A first encounter with siesta™

13/03/2024 - Federico Pedron

How do we run siesta?™

Tutorial files

Source this file:

/leonardo_work/EUHPC_TD02_030/softwares/siesta-5.0beta1/siestarc.sh

This will add SIESTA and all utilities to the path.

Tutorial files
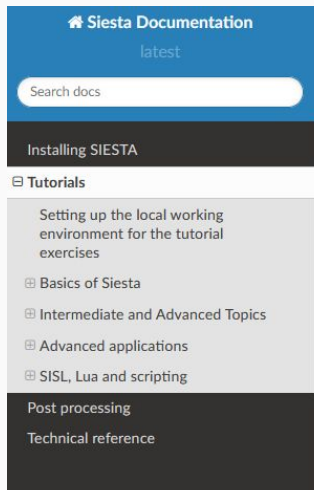
Each day, you should copy the tutorial folder available at

/leonardo_work/EUHPC_TD02_030/siesta-tutorials/day3-Wed

Each tutorial contains a sample submission script (run.sh). Edit it at your own convenience.

Tutorials themselves

You can find all tutorials at:

https://docs.siesta-project.org/projects/siesta/en/latest/tutorials/index.html

Submitting a job

```bash
#!/bin/bash
#SBATCH -J firstEncounterI
#SBATCH -n 8
#SBATCH -t 0:30:00
#SBATCH -o %x-%j.out
#SBATCH -e %x-%j.err
#SBATCH --partition=boost_usr_prod
#SBATCH --account=EUHPC_TD02_030
#SBATCH -D .

# DO NOT CHANGE THIS LINE
source /leonardo_work/EUHPC_TD02_030/softwares/siesta-5.0beta1/siestarc.sh

# EDIT THE CORRECT INPUT AND OUTPUT FILES.
srun -n 8 siesta < ch4.fdf > ch4.out
```

GPUs tomorrow!

The SIESTA method itself is very efficient, so for small systems (less than a 100 atoms), using a GPU is actually detrimental.

Tomorrow we will cover a few solvers for cases with **hundreds** or **thousands** of atoms, in which using accelerators becomes very beneficial.

A look at the inputs

Your first encounter!

Tutorial description:

https://docs.siesta-project.org/projects/siesta/en/latest/tutorials/basic/first-encounter/index.html

Tutorial folder:

day3-Wed/01-FirstEncounter_I

What are the main ingredients?

For most basic SIESTA calculations, we need at least two inputs:

- Pseudo potential files (e.g. available in PSML format from http://www.pseudo-dojo.org, or a PSF created with ATOM).

- An fdf file with the input options.

What's in the FDF?

The fdf file contains all relevant input options for our simulation: geometry information, atomic species information, level of theory, basis set information, and a plethora of fine-tuning options.

Let's have a look at the first fdf for this tutorial…

What's in the FDF? System information

```
#General system specifications
SystemName          CH4 molecule
SystemLabel         ch4
NumberOfAtoms       5
NumberOfSpecies     2

%block ChemicalSpeciesLabel
 1  6 C    # Species index, atomic number, species label
 2  1 H    # Species index, atomic number, species label
%endblock ChemicalSpeciesLabel
```

All output filenames will begin with "ch4."

Total number of atoms in the simulation box.

Different "kinds" of atoms present.

Note that we have to types of inputs:
single variables, and blocks.

What's in the FDF? System geometry

```
#Unit cell for the calculation
LatticeConstant 15 Ang
%block LatticeVectors
  1.000 0.000 0.000
  0.000 1.000 0.000
  0.000 0.000 1.000
%endblock LatticeVectors

#Atomic coordinates
AtomicCoordinatesFormat   Ang
%block AtomicCoordinatesAndAtomicSpecies
  0.000      0.000      0.000      1
  1.219     -0.284     -0.377      2
 -0.284      1.219     -0.377      2
 -0.140     -0.140      1.219      2
 -0.833     -0.833     -0.503      2
%endblock AtomicCoordinatesAndAtomicSpecies
```

Multiplies all lattice vectors by a constant. Note the units.

The lattice vectors themselves.

Unit for the atomic coordinates block. Can also be "fractional".

Atomic coordinates and species index (1 for C, 2 for H).

What's in the FDF? Other options

```
# Basis set definition
PAO.EnergyShift 250 meV
PAO.SplitNorm    0.15
PAO.BasisSize    SZ

#Real space grid
MeshCutoff 125.0 Ry

# Convergence of SCF
MaxSCFIterations 50
DM.MixingWeight 0.4
DM.NumberPulay  2

# Type of solution
SolutionMethod diagon
```

Basis Set Options

Relates to the amount of points for grid-based operations.

Options for SCF acceleration.

Solver options.

Let's have a look at the outputs…

What are all of these files???

```
0_NORMAL_EXIT           H.ion.nc              ch4.FA
BASIS_ENTHALPY          H.ion.xml             ch4.KP
BASIS_HARRIS_ENTHALPY   H.psf                 ch4.ORB_INDX
C.gga.psf               INPUT_TMP.45433       ch4.STRUCT_OUT
C.ion                   MESSAGES              ch4.XV
C.ion.nc                OUTVARS.yml           ch4.alloc
C.ion.xml               PARALLEL_DIST         ch4.bib
C.psf                   TIMES                 ch4.contrib.EPSIMG
CLOCK                   ch4.BONDS             ch4.fdf
FORCE_STRESS            ch4.BONDS_FINAL       ch4.out
H.gga.psf               ch4.DM                ch4.times
H.ion                   ch4.EIG               fdf.20230921T093943.534.log
```

What are all of these files???

Density Matrix Restart

Forces on atoms

Coordinate Restart

```
0_NORMAL_EXIT          H.ion.nc           ch4.FA
BASIS_ENTHALPY         H.ion.xml          ch4.KP
BASIS_HARRIS_ENTHALPY  H.psf              ch4.ORB_INDX
C.gga.psf              INPUT_TMP.45433    ch4.STRUCT_OUT
C.ion                  MESSAGES           ch4.XV
C.ion.nc               OUTVARS.yml        ch4.alloc
C.ion.xml              PARALLEL_DIST      ch4.bib
C.psf                  TIMES              ch4.contrib.EPSIMG
CLOCK                  ch4.BONDS          ch4.fdf
FORCE_STRESS           ch4.BONDS_FINAL    ch4.out
H.gga.psf              ch4.DM             ch4.times
H.ion                  ch4.EIG            fdf.20230921T093943.534.log
```

Forces and Stress

KS eigenvalues

Timing information

What are all of these files???

```
0_NORMAL_EXIT          H.ion.nc           ch4.FA
BASIS_ENTHALPY         H.ion.xml          ch4.KP
BASIS_HARRIS_ENTHALPY  H.psf              ch4.ORB_INDX
C.gga.psf              INPUT_TMP.45433    ch4.STRUCT_OUT
C.ion                  MESSAGES           ch4.XV
C.ion.nc               OUTVARS.yml        ch4.alloc
C.ion.xml              PARALLEL_DIST      ch4.bib
C.psf                  TIMES              ch4.contrib.EPSIMG
CLOCK                  ch4.BONDS          ch4.fdf
FORCE_STRESS           ch4.BONDS_FINAL    ch4.out
H.gga.psf              ch4.DM             ch4.times
H.ion                  ch4.EIG            fdf.20230921T093943.534.log
```

General Output file: log, out, *you* name it

Installation and run info, Start Time

```
Architecture    : ----
Compiler version: GNU-11.3.0
Compiler flags  : -fallow-argument-mismatch;-O3 -march=native
PP flags        : ----
Libraries       : ----
Parallelisations: MPI
GEMM3M support
NetCDF support
NetCDF-4 support
Lua support

Runtime information:
* Directory : /home/fnpedron/siesta-docs/work-files/tutorials/basic/first-encounter/CH4
* Running on 4 nodes in parallel.
>> Start of run:  21-SEP-2023    9:39:43

                   ***********************
                   *  WELCOME TO SIESTA  *
                   ***********************
```

Things we have in our FDF file

```
*************************** Dump of input data file ***************************
#General system specifications
SystemName           CH4 molecule
SystemLabel          ch4
NumberOfAtoms        5
NumberOfSpecies      2
%block ChemicalSpeciesLabel
  1  6 C   # Species index, atomic number, species label
  2  1 H   # Species index, atomic number, species label
%endblock ChemicalSpeciesLabel
#Unit cell for the calculation
LatticeConstant 15 Ang
%block LatticeVectors
   1.000 0.000 0.000
   0.000 1.000 0.000
   0.000 0.000 1.000
%endblock LatticeVectors
#Atomic coordinates
AtomicCoordinatesFormat   Ang
%block AtomicCoordinatesAndAtomicSpecies
   0.000       0.000       0.000    1
   1.219      -0.284      -0.377    2
  -0.284       1.219      -0.377    2
  -0.140      -0.140       1.219    2
  -0.833      -0.833      -0.503    2
%endblock AtomicCoordinatesAndAtomicSpecies
# Basis set definition
PAO.EnergyShift 250 meV
PAO.SplitNorm   0.15
PAO.BasisSize   SZ
#Real space grid
MeshCutoff 125.0 Ry
# Convergence of SCF
MaxSCFIterations 50
DM.MixingWeight 0.4
DM.NumberPulay  2
# Type of solution
SolutionMethod diagon
*************************** End of input data file ***************************
```

```
initatom: Reading input for the pseudopotentials and atomic orbitals ----------
Species number:   1 Atomic number:    6 Label: C
Species number:   2 Atomic number:    1 Label: H


---- Processing specs for species: C
Ground state valence configuration:   2s02  2p02
Reading pseudopotential information in formatted form from:
  C.psf

---- Processing specs for species: H
Ground state valence configuration:   1s01
Reading pseudopotential information in formatted form from:
  H.psf

---- Pseudopotential check for C

Pseudized shells:
2s( 2.00) rc: 1.29
2p( 2.00) rc: 1.29
3d( 0.00) rc: 1.29
4f( 0.00) rc: 1.29
Valence configuration for ps generation: (assumed as above)

---- Pseudopotential check for H

Pseudized shells:
1s( 1.00) rc: 1.25
2p( 0.00) rc: 1.25
3d( 0.00) rc: 1.25
4f( 0.00) rc: 1.25
Valence configuration for ps generation: (assumed as above)
For C, standard SIESTA heuristics set lmxkb to 2
 (one more than the basis l, including polarization orbitals).
Use PS.lmax or PS.KBprojectors blocks to override.
For H, standard SIESTA heuristics set lmxkb to 1
 (one more than the basis l, including polarization orbitals).
Use PS.lmax or PS.KBprojectors blocks to override.
```

Species and pseudopotential information

```
atom: -------------------------------------------------------------

atom: SANKEY-TYPE ORBITALS:

SPLIT: Orbitals with angular momentum L= 0

SPLIT: Basis orbitals for state 2s

SPLIT: PAO cut-off radius determined from an
SPLIT: energy shift=  0.018374 Ry

   izeta = 1
                 lambda =     1.000000
                     rc =     4.191849
                 energy =    -0.983900
                kinetic =     0.912099
    potential(screened) =    -1.895999
        potential(ionic) =    -5.500930

SPLIT: Orbitals with angular momentum L= 1

SPLIT: Basis orbitals for state 2p

SPLIT: PAO cut-off radius determined from an
SPLIT: energy shift=  0.018374 Ry

   izeta = 1
                 lambda =     1.000000
                     rc =     4.993604
                 energy =    -0.381878
                kinetic =     2.577411
    potential(screened) =    -2.959289
        potential(ionic) =    -6.460511
atom: Total number of Sankey-type orbitals:  4

atm_pop: Valence configuration (for local Pseudopot. screening):
 2s( 2.00)
 2p( 2.00)
Vna: chval, zval:     4.00000   4.00000

Vna:  Cut-off radius for the neutral-atom potential:   4.993604
```

Basis set generation (next session!)

Coordinates and selected options


```
coor:     Atomic-coordinates input format  =     Cartesian coordinates
coor:                                            (in Angstroms)

siesta: Atomic coordinates (Bohr) and species
siesta:      0.00000   0.00000   0.00000  1        1
siesta:      2.30358  -0.53668  -0.71243  2        2
siesta:     -0.53668   2.30358  -0.71243  2        3
siesta:     -0.26456  -0.26456   2.30358  2        4
siesta:     -1.57414  -1.57414  -0.95053  2        5

siesta: System type = molecule

initatomlists: Number of atoms, orbitals, and projectors:      5       8     25

siesta: ******************** Simulation parameters ****************************
siesta:
siesta: The following are some of the parameters of the simulation.
siesta: A complete list of the parameters used, including default values,
siesta: can be found in file out.fdf
siesta:
redata: Spin configuration                          = none
redata: Number of spin components                   = 1
redata: Time-Reversal Symmetry                      = T
redata: Spin spiral                                 = F
redata: Long output                                 =   F
redata: Number of Atomic Species                    =        2
redata: Charge density info will appear in .RHO file
redata: Write Mulliken Pop.                         = NO
redata: Matel table size (NRTAB)                    =    1024
redata: Mesh Cutoff                                 =   125.0000 Ry
redata: Net charge of the system                    =     0.0000 |e|
redata: Min. number of SCF Iter                     =        0
redata: Max. number of SCF Iter                     =       50
redata: SCF convergence failure will abort job
redata: SCF mix quantity                            = Hamiltonian
redata: Mix DM or H after convergence               =   F
redata: Recompute H after scf cycle                 =   F
redata: Mix DM in first SCF step                    =   T
redata: Write Pulay info on disk                    =   F
```

Outputs

```
                    ===================================
                         Single-point calculation
                    ===================================

outcell: Unit cell vectors (Ang):
        15.000000     0.000000     0.000000
         0.000000    15.000000     0.000000
         0.000000     0.000000    15.000000

outcell: Cell vector modules (Ang)   :    15.000000    15.000000    15.000000
outcell: Cell angles (23,13,12) (deg):     90.0000      90.0000      90.0000
outcell: Cell volume (Ang**3)        :    3375.0000
<dSpData1D:S at geom step 0
  <sparsity:sparsity for geom step 0
    nrows_g=8 nrows=2 sparsity=.2500 nnzs=16, refcount: 7>
  <dData1D:(new from dSpData1D) n=16, refcount: 1>
refcount: 1>
new_DM -- step:        1
Initializing Density Matrix...
DM filled with atomic data:
<dSpData2D:DM initialized from atoms
  <sparsity:sparsity for geom step 0
    nrows_g=8 nrows=2 sparsity=.2500 nnzs=16, refcount: 8>
  <dData2D:DM n=16 m=1, refcount: 1>
refcount: 1>
No. of atoms with KB's overlaping orbs in proc 0. Max # of overlaps:        5        8

InitMesh: MESH = 108 x 108 x 108 = 1259712
InitMesh: Mesh cutoff (required, used) =    125.000   143.274 Ry
New grid distribution [1]: sub = 2
New grid distribution [2]: sub = 2
New grid distribution [3]: sub = 2
Setting up quadratic distribution...

stepf: Fermi-Dirac step function
```

Type of run, cell information.

Sparsity information.

Mesh information (later!)

Outputs

```
siesta: Program's energy decomposition (eV):
siesta: Ebs     =       -86.773862
siesta: Eions   =       383.324493
siesta: Ena     =       115.426770
siesta: Ekin    =       143.738590
siesta: Enl     =       -16.728728
siesta: Eso     =         0.000000
siesta: Edftu   =         0.000000
siesta: DEna    =         1.592579
siesta: DUscf   =         0.349516
siesta: DUext   =         0.000000
siesta: Ex      =       -64.874822
siesta: Ec      =       -10.703118
siesta: Exc     =       -75.577940
siesta: EbV     =         0.000000
siesta: eta*DQ  =         0.000000
siesta: Emadel  =         0.000000
siesta: Emeta   =         0.000000
siesta: Emolmec =         0.000000
siesta: Ekinion =         0.000000
siesta: Eharris =      -223.671697
siesta: Etot    =      -214.523706
siesta: FreeEng =      -214.523706
```

Initial, non-SCF energy decomposition.

Outputs

```
        iscf      Eharris(eV)        E_KS(eV)       FreeEng(eV)      dDmax     Ef(eV) dHmax(eV)
  scf:    1       -223.671697     -214.523706     -214.523706  1.090911 -7.083002  1.436999
timer: Routine,Calls,Time,% = IterSCF        1        0.133  29.48
  scf:    2       -214.585551     -214.573147     -214.573147  0.040577 -6.647325  0.203018
  scf:    3       -214.573456     -214.573477     -214.573477  0.004139 -6.585363  0.150120
  scf:    4       -214.573442     -214.573493     -214.573493  0.002062 -6.424159  0.074339
  scf:    5       -214.573514     -214.573506     -214.573506  0.000928 -6.476298  0.003034
  scf:    6       -214.573506     -214.573506     -214.573506  0.000039 -6.474131  0.000389

SCF Convergence by DM+H criterion
max |DM_out - DM_in|           :       0.0000385344
max |H_out - H_in|      (eV) :       0.0003888059
SCF cycle converged after 6 iterations
|
Using DM_out to compute the final energy and forces
No. of atoms with KB's overlapping orbs in proc 0. Max # of overlaps:        5        8

siesta: E_KS(eV) =             -214.5735

siesta: E_KS - E_eggbox =       -214.5735

siesta: Atomic forces (eV/Ang):
-----------------------------------------
   Tot    0.000066    0.000066   -0.001085
-----------------------------------------
   Max    2.352006
   Res    1.126956    sqrt( Sum f_i^2 / 3N )
-----------------------------------------
   Max    2.352006    constrained

Stress tensor Voigt[x,y,z,yz,xz,xy] (kbar):        1.99      1.99      0.95      -0.20
(Free)E + p*V (eV/cell)      -218.0329
Target enthalpy (eV/cell)    -214.5735
```

SCF cycle information

Converged KS energy

Converged total forces and cell stress

Final energy decomposition


```
siesta: Program's energy decomposition (eV):
siesta: Ebs      =        -90.137390
siesta: Eions    =        383.324493
siesta: Ena      =        115.426770
siesta: Ekin     =        141.310823
siesta: Enl      =        -16.669337
siesta: Eso      =          0.000000
siesta: Edftu    =          0.000000
siesta: DEna     =          3.517376
siesta: DUscf    =          0.257037
siesta: DUext    =          0.000000
siesta: Ex       =        -64.416938
siesta: Ec       =        -10.674744
siesta: Exc      =        -75.091682
siesta: EbV      =          0.000000
siesta: eta*DQ   =          0.000000
siesta: Emadel   =          0.000000
siesta: Emeta    =          0.000000
siesta: Emolmec  =          0.000000
siesta: Ekinion  =          0.000000
siesta: Eharris  =       -214.573506
siesta: Etot     =       -214.573506
siesta: FreeEng  =       -214.573506

siesta: Final energy (eV):
siesta:  Band Struct. =      -90.137390
siesta:        Kinetic =      141.310823
siesta:        Hartree =      282.193258
siesta:          Edftu =        0.000000
siesta:            Eso =        0.000000
siesta:     Ext. field =        0.000000
siesta:          Exch. =      -64.416938
siesta:          Corr. =      -10.674744
siesta:      Bulk bias =        0.000000
siesta:     Exch.-corr. =     -75.091682
siesta:   Ion-electron =     -697.792327
siesta:        Ion-ion =      134.806422
siesta:        Ekinion =        0.000000
siesta: D3 dispersion =        0.000000
siesta:          Total =     -214.573506
siesta:          Fermi =       -6.474131
```

Final forces

Final stress/pressure

Electric dipole


```
siesta: Atomic forces (eV/Ang):
siesta:      1     0.152980     0.152980    -1.053682
siesta:      2    -2.352006     0.483512     0.761553
siesta:      3     0.483512    -2.352006     0.761553
siesta:      4     0.342189     0.342189    -0.971719
siesta:      5     1.373392     1.373392     0.501211
siesta: ----------------------------------------
siesta:    Tot     0.000066     0.000066    -0.001085

siesta: Stress tensor (static) (eV/Ang**3):
siesta:      0.001241    -0.000019    -0.000128
siesta:     -0.000019     0.001241    -0.000128
siesta:     -0.000128    -0.000128     0.000593

siesta: Cell volume =        3375.000000 Ang**3

siesta: Pressure (static):
siesta:                    Solid              Molecule  Units
siesta:           -0.00001116          0.00000003  Ry/Bohr**3
siesta:           -0.00102500          0.00000251  eV/Ang**3
siesta:           -1.64224685          0.00402704  kBar
(Free)E+ p_basis*V_orbitals  =         -214.071102
(Free)Eharris+ p_basis*V_orbitals  =           -214.071102

siesta: Electric dipole (a.u.)  =   -0.011992    -0.011992     0.008053
siesta: Electric dipole (Debye) =   -0.030480    -0.030480     0.020469
```

Primary bibliography, and end-of-run time


```
cite: Please see "ch4.bib" for an exhaustive BiBTeX file.
cite: Please clearly indicate Siesta version in published work:
cite: This calculation has made use of the following articles
cite: which are encouraged to be cited in a published work.
        Primary SIESTA paper
            DOI: www.doi.org/10.1088/0953-8984/14/11/302


>> End of run:  21-SEP-2023   9:39:44
Job| completed
```

Try it!

A look into the level of theory

General considerations

Before doing a production run with SIESTA:

- Choose an appropriate XC functional.
- Test the pseudopotentials (not covered here)
- Test the basis sets
- Converge the mesh cut-off for calculations
- Converge the k-point sampling

General considerations

Before doing a production run with SIESTA:

- Choose an appropriate XC functional.
- Test the pseudopotentials (not covered here)
- Test the basis sets
- Converge the mesh cut-off for calculations
- Converge the k-point sampling

This tutorial

Tutorial description:

https://docs.siesta-project.org/projects/siesta/en/latest/tutorials/basic/first-encounter-theorylevel/index.html

Tutorial folder:

day3-Wed/02-FirstEncounter_II

Choosing an XC functional

Which functional to choose will depend on what you want to do and the system you are running. **Read the literature!**

Telling SIESTA what functional to use is easy enough:

```
XC.functional GGA
XC.authors     PBE
```

Choosing an XC functional

XC.functional and XC.authors must be consistent!

```
XC.functional GGA
XC.authors     PBE
```

You should also choose appropriate pseudopotentials.

Key concepts on basis sets

SIESTA can automatically generate basis sets, or you can provide whatever radial function you want.

Basis functions become strictly zero beyond a certain radius, *rcut*.

For multiple-z basis, the second-z orbital is equal to the first-z orbital beyond a matching radius *rmatch*.
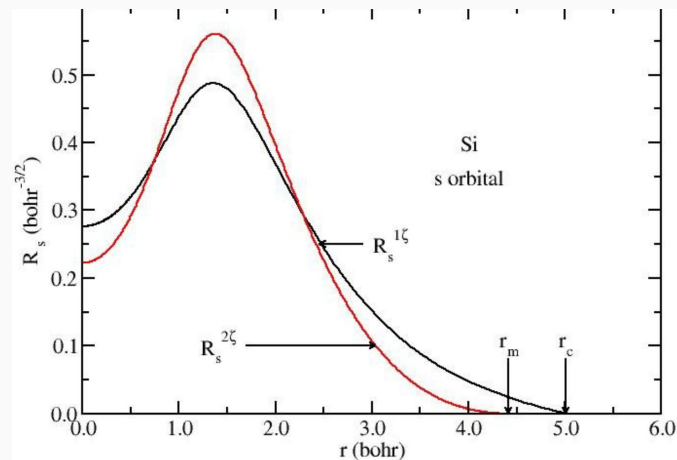
Global basis set options

Cardinality: amount of basis functions per atom (SZ, SZP, DZ, DZP, TZP).

```
PAO.BasisSize    DZP
```

Energy-shift: controls the cut-off radii of all atoms in a cohesive way.

```
PAO.EnergyShift 0.01 Ry
```

Global basis set options

Split Norm: controls the matching radii for all multiple-zeta orbitals.

```
PAO.Splitnorm    0.15
```

Soft confinement: Use a soft confinement potential when creating the basis orbitals.

```
PAO.SoftDefault      T
```

First tests with methane (geometry optimization)

For XC functional options, check CH4-XC-Functional

For the basis sets, go into CH4-Basis

Take note of how the **total energy (from output)**, **bond lengths (ch4.BONDS file)**, and **total time (from ch4.times)** change in this cases:

1) When changing the basis set between **SZ, SZP, and DZP**. Use an energy shift of **100 meV**.
2) For **DZP,** changing the energy shift between **10 meV, 50 meV, and 100 meV**.

Budgeting

Both increasing the cardinality (SZ -> SZP -> DZP -> TZP) and reducing the energy-shift increase the quality of results and the computational costs.

Costs are affected differently though:

- **Cardinality** increases the cost of **diagonalization** ($\sim N^3$)
- **Energy shift**/Cut-off radius increase the cost of **grid operations** ($\sim N$).

For s**mall systems** (a few N), **grid operations** are the dominant part of the calculation. Meanwhile, for **large systems** (large N), **diagonalization** becomes dominant.

This means that for very large systems (300-800 atoms), increasing the cut-off radii does not greatly increase computational costs.

Optimizing the basis set?

SIESTA can benefit greatly from manually optimizing our basis set a bit further. This increases the **quality** of our result **without really sacrificing computational cost**.

The only real issue here is the **time spent optimizing** the basis set.