MAX DRIVING THE EXASCALE TRANSITION

siesta

ICN2 Institut Català de Nanociència i Nanotecnologia

EXCELENCIA SEVERO OCHOA

Convergence Issues

Roberto Robles

ICN2, Barcelona, Spain

Thanks to: D. Sánchez-Portal, E. Anglada, J.M. Soler, A. García, E. Artacho

You have to do TESTS!

Density functional theory

Different basis sets

Richard Martin's classification

1) **Plane waves and grids:**

Pros: Expansion in plane waves is simple and easy to implement.
Systematic way of improving the calculation

Cons: Large number of plane waves → memory requirements
Vacuum as expensive as matter

Codes: Quantum Espresso, VASP, Abinit, CASTEP...

2) **Localized atomic-like orbitals:** *gaussians, numerical atomic-like orbitals...*

Pros: Very efficient, no cost for vacuum

Cons: Harder to reach basis convergence, greater care in the election of the basis

Codes: **SIESTA**, Gaussian, CRYSTAL, OpenMX...

3) **Atomic sphere methods:** *FP-LAPW, LMTO, KKR...*

Atomic-like features near the nucleus, smooth functions between atoms

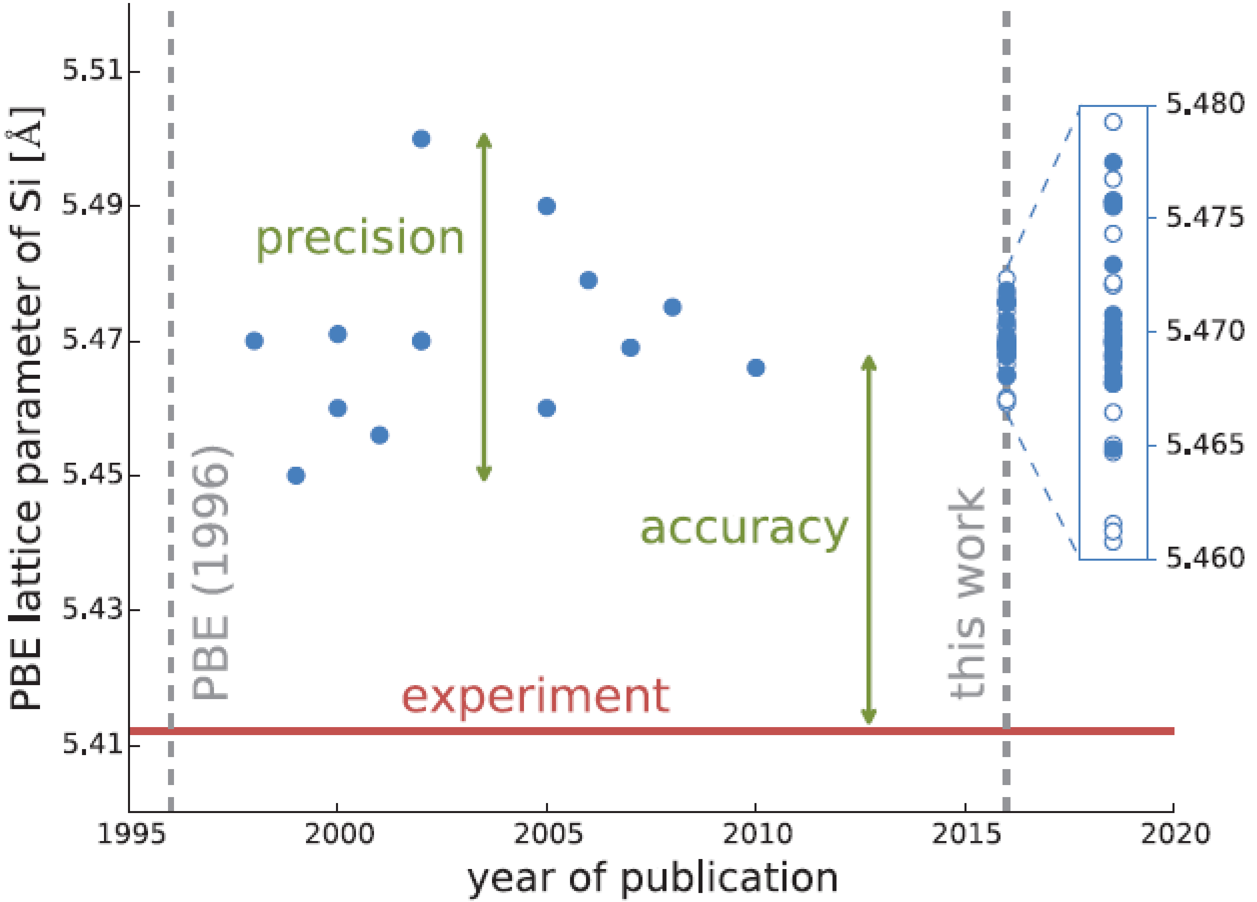
Pros: Very precise, best of both worlds

Cons: Computationally expensive, difficult to implement

Codes: FLEUR, Wien2k, RSPt, EXCITING...

If applied carefully, all of them agree after convergence

Convergence: precision vs accuracy



→ Ideal DFT-PBE value

Fig. 1. Historical evolution of the predicted equilibrium lattice parameter for silicon. All data points represent calculations within the DFT-PBE framework. Values from literature (data points before 2016) (*15*, *16*, *18*, *56*–*62*, *63*–*65*) are compared with (i) predictions from the different codes used in this study (2016 data points, magnified in the inset; open circles indicate data produced by older methods or calculations with lower numerical settings) and (ii) the experimental value, extrapolated to 0 K and corrected for zero-point effects (red line) (*26*). The concepts of precision and accuracy are illustrated graphically.

Kurt Lejaeghere et al., *"Reproducibility in density functional theory calculations of solids". Science* **351,** aad3000 (2016)

Precision and accuracy of the calculation

- DFT
- XC functional

Accuracy

- Pseudopotentials
- Basis set

- Reciprocal space grid: K-sampling
- Real-space grid
- Electronic temperature
- ...

Precision

Basic strategy

Typical steps in a research project:

1) Exploratory and feasibility tests

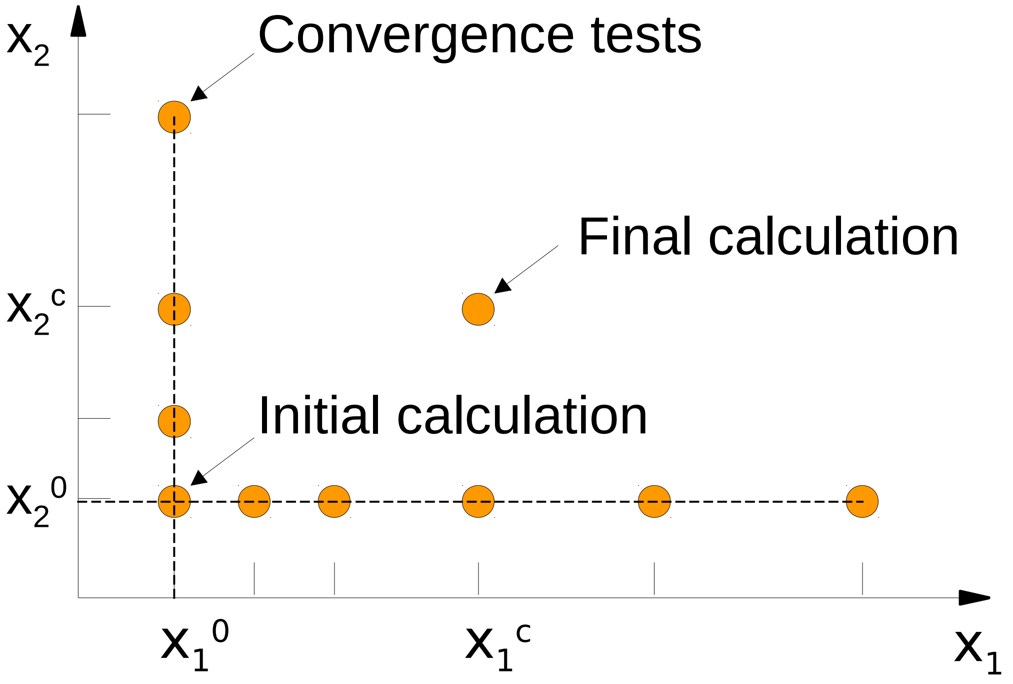
2) Convergence tests

3) Final converged calculations

It is impossible to obtain fully converged calculations without convergence tests

Convergence tests

- Choose a relevant magnitude *A* of the problem
  (e.g. lattice constant, binding energy, magnetic moment...)
- Choose a set of parameters x_i
  (e.g. size of your basis set, number of k-points, grid mesh cutoff...)



Convergence tests

Final calculation

Initial calculation

Goal:

$$\frac{\delta A}{\delta x_i} < \text{our tolerance}$$

Monitor:
- Convergence
- CPU time
- Memory

Convergence tests

Practical hints

- Objective: qualitative or quantitative results?
- Do not try a converged calculation from the start
- Start with minimum values of all x_i
- Do not assume convergence for any x_i
- Choose a simpler reference system for some tests
- Refrain from stopping tests when results are “good”

Parameter list

- **XC functional:** LDA, GGA, vdW, LDA+U...
- **Pseudopotential:**
  - Method of generation
  - Number of valence states
  - Number of angular momenta
  - Core matching radii
  - Nonlinear core corrections
- **Basis sets:**
  - Shape
  - Number of functions: highest angular momentum, number of zetas
  - Range
- K-point sampling
- Real-space grid: mesh cutoff
- Self-consistent cycle tolerance
- Supercell size (system and vacuum size)
- Electronic temperature
- Geometry relaxation tolerance

Parameter interactions

$$\frac{\delta^2 A}{\delta x_i \delta x_j} \neq 0$$

Number of k-points:

- Supercell size
- Geometry
- Magnetic moments

Mesh cutoff:

- Pseudopotential
- Nonlinear core corrections
- Basis set
- GGA

Insight into relevant parameters

- **Pseudopotential**
- Basis set
- K-point sampling
- Real-space grid
- Self-consistent cycle

Pseudopotentials

SIESTA

- Norm-conserving pseudopotentials:
  - Typically Troullier-Martins
  - Soon ONCV → Better transferability [D.R. Hamann, *Phys Rev B* **88**, 085117 (2013)] https://code.launchpad.net/~siesta-pseudos-bases/siesta/trunk-psml (advanced users)

- The choice of the pseudopotential is critical to obtain reliable results
  - Number of valence states
  - Which atomic reference configuration
  - Non linear core corrections?
  - Core radii

- Always generate pseudopotentials with great care

- Always test the pseudopotential in a well-known situation as close as possible to the systems you will be considering

Insight into relevant parameters

- Pseudopotential
- **Basis set**
- K-point sampling
- Real-space grid
- Self-consistent cycle

Basis sets

SIESTA

The only requirements:

1. $\phi_{Ilmn}(\vec{r}) = R_{Iln}(|\vec{r_I}|)Y_{lm}(\hat{r_I})$

2. Finite support (zero beyond a given cutoff radius, r_c)

They can be:

- As many as you want (both l-channels and z’s)
- Of any (radial) shape
- Of any cutoff radius
- Centered anywhere (not necessarily on atoms)

Basis sets

Convergence of the basis set size

- **Single-ζ (minimal or SZ)**

One **single radial function** per angular momentum shell occupied in the free atom

Improving the quality:

- **Radial flexibility:**

  Add more than one radial function within the same angular momentum than SZ

  → Multiple-ζ

  + Diffuse orbitals, floating orbitals...

- **Angular flexibility:**

  Add shells of different atomic symmetry (different l)

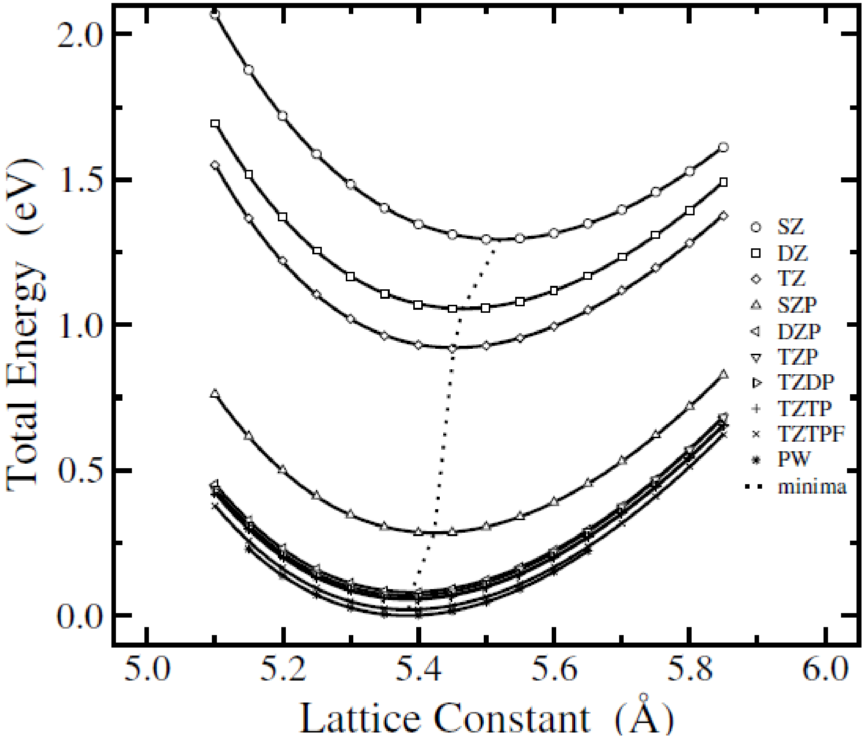
  → Polarization

Basis sets

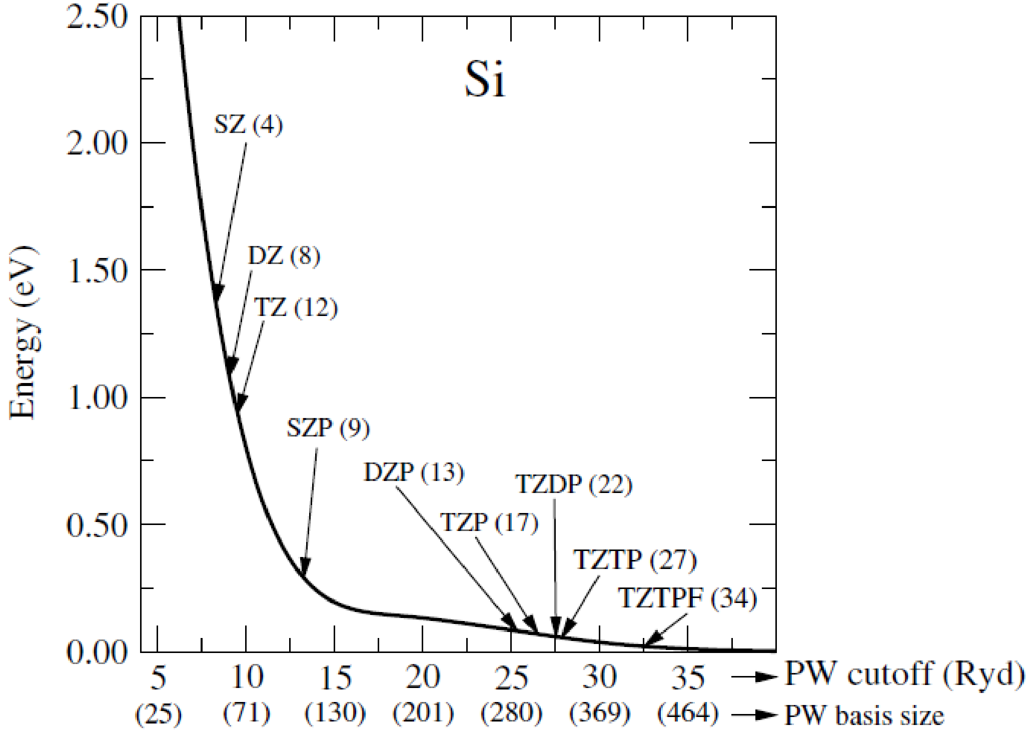
Convergence of the basis set size

Bulk Si

Cohesion curves



PW vs NAO convergence



J.M. Soler et al., *"The SIESTA method for ab-initio order-N materials simulation". J. Phys.: Condens. Matter.* **14,** 2745 (2002)

Basis sets

SZ basis

Bad results... but not as bad as you might expect:

- Energetics changes considerably; however, energy differences might be reasonable enough
- Charge transfer and other basic chemistry is usually OK
  (at least in simple systems)
- If the geometry is set to the experiment, we typically have a nice band structure for occupied and lowest unoccupied bands

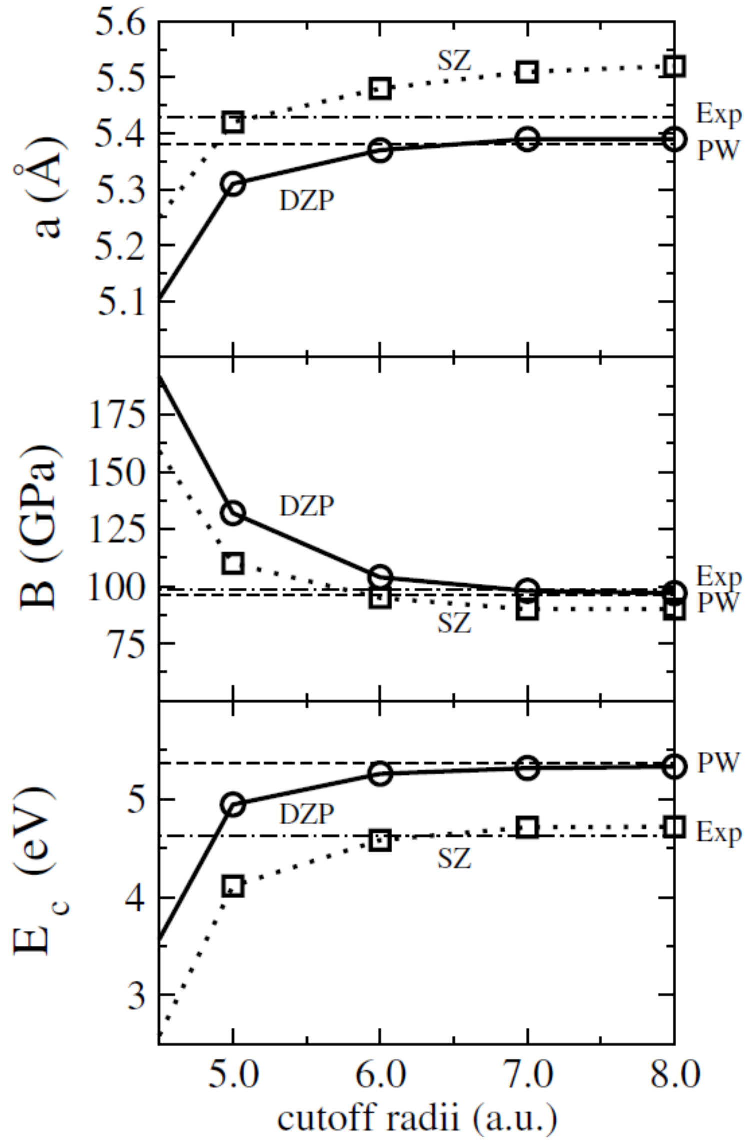
When can SZ basis sets be used:

- Long molecular dynamics simulations
  (once we are sure that energetics is reasonable)
- Exploring very large systems and/or systems with many degrees of freedom (complicated energy landscape)

Basis sets

Convergence with the range (r_c)

Bulk Si
equal s,p
orbital radii



Remarks:

- For incomplete bases (SZ), increasing radii may not produce better properties
- Affects total energies, but energy differences converge better
- More relevant for surfaces, small molecules, adsorbates
- Increasing the range makes the calculation more expensive!

J.M. Soler et al., *"The SIESTA method for ab-initio order-N materials simulation". J. Phys.: Condens. Matter.* **14,** 2745 (2002)

Basis sets

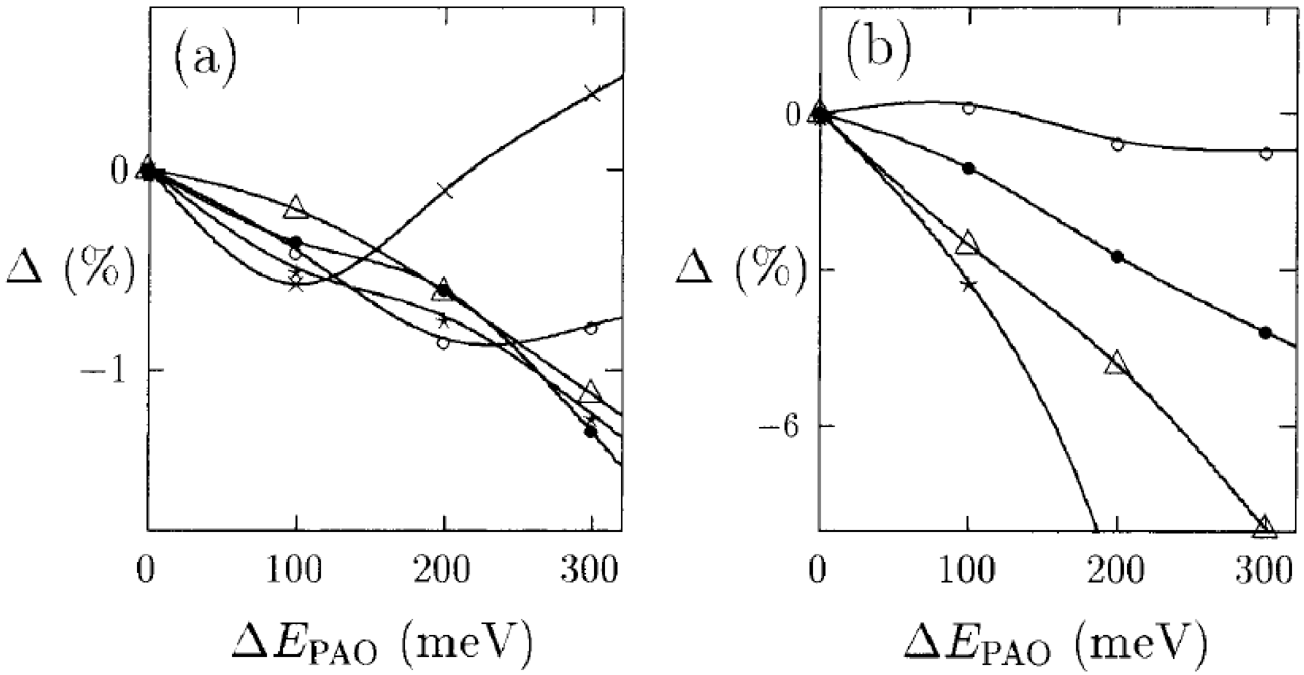
Range: energy shift

A single parameter for all cutoff radii → ΔE_{PAO}

Convergence vs Energy shift of

Bond lengths

Bond energies



Reasonable values
for practical calculations
ΔE_{PAO}~50-200 meV

E. Artacho et al., *Phys. Stat. Solidi (b)* **215,** 809 (1999)

Basis sets

Procedure

- SZ: (energy shift)
  Semiquantitative results and general trends

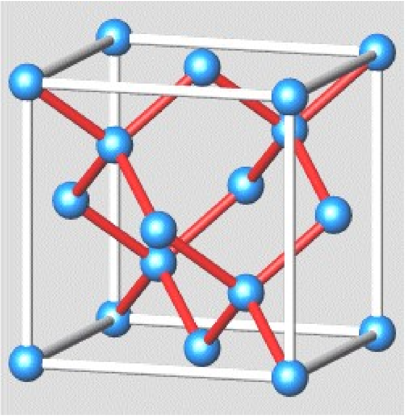
- DZP: automatically generated (energy shift, split valence)
  Good quality for most systems
  Good balance: converged results *vs* computational cost
  Default, but tests are always needed

- Variational optimization:
  Diffuse orbitals
  Floating orbitals
  → Example: noble metal surfaces (S.García-Gil, *Phys. Rev B* **79**, 075441 (2009))

Insight into relevant parameters

- Pseudopotential
- Basis set
- **K-point sampling**
- Real-space grid
- Self-consistent cycle

K-point sampling



$$V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R})$$

Bloch Theorem: The eigenstates of the one-electron Hamiltonian in a periodic potential can be chosen to have the form of a plane wave times a function with the periodicity of the Bravais lattice.

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}) \qquad u_{n,\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{n,\mathbf{k}}(\mathbf{r})$$

Periodicity in reciprocal space
$\mathbf{G} \equiv$ Reciprocal lattice vector

$$\begin{cases} \psi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}) = \psi_{n,\mathbf{k}}(\mathbf{r}) \\ \varepsilon_{n,\mathbf{k}+\mathbf{G}} = \varepsilon_{n,\mathbf{k}} \end{cases}$$

Instead of computing an infinite number of electronic wave functions → Finite number of wave functions at an infinite number of k-points in the 1BZ

K-point sampling

Many quantities require integration of Bloch functions over the first Brillouin zone (BZ):

$$n(\mathbf{r}) = \sum_i^{occ} \int_{BZ} |\psi_{i\mathbf{k}}(\mathbf{r})|^2 d\mathbf{k} \qquad\qquad E_{BS} = \sum_i \int_{BZ} n_i(\mathbf{k})\varepsilon_i(\mathbf{k}) d\mathbf{k}$$

In principle: we should know the eigenvalues and/or eigenvectors at all the k-points in the first BZ

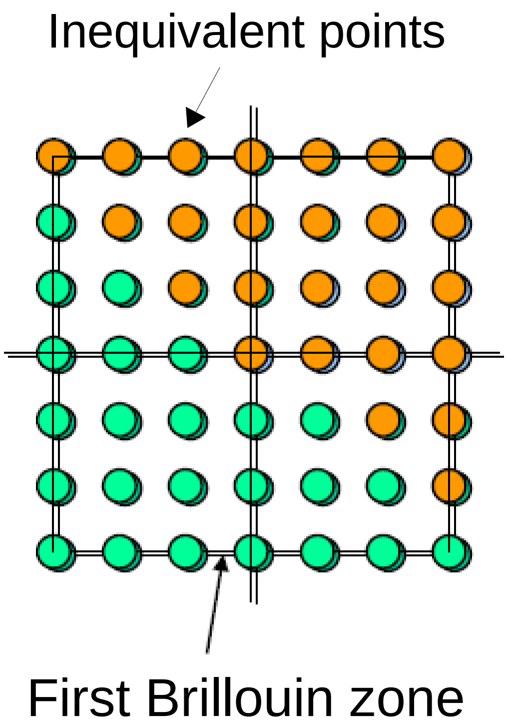
In practice: electronic wave functions at k-points that are very close together will be almost identical

→ It is possible to represent electronic wave functions over a region of reciprocal space by the wave function at a single k-point

→ Approximated by sums over selected k-points

$$\int d\mathbf{k} \longrightarrow \sum_{\mathbf{k}} \Delta\mathbf{k}$$

K-point sampling

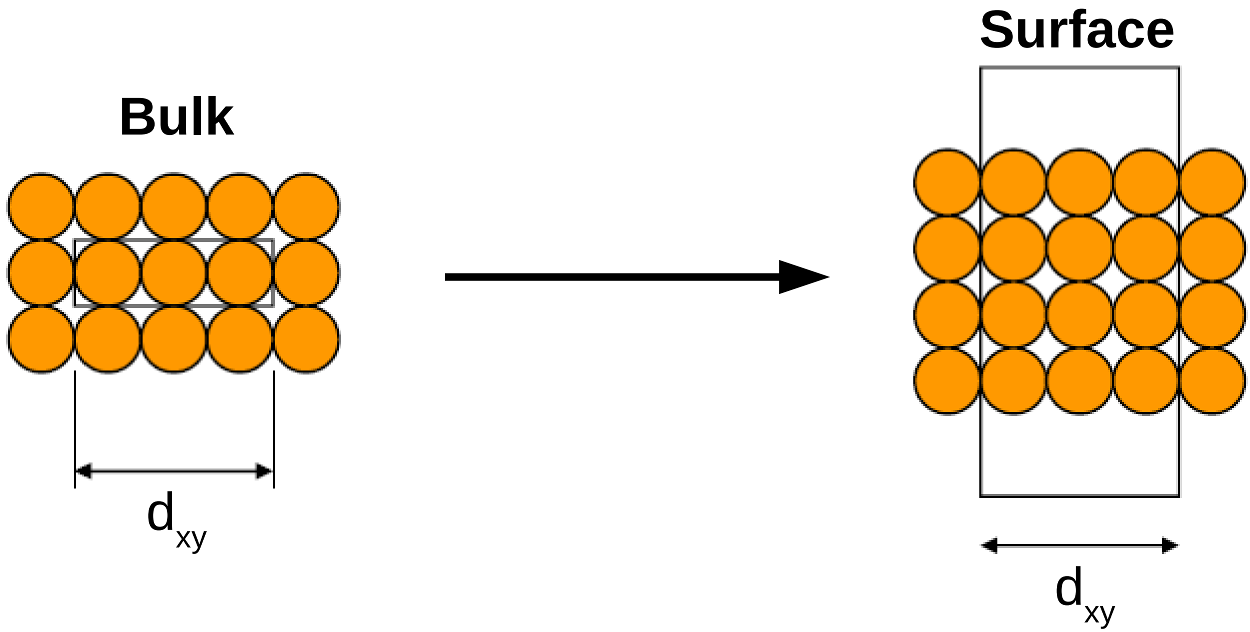


- Controlled by *kgrid.Cutoff* or *kgrid.MonkhorstPack*
- Only time reversal symmetry is used in SIESTA
- Convergence is equal to other codes:
  - Metals require a lot of k-points
  - Insulators require less k-points
- An incremental procedure is the best approach:
  - Density matrix and geometry are good with a "*reasonable*" number of k-points
  - Use a *"reasonable"* number for most calculations, then increase the number to refine the results

K-point sampling

The BZ sampling must be similar for all the considered structures:
→ Try to keep the same density of k-points

Example: surface (slab) calculations



For the same d_{xy} use the same K_{xy} points

Insight into relevant parameters

- Pseudopotential
- Basis set
- K-point sampling
- **Real-space grid**
- Self-consistent cycle

Real-space grid

It is used to project ρ(**r**) in order to calculate:

- XC potential
- Solve Poisson equation to get Hartree potential
- Calculate three center integrals (difficult to tabulate and store)

The real-space grid is not part of the basis set, is an auxiliary grid and, therefore, convergence of energy is not necessarily variational with respect to its fineness.

Mesh cutoff

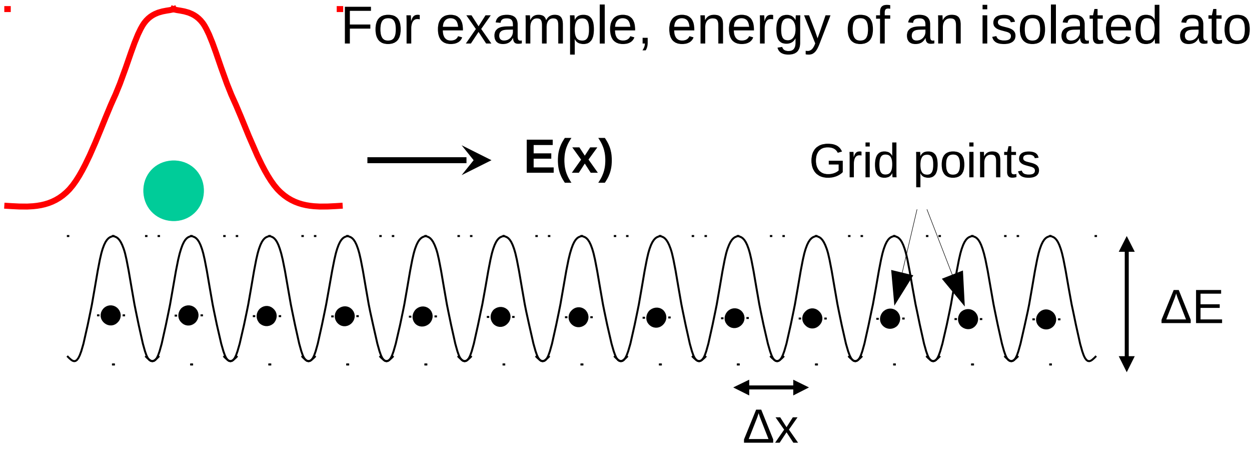
- Single parameter which controls the fineness of the grid
- It is the maximum kinetic energy of the plane waves that can be represented in the grid without aliasing

Real-space grid

“Eggbox” effect

The real-space grid breaks translational symmetry: “eggbox” effect

For example, energy of an isolated atom moving along the grid:



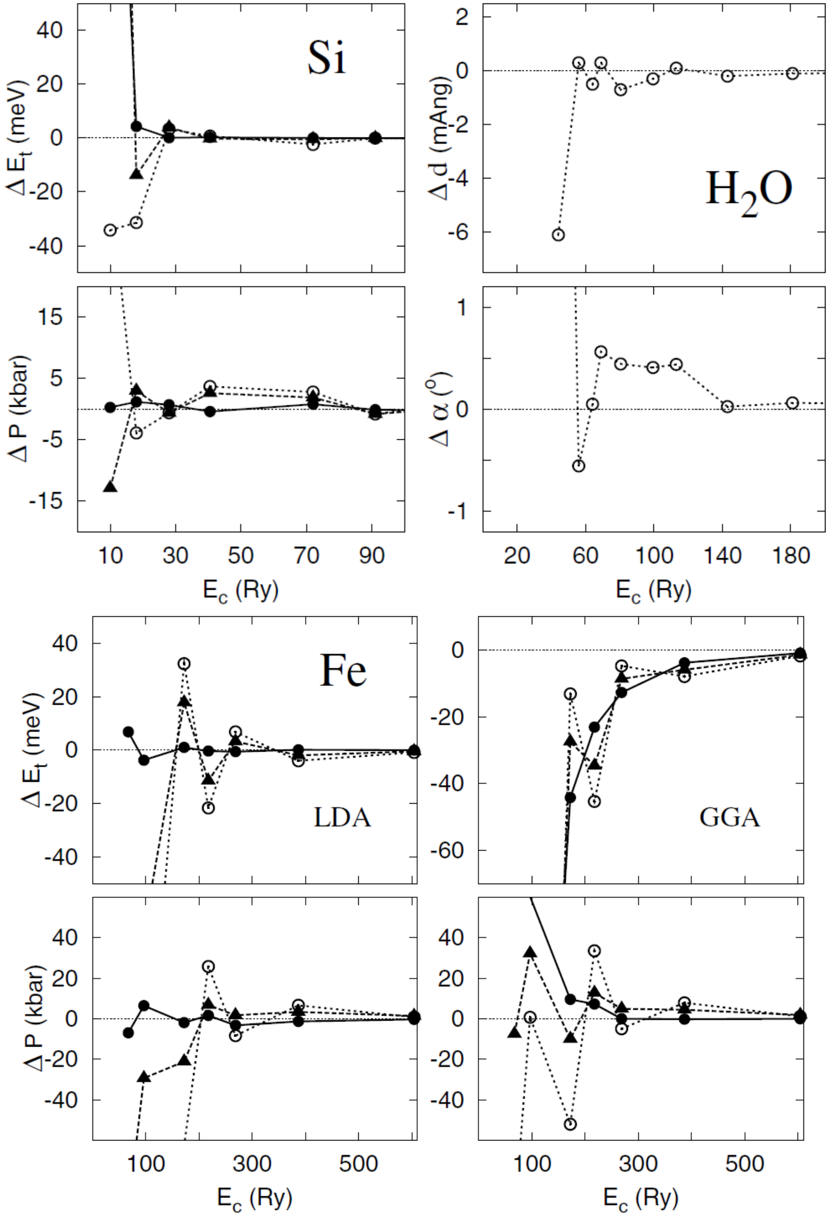
- Forces are more affected than energy
- Important for very precise relaxations and phonon calculations

Solutions:

- Increase Mesh cutoff
- *Grid.CellSampling*
- Filtering of atomic orbitals and potentials
  (E. Anglada et al, *Phys. Rev B* **73**, 115122 (2006))

Real-space grid

Mesh cutoff convergence

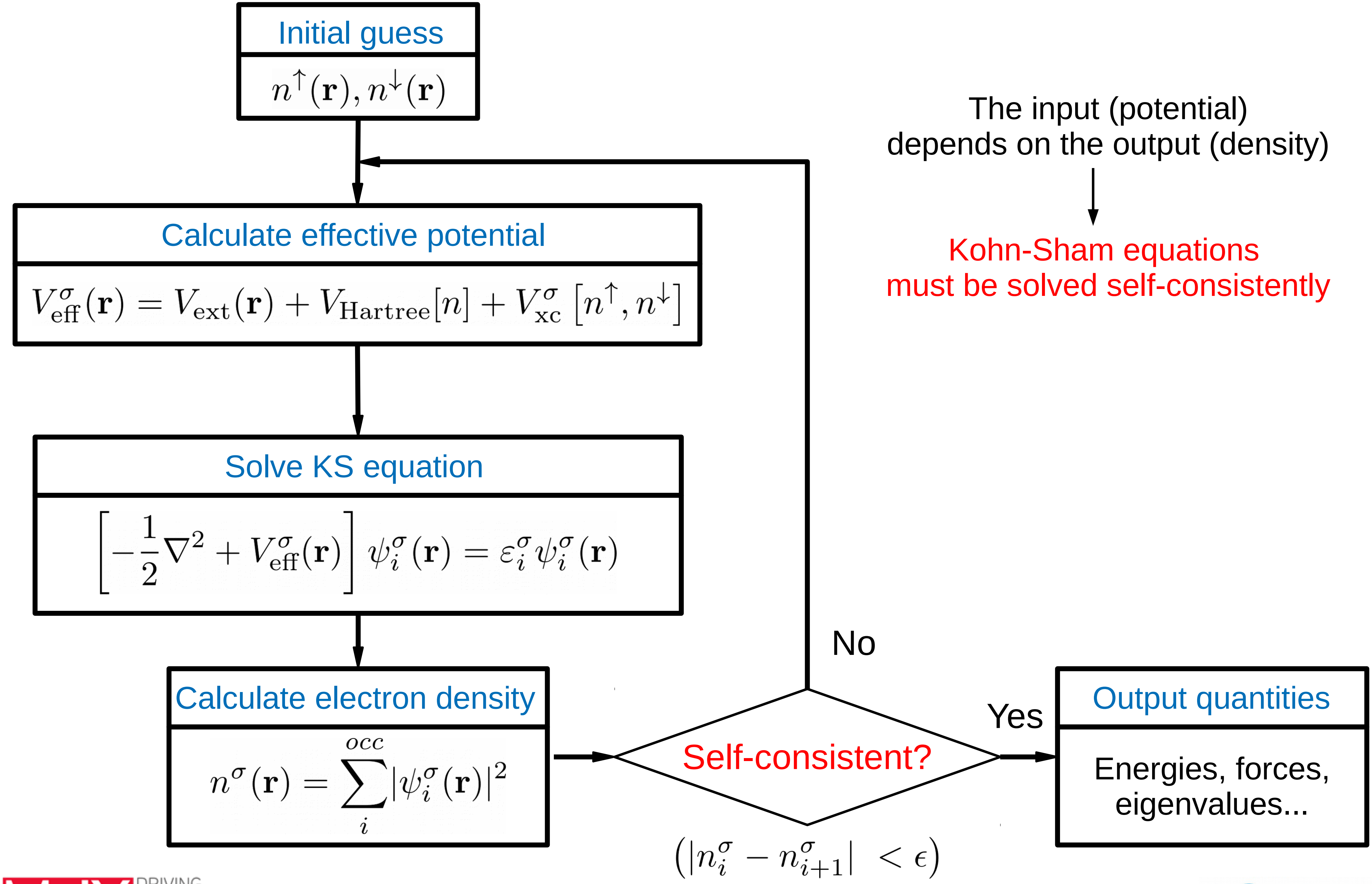


- Convergence is rarely achieved below 100 Ry
- Starting value around 150-200 Ry
- GGA and non-linear core corrections require higher values
- Options to improve:
  - *Grid.CellSampling*
  - Filtering of orbitals and potentials
    (E. Anglada et al, *Phys. Rev B* **73**, 115122 (2006))

Insight into relevant parameters

- Pseudopotential
- Basis set
- K-point sampling
- Real-space grid
- **Self-consistent cycle**

Self-consistent cycle



Initial guess

$n^{\uparrow}(\mathbf{r}), n^{\downarrow}(\mathbf{r})$

Calculate effective potential

$$V_{\text{eff}}^{\sigma}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + V_{\text{Hartree}}[n] + V_{\text{xc}}^{\sigma}\left[n^{\uparrow}, n^{\downarrow}\right]$$

Solve KS equation

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}^{\sigma}(\mathbf{r})\right]\psi_i^{\sigma}(\mathbf{r}) = \varepsilon_i^{\sigma}\psi_i^{\sigma}(\mathbf{r})$$

Calculate electron density

$$n^{\sigma}(\mathbf{r}) = \sum_i^{occ} |\psi_i^{\sigma}(\mathbf{r})|^2$$

Self-consistent?

$$\left(\left|n_i^{\sigma} - n_{i+1}^{\sigma}\right| < \epsilon\right)$$

No

Yes

Output quantities

Energies, forces, eigenvalues...

The input (potential)
depends on the output (density)

Kohn-Sham equations
must be solved self-consistently

Self-consistent cycle

“How much do I need to converge?”: value of ε

$$\left|n_i^{\sigma} - n_{i+1}^{\sigma}\right| \; < \epsilon$$

- Density matrix as monitor for self-consistency: *SCF.DM.Tolerance* (default: 10^{-4} eV)
  - Stick to this value or smaller
  - Higher values could be considered in special situations, like preliminary relaxations
  - Above 10^{-3} eV the errors are too large

- Hamiltonian matrix elements as monitor for self-consistency: *SCF.H.Tolerance* (default: 10^{-3} eV)
  - Default value is good in most situations
  - Could be difficult to achieve in some circumstances, higher values can be acceptable if DM is well converged

- Requirements depend on which properties we are interested in

- Always check that things make sense

Self-consistent cycle

Mixing

- Mixing of input and output quantities is needed to achieve self-consistency

$$\rho_{\mathrm{in}}^{n+1} = \alpha\rho_{\mathrm{out}}^{n} + (1-\alpha)\rho_{\mathrm{in}}^{n}$$

- We can mix:
  - Hamiltonian elements (default, usually faster)
  - Density matrix (old default)

- Many different options, only by testing you can find good parameters for your system:
  - *DM.MixingWeight*: α is system dependent and difficult to guess:
    0.1-0.15 for insulators and semiconductors, smaller for metals
  - Type of mixing: linear, Pulay, Broyden
  - *DM.NumberPulay*: 3-7

- Self-consistency can be quite difficult to achieve for some systems
  - Magnetic systems (specially non-collinear)
  - Spin-orbit coupling

Conclusions

- We have different choices to make and parameters to determine, every one has to be carefully checked:
  - XC functional
  - Pseudopotential
  - Basis set
  - K-points
  - Grid cutoff
  - ...

- The best way to handle it is an incremental approach: start from a reasonable guess and converge the parameters until the desired precision is reached

- All the parameters are system dependent, everything has to be tested

TEST, TEST, TEST

You have to do TESTS!