An introduction to the plane-wave pseudopotential method

P. Giannozzi
Università di Udine and CNR-IOM Democritos, Trieste, Italy

Summer School on Advanced Materials and Molecular Modelling
September 16-20, Ljubljana, Slovenija

Many pictures courtesy of Stefano de Gironcoli and Shobhana Narasimhan, JNCASR

Outline

1. Density-functional theory (DFT): Hohenberg-Kohn theorem, Kohn-Sham equations
2. Solution of the DFT problem: Self-consistency, global minimization, basis sets
3. Periodicity, supercells, plane waves
4. Pseudopotentials
5. Brillouin-Zone sampling: k-points

Born-Oppenheimer approximation

Time-dependent Schrödinger equation for electrons and nuclei:

$$i\hbar\frac{\partial\Phi(r,R;t)}{\partial t} = \left(-\sum_{\mu}\frac{\hbar^2}{2M_\mu}\nabla^2_{\mathbf{R}_\mu} - \sum_{i}\frac{\hbar^2}{2m}\nabla^2_{\mathbf{r}_i} + V(r,R)\right)\Phi(r,R;t)$$

where

$$V(r,R) = -\sum_{i,\mu}\frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_\mu|} + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2}\sum_{\mu\neq\nu}\frac{Z_\mu Z_\nu e^2}{|\mathbf{R}_\mu - \mathbf{R}_\nu|}$$

Notation: $r \equiv (\mathbf{r}_1, \ldots, \mathbf{r}_n)$ (electrons); $R \equiv (\mathbf{R}_1, \ldots, \mathbf{R}_N)$ (nuclei).

Born-Oppenheimer approximation ($M >> m$): solve the electronic problem

$$\left(-\frac{\hbar^2}{2m}\sum_{i}\nabla^2_{\mathbf{r}_i} + V(r,R)\right)\Psi_R(r) = E(R)\Psi_R(r)$$

for given R. One obtains $E(R)$, the *total energy* or *potential energy surface* (PES). $E(R)$ determines the nuclear motion, its local minima correspond to meta-stable states, its global minimum to the ground state of the system.

1. Solving the electronic problem: Hohenberg-Kohn theorem

Let us introduce the ground-state *charge density* $n(\mathbf{r})$. For N electrons:

$$n(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r}_2, ...\mathbf{r}_N)|^2 d\mathbf{r}_2 ... d\mathbf{r}_N.$$

The *Hohenberg-Kohn theorem* (1964) can be demonstrated: there is a *unique* potential $V(r, R)$ having $n(\mathbf{r})$ as ground-state charge density. Consequences:

- The electronic part of the energy can be written as a *functional* of $n(\mathbf{r})$:

$$E[n(\mathbf{r})] = F[n(\mathbf{r})] + \int n(\mathbf{r}) V(\mathbf{r}) d\mathbf{r}$$

  where $F[n(\mathbf{r})]$ is a *universal* functional of the density, $V(\mathbf{r})$ is the external (nuclear) potential acting on each electron:

$$V(\mathbf{r}) = -\sum_{\mu} \frac{Z_\mu e^2}{|\mathbf{r} - \mathbf{R}_\mu|}.$$

- $E[n(\mathbf{r})]$ is *minimized* by the ground-state charge density $n(\mathbf{r})$.

Density-Functional Theory: Kohn-Sham approach

Let us introduce the orbitals ψ_i for an auxiliary set of *non-interacting electrons* whose charge density is the same as that of the true system:

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2, \qquad \langle \psi_i | \psi_j \rangle = \delta_{ij}$$

Let us rewrite the energy functional in a more manageable way as:

$$E = T_s[n(\mathbf{r})] + E_H[n(\mathbf{r})] + E_{xc}[n(\mathbf{r})] + \int n(\mathbf{r}) V(\mathbf{r}) d\mathbf{r}$$

where $T_s[n(\mathbf{r})]$ is the kinetic energy of the non-interacting electrons:

$$T_s[n(\mathbf{r})] = -\frac{\hbar^2}{2m} \sum_i \int \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) d\mathbf{r},$$

$E_H[n(\mathbf{r})]$ is the *Hartree energy*, due to electrostatic interactions:

$$E_H[n(\mathbf{r})] = \frac{e^2}{2} \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}',$$

$E_{xc}[n(\mathbf{r})]$ is called *exchange-correlation energy* (a reminiscence from the Hartree-Fock theory) and includes all the remaining (unknown!) energy terms.

Minimization of the energy with respect to ψ_i yields the *Kohn-Sham* (KS) equations:

$$\underbrace{\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r})\right)}_{H_{KS}}\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}),$$

where the Hartree and exchange-correlation potentials:

$$V_H(\mathbf{r}) = \frac{\delta E_H[n(\mathbf{r})]}{\delta n(\mathbf{r})} = e^2\int\frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}d\mathbf{r}', \qquad V_{xc}(\mathbf{r}) = \frac{\delta E_{xc}[n(\mathbf{r})]}{\delta n(\mathbf{r})}$$

depend self-consistently upon the ψ_i via the charge density.

The energy can be rewritten in an alternative form using the KS eigenvalues ϵ_i:

$$E = \sum_i \epsilon_i - E_H[n(\mathbf{r})] - \int n(\mathbf{r})V_{xc}(\mathbf{r})d\mathbf{r} + E_{xc}[n(\mathbf{r})]$$

Exchange-correlation functionals: simple approximations

What is $E_{xc}[n(\mathbf{r})]$? Viable approximations are needed to turn DFT into a useful tool.

- **Local Density Approximation** (LDA): First, "historical" approach (1965). Replace the energy functional with a *function* of the local density $n(\mathbf{r})$:

$$E_{xc} = \int n(\mathbf{r})\epsilon_{xc}(n(\mathbf{r}))d\mathbf{r}, \qquad V_{xc}(\mathbf{r}) = \epsilon_{xc}(n(\mathbf{r})) + n(\mathbf{r})\left.\frac{d\epsilon_{xc}(n)}{dn}\right|_{n=n(\mathbf{r})}$$

  where $\epsilon_{xc}(n)$ is calculated for the homogeneous electron gas of density n (using Quantum Monte Carlo techniques) and fitted to some analytic form

- **Generalized Gradient Approximation** (GGA). The next step: a class of functionals depending upon the local density and the *local gradient* $|\nabla n(\mathbf{r})|$ of the density:

$$E_{xc} = \int n(\mathbf{r})\epsilon_{GGA}\left(n(\mathbf{r}), |\nabla n(\mathbf{r})|\right) d\mathbf{r}$$

  There are many flavors of GGA, yielding similar (but slightly different) results. These are by now the "basic" functionals in most present-day calculations, with excellent price-to-performance ratio, but some noticeable shortcomings.

Spin-polarized extension: LSDA

Simplest case: assume a unique quantization axis for spin. Energy functional:

$$E \equiv E[n_+(\mathbf{r}), n_-(\mathbf{r})] = T_s + \int n(\mathbf{r}) V(\mathbf{r}) d\mathbf{r} + E_H + E_{xc}[n_+(\mathbf{r}), n_-(\mathbf{r})]$$

$n_\sigma(\mathbf{r}) =$ charge density with spin polarization σ
$n(\mathbf{r}) = n_+(\mathbf{r}) + n_-(\mathbf{r})$ total charge density.

Minimization of the above functional yields the Kohn-Sham equations:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r}')}{\mid \mathbf{r} - \mathbf{r}' \mid} d\mathbf{r}' + V_{xc}^\sigma(\mathbf{r}) \right] \psi_i^\sigma(\mathbf{r}) = \epsilon_i^\sigma \psi_i^\sigma(\mathbf{r})$$

Exchange-correlation potential and charge density:

$$V_{xc}^\sigma(\mathbf{r}) = \frac{\delta E_{xc}}{\delta n_\sigma(\mathbf{r})}, \qquad n_\sigma(\mathbf{r}) = \sum_i f_i^\sigma |\psi_i^\sigma(\mathbf{r})|^2$$

Note the extension to *fractional occupancies* (i.e. metallic systems): $0 \leq f_i^\sigma \leq 1$.
Noncolinear magnetism (no fixed axis for magnetization) can also be described.

"Basic" DFT: advantages and shortcomings

+ **Computationally convenient**: calculations in relatively complex condensed-matter systems become affordable (GGA marginally more expensive than LDA)

+ **Excellent results** in terms of prediction of atomic structures, bond lengths, lattice parameters (within $1 \div 2\%$), binding and cohesive energies (5 to 10% GGA; LDA much worse, strongly overestimates), vibrational properties. Especially good for $sp-$bonded materials, may work well also in more "difficult" materials, such as transition metal compounds

– The infamous **band gap problem**: $\epsilon_c - \epsilon_v$ (or HOMO-LUMO in quantum chemistry parlance) wildly underestimates the true band gap, $\Delta = I - A$, where $I = E(N) - E(N-1)$, ionization potential, $A = E(N+1) - E(N)$, electron affinity

– Serious trouble in dealing with **strongly correlated materials**, such as e.g. magnetic materials (trouble mostly arising from spurious self-interaction)

– **No van der Waals interactions** in any functional based on the local density and gradients: van der Waals is nonlocal, cannot depend upon charge overlap

2. Towards electronic ground state (fixed nuclei)

Possible methods to find the DFT ground state:

1. By the *self-consistent* solution of the Kohn-Sham equations

$$H_{KS}\psi_i \equiv \left(T + V + V_H[n] + V_{xc}[n]\right)\psi_i = \epsilon_i \psi_i$$

where

- $n(\mathbf{r}) = \sum_i f_i |\psi_i(\mathbf{r})|^2$ is the charge density, f_i are occupation numbers
- V is the nuclear (pseudo-)potential acting on electrons (may be nonlocal)
- $V_H[n]$ is the Hartree potential, $V_H(\mathbf{r}) = e^2 \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'$
- $V_{xc}[n]$ is the exchange-correlation potential. For the simplest case, LDA, $V_{xc}[n]$ is a *function* of the charge density at point $\mathbf{r}$: $V_{xc}(\mathbf{r}) \equiv \mu_{xc}(n(\mathbf{r}))$

Orthonormality constraints $\langle \psi_i | \psi_j \rangle = \delta_{ij}$ automatically hold.

2. By *constrained global minimization* of the energy functional

$$E[\psi] = \sum_i f_i \langle \psi_i | T + V | \psi_i \rangle + E_H[n] + E_{xc}[n]$$

under orthonormality constraints $\langle \psi_i | \psi_j \rangle = \delta_{ij}$, i.e. minimize:

$$\widetilde{E}[\psi, \Lambda] = E[\psi] - \sum_{ij} \Lambda_{ij} \left(\langle \psi_i | \psi_j \rangle - \delta_{ij} \right)$$

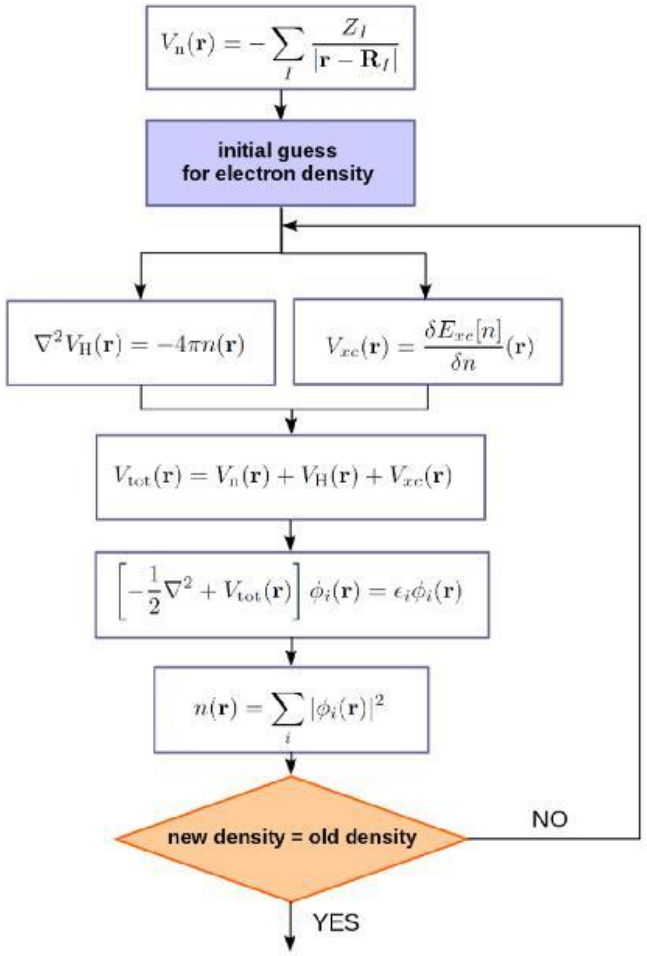
where

- V, $n(\mathbf{r})$ are defined as before, $\psi \equiv$ all occupied Kohn-Sham orbitals
- Λ_{ij} are Lagrange multipliers, $\Lambda \equiv$ all of them
- $E_H[n]$ is the Hartree energy, $E_H = \dfrac{e^2}{2} \displaystyle\int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'$
- $E_{xc}[n]$ is the exchange-correlation energy. For the simplest case, LDA, $E_{xc} = \displaystyle\int n(\mathbf{r}) \epsilon_{xc}(n(\mathbf{r})) d\mathbf{r}$ where ϵ_{xc} is a function of $n(\mathbf{r})$.

Towards electronic ground state II

In a self-consistent approach, we need to find the self-consistent charge density (or potential), performing the following operations:

- Calculate the potential from the charge density
- Solve (*diagonalize*) the Kohn-Sham equations at fixed potential
- Calculate the charge density from Kohn-Sham orbitals



In a global-minimization approach, the operations are basically the same, since one needs the *gradients* of the energy functional, that is, $H_{KS}\psi$ products:

$$\frac{\delta \widetilde{E}[\psi, \Lambda]}{\delta \psi_j^*} = H_{KS}\psi_j - \sum_i \Lambda_{ji}\psi_i$$

Towards electronic ground state III

Let us start from some guess of the input charge density $n^{in}(\mathbf{r})$:

$$n^{in} \longrightarrow (V_H + V_{xc})[n^{in}] \longrightarrow \psi_i(\mathbf{r}) \longrightarrow n^{out}(\mathbf{r}) = \sum_i f_i |\psi_i(\mathbf{r})|^2$$

Such procedure defines the output charge density as a *functional* of the input one: $n^{out} \equiv F[n^{in}]$. Assuming we have a black box producing $F[n]$, we have to reach self-consistency, that is, find n^{gs} such that $n^{gs} = F[n^{gs}]$

Simply re-inserting n^{out} as n^{in} *is not guaranteed to converge* (it seldom does!). Reason: there is no guarantee that such procedure leads to a reduction of all component of the error (in particular, in typical condensed-matter systems low-frequency, small-**G** components of the error are not reduced). One can use a simple **mixing algorithm**:

$$n^{new} = \alpha n^{out} + (1 - \alpha) n^{in}, \qquad 0 < \alpha < 1$$

guaranteed to converge if α is small enough.

Practical, more sophisticated algorithms (Anderson, Broyden, DIIS) use the input and output of several preceding steps to determine the next optimal input combination.

Calculation of the total energy

Once self-consistency (or the minimum) is reached, the **total energy** of the system can be calculated:

$$E = \sum_i f_i \langle \psi_i | T + V | \psi_i \rangle + E_H[n] + E_{xc}[n] + E_{ion-ion}$$

where $E_{ion-ion}$ is the repulsive contribution from nuclei to the energy:

$$E_{ion-ion} = \frac{e^2}{2} \sum_{\mu \neq \nu} \frac{Z_\mu Z_\nu}{\mid \mathbf{R}_\mu - \mathbf{R}_\nu \mid}$$

Equivalent expression for the energy, using Kohn-Sham eigenvalues:

$$E = \sum_i f_i \epsilon_i - E_H[n] + E_{xc}[n(\mathbf{r})] - \int n(\mathbf{r}) V_{xc}[n(\mathbf{r})] d\mathbf{r} + E_{ion-ion}$$

The total energy depends upon all atomic positions $\mathbf{R}_\mu$.

Hellmann-Feynman Forces

Forces on atoms are the derivatives of the total energy wrt atomic positions. The *Hellmann-Feynman theorem* tells us that forces are the expectation value of the derivative of the external potential only:

$$\mathbf{F}_\mu = -\frac{\partial E}{\partial \mathbf{R}_\mu} = -\sum_i f_i \langle \psi_i | \frac{\partial V}{\partial \mathbf{R}_\mu} | \psi_i \rangle = -\int n(\mathbf{r}) \frac{\partial V}{\partial \mathbf{R}_\mu} d\mathbf{r}$$

the rightmost expression being valid only for *local* potentials, $V \equiv V(\mathbf{r})$ (the one at the left is more general, being valid also for nonlocal potentials $V \equiv V(\mathbf{r}, \mathbf{r}')$).

Demonstration (simplified). In addition to the explicit derivative of the external potential (first term), there is an implicit dependency via the derivative of the charge density:

$$\frac{\partial E}{\partial \mathbf{R}_\mu} = \int n(\mathbf{r}) \frac{\partial V}{\partial \mathbf{R}_\mu} d\mathbf{r} + \int \frac{\delta E}{\delta n(\mathbf{r})} \frac{\partial n(\mathbf{r})}{\partial \mathbf{R}_\mu} d\mathbf{r}$$

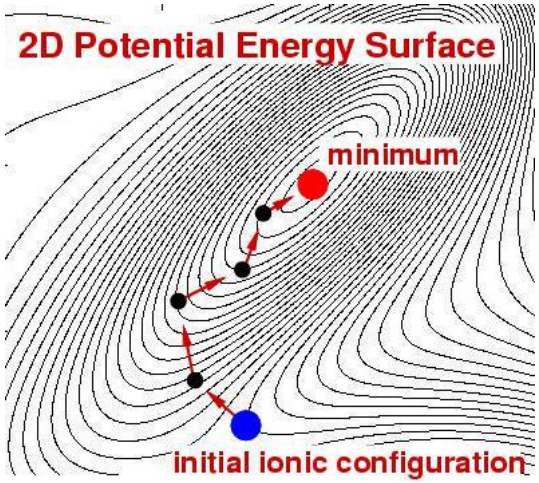
The red term cancels due to the *variational character* of DFT: $\delta E / \delta n(\mathbf{r}) = \mu$, constant.

The calculation of the Hellmann-Feynman forces is straightforward (in principle, not necessarily in practice!) once the self-consistent electronic structure is calculated.

Structural Optimization and Molecular Dynamics

Within the *Born-Oppenheimer*, or *adiabatic* approximation, the total energy as a function of atomic positions, or *Potential Energy Surface* (PES), determines the behaviour of nuclei.

The *global* ground state can be found by minimizing the function $E(\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_N)$, depending upon the $3N$ atomic coordinates for a system of N atoms. This is a "standard" mathematical problem: finding the minimum of a function, knowing its derivatives, that is, the Hellmann-Feynman forces (in the picture, a cartoon of a PES in two dimensions with the path to the minimum).



Once forces are calculated, one can perform not only structural optimization, but also *molecular dynamics*. If a classical behaviour of the nuclei is assumed, all the machinery of classical MD can be recycled, with forces calculated from *first principles*.

Diagonalization of the KS Hamiltonian

The solution of the Kohn-Sham problem $H_{KS}\psi = \epsilon\psi$ at fixed potential is (usually) the toughest problem. How to proceed? By expanding ψ into some suitable **basis set** $\{\phi_i\}$ as

$$\psi(\mathbf{r}) = \sum_i c_i \phi_i(\mathbf{r}).$$

For an orthonormal basis set, we solve the secular equations

$$\sum_j (H_{ij} - \epsilon\delta_{ij})c_j = 0$$

where $H_{ij} = \langle\phi_i|H_{KS}|\phi_j\rangle$ are the *matrix elements* of the Hamiltonian.
For a non-orthonormal basis set, we solve the generalized problem:

$$\sum_j (H_{ij} - \epsilon S_{ij})c_j = 0$$

where $S_{ij} = \langle\phi_i|\phi_j\rangle$ is the *overlap matrix*.

Diagonalization algorithms are well known in linear algebra, but in practice, one has to resort to smarter *iterative* algorithms, allowing *not* to store those matrices.

Most popular basis sets

We have to choose now a suitable basis set. Typical candidates include

- *Localized* basis sets:
  atom-centred functions such as
  - Linear Combinations of Atomic Orbitals (LCAO)
  - Gaussian-type Orbitals (GTO)
  - Linearized Muffin-Tin Orbitals (LMTO)

- *Delocalized* basis sets:
  - Plane Waves (PW)

One could also consider *mixed* basis sets. The Linearized Augmented Plane Waves (LAPW) could be classified in this category.

Advantages and disadvantages of various basis sets

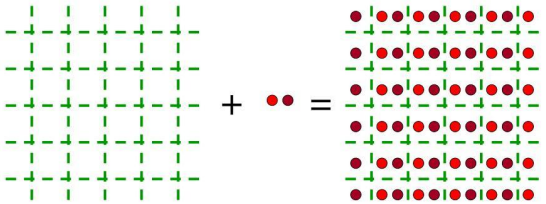
- Localized basis sets:
  - \+ fast convergence with respect to basis set size (just a few functions per atom needed)
  - \+ can be used in finite as well as in periodic systems (as Bloch sums: $\phi_{\mathbf{k}} = \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot\mathbf{R}}\phi(\mathbf{r}-\mathbf{R})$)
  - – difficult to evaluate convergence quality (no systematic way to improve convergence)
  - – difficult to use (two- and three-centre integrals)
  - – difficult to calculate forces (*Pulay forces* if basis set is not complete)

- Plane Waves:
  - – slow convergence with respect to basis set size (many more PWs than localized functions needed)
  - – require periodicity: in finite systems, *supercells* must be introduced
  - \+ easy to evaluate convergence quality (just increase a single parameter, the *cutoff*)
  - \+ easy to use (Fourier transform)
  - \+ easy to calculate forces (no Pulay forces even if the basis set is incomplete)

3. Periodicity

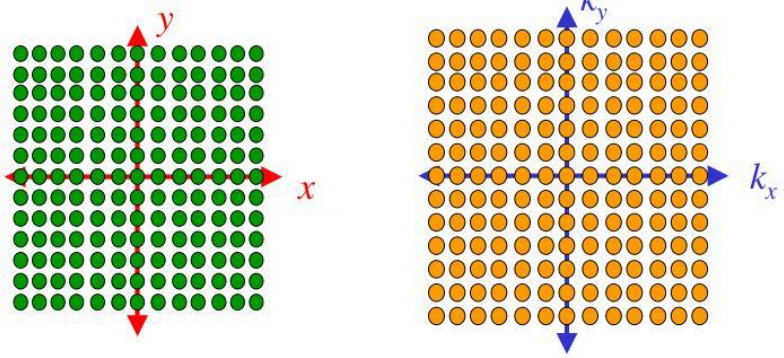
Let us focus on the case of the *infinite perfect crystals*, having translation symmetry. A perfect crystal is described in terms of

- a periodically repeated **unit cell** and a **lattice** of translation vectors, defined via three primitive vectors $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ and integer coefficients n_1, n_2, n_3:
$$\mathbf{R} = n_1\mathbf{R}_1 + n_2\mathbf{R}_2 + n_3\mathbf{R}_3.$$



- a **basis** of atomic positions $\mathbf{d}_i$ into the unit cell

- a **reciprocal lattice** of vectors $\mathbf{G}$ such that $\mathbf{G} \cdot \mathbf{R} = 2\pi l$, with l integer:
$$\mathbf{G} = m_1\mathbf{G}_1 + m_2\mathbf{G}_2 + m_3\mathbf{G}_3,$$
with $\mathbf{G}_i \cdot \mathbf{R}_j = 2\pi\delta_{ij}$ and m_1, m_2, m_3 integer.

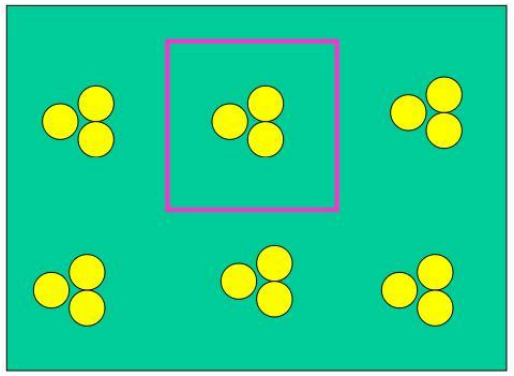


Non periodic systems: supercells

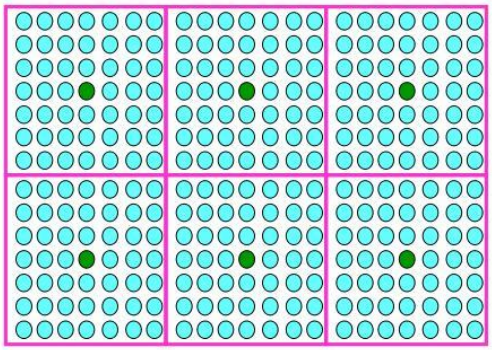
What about e.g. defects in crystals, surfaces, alloys, amorphous materials, liquids, molecules, clusters? none of these has perfect periodicity! One can use **supercells**, introducing an artificial periodicity.

The supercell geometry is dictated by the type of system under investigation:

- Molecules, clusters:
  the supercell must allow a minimum distance of at least a few A (~ 6) between the closest atoms in different periodic replica.

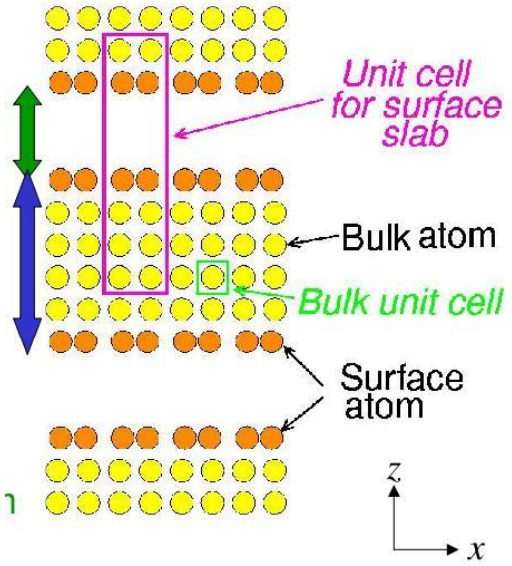


- Defects in crystals:
  the supercell is commensurate with the perfect crystal cell. The distance between periodic replica of the defect must be "big enough" to minimize spurious defect-defect interactions.



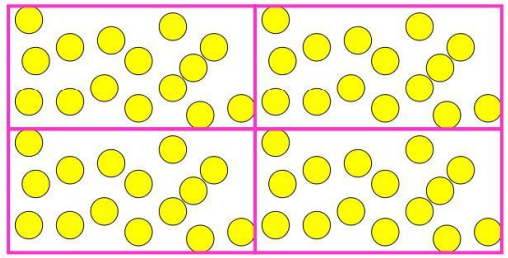
Surfaces:

- *slab geometry*. The number of layers of the materials must be "big enough" to have "bulk behaviour" in the furthest layer from the surface. The number of empty layers must be "big enough" to have minimal interactions between layers in different regions.



Alloys, amorphous materials, liquids:

- the supercell must be "big enough" to give a reasonable description of physical properties.



Conceptually there is no difference between a "supercell" and an ordinary unit cell: typically, "supercell" is used when the periodicity is not perfect or non-existent

Band Structure, Bloch states

The one-electron states $\psi(\mathbf{r})$ of a perfect crystal Hamiltonian $H = T + V$ are described by a **band index** i and a **wave vector** $\mathbf{k}$.

It is convenient to consider the *thermodynamic limit*: a slab of crystal composed of $\mathcal{N} = \mathcal{N}_1\mathcal{N}_2\mathcal{N}_3$ unit cells, $\mathcal{N} \to \infty$, obeying Periodic Boundary Conditions:

$$\psi(\mathbf{r} + \mathcal{N}_1\mathbf{R}_1) = \psi(\mathbf{r} + \mathcal{N}_2\mathbf{R}_2) = \psi(\mathbf{r} + \mathcal{N}_3\mathbf{R}_3) = \psi(\mathbf{r}).$$

There are $\mathcal{N}$ wave vectors $\mathbf{k}$ in the unit cell of the reciprocal lattice, called the **Brillouin Zone**. The one-electron states (energy bands) can be written as

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

where $u_{i,\mathbf{k}}(\mathbf{r})$ is translationally invariant:

$$u_{i,\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{i,\mathbf{k}}(\mathbf{r}).$$

Plane-wave basis set

A PW basis set for states of wave vector $\mathbf{k}$ is defined as

$$\langle \mathbf{r}|\mathbf{k}+\mathbf{G}\rangle = \frac{1}{\sqrt{\mathcal{N}\Omega}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \quad \frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2 \leq E_{cut}$$

Ω = unit cell volume, $\mathcal{N}\Omega$ = crystal volume, E_{cut} = cutoff on the kinetic energy of PWs (in order to have a finite number of PWs!). The PW basis set is *complete* for $E_{cut} \to \infty$ and *orthonormal*: $\langle \mathbf{k}+\mathbf{G}|\mathbf{k}+\mathbf{G}'\rangle = \delta_{\mathbf{G}\mathbf{G}'}$

In a PW basis set one works with *Fourier components* ($c_{i,\mathbf{k}+\mathbf{G}}$ below):

$$|\psi_i\rangle = \sum_{\mathbf{G}} c_{i,\mathbf{k}+\mathbf{G}} |\mathbf{k}+\mathbf{G}\rangle$$

$$c_{i,\mathbf{k}+\mathbf{G}} = \langle \mathbf{k}+\mathbf{G}|\psi_i\rangle = \frac{1}{\sqrt{\mathcal{N}\Omega}} \int \psi_i(\mathbf{r}) e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} d\mathbf{r} = \widetilde{\psi}_i(\mathbf{k}+\mathbf{G}).$$

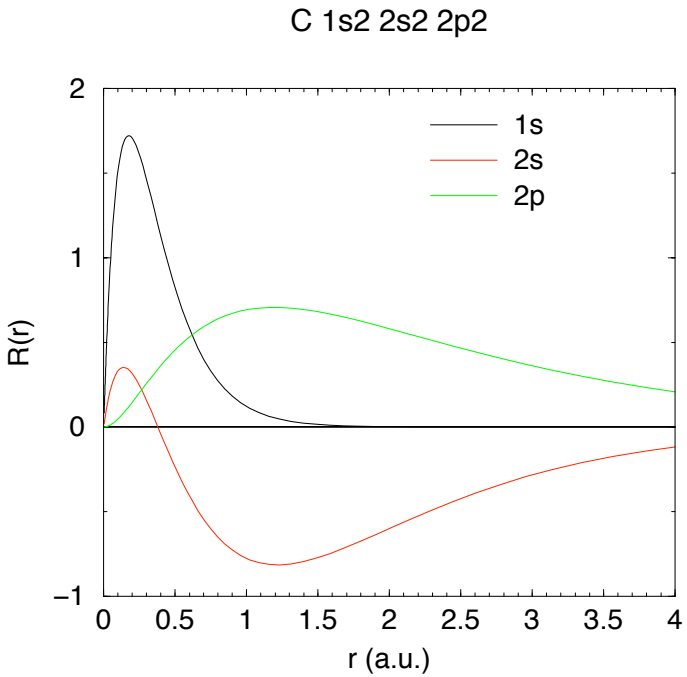
Real-space quantities can be obtained on a grid using Fast Fourier Transform.

4. The need for Pseudopotentials

Are PWs a practical basis set for electronic structure calculations? Not really!
From simple Fourier analysis: length scale $\delta \longrightarrow$ Fourier components up to $q \sim 2\pi/\delta$.
In a solid, this means $\sim 4\pi(2\pi/\delta)^3/3\Omega_{BZ}$ PWs (volume of the sphere of radius q divided by $\Omega_{BZ} = 8\pi^3/\Omega$, volume of the Brillouin Zone).

Estimate for diamond: $1s$ wavefunction has $\delta \simeq 0.1$ a.u., $\Omega = (2\pi)^3/(a_0^3/4)$ with lattice parameter $a_0 = 6.74$ a.u. $\longrightarrow 250,000$ PWs! We need to:

- get rid of core states
- get rid of orthogonality wiggles close to the nucleus



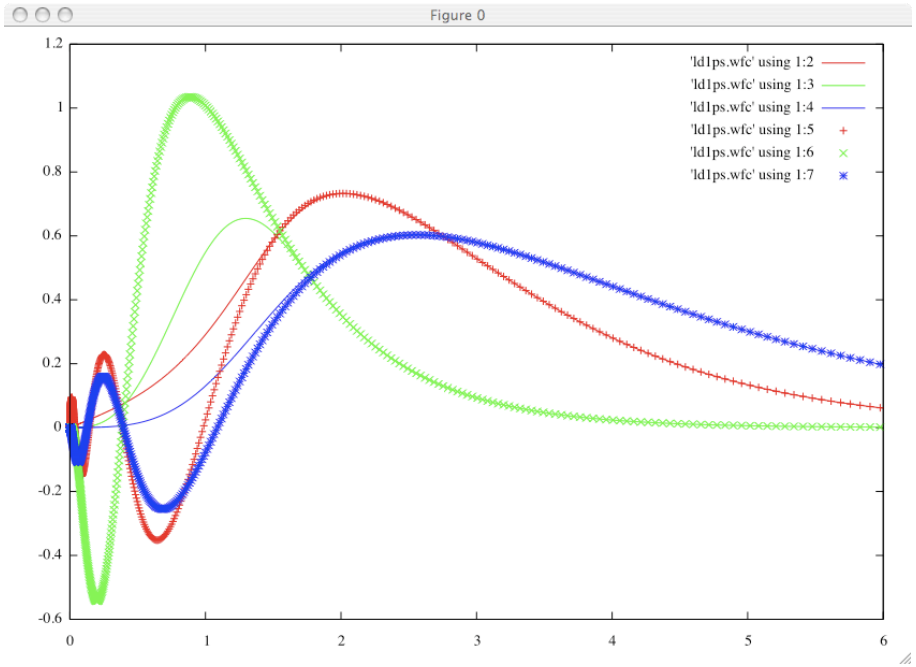
Solution: **Pseudopotentials** (PP). A smooth effective potential that reproduces the effect of the nucleus plus core electrons on valence electrons.

Understanding Pseudopotentials

Smoothness and *transferability* are the relevant keywords:

- We want our pseudopotential and pseudo-orbitals to be as *smooth* as possible so that expansion into plane waves is convenient (i.e. the required kinetic energy cutoff is small)

- We want our pseudopotential to produce pseudo-orbitals that are as close as possible to true ("all-electron") orbitals outside the core region, *for all systems containing a given atom* (in the figure: all-electron and pseudo-orbitals for Si)



Of course, the two goals are usually conflicting!

Pseudopotentials have a long story: let's start from the end.

Understanding PP: Projector-Augmented Waves

Let us look for a linear operator $\widehat{T}$ connecting all-electron orbitals $|\psi_i\rangle$ to pseudo-orbitals $|\widetilde{\psi}_i\rangle$ as in: $|\psi_i\rangle = \widehat{T}|\widetilde{\psi}_i\rangle$. Pseudo-orbitals will be our variational parameters. We write the charge density, energy, etc. using pseudo-orbitals and $\widehat{T}$ instead of all-electron orbitals.

The operator $\widehat{T}$ can be defined in terms of its action on atomic waves (i.e. orbitals at a given energy, not necessarily bound states):

- $|\phi_l\rangle$: set of atomic all-electron waves (bound or unbound states)
- $|\widetilde{\phi}_l\rangle$: corresponding set of atomic pseudo-waves. Beyond some suitable "core radius" R_l, $\widetilde{\phi}_l(r) = \phi_l(r)$; for $r < R_l$, $\widetilde{\phi}_l(r)$ are smooth functions.

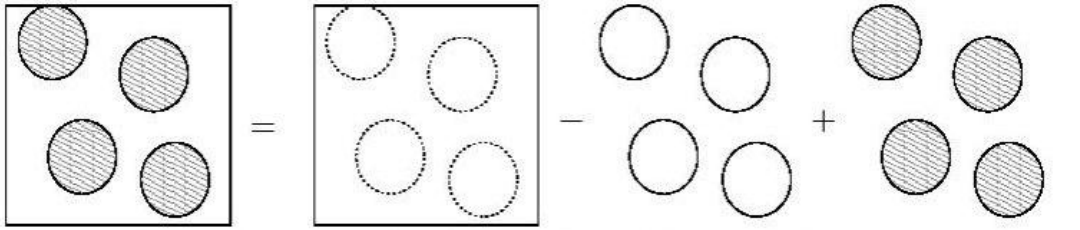
(P. E. Blöchl, PRB **50**, 17953 (1994))

Understanding PP: the PAW transformation

If the above sets are complete in the core region, the operator $\widehat{T}$ can be written as

$$|\psi_i\rangle = \widehat{T}|\widetilde{\psi}_i\rangle = |\widetilde{\psi}_i\rangle + \sum_l \left(|\phi_l\rangle - |\widetilde{\phi}_l\rangle\right)\langle\beta_l|\widetilde{\psi}_i\rangle$$

where the β_l "projectors" are atomic functions, having the properties $\langle\beta_l|\widetilde{\phi}_m\rangle = \delta_{lm}$ and $\beta_l(r) = 0$ for $r > R_l$. The logic is described in the picture below:



The pseudopotential itself is written as a *nonlocal* operator, $\hat{V}$, in terms of the β_l projectors:

$$\hat{V} \equiv V_{loc}(r) + \sum_{lm} |\beta_l\rangle D_{lm}\langle\beta_m|$$

(V_{loc} contains the long-range Coulomb part $-Ze^2/r$)

Understanding PP: Charge in PAW

The (valence) charge density is no longer the simple sum of $|\widetilde{\psi}_i|^2$:

$$n(\mathbf{r}) = \sum_i f_i |\widetilde{\psi}_i(\mathbf{r})|^2 + \sum_i f_i \sum_{lm} \langle \widetilde{\psi}_i | \beta_l \rangle Q_{lm}(\mathbf{r}) \langle \beta_m | \widetilde{\psi}_i \rangle,$$

and

$$Q_{lm}(\mathbf{r}) = \phi_l^*(\mathbf{r})\phi_m(\mathbf{r}) - \widetilde{\phi}_l^*(\mathbf{r})\widetilde{\phi}_m(\mathbf{r}).$$

The *augmentation charges* $Q_{lm}(\mathbf{r})$ are zero for $r > R_l$. A generalized orthonormality relation holds for pseudo-orbitals:

$$\langle \widetilde{\psi}_i | S | \widetilde{\psi}_j \rangle = \int \widetilde{\psi}_i^*(\mathbf{r}) \widetilde{\psi}_j(\mathbf{r}) d\mathbf{r} + \sum_{lm} \langle \widetilde{\psi}_i | \beta_l \rangle q_{lm} \langle \beta_m | \widetilde{\psi}_j \rangle = \delta_{ij}$$

where $q_{lm} = \int Q_{lm}(\mathbf{r}) d\mathbf{r}$. The D_{lm} quantites and β_l, Q_{lm} functions are atomic quantities that define the PP (or PAW set).

PP taxonomy: PAW, Ultrasoft, norm-conserving

- In the full PAW scheme, the augmentation functions are calculated and stored on a *radial grid*, centred at each atom. The charge density is composed by a "smooth" term expanded into plane waves, and an "augmentation" term calculated on the radial grid (Kresse and Joubert, PRB59, 1759 (1999))

- In the Ultrasoft PP scheme (D.Vanderbilt, B 41, R7892 (1990)), the augmentation functions $Q_{lm}(r)$ are *pseudized*, i.e. made smoother: both "smooth" and "augmentation" terms can be calculated on a FFT grid, in either reciprocal or real space. The latter term usually requires a larger grid in G-space than the former

- If we set $Q_{lm}(r) = 0$, we obtain good old norm-conserving PPs (Hamann, Schlüter, Chiang 1982) in the separable, nonlocal form.

Which pseudopotentials are good for me?

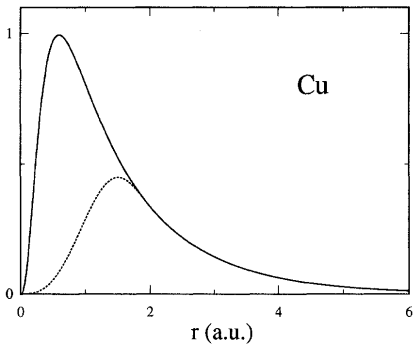
- **Norm-conserving**:
  - \+ are simple to generate and to use. Theory and methodological improvements are invariably implemented first (and often only) for norm-conserving PPs
  - – are relatively *hard*: core radii R_l cannot exceed by much the outermost maximum of the valence atomic orbitals, or else the loss of transferability is large. For some atoms: 2p elements C, N, O, F, 3d transition metals, 4f rare earths, this restriction may lead to very high plane-wave cutoffs (70 Ry and up)
  - – do not give any sensible information about the orbitals close to the nucleus (all-electron orbitals can be "reconstructed" using the PAW transformation)

This is usually your first choice and starting point.

Which pseudopotentials are good for me? (II)

- **Ultrasoft**:
  - \+ can be made *smooth* with little loss of transferability: core radii R_l can be pushed to larger values, even for "difficult cases". Cutoffs of 25 to 35 Ry are usually good for most cases. Note that you may need a second FFT grid for augmentation charges, with typical cutoff $8 \div 12\times$ orbital cutoff (instead of 4)
  - \- are not simple to generate: the pseudization of augmentation charges is often a source of trouble (e.g. negative charge)
  - \- introduce additional terms in the formalism: some property calculations may not implemented
  - – give even less information about the orbitals close to the nucleus (all-electron orbitals can be "reconstructed")

Ultrasoft PPs are typically used in all cases where norm-conserving PPs are too hard: C, N, O, F, 3d elements, "semicore" states



Which pseudopotentials are good for me? (III)

- **PAW**:
  - \+ most transferrable, even for atoms that are "difficult" for Ultrasoft PPs (e.g. magnetic materials): accuracy is comparable to all-electron techniques (e.g. FLAPW)
  - \+ give information about the orbital close to the nucleus
  - \- as complex to generate as Ultrasoft PPs
  - \- introduce even more additional terms in the formalism: some property calculations may not implemented

Which pseudopotentials are good for me? (IV)

There are a few more aspect to be considered in the choice of a pseudopotential:

- *PPs are bound to a specific XC functional*, at least in principle. Exception: Hybrid, nonlocal (vdW-DF), meta-GGA functionals, for which very few (or no) PPs are available. Typically, PPs computed from the most similar GGA are used instead.

- The distinction between "core" and "valence" electrons is not always clear-cut. In some cases you may need to extend "valence" to include the so-called *semicore states* in order to achieve better (or less lousy) transferability. E.g.: 3d states in Zn and Ga; 3s and 3p states in 3d transition metals Fe, Co, Ni, ...

Inclusion of semicore states adds considerable complexity to both the generation and the practical usage of a PP: to be done only if needed.

Where do I find pseudopotentials?

There are many ready-to-use PPs tables around, but *there is not a single standard PP file format*: each code has its own format.

QUANTUM ESPRESSO accepts an XML-like format called UPF, plus some old formats. See the pseudopotential page on the web site `pseudopotentials.quantum-espresso.org`, for more on

- **PSlibrary**: a project by A. Dal Corso to set up verified PPs for most elements
- other available PPs in UPF format (and their naming convention)
- other pseudopotential repositories, conversion from other formats

If everything else fails, PPs have to be generated.

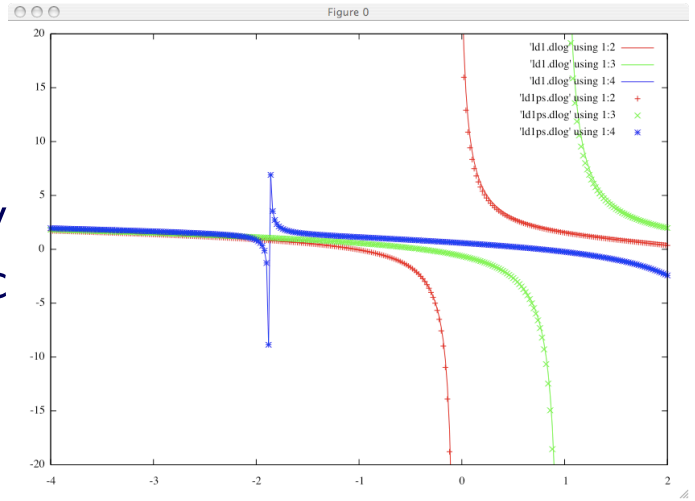
Pseudopotential testing

PPs must be *always* tested to check for

- absence of *ghost states*: spurious unphysical states in the valence region of energies, or close to it. All nonlocal PPs can be affected
- poor transferability

Testing can be performed

- by the PP generation code itself, by comparing *energy differences* between electronic configurations, and *logarithmic derivatives*;



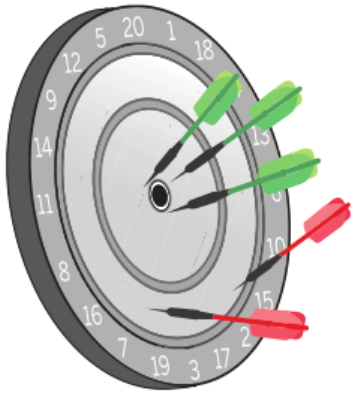
- in simple molecular or solid-state systems, *ideally* by comparing with accurate all-electron results; less ideally, with other PP results; even less ideally, with experimental data

Pseudopotential vs all-electron calculations

Systematic comparisons of different pseudopotential and all-electron DFT codes: *Reproducibility in density-functional theory calculations of solids*, K. Lejaeghere *et multis aliis*, Science **351** (6280), aad3000 (2016), DOI 10.1126/science.aad3000

Tests *precision* of the computational methods, not physical *accuracy* of results. Main outcomes: 1) PPs and all-alectron calculations have comparable precision, and 2) everybody is converging towards the same set of results,

Recent DFT methods yield reproducible results. Whereas older DFT implementations predict different values (red darts), codes have now evolved to mutual agreement (green darts). The scoreboard illustrates the good pairwise agreement of four classes of DFT implementations (horizontal direction) with all-electron results (vertical direction). Each number reflects the average difference between the equations of state for a given pair of methods, with the green-to-red color scheme showing the range from the best to the poorest agreement.

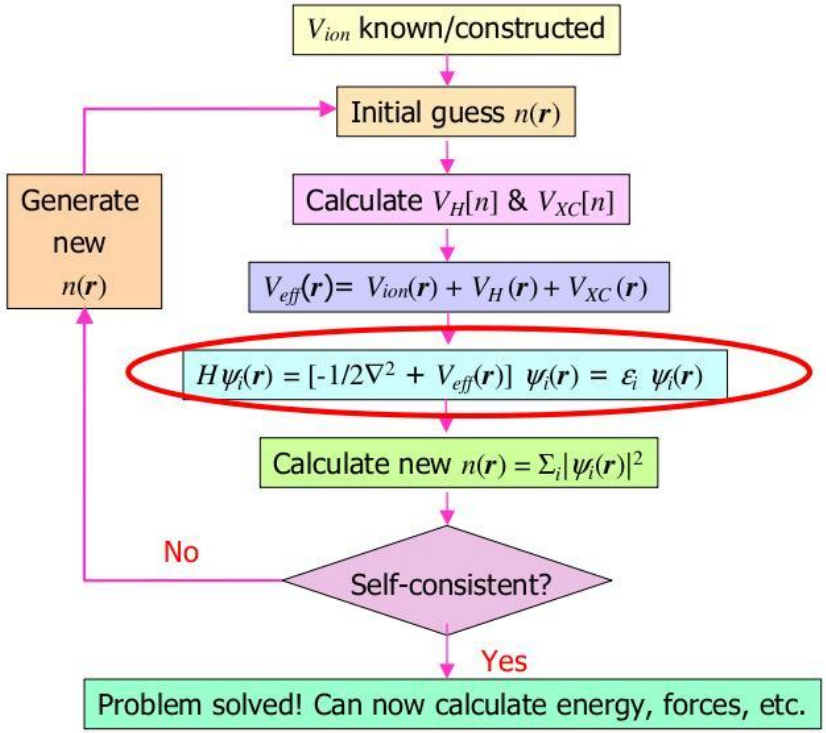


Scorecard

	0.3	0.3	0.6	1.0	0.9	0.3	1.5	0.6	0.9	0.4	1.0	0.4	0.4	6.3	13.5	1.1	2.1	0.7	1.4
0.3		0.1	0.5	0.9	0.8	0.2	1.5	0.6	0.8	0.4	1.0	0.5	0.3	6.3	13.4	1.1	2.1	0.7	1.4
0.3	0.1		0.5	0.9	0.8	0.2	1.5	0.6	0.8	0.4	0.9	0.5	0.3	6.3	13.4	1.1	2.1	0.7	1.4
0.6	0.5	0.5		0.8	0.6	0.4	1.5	0.6	0.8	0.6	1.0	0.7	0.5	6.3	13.2	1.0	1.9	0.6	1.3
1.0	0.9	0.9	0.8		0.9	0.9	1.8	0.9	1.3	1.0	1.4	1.0	0.9	6.4	13.0	1.2	1.8	1.0	1.6
0.9	0.8	0.8	0.6	0.9		0.8	1.7	0.7	1.1	0.8	1.3	1.0	0.8	6.5	13.2	1.1	1.8	0.8	1.5
0.3	0.2	0.2	0.4	0.9	0.8		1.5	0.5	0.8	0.3	1.0	0.5	0.3	6.2	13.4	1.0	2.0	0.6	1.4

PW-PP method, iterative diagonalization

The solution of $(H - \epsilon)\psi_i = 0$ for a large $N \times N$ matrix costs $T_{CPU} = \mathcal{O}(N^3)$. Too much for most applications: N, the number of PWs, can be very large for large supercells...
...but we actually need only the lowest occupied $M << N$ eigenvectors. Solution:
Iterative diagonalization. Based on iterative refinement of a trial solution. Refinement is stopped when the reached accuracy is deemed sufficient. Typical algorithm: *Block Davidson*.



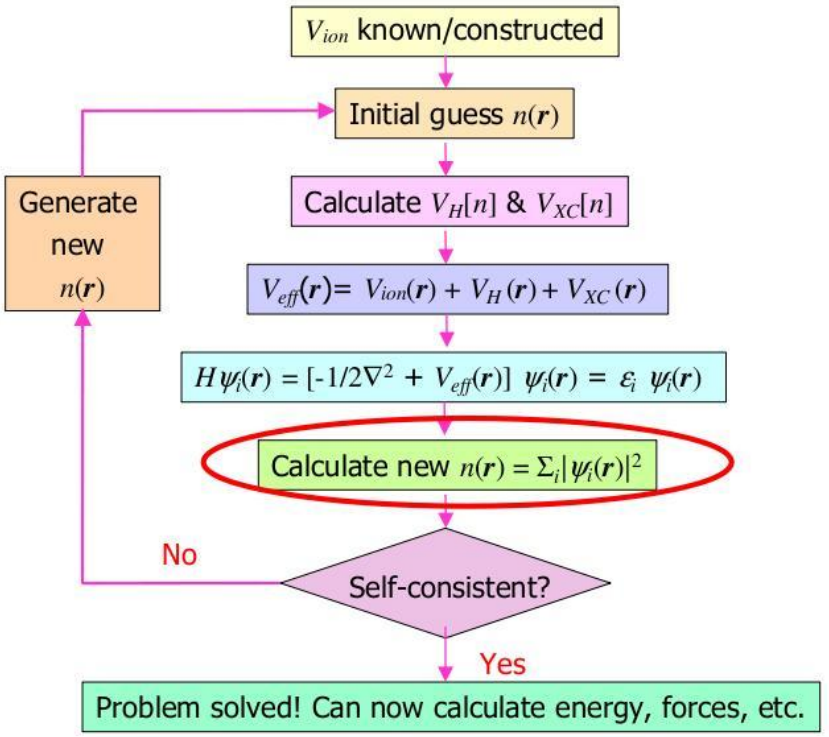
Iterative diagonalization is very convenient in conjunction with SCF iteration:

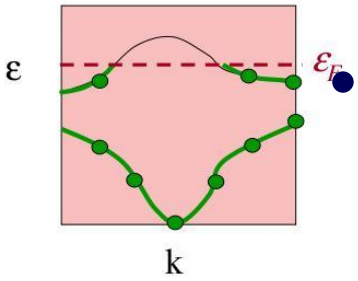
- high accuracy not needed in the first iterations
- starting trial wavefunctions available from previous iteration
- needed approximate inverse matrix easily calculated (H is diagonally dominated)

5. Brillouin Zone sampling

- The calculation of the charge density (and of many other quantities) requires sums over all **k**-points in the Brillouin Zone (BZ); in practice, some form of *BZ sampling* is needed. Convergence wrt **k**-point sampling *must be tested*!

- For insulators in large supercells, amorphous systems, liquids, molecules, sampling with Γ ($\mathbf{k} = 0$) only is fine. For other insulators, a small number of **k**-points is usually sufficient.





- For metals, a very fine sampling of the *Fermi surface*, together with some *broadening*, or *smearing*, technique, is needed. One could in principle use Fermi-Dirac occupations at finite T, but this would require either very high T or too fine sampling

Metals: broadening technique

The practical way to deal with metals uses a broadening σ in the following way:

$$\sum_i f_i \epsilon_i \to \int_{-\infty}^{\epsilon_F} \delta(\frac{\epsilon - \epsilon_i}{\sigma}) \epsilon d\epsilon = \sum_i \theta_i \epsilon_i + \sum_i \delta_i,$$

where $\delta(x)$ is a gaussian or similar function centered around $x = 0$,

$$\theta_i = \int_{-\infty}^{\epsilon_F} \delta(\frac{\epsilon - \epsilon_i}{\sigma}) d\epsilon, \quad \delta_i = \int_{-\infty}^{\epsilon_F} \delta(\frac{\epsilon - \epsilon_i}{\sigma})(\epsilon - \epsilon_i) d\epsilon = \sigma^2 \int_{-\infty}^{(\epsilon_F - \epsilon_i)/\sigma} x\delta(x) dx,$$

and ϵ_F is determined by the condition $\sum_i f_i =$ number of electrons. It is equivalent to introduce a fictitious "temperature" σ/k_B and the corresponding "free energy".

Specially tailored broadening functions (Marzari-Vanderbilt, Methfessel-Paxton), ensure fast convergence. This must be tested by performing several runs at different σ for increasingly dense k-point grids, until a suitable k-point grid and σ are found yielding satisfactorily converged results.

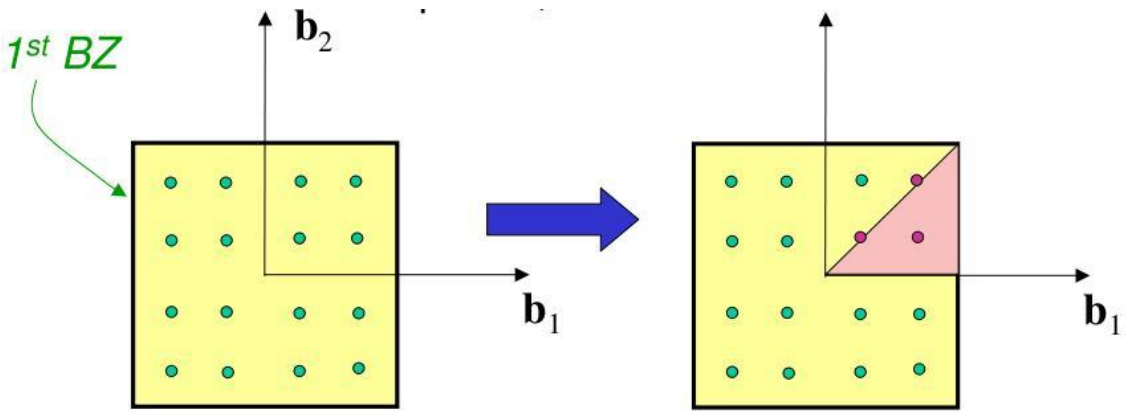
Grids of k-points for Brillouin Zone sampling

Typical ways of sampling the Brillouin Zone (BZ):

- **Special Points** (e.g. Baldereschi, Chadi and Cohen)
  Points designed to give quick convergence for particular crystal structures.

- **Uniform Grids** (e.g. Monkhorst-Pack)
  Equally spaced in reciprocal space. May be centred on origin ("non-shifted") or not ("shifted").

In presence of *symmetry*, only **k**-points in the *Irreducible* BZ, or IBZ, need to be computed: the charge density is reconstructed using symmetry. Appropriate *weights* for **k**-points must be specified (or can be calculated).



Plane-Wave Pseudopotential method, summary

- Supercell geometry: lattice vectors + atoms in the unit cell
- Plane-wave basis set, determined by the lattice and by a single parameter (*cutoff*)
- Atom-based pseudopotentials representing the electron-nuclei potential (V_{nuc} in the figure)
- Charge density computed with valence electrons only, on a suitable grid of **k**-points.