Module 2: Quantum Espresso Walkthrough –
Energy and Geometry Optimization of the H_2 Molecule

We will be using the **PWSCF** code for quantum mechanical calculations of extended systems. The PWSCF program is part of the **Quantum Espresso** package. It is a full ab-initio package implementing electronic structure and energy calculations, linear response methods (to calculate phonon dispersion curves, dielectric constants, and Born effective charges) and third-order anharmonic perturbation theory. PWSCF can use *norm-conserving pseudopotentials* (PP), *ultrasoft pseudopotentials* (US-PP) and *PAW* potentials within *density functional theory* (DFT). The PWSCF code and the Quantum-Espresso package are freely available under the conditions of the GNU GPL. Further information (including online manual) can be found at the Quantum-Espresso website http://www.quantum-espresso.org.

❖ EWS PRE-INSTALLATION QE 5.0.3

A full installation of QE 5.0.3 (patched version of 5.0.2) is available on EWS Linux at:

`/class/mse498af/software/2_espresso/espresso-5.0.2/`

The `/class/mse498af/software/2_espresso/espresso-5.0.2/Doc` subdirectory contains *all kinds of documentation about the code*. As we construct input files, you can look in that directory to get more information about all of the options.

❖ QE 5.0.3 LOCAL INSTALLATION INSTRUCTIONS

You should have received instructions to perform a local install of QE in your EWS space or your personal machine.

❖ QE UNITS

- 1 bohr = 1 a.u. (atomic unit) = 0.529177249 Å
- 1 Rydberg (Ry) = 13.6056981 eV
- 1 kcal/mol = 43.36 meV
- 1 kJ/mol = 10.36 meV
- 1 eV =1.60217733 $\cdot$ 10^{-19} J

❖ QE WALKTHROUGH - THE H_2 MOLECULE

This walkthrough tutorial describes how to calculate energies and perform geometry relaxation using the PWSCF code in Quantum Espresso.

First create a directory for your runs, we will then copy over the input file and scripts that we will use. Create this directory somewhere convenient in your home directory. **Do NOT create a subdirectory within the directory hierarchy holding the QE installation.** For example:

```
$ cd ~
$ mkdir MSE498_QE
$ cd MSE498_QE
$ mkdir Runs
$ cd Runs
$ mkdir H2
$ cd H2
```

1. The input files

1a. Energy calculation

We will look at the `H2.scf.inp` input file; **download this file from COMPASS and place this file in Runs/H2**. If you are logged in remotely, you will need to use `scp` or `sftp` to perform the upload. This input file is for an H_2 molecule in a cubic simulation cell. Look at the input file, which you have just copied over to your directory:

```
$ less H2.scf.inp
```

You can scroll through the file by typing space (forward), b (backwards); hit q to quit.

The file will look something like this:

```
(1)  &CONTROL
(2)   calculation = 'scf' ,
(3)   restart_mode = 'from_scratch' ,
(4)   prefix = 'H2' ,
(5)   tstress = .true. ,
(6)   tprnfor = .true. ,
(7)   pseudo_dir = '$HOME/QE/espresso-5.0.2/pseudo',
(8)   outdir = '/tmp/$USER',
(9)  /
(10)  &SYSTEM
(11) ibrav = 1,
(12)  celldm(1) = 10.0,
(13)  nat = 2,
(14)  ntyp = 1,
(15)  ecutwfc = 20,
(16) /
(17) &ELECTRONS
(18)  conv_thr = 1.0d-8 ,
(19)  mixing_mode = 'plain' ,
(20)  diagonalization = 'david' ,
(21) /
(22) ATOMIC_SPECIES
```

```
(23)   H   1.00794  H_US.van
(24) ATOMIC_POSITIONS bohr
(25)    H     -0.7  0.0  0.0
(26)    H      0.7  0.0  0.0
(27) K_POINTS automatic
(28)   1 1 1   0 0 0
```

Lines numbers are added for reference.

Line 1: `control` declares the control block.

Line 2: `calculation = 'scf'` tells PWSCF that this will be a self-consistent field calculation (DFT).

Line 3: `restart_mode = 'from_scratch'` declares that we will be generating a new structure.

Line 4: `prefix='H2'` declares the filename prefix to be used for temporary files.

Line 5: `tstress = .true.` is a flag to calculate the stress tensor.

Line 6: `tprnfor = .true.` is a flag to calculate the forces.

Line 7: `pseudo_dir = '/class/mse498af/software/2_espresso/espresso-5.0.2/pseudo'` defines the location of the directory where you store the pseudo-potentials. If you are using a local install of QE will need to **edit this line** to reflect the correct path to your local pseudo directory. You can alternatively remove this line if you set the environment variable `ESPRESSO_PSUEDO` via

`$ export ESPRESSO_PSEUDO=/class/mse498af/software/2_espresso/espresso-5.0.2/pseudo`

Line 8: `outdir='/tmp/$USER'` defines the location of the temporary files. This should always be a local scratch disk so that large I/O operations do not occur across the network. You will need to **edit this line** changing `$USER` to your user name, to ensure that you don't accidentally try to overwrite your colleagues' directories.

Line 9: `/` denotes the end of a block.

Lines 10-13: the `system` block

`ibrav` – gives the crystal system. ibrav=1 is a simple cubic structure.

This is used because the symmetry of the structure can reduce the number of calculations you need to do. If you need other crystal systems, consult the `INPUT_PW` file (txt or html) in the `espresso-5.0.2/Doc` directory.

`celldm` – defines the dimensions of the cell. You will be changing this parameter. celldm is in atomic units, or bohrs. Remember that 1 bohr = 0.529177249 Å. The value will depend on the Bravais lattice of the structure. For simple cubic, celldm(1) = a_0. In cubic systems, $a = b = c = a_0$. Consult `INPUT_PW` for other crystal systems.

`nat` – number of atoms (each individual unique atom).

`ntyp` – number of types of atoms

`ecutwfc` – Energy cutoff for pseudo-potentials in Rydbergs. This one is important; you will be changing this parameter.

Lines 14-19: The `electrons` block

`diagonalization` – diagonalization method. Use the default for now.

`mixing_mode` – Mixing method. Don't worry about this for now.
`conv_thr` – Convergence threshold. Use the default for now.
Lines 20-21: Atomic species declaration
After the keyword `ATOMIC_SPECIES`, for each `ntyp` enter
`"atomic symbol" "atomic weight" "pseudo-potential"`
Lines 22-24: Atomic positions
After the keyword `ATOMIC_POSITIONS units`, for each `nat` enter
`atomic symbol x y z`
where x, y, z are given in the units specified by `units` (alat, bohr, crystal, angstrom).
Lines 25-26: k-point selection
after the keywork `K_POINTS`, "automatic" tells PWSCF to automatically generate a k-point grid.
The format of the next line is
nkx nky nkz offx offy offz
where nk* is the number of intervals in a direction and off* is the offset of the origin of the grid.
You can read the documentation for the input file in `INPUT_PW` file (txt or html) in the `espresso-5.0.2/Doc` directory.

1b. Geometry optimization

To relax the atomic positions of the hydrogen atoms using the forces, we use the input file `H2.relax.inp`; **download this file from COMPASS and place this file in Runs/H2**.

The only difference between this input file and the previous one is the second line and the added section `IONS` in which we use the default values for all parameters for now.

```
 calculation = 'relax' ,

 &IONS
 /
```

2. Running PWSCF

To run the PWSCF code and calculate the energy of the H_2 molecule, type

```
$ /class/mse498af/software/2_espresso/espresso-5.0.2/bin/pw.x <
H2.scf.inp > H2.scf.out
```

To relax the bondlength of the H_2 dimer use the second input file

```
$ /class/mse498af/software/2_espresso/espresso-5.0.2/bin/pw.x <
H2.relax.inp > H2.relax.out
```

2a. Energy calculation

```
$ less H2.scf.out
```

Scroll through this file. The beginning will just recap the configuration that is being calculated. Then there is some information about the pseudo-potentials that PWSCF just read in. The next part tells you about intermediate energies that PWSCF calculates, before the calculation is converged. Near the end, there will be something like:

```
! total energy = -2.31089645 ryd
```

This is your total energy, as calculated by PWSCF. Your number may be slightly different. The final occurrence of "total energy" will have an exclamation point by it, something you can use to hunt for it. You can skip to this right away by using search functions in `vi` or `grep`, or just scroll down a lot. For example:

```
$ grep '! *total energy' H2.scf.out
```

To make sure you include the *; it tells grep to allow as many spaces as needed between the ! and "total energy". The quotes are also required, as both ! and * are special characters in unix.

Scrolling down further, we come to the computed forces experienced by the ions given the DFT calculated electronic structure computed around their (fixed) positions:

```
Forces acting on atoms (Ry/au):

     atom    1 type  1   force = -0.03840675 0.00000000 0.00000000
     atom    2 type  1   force =  0.03840675 0.00000000 0.00000000

     Total force =  0.054315     Total SCF correction =  0.000002
```

In the energy calculation the atoms are not allowed to move and these are the forces they experience. In the geometry optimization calculation, these forces will be used in an optimization protocol to adjust the ion position to minimize the total system energy.

At the end of the file, it will tell you how long your program took to run in terms of CPU time (total time of all processors) and "wallclock" (which is the time that you experience). The last line is something like

```
     PWSCF        :     0.32s CPU         0.40s WALL
```

and all of the preceding lines are breakdowns of time for each routine

```
     init_run     :      0.05s CPU      0.05s WALL (       1 calls)
     electrons    :      0.19s CPU      0.22s WALL (       1 calls)
     forces       :      0.01s CPU      0.01s WALL (       1 calls)
     stress       :      0.02s CPU      0.02s WALL (       1 calls)

     Called by init_run:
     wfcinit      :      0.00s CPU      0.00s WALL (       1 calls)
     potinit      :      0.03s CPU      0.03s WALL (       1 calls)

     Called by electrons:
     c_bands      :      0.02s CPU      0.02s WALL (       6 calls)
```

```
     sum_band     :      0.02s CPU      0.02s WALL (       6 calls)
     v_of_rho     :      0.14s CPU      0.15s WALL (       7 calls)
     newd         :      0.01s CPU      0.01s WALL (       7 calls)
     mix_rho      :      0.01s CPU      0.01s WALL (       6 calls)
```

… etc. …

Your numbers may be different from these. You should start to develop a feel for how long your runs take, and how much memory they will use.

2b. Geometry optimization

```
$ less H2.relax.out
```

When relaxing the structure, there will also be lines which say:

```
     End of BFGS Geometry Optimization

     Final energy   =      -2.3118661236 Ry
Begin final coordinates

ATOMIC_POSITIONS (bohr)
H       -0.725970736   0.000000000   0.000000000
H        0.725970736   0.000000000   0.000000000
End final coordinates
```

And you will see that the forces on the nuclei are decreasing towards zero as their positions are optimized. The last lines show the relaxed atom coordinates.

3. Convergence issues in first-principles calculations.

The PWSCF code expands the one-electron wave functions in *basis functions* that are plane waves. They are called "plane waves" because surfaces of constant phase are parallel planes perpendicular to the direction of propagation. The plane waves are chosen to have a periodicity compatible with the periodic boundary conditions of the simulation cell, *i.e.* the set of *G* vectors are integer multiples of the three primitive lattice vectors. In actual calculations, we use plane waves up to a cutoff value to make the plane wave expansion finite. The *cutoff* is always given in energy units (such as Rydberg or eV) corresponding to the kinetic energy of the highest *G*.

N.B. *The units of reciprocal lattice are the inverse of the direct lattice, or 1/length. To convert 1/length to energy units we use* $E = h\nu = h/\lambda$.

We will determine the convergence of the total energy of the H_2 molecule with respect to the energy cutoff of the plane wave basis set. Open the file `H2.scf.inp` using `vi` (or some other text editor) and look for the parameter `ecutwfc`. Double this parameter from 20 to 40 Ry and assess how the total energy of the H_2 molecule changes when we increase the energy cutoff of the plane wave basis.

We would like to identify the required cutoff energy for a convergence of the energy to within 2 mRy/atom (i.e. 0.004 Ry for our 2 atom system), and determine the accuracy of the bond length at this cutoff energy. To speed up the convergence calculations, we will use a script. Following is the script we will be using to check the convergence with respect to the plane wave cutoff. **Download this file from COMPASS and place this file in Runs/H2.**

`$ less Run_Ecut.bash`

```
#!/bin/bash

INPUTFILE=H2.scf.inp
PWSCF=/class/mse498af/software/2_espresso/espresso-5.0.2/bin/pw.x

for ecut in 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
do
  sed "/ecutwfc/s/.*/ ecutwfc=$ecut/" $INPUTFILE | $PWSCF > out
  e=$(grep '! *total energy' out | tail -n 1 | awk '{print $5}')
  echo $ecut $e

done
exit
```

The script loops over the values ecut from 10 to 80 Ry in steps of 5 Ry. It creates a new input file based on `H2.scf.inp` by changing the entry in the line `ecutwfc` using the bash utility `sed` (kind of like `awk`) and then *pipes this as input* to the PWSCF code. Finally, it extracts the energy and bondlength from the file "out" using `grep` and prints them to the terminal. It rewrites the file "out" each time.

Run the script to determine the convergence of the total energy of the H_2 molecule with respect to the energy cutoff.

`$ chmod 755 Run_Ecut.bash`
`$ ./Run_Ecut.bash`

Use MATLAB to fit your data of total energy vs. cutoff energy to a decaying exponential to extrapolate out to the energy at an *infinite cutoff*. Consider fitting a function of the form $y = Ae^{-Bx} + C$. Since this is cannot be formulated as linear regression problem, let's use the MATAB curve fitting toolbox GUI by typing `>> cftool` into the command prompt.

(If the curve fitting toolbox is unavailable, you can use the online curve fitting server available at http://zunzun.com/Equation/2/Exponential/Exponential%20With%20Offset/)

Load the y data (E_{total}) and x data (E_{cut}) from your workspace into the tool, and specify a custom function of the form `y=a*exp(-b*x)+c`. Using the fitted values for a, b, and c, what is:

a) $\lim_{E_{cut}\to\infty} E_{total}$

b) the minimum value of E_{cut} for the desired accuracy of 0.004 Ry?

When you have found the cutoff that you will use for hydrogen, **rerun the relaxation calculation using `H2.relax.inp` with your converged cutoff** to make a DFT prediction for the hydrogen molecule bond length. The experimentally determined value is ~74 pm.

In this example, we have considered a single H_2 molecule in a large real space unit cell, meaning the interactions between periodic images of the molecule are effectively zero. For atoms and molecules with no periodic interactions, the Bloch Theorem does not apply (i.e. there is no translational invariance of the Hamiltonian for isolated entities). A molecule in a large box is effectively isolated, allowing us to perform a single k-point calculation. (Recall that if a lattice vector is longer in real space, it is shorter in *k*-space.) Including more k-points just captures the intermolecular interactions more accurately, and for large real space cells these are effectively zero. In your project, you will consider a condensed system, and so will need to look at the convergence of the energy with respect to both *the cutoff energy* **and** *the k-point mesh*. You can modify the automation script and input files used in this walkthrough for your project.

❖ Organizing your runs

You can organize your runs any way you like, or you don't have to organize them at all. One way is to make directories for each problem you do, and name your output files accordingly. Good organization may save you headache in the long run but this is really up to you. Be cognizant especially of the working directories (outdir) that you specify in your input files. Keep in mind also that—as we saw here—input files are just text files, and so you can write scripts that create and manipulate those input files in an automated way. This significantly improves the efficiency and reduces human error in running computational simulations.

❖ FAQ

I do not understand quantum mechanics at all.

- General introductions to Quantum Mechanics:
  http://walet.phy.umist.ac.uk/QM/LectureNotes/QM.html.
  In particular look at Chapters 1, 2, 3, 8, 11.
- Operators, expectation values, bra-kets: Paragraphs 1.1 and 1.2 of
  http://www-keeler.ch.cam.ac.uk/lectures/quant_letter.pdf
- Or, more complete:
  http://www.math.utah.edu/~gold/doc/quantum.pdf
- Very simple primer:
  http://www.chem1.com/acad/webtut/atomic/

N.B. It is not essential know every detail of quantum mechanics to run quantum mechanics code!

How precisely do I need to get the lattice parameter?

Lattice parameters are typically listed to within 0.01 Angstroms. There are applications when higher precision is required; this is not one of them.

My E vs. lattice constant plot is jagged.

There are a number of solutions to this; the easiest is to raise the energy cutoff.

I don't like scripts.

Use scripts! They will save you a lot of time in the long run.

The weights of the k-points add up to 2, not 1.

Yes. This is a "feature" of the code. Don't worry about it.

How is "convergence of energy" defined?

You say that your energy is converged to X Rydbergs when $E_{true} - E_n = X$ (n is the current energy). How do you know E_{true}? In practice you might take your energy at the highest cutoff (or k-grid, or whatever) that you calculated – if that seems converged, you might call that E_{true}. Alternatively, you might extrapolate to an infinite cutoff energy using curve fitting. The most important thing is that you do **NOT** define convergence as $E_{n+1} - E_n$, where n is a step in k-point, energy cutoff, or whatever.

You do need to be careful though. It is possible to get "false" or "accidental" convergence as well. That is, your energy at a 2×2×2 k-grid may be the same as the energy at a 8×8×8 k-grid, but the energy at a 4×4×4 might be very different from both of these. In this case, you aren't really converged at a 2×2×2 k-grid.

I don't understand convergence of energy and forces.

From experience, we know that a good error on energy differences is ~5 meV/atom and on forces is ~10 meV/Angstrom. These are just values are just a general guide derived from many first-principles calculations in the past. Depending on the system and property you may have to use different convergence criteria.

I am having problems both converting Ryd to eV and Bohr to Å.

These units page may help you:
http://physics.nist.gov/cuu/Constants/index.html
http://www.chemie.fu-berlin.de/chemistry/general/units_en.html

Does PWSCF use LDA or GGA? DFT or Hartree Fock?

PWSCF uses DFT. It has both LDA and GGA.

Why do I take symmetric k-point grids? Can I take asymmetric k-point grids?

For this material, we take a symmetric k-point grid because it is the same in all three lattice directions. This is not always true. The 'best' *k*-point meshes are those that sample *k*-space evenly in all directions. Thus, for a tetragonal cell (with a=b, c=2a, and all angles 90 degrees) we might take a 8×8×4 mesh. (Note that if a lattice vector is longer in real space, it is shorter in *k*-space.)

When I try to run QE I get a mysterious IOTK error. What's happening?

The IOTK library is used by QE to write and read XML files. QE will periodically throw an IOTK runtime error that is not easily reproducible and poorly understood even by the QE developers (see http://www.quantum-espresso.org/faq/frequent-errors-during-execution/#5.3). The easiset workaround is to just use a different workstation, or -- if you are working remotely -- re-login with a new ssh shell.