Some Reminders

12/11/2024 - Federico Pedron

ICN2 Institut Català de Nanociència i Nanotecnologia
Barcelona Supercomputing Center Centro Nacional de Supercomputación
CSIC Consejo Superior de Investigaciones Científicas
MAX Driving the Exascale Transition
Ministerio de Ciencia, Innovación y Universidades
Financiado por la Unión Europea NextGenerationEU
Plan de Recuperación, Transformación y Resiliencia
Agencia Estatal de Investigación

Reminders

Fill the survey!!!

https://siesta-project.org/siesta/events/SIESTA_School-2024/Sessions.html

https://siesta-project.org/siesta/events/SIESTA_School-2024/MN5.html

https://docs.siesta-project.org/projects/siesta/en/school-2024/tutorials/index.html

Reminders

Before doing a calculation with SIESTA, there are a few things we should do:

- Verify the pseudopotentials
- Test the basis sets (today!)
- Converge the mesh cut-off for calculations (today!)
- Converge the k-point sampling (today!)

Basis set optimization

12/11/2024 - Federico Pedron

ICN2 10 YEARS
Institut Català
de Nanociència
i Nanotecnologia

BSC
Barcelona
Supercomputing
Center
Centro Nacional de Supercomputación

CSIC
Consejo Superior de Investigaciones Científicas

MAX
DRIVING
THE EXASCALE
TRANSITION

MINISTERIO
DE CIENCIA, INNOVACIÓN
Y UNIVERSIDADES

Financiado por
la Unión Europea
NextGenerationEU

Plan de Recuperación,
Transformación y
Resiliencia

AGENCIA
ESTATAL DE
INVESTIGACIÓN

Key concepts

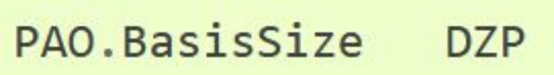
siesta™ can automatically generate basis sets, or you can provide whatever radial function you want.

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*.

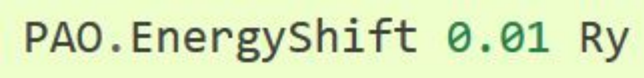
Global basis set options

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



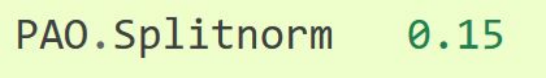
```
PAO.BasisSize   DZP
```

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



```
PAO.EnergyShift 0.01 Ry
```

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



```
PAO.Splitnorm   0.15
```

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** (~N^3)
- **Energy shift**/Cut-off radius increase the cost of **grid operations** (~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.

Without increasing the cardinality, we want to get the best orbitals we can!

Optimizing a Basis Set

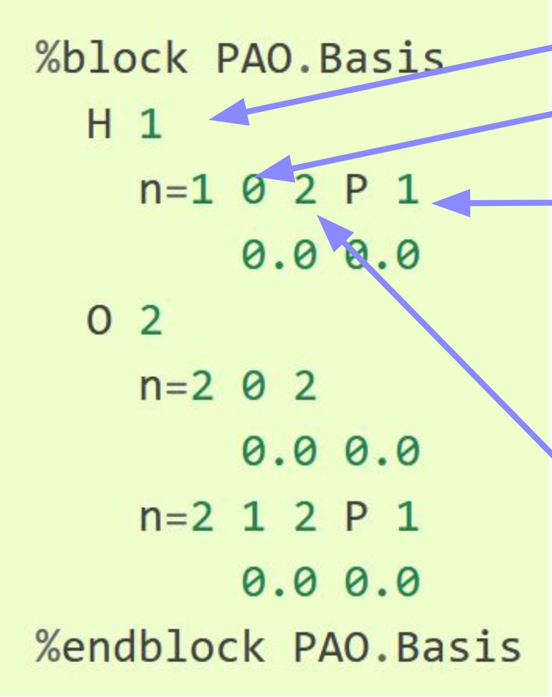
Exploring the PAO.Basis block

We have each species and all orbitals with different (n,l) separated.

For water:

H -> 1s

O -> 2s, 2p



```
%block PAO.Basis
  H 1
    n=1 0 2 P 1
        0.0 0.0
  O 2
    n=2 0 2
        0.0 0.0
    n=2 1 2 P 1
        0.0 0.0
%endblock PAO.Basis
```

Number of different orbital (n,l)

n and l: n=1, l=0, which means 1s

P 1 = add one set of polarization orbitals. DZP basis.

Number of zetas: DZ basis?

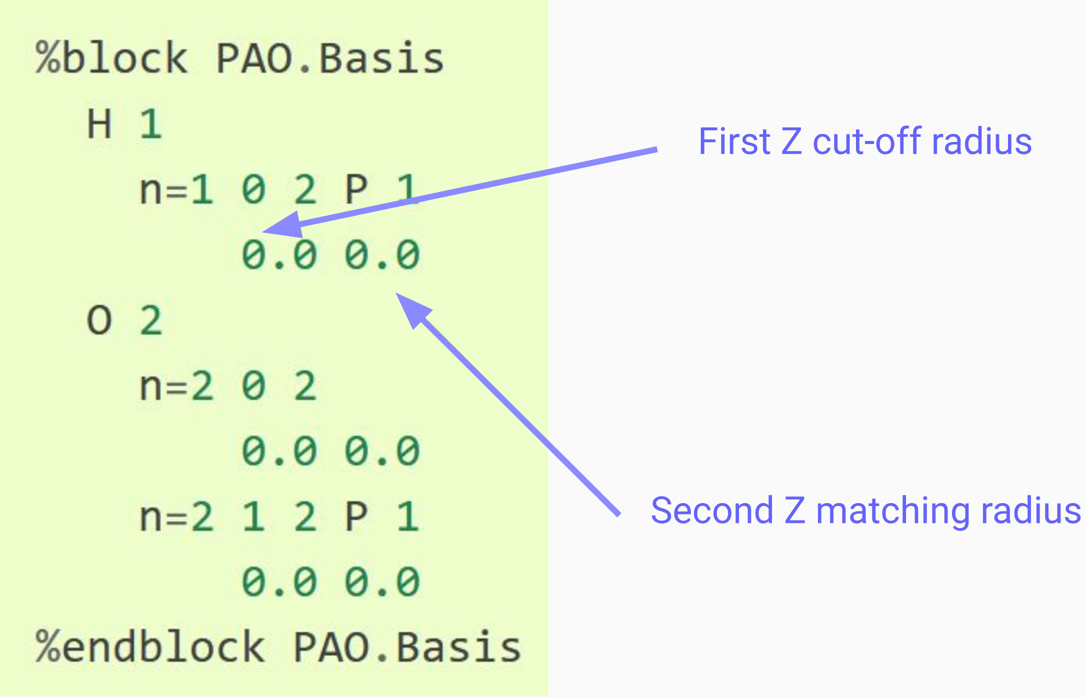
Exploring the PAO.Basis block

We have each species and all orbitals with different (n,l) separated.

For water:

H -> 1s

O -> 2s, 2p



```
%block PAO.Basis
  H 1
    n=1 0 2 P 1
        0.0 0.0
  O 2
    n=2 0 2
        0.0 0.0
    n=2 1 2 P 1
        0.0 0.0
%endblock PAO.Basis
```

First Z cut-off radius

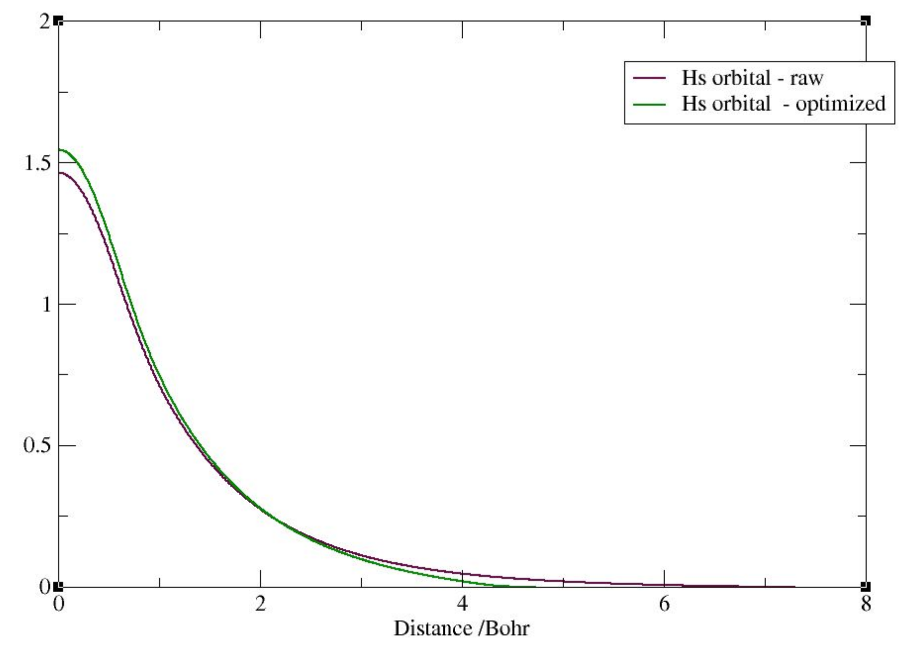
Second Z matching radius

Basis Enthalpy

- We want to get a good energy for a set of orbitals.
- We don’t want those orbitals to get needlessly large.

$$\text{Basis Enthalpy} = E_{\text{total}} + \text{“}\mathbf{P}_{\mathbf{basis}} \cdot V_{\text{orbitals}}\text{”}$$

Basis Enthalpy



Basis Enthalpy

- Not a real physical magnitude, we choose it as a input value.

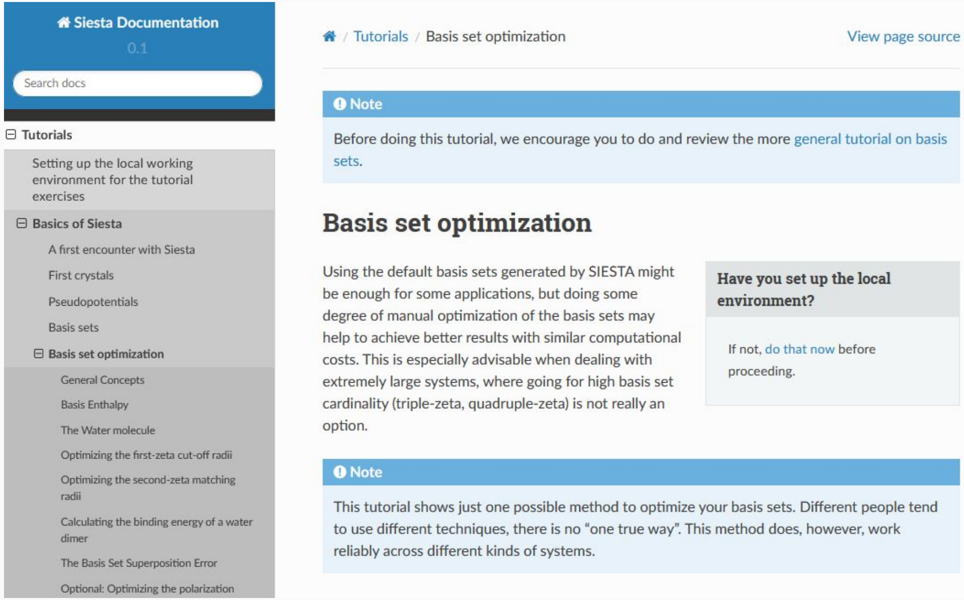
```
BasisPressure  0.2 GPa
```

- The **0.2 GPa** default works well for most cases, but for first- and second-row elements, it might result in very short orbitals. Use **0.02 GPa** instead.

Getting practice…

In the day2, go to **01-BasisSet-Optimization**.

Get to the **Basis set optimization** tutorial, and follow the first two practical sections: **Optimizing the First-Zeta cutoff radii**, and **Optimizing the Second-Zeta matching radii**.



Note

Before doing this tutorial, we encourage you to do and review the more general tutorial on basis sets.

Basis set optimization

Using the default basis sets generated by SIESTA might be enough for some applications, but doing some degree of manual optimization of the basis sets may help to achieve better results with similar computational costs. This is especially advisable when dealing with extremely large systems, where going for high basis set cardinality (triple-zeta, quadruple-zeta) is not really an option.

Have you set up the local environment?

If not, do that now before proceeding.

Note

This tutorial shows just one possible method to optimize your basis sets. Different people tend to use different techniques, there is no "one true way". This method does, however, work reliably across different kinds of systems.

Testing the Basis Set

Testing the optimized basis

How do we know if we effectively have a better basis set than the default?

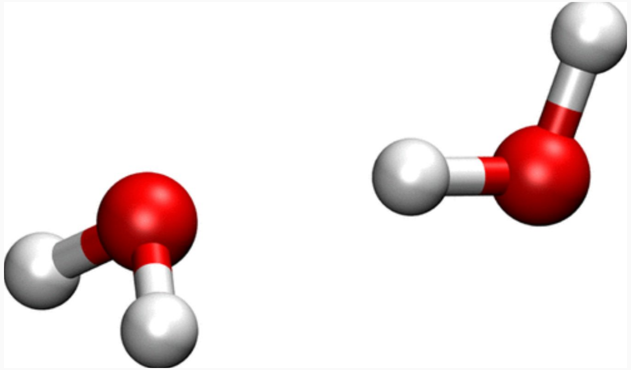
At least, three things are important to check:

- Costs
- Quality
- **Transferability**

We need to test this in a **slightly different system**!

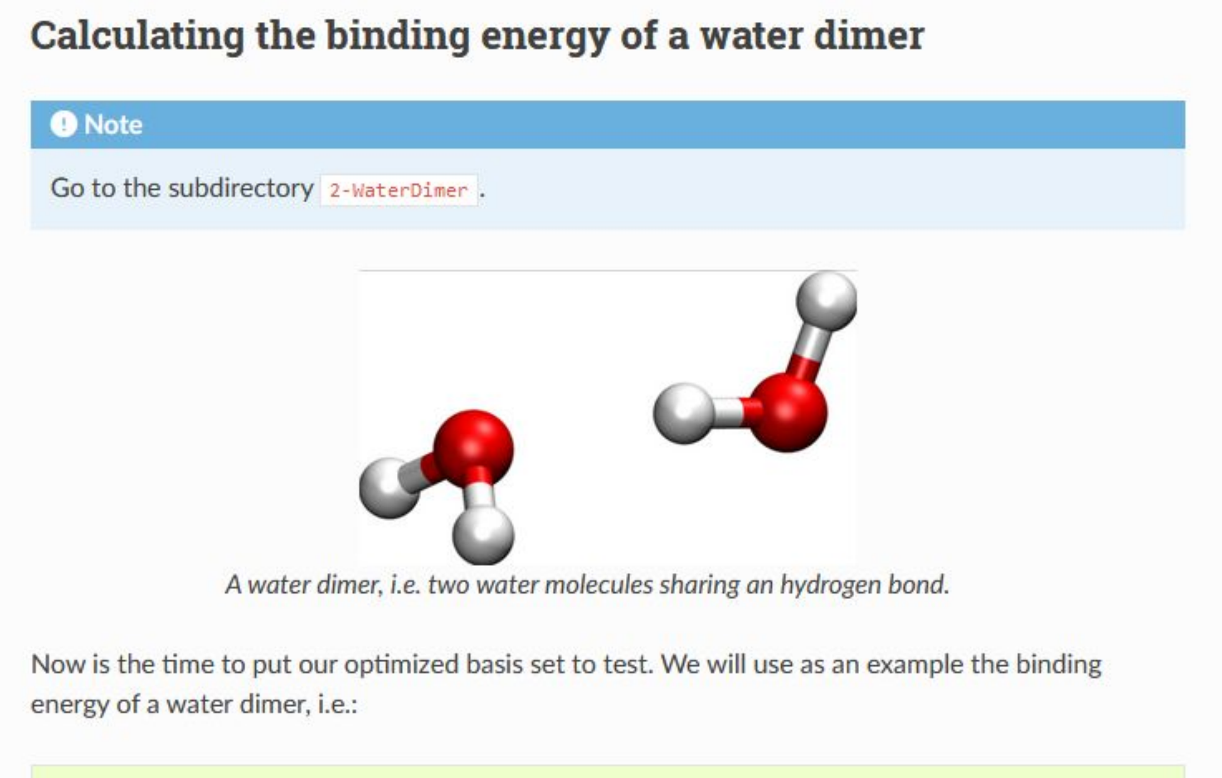
Binding energy of a water dimer

$$E_{binding} = E_{dimer} - 2.E_{monomer}$$



Getting practice…

Get to the **Basis set optimization** tutorial, and follow the third practical section: **Calculating the energy of a water dimer**.



Calculating the binding energy of a water dimer

Note

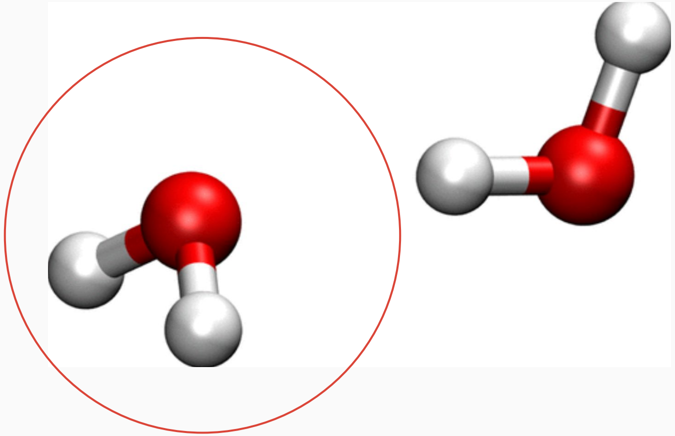
Go to the subdirectory `2-WaterDimer`.

A water dimer, i.e. two water molecules sharing an hydrogen bond.

Now is the time to put our optimized basis set to test. We will use as an example the binding energy of a water dimer, i.e.:

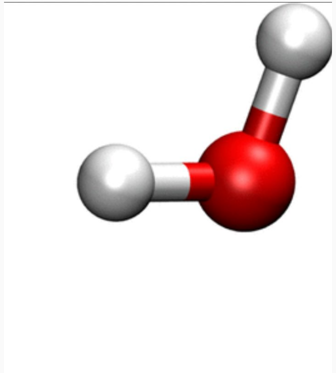
Basis Set Superposition Error

Basis Set Superposition Error



Basis Set Superposition Error

POOF!



We lost the basis functions for the second molecule! What if they are important?

Basis Set Superposition Error

Who you gonna call?

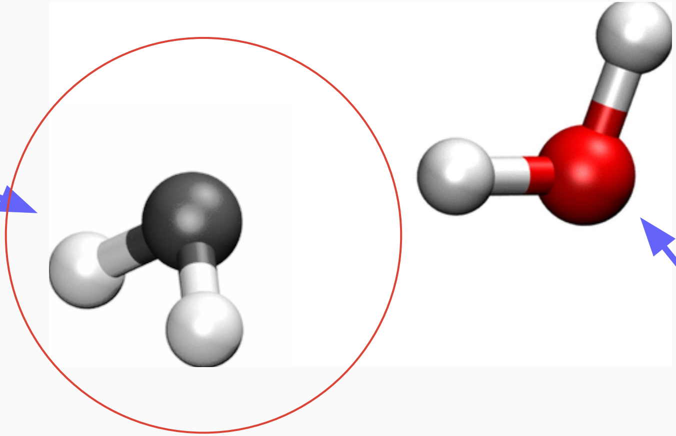
Ghost atoms!

Ghost atoms

We add the basis functions that would belong to an atom, as if the atom were there.

We **do not add electrons or nuclei** to the calculation!

Ghost atoms

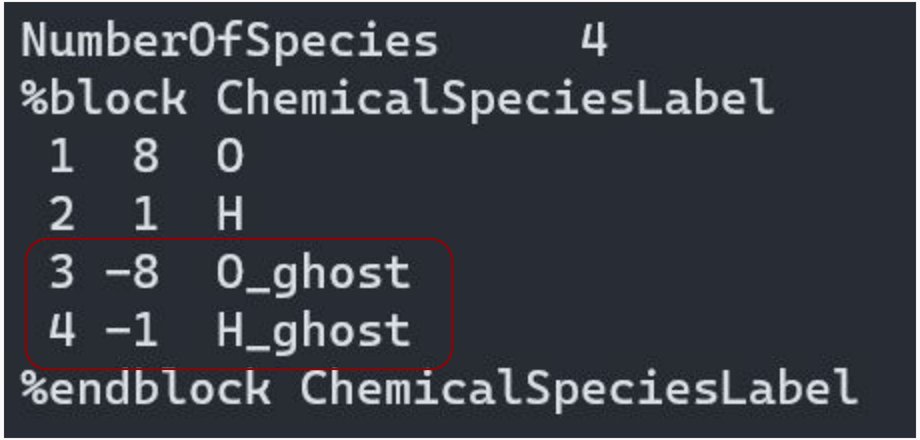


Ghost

Afraid of no ghosts

Ghost atoms

To add ghost atoms, we just create a new species with **negative atomic number**.

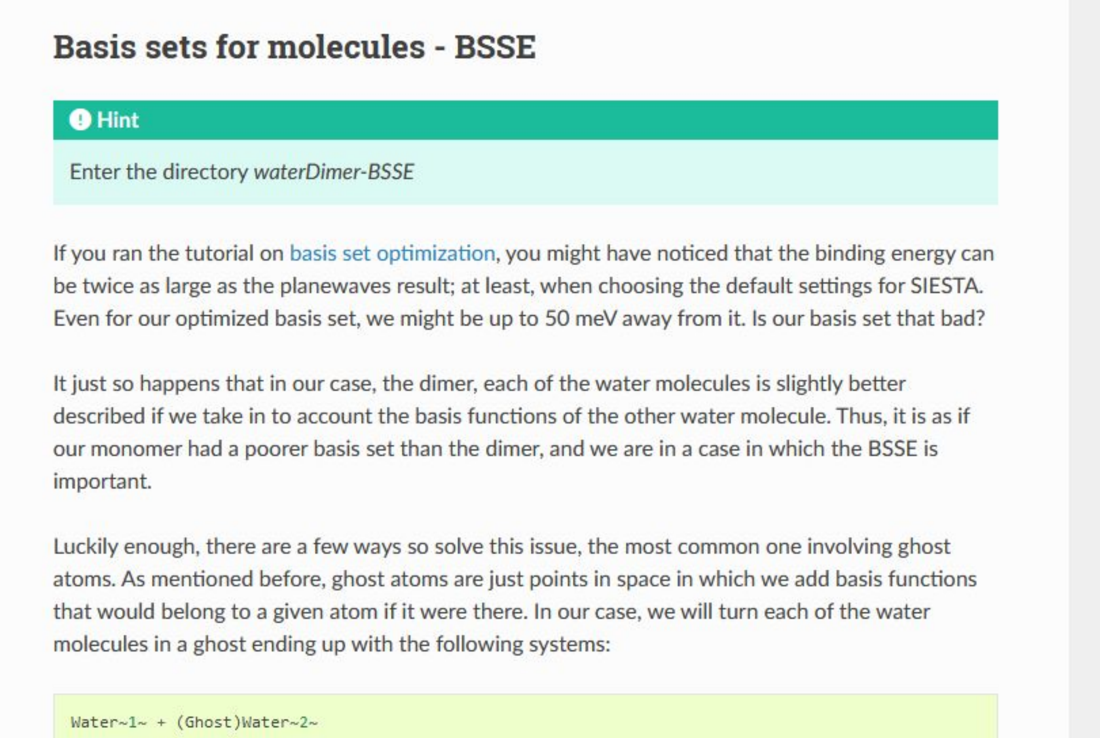


```
NumberOfSpecies     4
%block ChemicalSpeciesLabel
 1  8  O
 2  1  H
 3 -8  O_ghost
 4 -1  H_ghost
%endblock ChemicalSpeciesLabel
```

Yes, this means **we have to duplicate the pseudopotential files and add extra terms to the PAO.Basis block**.

Back to practice

Go now to the **Basis Sets - Tips and Tricks** tutorial, and follow the first practical section: **Basis sets for molecules - BSSE**.



Basis sets for molecules - BSSE

Hint

Enter the directory *waterDimer-BSSE*

If you ran the tutorial on basis set optimization, you might have noticed that the binding energy can be twice as large as the planewaves result; at least, when choosing the default settings for SIESTA. Even for our optimized basis set, we might be up to 50 meV away from it. Is our basis set that bad?

It just so happens that in our case, the dimer, each of the water molecules is slightly better described if we take in to account the basis functions of the other water molecule. Thus, it is as if our monomer had a poorer basis set than the dimer, and we are in a case in which the BSSE is important.

Luckily enough, there are a few ways so solve this issue, the most common one involving ghost atoms. As mentioned before, ghost atoms are just points in space in which we add basis functions that would belong to a given atom if it were there. In our case, we will turn each of the water molecules in a ghost ending up with the following systems:

```
Water~1~ + (Ghost)Water~2~
```

Compared to just picking a low Energy Shift, how does the optimized basis set perform in terms of..?

- Overall result ($E_{binding}$)
- Time
- Basis set superposition error