Recall the HCN transition structure you found earlier.
- Determine the activation barrier for this reaction, taking into account the ZPVE.
- How does the activation energy change if you deuterate the system?
Exercise: The Diels–Alder Reaction

The simplest example of a Diels–Alder reaction is the addition of ethylene to 1,3-butadiene to give cyclohexene.

Use the transition structure provided to create a reaction profile at the B3LYP/3–21G level of theory.

The IRC following may not take you all the way to the product. Use the final IRC structure as an initial starting point for an optimisation to obtain cyclohexene.

Calculate the energy released as the TS is converted to cyclohexene.
If a molecule has a dipole of sufficient strength, it may be possible for it to capture an electron in a dipole-bound state.

The extra electron is very weakly bound, occupying an orbit that can stretch past 50 Å from the centre of the molecule.

Dipole–bound anions have been conjectured to exist in space, and contribute to the diffuse interstellar bands in absorption spectra.

The minimum strength for a point dipole to support a DBS is 1.625 debye. This value is an underestimate for real systems as it ignores the effects of molecular rotation and vibration.

$\text{H}_2\text{CCN}^-$, which is known to exist in space, has a dipole of around 3.5 debye and therefore should support a DBS.
Exercise: $\text{H}_2\text{CCN}^-$

- Create and input deck for $\text{H}_2\text{CCN}$ and optimise your guess structure at the HF/6–311G(d,p) level of theory. Note the molecule has $C_{2v}$ symmetry.

- Add a ghost atom (symbol Gh) about 3Å from the C atom at the positive end of the molecule. This center will support the very diffuse functions required to model the DBS.

- Use a diffuse $sp$ basis with exponents:
  
  $\begin{align*}
  5 \times 10^{-3} & \quad 5 \times 10^{-4} & \quad 5 \times 10^{-5} & \quad 5 \times 10^{-6} \\
  1 \times 10^{-3} & \quad 1 \times 10^{-4} & \quad 1 \times 10^{-5}
  \end{align*}$

- Attempt to determine the binding energy of the electron in the singlet and triplet DBS. (Hint: use a $\text{occupied}$ block and $\text{mom_start}$ if required.)

- Compare your values with the experimental bound of $\leq 7$meV ($1 \text{E}_h = 27.211 \text{ eV}$).

- What 3 bits of information from your output suggest you have a DBS?
Exercises 2
Exercise: Binding Energy Of Benzene Dimer

- Recall the benzene dimer you looked at earlier. Calculate the binding energy of this system using MP2 and Local MP2 theory.
- What fraction of the binding energy is recovered using the Local MP2 approximation?
- What effect does the BSSE have on these binding energies?
- If you have enough time, you may like to consider a larger basis set than 3–21G.
You will need to be familiar with the following rem variables:

<table>
<thead>
<tr>
<th>Rem Variable</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>correlation</td>
<td>Set to ccsd, od, vod,…</td>
</tr>
<tr>
<td>cc_nhighspin</td>
<td>([i, j, \ldots]) \Rightarrow Number of high–spin states to compute in each symmetry.</td>
</tr>
<tr>
<td>cc_nlowspin</td>
<td>([i, j, \ldots]) \Rightarrow Number of low–spin states to compute in each symmetry.</td>
</tr>
<tr>
<td>cc_state_deriv</td>
<td>(n) \Rightarrow Determines the state to be considered for properties or geometries.</td>
</tr>
<tr>
<td>cc_refsym</td>
<td>(n) \Rightarrow Determines the symmetry of the state to be considered for properties or geometries.</td>
</tr>
<tr>
<td>cc_symmetry</td>
<td>true/false \Rightarrow Default is false, but in most cases you will probably want it on.</td>
</tr>
</tbody>
</table>
Exercise: Excited States of CH$_2$O

- Construct an input deck for the formaldehyde molecule. Using the following geometric parameters:
  \[
  \begin{align*}
  R(\text{C–O}) &= 1.220 \text{ Å} \\
  R(\text{C–H}) &= 1.104 \text{ Å} \\
  (\text{O–C–H}) &= 122.19^\circ
  \end{align*}
  \]

- Using the aug-cc-pVDZ basis, compute the EOM–CCSD and EOM–OD excitation energies for the $^1A_1$, $^1A_2$ and $^1B_1$ excited states. (You will need to use \textit{unrestricted = false}).

- Compare the results with the experimental values of 8.14, 4.07 and 7.11 respectively.

- Is CIS able to predict reasonable values for these energies?