Special Course in Theoretical Physics PHYS-E0546: DFT for Practitioners

Aalto University, September 2022



Tutorial 1: Basics of Electronic Structure Theory Manuscript for Exercise Problems

Prepared by Ondřej Krejčí, Orlando Silveira and Adolfo Fumega Developed by Patrick Rinke, Milica Todorović, Department of Applied Physics Aalto University, 2022

A quick summary of the exercises

This tutorial will give you a basic introduction to electronic structure calculations for very simple systems. As every DFT code has its own philosophy, this work should also familiarize you with fundamental aspects of using FHI-aims. This tutorial consists of five exercises:

Problem I:The hydrogen atomProblem II:Ground state of atomsProblem III:Bonds and binding in moleculesProblem IV:Hydrofluoric acid (HF): bond length and dipole momentProblem V:Knallgas reaction: reaction energy

For every exercise, we also provide solutions and sample input files. They can be found in \$TU-TORIALS/ directory. However, we strongly recommend to use the provided input files only in case of time shortage. You will maximize your learning progress by trying to generate the input files on your own. In case you get stuck with a particular problem, do not hesitate to ask one of the tutors. An executable of FHI-aims, as well as species files and utility scripts, are provided in the /work/courses/unix/PHYS/E0546/CODE/ directory.

Note: Please do not copy and paste the description in this pdf into your input files. Typically, invisible characters from the formatting are copied, too. FHI-aims will revoke these characters and your calculations will not start.

The very basics of FHI-aims

Each calculation should be done in a *separate directory* containing the two mandatory input files control.in and geometry.in. FHI-aims is then called in this directory.

In short:

- Each calculation in a separate directory
- 2 input files:
 - control.in
 - geometry.in
- Start calculation aims.x | tee output

The above starts a calculation on a single processor core, shows the main output on the screen and, at the same time, pipes it to the output file. The output file contains the basic information and results of the calculation such as the total energy, atomic forces, and so forth. Additional output files might be generated according to the specified settings.

Nuclear positions: geometry.in

The geometry.in file contains all information about the atomic structure of the system. This includes the nuclear coordinates, which are specified by the keyword atom, followed by cartesian coordinates (in units of Å) and the element symbol, called "species" in FHI-aims. In addition, comments can be added using a preceding hash symbol. Fig. 1 shows an example geometry.in file for a hydrogen atom. For periodic calculations, lattice vectors can also be given in this file. This will be covered in one of the next tutorials. At present, we will stick to non-periodic systems.

```
#The hydrogen atom atom 0.0 0.0 0.0 H
```

Figure 1: An example geometry.in file for a hydrogen atom positioned at the origin.

Choosing the method: control.in

This file contains all physical and computational settings for the calculation. Fig. 2 shows a minimal example of a control.in file, which can be used as a template. These basic settings should be used as default for part I of this tutorial, unless specified otherwise.

Figure 2: Default physical and computational settings for control.in.

A slightly more extended In this example, the following options are set:

• xc

This keyword sets the method to be used, in this case the exchange-correlation functional. For example you can choose the option pbe, which requests the Perdew-Burke-Ernzerhof (PBE) functional [1] for the DFT calculation.

• charge

Set the total charge of the system in units of |e|. For a neutral system, this is zero.

• spin

This keyword governs the spin treatment. It can be set to **none**, which requests a spin-restricted calculation, or to **collinear**, which requests a spin-unrestricted (polarized) calculation. In a spin-restricted calculation, α and β spins are assumed to be equal. Only one spin-channel is treated and hence, the number of electrons is effectively halved. This accelerates the calculations significantly.

• default_initial_moment hund

Sets the initial spin of the atoms. The value "hund" requests an initial moment taken from Hund's rule, but it is often more sensible to provide a reasonable guess by hand for a single atom. For multi-atomic system it is now required to specify initial moments for selected atoms in the geometry.in file. Only necessary for spin collinear calculations.

The other lines determine the convergence criteria of the self-consistency cycle regarding

- sc_accuracy_rho the change of the density
- the sum of orbital eigenvalues: $\verb"sc_accuracy_eev"$, and

• the total energy sc_accuracy_etot between two consecutive cycles.

In addition to these keywords, the control.in file must contain the definition of the computational parameters for each species that is specified in geometry.in. The order of the species in the listing is irrelevant. FHI-aims is shipped with pre-defined settings for all species which govern the key parameters regarding the numerical accuracy. They include, *inter alia*, the specification of all real-space integration grids, the accuracy of the Hartree potential and, of course, the basis set. For all elements, defaults are provided for three different levels of accuracy: *light, tight*, and *really tight*. They can be found in the directory

\$SPECIES

and should be copied and pasted into control.in, e.g. via the command

cat \$SPECIES/really_tight/01_H_default > control.in

which pastes the really_tight settings of the H-atom into the control.in file. Already the *tight* species_defaults are rather safe and *really tight* settings are overconverged for most purposes. In addition the number of basis functions can be varied, as well as the basis functions themselves. The basis functions associated with a given species are tabulated at the end of these default settings, as shown in Fig. 3

The idea of keeping the species defaults out in the open is that, even if they are not modified, these are the critical accuracy parameters which one might look at to ensure numerical convergence. Each line denotes a specific basis function. They can be read as follows: The first keyword denotes the "type" of the basis function. *Hydro* means that this is a hydrogen-like basis functions. Some basis functions are of the type *ionic*. They are described in more detail in the manual. The next two symbols correspond to the first two quantum numbers (principal quantum number n, orbital quantum number l) of the basis functions, and the final number corresponds to the "effective nuclear charge" for which this basis function is created. hydro 1 s 0.85 corresponds to the exact solution for the 1 s basis function of a hydrogen atom if it had a nuclear charge of only 0.85.

The basis functions are classified in "tiers". Not all basis functions are enabled by default. Rather, some are commented out using the "#" symbol. They can be included in the calculation by removing the hash symbol from the corresponding lines. Systematically improved calculations can be performed by enabling additional tiers one after another.

Additional tools and programs

Linux:

A compilation of useful Linux commands can be found on MyCourses: https://shorturl.at/ CFJM7. We printed for you the same version of Linux and Bash shell commands.

Scripts:

For some exercises, scripts are required for dedicated tasks. All scripts you will need for this tutorial can be found in **\$TUTORIALS**/

Visualization tools:

To visualize structures, vibrational modes, charge density plots, etc., several programs including molden, vmd, jmol and python-ase are installed on your work stations.

Plotting and Editing:

Gnuplot and xmgrace are installed on your workstation to visualise and plot data. Also several editors are available, including gedit, emacs, and vi.

Interactive python shell:

python3, spyder and ipython development environments for python are installed on all the computers.

```
#
# FHI-aims code project
 VB, Fritz-Haber Institut, 2007
#
#
 Suggested "safe" defaults for H atom
#
#
 (to be pasted into control.in file)
*****************
 species
            Η
#
    global species definitions
  nucleus
                 1
                 1.00794
  mass
#
  l_hartree
                 8
#
                 4.0 2.0 1.0
  cut_pot
  basis_dep_cutoff
                0.d0
#
  radial_base
                 24 7.0
  radial_multiplier
                 2
  angular_grids
                 specified
    division 0.2783 110
    division 0.3822 194
    division 0.5626 302
    division 0.5922 434
    division 0.6227 590
#
    division 0.7206 770
    outer_grid 770
#
    outer_grid 590
****************
#
# Definition of "minimal" basis
#
#
   valence basis states
  valence 1 s
                1.
#
    ion occupancy
  ion_occ
          1 s
                 0.5
****************
#
# Suggested additional basis functions. For production calculations,
 uncomment them one after another (the most important basis functions
#
# are listed first).
#
# Basis constructed for dimers: 0.5 A, 0.7 A, 1.0 A, 1.5 A, 2.5 A
#
"First tier" - improvements: -1014.90 meV to -62.69 meV
#
   hydro 2 s 2.1
   hydro 2 p 3.5
#
 "Second tier" - improvements: -12.89 meV to -1.83 meV
#
   hydro 1 s 0.85
#
   hydro 2 p 3.7
#
   hydro 2 s 1.2
#
    hydro 3 d 7
#
  "Third tier" - improvements: -0.25 meV to -0.12 meV
#
   hydro 4 f 11.2
#
    hydro 3 p 4.8
#
    hydro 4 d 9
#
    hydro 3 s 3.2
```

Figure 3: Tabulated basis functions for hydrogen. The basis functions are classified in "tiers". In this example only the tier 1 and minimal basis functions are enabled.

Problem I: The hydrogen atom

In this exercise, we aim to convey the basics of FHI-aims using the hydrogen atom. The hydrogen atom is the simplest non-trivial system possible and the only one for which the exact analytic solution is known. By the end of the first exercise, we will see how various computational methods compare to each other and to the exact solution. From a technical perspective, we will learn how to generate input files, read the standard FHI-aims output, and perform basis set convergence tests.

Getting started - the hydrogen atom

Educational Objectives:

- Become acquainted with running FHI-aims calculations
- Learn how to do systematic basis set convergence

Tasks

- 1. Create a new directory, e.g. called h_atom or problem_1.
- 2. Generate a simple geometry.in file by hand, which contains only a single hydrogen atom, using the example shown in Fig. 1. This corresponds to a single hydrogen atom in a hypothetical ideal gas phase. It is located at the origin of the coordinate system, although its position does not matter here.
- 3. Generate a simple control.in file by hand, using the example control.in file given in Fig. 2. Keep using pbe for the xc method. Finally, append the "really_tight" species data of H to the end of the control.in file, e.g. via the command cat \$SPECIES/really_tight/01_H_default >> control.in
- Now, run FHI-aims: aims.x | tee output

Once the calculation has finished, open the **output** file. If you find the line "Self-consistency cycle converged." near the end, then your calculation is converged. We are now interested in the total energy. Search for the block:

Total energy uncorrected	:	-0.135697361759588E+02	eV
Total energy corrected	:	-0.135697361759588E+02	eV
Electronic free energy	:	-0.135697361759588E+02	eV

Or easy to read block closer to the end of the output file:

Total energy of the DFT / Hartree-Fock s.c.f. calculation	n :	-13.569736176	eV
Final zero-broadening corrected energy (caution - metals	only) :	-13.569736176	eV
For reference only, the value of 1 Hartree used in FHI-air	ms is :	27.211384500	eV

In this special case, all energies are equal, but this will not be the case if fractionally occupied orbitals were found! For non-metallic systems, as in this tutorial, always use the **Total** energy uncorrected value. (You will learn what the other two values mean in a future lecture.) Compare it with the exact result for the hydrogen atom (0.5 Hartree ≈ 13.6057 eV).

TIP:

In later exercises, to find this value fast and efficiently, use the command grep 'Total_energy_uncorrected' output

5. We will learn more about basis functions and basis convergence tests in the next lecture. Here you can already gather some first-hand experience on the basis set dependence of your calculations. To do so, redo the calculation with different basis sets (minimal, tier1, tier2, tier3) by (un)commenting the basis functions at the end of the control.in file. Calculations with minimal basis set are performed by removing all the additional basis functions that are listed in the file. Search the output file to find out how many basis functions are actually used in the calculations. Then, plot the total energy as function of the basis set size. At which tier does the total energy converge?

TIP:

```
To plot the results, simply create a text file (e.g., results.dat) with two
columns, the number of basis functions and the obtained total energy.
This file can be plotted directly using the command:
xmgrace results.dat
or:
gnuplot
> plot 'results.dat' w lp
```

Optional: Method performance

To learn about the performance of difference exchange-correlation functionals, repeat for different methods. Replace **pbe** in control.in with

- pw-lda
- pbe0

Do all methods converge with basis set size? Do all converge to the same result?

Optional: An optimal method and basis set for the H atom

Optional part: The Hartree-Fock method is exact for single electron systems (within the Born-Oppenheimer approximation). Run the above calculations using the Hartree-Fock method (keyword hf). How close do you get to the exact result?

Problem II: Ground state of atoms

Educational Objectives:

- Learn about the DFT ground state for different chemical species
- Compare DFT results to a high-accuracy simulation method

In this exercise, you will perform PBE total energy calculations for different atoms. Please enter the numbers in the table and compare them to the essentially exact configuration interaction (CI) results from Chakravorty et al. [2].

total energy (all energies in Ha)	Н	Не	Li	Be	С
PBE					
full CI*	-0.500	-2.902	-7.478	-14.667	-37.845
deviation in Ha deviation in %					

*Chakravorty et al. [2]

Tasks

- 1. Follow steps 1 to 3 from the previous exercise for the atoms in the table above. Do not forget to also replace the hydrogen basis set with the one for the new atom. Enter the total energies into the table.
- 2. In principle you would have to repeat step 4 (the basis set test) for every atom again. Here it is enough, if you run the calculation for the basis set level that you deemed sufficient in the previous exercise, i.e., tier1, tier2, or tier3.
- 3. How well do the DFT total energies produced with the pbe exchange-correlation functional compare to the essentially exact CI results?

Problem III: Bonds and binding in molecules

Educational Objectives:

- Learn about total energy difference
- Compute your first observables

Let us now look at molecules, molecular binding and bond strength. An instructive example are the three hydrocarbons ethyne (C_2H_2 , also known as acetylene) ethene (C_2H_4 , also known as ethylene), and ethane (C_2H_6) (see Figure 4).

A major application of acetylene is the conversion to ethylene for making a variety of polyethylene plastics. Acetylene is mainly manufactured by the partial combustion of methane or appears as a side product in the ethylene stream from cracking of hydrocarbons. Approximately 400,000 tonnes are produced by this method annually [3]. Ethylene is widely used in the chemical industry, and its worldwide production (over 109 million tonnes in 2006) exceeds that of any other organic compound. Polyethylenes of various types consume more than half of the world's ethylene supply. Polyethylene, also called polythene, is the world's most widely used plastic (primarily to make films used in packaging, carrier bags and trash liners). Linear alpha-olefins, produced by oligomerization (formation of short polymers) are used as precursors, detergents, plasticisers, synthetic lubricants, additives, and also as co-monomers in the production of polyethylenes. Ethylene is also an important natural plant hormone, used in agriculture to force the ripening of fruits [4].

The chief use of ethane is also in the chemical industry in the production of ethene (ethylene) by steam cracking. When diluted with steam and briefly heated to very high temperatures (900 °C or more), heavy hydrocarbons break down into lighter hydrocarbons, and saturated hydrocarbons become unsaturated. Ethane is favoured for ethene production because the steam cracking of ethane is fairly selective for ethene, while the steam cracking of heavier hydrocarbons yields a product mixture poorer in ethene, and richer in heavier alkenes (olefins) such as propene (propylene) and butadiene, and in aromatic hydrocarbons [5].

To study molecular properties useful quantities are the atomisation energy

$$E_{\text{atomization}} = E_{tot}(\text{molecule}) - \sum_{i} E_{tot}(\text{atom}_i)$$
(1)

and the energy stored in the carbon-carbon bond

$$E_{\text{C-C bond}} = E_{tot}(\text{molecule}) - E_{tot}(\text{left fragment}) - E_{tot}(\text{right fragment})$$
(2)

where E_{tot} denotes the total energy of the corresponding parts.

Tasks

1. In the folder \$TUTORIALS/ you will find the geometries of the different molecules. Alternatively, you can look up the experimental geometries for the molecules on the NIST database (http://cccbdb.nist.gov/) and then set-up your own geometry.in file. Then calculate the total energy of the three molecules, as you did in the previous exercises for the isolated atoms.



Figure 4: Ball and stick model of acetylene, ethylene and ethane. Carbon atoms are in grey and hydrogen atoms in white.

- 2. To determine the atomization energy according to Equation 1 you will also need the energies of the individual atoms, which you already calculated in exercise 2. Make sure that you use converged or at least consistent basis sets.
- 3. To determine the C-C bond energy according to Equation 2 you will need the energies of the two fragments that result from cleaving the C-C bond. For the fragment calculations, modify the geometry and control files accordingly.
- 4. What do you observe and how can this be explained?

	C_2H_2	C_2H_4	C_2H_6
atomisation energy			
$experiment^*$	$16.86~{\rm eV}$	$23.07 \ {\rm eV}$	28.89 eV
deviation in eV			
deviation in $\%$			
C-C bond energy			

*NIST Chemistry WebBook http://webbook.nist.gov/chemistry/

Problem IV: Hydrofluoric acid (HF): bond length and dipole moment

Hydrofluoric acid (HF)

One of the most influential papers in chemistry for the systematic investigation of the performance of DFT was Johnson, Gill, and Pople[6], in which several properties of a large number of diatomic systems were consistently computed and compared to experimental values. In the style of this work, we will calculate the binding curve, atomization energy (ΔH_{at}), and the dipole moment for hydrogen fluoride (HF) with two methods. From a technical perspective, this exercise teaches you how simple shell scripting can be used to make your (computational) life easier.

Educational Objectives:

- Become acquainted with shell scripting.
- Learn how to compute dipoles and atomic charges.
- Find the equilibrium bond distance of a simple diatomic molecule.
- Assess the reliability of charge partition schemes.

Tasks:

1. The first task of this exercise will be to find the equilibrium bond distance of HF from a series of calculations. Start by creating a geometry.in file which contains a H and a F atom, as shown in Fig. 5.

```
#HF at variable bond distances
atom 0.0 0.0 0.0 H
initial_moment 1
atom 0.0 0.0 Dist F
initial_moment 1
```

Figure 5: The geometry data for a HF molecule. H is put on the origin and F is located Dist Å away from the origin along the z-axis.

In this example, H is put on the origin and F is located Dist Å away from the origin along the z-axis. Hereby, Dist is a placeholder which will be replaced by the actual distance later using a script.

- 2. Create a control.in file, and specify a pbe calculation for a neutral system as in the previous exercises. Feel free to copy the file from the previous exercise, but remove the *really tight* species setting for hydrogen. Instead, please paste the *tight* setting for H and F into the control.in file. We also want to compute the electronic dipole moment, which is requested using the keyword output dipole. Note that although it does not matter where you put the keyword, it is recommended to put all manually set keywords *before* the species tags.
- 3. Next, create a bash script which runs FHI-aims for a series of bond distances (ideally between 0.7Å and 1.3Å with 0.1Å steps, and a denser step width of 0.02 Å between 0.85 Å and 0.95 Å).

Alternatively, you may also use the bash or python scripts we provide in \$TUTORIALS/

For each distance, it should

- create a unique directory
- create the control.in and geometry.in file from templates
- replace the bond distance place holder **Dist** with the bond distance and
- start FHI-aims.

Run the script (python run.py or ./Run.sh) and plot the total energies vs. the bond length. This can be scripted as well. Example scripts (PP_Energies.*) are also provided in \$TU-TORIALS/

Which bond length corresponds to the lowest energy? How does the bond length compare to the experimental bond length of 0.917Å?

4. To compare with experimental values, we compute the atomization energy (ΔH_{at}) . In order to calculate ΔH_{at} , we will also need the total energy of the isolated H and F atoms. Compute the total energies for the single atoms using the **pbe** exchange-correlation functional.

Note:

Atoms are highly symmetric systems, often with multiple degenerate solutions. In the case of fluorine, for example, the unpaired p-electron might sit in the p_x , p_y , or p_z orbital. All three solutions are equivalent. If the calculation is started unbiased, it might converge to a superposition of these three cases, which is a saddle point on the potential energy surface and results in partial electron occupations. Although in DFT non-integer occupations are in principle allowed, one should be very suspicious when obtaining such a solution for non-metallic systems. Typically, solutions exist that are lower in energy. They can be found by breaking the inherent symmetry of the problem, for example by applying a small external field at the beginning of the SCF cycle.

To break the inherent symmetry of an atom and ensure integer occupation, set the keyword switch_external_pert 10 safe. This means that for 10 iterations, a small external field in the z-direction is applied and then switched off. Usually, this is sufficient to perturb the SCF out of the symmetric solution and towards the correct electronic structure. Calculate the atomization energy (ΔH_{at}) of HF by subtracting the free-atom energies from the predicted total energy of HF (i.e. the minimum total energy found when varying bond distances).

$$\Delta H_{at} = E_{tot}^{HF} - E_{atom}^{H} - E_{atom}^{F} \tag{3}$$

How does this compare to the experimental value of $\Delta H_{at} = 135.2$ kcal mol⁻¹ (5.86 eV)?

- 5. Now, let us look at the dipole moment. Search for the corresponding line in the output file. How does the dipole at the equilibrium distance compare with the experimental value of 1.82 Debye? Plot the dipole moment vs. the bond distance. You will find a (mostly) linear correspondence. Do you expect this trend to continue at large distances? Why or why not?
- 6. **Optional:** Repeat the bond length determination using the exchange-correlation functional **pbe0**. How does the optimal bond length change? How much does the total energy change? Are the dipole moments the same?

Optional: Charge partition schemes.

Chemical reactivity and many physical properties are often explained in terms of atomic charges. However, atomic charges are not physical observables, since no unique operator exists to determine this quantity. They rather depend on the chosen charge partition scheme. The charge partition schemes that is probably most commonly used is Mulliken [7]. In FHI-aims, you can request it by specifying output mulliken in control.in.

For the equilibrium structure of HF, compute the atomic charges with this scheme using the exchange-correlation functional **pbe0**. Use the charges to calculate the dipole moment μ in the point dipole approximation. In this approximation, for a two atom system, the dipole moment is defined as:

$$\mu = q \cdot |\vec{r_H} - \vec{r_F}| \tag{4}$$

where q is the atomic partial charge and $\vec{r_H}$ and $\vec{r_F}$ are the atomic positions of the atoms. The absolute value of the difference $|\vec{r_H} - \vec{r_F}|$ is the distance between hydrogen and fluorine. Compare the dipole moment to the one computed by FHI-aims. How do they compare?

Problem V: Knallgas reaction: reaction energy

Educational Objectives:

- Become acquainted with DFT in chemistry.
- Learn how to calculate reaction energies.

A simple, yet effective, chemical reaction is oxyhydrogen combustion (or Knallgas reaction in German):

 $2 \ \mathrm{H_2} + \mathrm{O_2} \rightarrow 2 \ \mathrm{H_2O}$

Oxyhydrogen is a mixture of H_2 and O_2 . It will combust when brought to its autoignition temperature. For the stoichiometric mixture, 2:1 hydrogen:oxygen, at normal atmospheric pressure, autoignition occurs at about 570 °C. The mixture is quite volatile and can be ignited with a spark. A fact that led to the Hindenburg zeppelin disaster in 1937.

We will use DFT to calculate the energy of the reaction. Note, chemists speak of a reaction enthalpy ΔH_R , which is the difference between the enthalpy of the product(s) and the reactants. Since we do not consider entropy in this exercise, we only calculate the reaction energy ΔE_R

$$\Delta E_R = E_{tot}(\text{products}) - E_{tot}(\text{reactants}). \tag{5}$$

Tasks:

- 1. Create control.in and geometry.in files for the product and the reactant. You can find the experimental geometries for H₂O, H₂ and O₂ on the NIST database (http://cccbdb.nist.gov/). Alternatively, we have provided geometries for you in the \$TUTORIALS/ directory.
- 2. Use the knowledge you have acquired thus far to launch pbe calculations and then calculate ΔE_R according to Equation 5.
- 3. Is the reaction exothermic (i.e. it releases energy) or endothermic (i.e. it requires energy to proceed)?

References

- John P. Perdew, Kieron Burke, and Matthias Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Phys. Rev. Lett.* 77 (18 Oct. 1996), pp. 3865–3868. DOI: 10.1103/ PhysRevLett.77.3865. URL: http://link.aps.org/doi/10.1103/PhysRevLett.77.3865.
- [2] Subhas J. Chakravorty et al. "Ground-state correlation energies for atomic ions with 3 to 18 electrons". In: *Phys. Rev. A* 47 (5 May 1993), pp. 3649–3670. DOI: 10.1103/PhysRevA.47. 3649. URL: https://link.aps.org/doi/10.1103/PhysRevA.47.3649.
- [3] http://en.wikipedia.org/wiki/Alkyne.
- [4] http://en.wikipedia.org/wiki/Ethylene.
- [5] http://en.wikipedia.org/wiki/Ethane.
- Benny G. Johnson, Peter M. W. Gill, and John A. Pople. "The performance of a family of density functional methods". In: *The Journal of Chemical Physics* 98.7 (1993), pp. 5612–5626.
 DOI: 10.1063/1.464906. URL: http://aip.scitation.org/doi/abs/10.1063/1.464906.
- [7] R. S. Mulliken. "Electronic Population Analysis on LCAO-MO Molecular Wave Functions. II. Overlap Populations, Bond Orders and Covalent Bond Energies". In: *The Journal of Chemical Physics* 23.10 (Oct. 1955), pp. 1841–1846. DOI: 10.1063/1.1740589.