

# Electronic Structure Theory

ab-initio simulations of materials properties

Davide Romanin

Department of Applied Science (DISAT)  
Politecnico di Torino, Torino

May 6, 2020

# Table of Contents

- Introduction
- Density Functional Theory: the Hohenberg-Kohn theorems
- Density Functional Theory: the Kohn-Sham equations
- The exchange-correlation functional
- Plane-wave basis set
- Pseudopotentials
- Density Functional Perturbation Theory
- Introduction to Wannier-interpolation (maybe)

# The Exchange-Correlation Hole

We would like to give a more physical interpretation to the exchange-correlation functional  $E_{XC}[n]$ . This term arises from the interaction among particles, i.e. due to electron-electron correlation. We consider a Coulomb interaction which is adiabatically turned on:

$$U_{\lambda} = \lambda \frac{e^2}{2} \sum_{ij} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} = \lambda U \quad (1)$$

where  $\lambda \in [0; 1]$ , which can varies continuously such that:

- $\lambda = 0 \rightarrow$  independent particle system;
- $\lambda = 1 \rightarrow$  fully interacting particle system;

Consider now the following hamiltonian:

$$\hat{H}(\lambda) = \hat{T} + \lambda U + \int V(\mathbf{r}, \lambda) n(\mathbf{r}) d\mathbf{r} \quad (2)$$

such that  $n(\mathbf{r})$  is always the prescribed groundstate density for every choice of  $V(\mathbf{r}, \lambda)$  (V-representability). We also assume that the groundstate wavefunction  $|\psi(\lambda)\rangle$  is a continuous, differentiable function of  $\lambda$ .

It is clear that:

- $\hat{H}(\lambda = 0) \rightarrow$  Kohn-Sham hamiltonian and  $V(\mathbf{r}, \lambda = 0) = V_{KS}$ ;
- $\hat{H}(\lambda = 1) \rightarrow$  true hamiltonian of the system;

Starting from the Hellman-Feynman theorem, it is possible to derive an integral expression (the *adiabatic connection formula*), which provides a way to compute finite energy differences between any two states connected by continuous variations of the hamiltonian:

$$\Delta E = \int_{\lambda_1}^{\lambda_2} d\lambda \frac{\partial E}{\partial \lambda} = \int_{\lambda_1}^{\lambda_2} d\lambda \langle \psi(\lambda) | \frac{\partial \hat{H}}{\partial \lambda} | \psi(\lambda) \rangle \quad (3)$$

If  $[\lambda_1; \lambda_2] = [0 : 1]$ , we have a smooth interpolation of the Hamiltonian between the Kohn-Sham Hamiltonian and the true Hamiltonian of the system.

# The Exchange-Correlation Hole: derivation I

# The Exchange-Correlation Hole: derivation II

# The Exchange-Correlation Hole: derivation III



- the exchange-correlation hole is related to the pair-correlation function:

$$f_{XC}(\mathbf{r}, \mathbf{r}') = \int_0^1 d\lambda [g_\lambda(\mathbf{r}, \mathbf{r}') - 1] \quad (4)$$

- It obeys the sum rule:

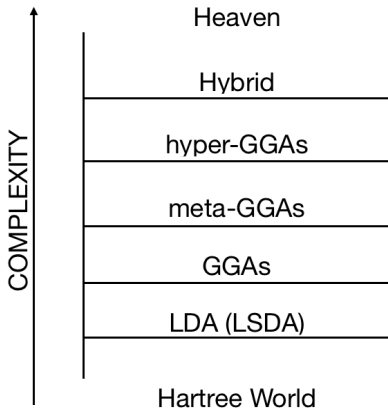
$$\int f_{XC}(\mathbf{r}, \mathbf{r}') d\mathbf{r}' = -1 \quad (5)$$

- it is responsible for the self-interaction problem in DFT

# Jacob's Ladder

Perdew-Schmidt 2011, Perdew et al. 2005.

It gives a classification of XC functionals based on their accuracy and on their computational complexity.



# Local (spin) density approximation L(S)DA

Solids can often be considered as close to the limit of the homogeneous electron gas. In that limit the effects of exchange and correlation are **local** in character: we assume that the exchange-correlation energy density at each point is the same as in a homogeneous electron gas with that density.

$$\begin{aligned}
 E_{XC}^{LSDA}[n^\uparrow, n^\downarrow] &= \int n(\mathbf{r}) \epsilon_{XC}^{hom}(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r})) d^3r = \\
 &= \int n(\mathbf{r}) [\epsilon_X^{hom}(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r})) + \epsilon_C^{hom}(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r}))] d^3r
 \end{aligned} \tag{6}$$

For the unpolarized system we have the LDA:  $n^\uparrow(\mathbf{r}) = n^\downarrow(\mathbf{r}) = n(\mathbf{r})/2$   
 Moreover  $\epsilon_X^{hom}(n(\mathbf{r})) \propto n(\mathbf{r})^{1/3}$  and  $\epsilon_C^{hom}(n(\mathbf{r}))$  can be "accurately" computed via Quantum Monte Carlo simulations.

LDA works very well for elements that can be described by the nearly-free electron model (s and p elements), and is good for the 3d elements if we consider LSDA (Local Spin Density Approximation).

## PROBLEMS:

- underestimates stability of an atom;
- too high binding energy;
- OVERBIND in solids;
- STIFFNESS in solids;

# Generalized Gradient Approximations GGA

Real systems are clearly inhomogeneous, with spatially varying electric fields due to nuclei and screening. As a first step beyond LDA, we may consider an expansion of the functional in terms of the gradient of the electronic density, i.e.  $|\nabla n(\mathbf{r})|$ ,  $|\nabla n(\mathbf{r})|^2$ ,  $\nabla^2 n(\mathbf{r})$ , ... (Gradient expansion approximation, GEA). However, gradients may become so large that an expansion is no longer justified.

Therefore, the idea is to build a functional which depends on gradients of the density modifying the behavior of large gradients in order to preserve desired properties (GGA):

$$E_{XC}^{GGA}[n^\uparrow, n^\downarrow] = \int n(\mathbf{r}) \epsilon_{XC}(n^\uparrow(\mathbf{r}), n^\downarrow(\mathbf{r}), |\nabla n^\uparrow(\mathbf{r})|, |\nabla n^\downarrow(\mathbf{r})| \dots) \quad (7)$$

Different GGA make various assumptions on  $\epsilon_{XC}^{GGA}$ , the most used is the PBE-GGA (Perdew-Burke-Ernzerhof) [?]. GGA gives good results for transition metals.

## PROBLEMS:

- a little bit more expensive than LDA;
- better stability of an atom w.r.t. LDA;
- better binding energy w.r.t. LDA;
- UNDERBIND in solids;
- SOFTNESS in solids;

# Hybrid functionals

They are combination of the exact Hartree-Fock exchange functional  $E_x^{HF}$  and a density functional approximation exchange-correlation functional  $E_x^{DFA}$  (i.e. LDA, GGAs,...). They are the most accurate functionals available, as far as energetics is concerned., however they are costly from the computational point of view.

$$\text{HF: } E = T + \int n(\mathbf{r})V(\mathbf{r})d\mathbf{r} + E_{\text{hartree}} - \underbrace{\frac{e^2}{2} \sum_{ij} \int \frac{\psi_i^*(\mathbf{r})\psi_j(\mathbf{r})\psi_j^*(\mathbf{r}')\psi_i(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'}_{E_x^{HF}} \quad (8)$$

$$\text{Hybrid-DFT: } E = T + \int n(\mathbf{r})V(\mathbf{r})d\mathbf{r} + E_{\text{hartree}} + \underbrace{\alpha_x E_x^{HF} + (1 - \alpha_x) E_x^{DFT}}_{E_{xc}^{DFT}} + E_c^{DFT} \quad (9)$$

The most famous hybrid functional is the B3LYP functional developed by Becke, which is an empirical hybrid functional given by:

$$E_{XC} = E_{XC}^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x E_x^{Becke} + a_c E_c^{LYP} \quad (10)$$

where coefficients are empirically adjusted to fit atomic and molecular data.

The hybrid functionals corrects most of the LDA/GGA errors. In a plane-wave implementation,  $T_{CPU,hybrid} \propto 10 \times T_{CPU,LDA/GGA}$ .



# Diagonalizing the KS hamiltonian

Problem: we want to diagonalize the KS hamiltonian Solution: expand the wavefunction into a suitable basis set, orthonormal and complete

$$\psi(\mathbf{r}) = \sum_i c_i \phi_i(\mathbf{r}) \Rightarrow \sum_j (H_{ij} - \epsilon_j \delta_{ij}) c_j = 0 \quad (11)$$

which is the SECULAR EQUATION and where  $H_{ij} = \langle \phi_i | \hat{H} | \phi_j \rangle$ .

In a solid it is convenient to use PBC for a volume  $\Omega \rightarrow \infty$ : plane waves are the natural choice for periodic systems and provide simple algorithms for practical calculations. Indeed, any periodic functions can be expanded in the complete set of Fourier components.

# Plane-wave expansion

A PW basis set is such that:

$$\langle \mathbf{r} | \mathbf{k} + \mathbf{G} \rangle = \frac{1}{\sqrt{N\Omega_{\text{cell}}}} e^{i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}} \quad (12)$$

For a given  $\mathbf{k}$ :

$$|\psi_{i\mathbf{k}}\rangle = \sum_{\mathbf{G}} c_{i,\mathbf{k}+\mathbf{G}} |\mathbf{k} + \mathbf{G}\rangle \quad (13)$$

$$\langle \mathbf{k} + \mathbf{G} | \mathbf{k} + \mathbf{G}' \rangle = \delta_{\mathbf{G}\mathbf{G}'} \Rightarrow \text{ORTHONORMALITY} \quad (14)$$

$$c_{i,\mathbf{k}+\mathbf{G}} = \langle \mathbf{k} + \mathbf{G} | \psi_i \rangle = \tilde{\psi}_i(\mathbf{k} + \mathbf{G}) = \frac{1}{\sqrt{N\Omega_{\text{cell}}}} \int \psi_i(\mathbf{r}) e^{i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}} d\mathbf{r} \quad (15)$$

# Energy and Density Cutoffs

In PW-DFT codes we have to specify an energy cutoff. Indeed, we cannot have a complete expansion in terms of infinitely many planewaves. It can be shown that contribution of higher frequency planewaves are less than those of lower frequency. Thus we consider only wavevectors that satisfy:

$$\frac{1}{2}|\mathbf{k} + \mathbf{G}|^2 < E_{\text{CUT}} \quad (16)$$

By doing so we are limiting the number of plane-waves allowed for our problem. This cutoff depends on the system and in principle should be chosen by a convergence study. Moreover, the density requires Fourier components that extend twice as far in each direction as those needed for the wavefunction because  $n \propto |\Psi|^2$ . As a consequence, the cutoff for the electronic density is  $4 \cdot E_{\text{CUT}}$ .

# Consequences of the energy cutoff

# Derivation of the density cutoff

# Brillouin zone integration

For many quantities, it is necessary to perform integrals in the  $k$ -space. For a general function  $f(\mathbf{k})$ , the average value per cell is just an integral over the Brillouin zone (BZ):

$$\bar{f} = \frac{1}{\Omega_{BZ}} \int_{BZ} f(\mathbf{k}) d^3k \quad (17)$$

where  $\Omega_{BZ}$  is the volume of the Brillouin zone.

Numerically, this integral is converted into a sum over a discretized  $\mathbf{k}$ -space:

$$\bar{f} = \frac{1}{N_k} \sum_{\mathbf{k}} f(\mathbf{k}) \quad (18)$$

where  $N_k$  is the number of  $\mathbf{k}$  points in the discretized space.

The question is: how can we choose this discretization in an optimal way?

The general method proposed by Monkhorst and Pack, leads to a uniform set of points determined by a simple formula valid for any crystal:

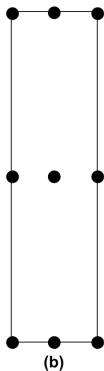
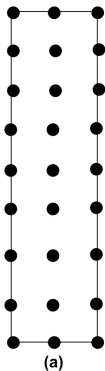
$$\mathbf{k}_{n1,n2,n3} = \sum_i^3 \frac{2n_i - N_i - 1}{2N_i} \mathbf{G}_i \quad (19)$$

where  $\mathbf{G}_i$  are the primitive reciprocal lattice vectors,  $n_i = 1, \dots, N_i$  and  $N_i$  is the number of  $\mathbf{k}$  points to be chosen in each direction.

For many lattices, especially cubic, it is preferable to choose  $N_i$  to be even. In this case the set doesn't involve the highest symmetry points, it omits the  $\mathbf{k} = 0$  point and points on the BZ boundary.

Since each new  $\mathbf{k}$ -point means a new self-consistent Kohn-Sham system, it's crucial to maximize the accuracy while reducing the size of the  $\mathbf{k}$ -point set as much as possible.

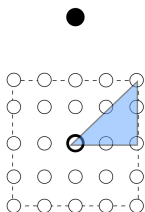
# Uniformity of the k-grid



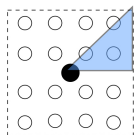
Mesh points that are uniformly distributed are preferred over those that have different densities in different dimensions. It is all about the accuracy with which we resolve things in different dimensions. As we can see from the figure on the left it is preferable the first situation (a) instead of the second one (b).



# Symmetries and the irreducible Brillouin zone



We can gain computational time if we use the symmetries of the system. In this way not all k-points would be used when performing the self-consistent cycle: k-points are connected to each other by symmetry operations! We can therefore define a irreducible Brillouin zone (IBZ), which is the minimum set of points through which we can generate the entire k-grid by symmetry operations.

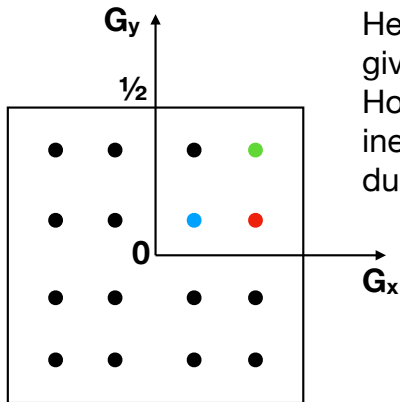


Therefore, the integrals/sums over the BZ will be converted into integrals/sums over the IBZ:

$$\bar{f} = \sum_{\mathbf{k}}^{\text{IBZ}} w_{\mathbf{k}} f(\mathbf{k}) \quad (20)$$

where  $w_{\mathbf{k}}$  are the relative weights.

# Example: sampling the BZ



Here we consider  $N_x = N_y = 4$ , giving a total of 16  $k$  points. However there are only 3 inequivalent points in the IBZ due to symmetries of the square.

$$k_1, \omega_{k_1} = 1/4$$

$$k_2, \omega_{k_2} = 1/4$$

$$k_3, \omega_{k_3} = 1/4$$

# Insulators vs metals

In terms of  $k$ -integration, the insulators are well behaved: the density of states goes to zero smoothly before the gap and the integration of a smooth function does not cause any trouble for numerical implementations (generally). On the other hand, metals have a sharp discontinuity given by the presence of the Fermi function, which is a sharp step function at zero temperature.

In the smearing approach the local density of states is convoluted with a smearing function:

$$f(\epsilon) = \frac{1}{\sigma} \tilde{\delta}\left(\frac{\epsilon}{\sigma}\right) \quad (21)$$

In order to overcome this numerical problem, we introduce an artificial smearing which makes the integrand smoother. In principle, as  $N_k \rightarrow \infty$  the smearing should tend to zero, however this cannot be achieved in practice. Therefore:

$$n(\mathbf{r}, \epsilon) = \sum_{i\mathbf{k}} \frac{1}{\sigma} \tilde{\delta}\left(\frac{\epsilon - \epsilon_{i\mathbf{k}}}{\sigma}\right) |\psi_i(\mathbf{r})|^2 \Rightarrow N = \int_{-\infty}^{\epsilon_F} n(\epsilon) d\epsilon = \sum_{i\mathbf{k}} \tilde{\theta}\left(\frac{\epsilon_F - \epsilon_{i\mathbf{k}}}{\sigma}\right) \quad (22)$$

where  $\tilde{\theta}(x) = \int_{-\infty}^x \tilde{\delta}(x') dx'$

# Smearing function

- **Fermi-Dirac function:** We introduce an artificial temperature in the system, so that the step function can be approximated by a Fermi-Dirac distribution:

$$\tilde{\delta}\left(\frac{\epsilon_F - \epsilon}{\sigma}\right) = \frac{1}{e^{(\epsilon - \epsilon_F)/\sigma} + 1} \quad (23)$$

- **Gaussian:** The smearing is obtained by approximating the delta function with a Gaussian;
- **Methfessel-Paxton:** The delta function is approximated with a linear combinations of Hermite polynomials and Gaussians:

$$\tilde{\delta}(x) \approx \sum_n^N A_n H_{2n} e^{-x^2} \quad (24)$$

- **Marzari-Vanderbilt:** The delta function is approximated by a Gaussian multiplied by a first order polynomial:

$$\tilde{\delta} \approx \frac{1}{\sqrt{\pi}} e^{-[x - (1/\sqrt{2})]^2} (2 - \sqrt{2}x) \quad (25)$$

# Pseudopotential method

In a plane-wave expansion we have the problem of representing localized functions  $\Rightarrow$  high number of plane-waves. Indeed:

- CORE wf: sharply peaked close to nuclei;
- VALENCE wf: oscillate rapidly close to the core regions (orthogonality);

The basic idea is to replace the strong Coulomb potential of the nucleus and the effects of the tightly bound core electrons by an effective ionic potential acting on valence electrons.

Pseudopotentials are not unique, there is the freedom to choose forms that simplify the calculations and the interpretation of the resulting electronic structure.

# Full-potential or all-electron method

It is an accurate method but slow from the computational time point of view.

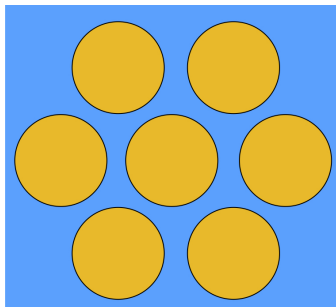
The only part which is approximated is the Coulomb potential due to the nucleus, which is usually treated as a spherically-symmetric potential.

Then we take fully into account both core and valence electrons. The effective potential is split into an interstitial part near the atom and a "free-electron"-like for the regions between them.

Some methods are:

- Augmented plane-waves (APW);
- Linearized augmented plane-waves (LAPW);
- Full-potential LAPW (FLAPW);
- Projector-augmented waves (PAW), which is a mix between FP and PP;

# Augmented plane-wave expansion



**Muffin-tin potential:** The total effective potential is approximated as spherically symmetric  $V_{KS}(\mathbf{r}) \rightarrow V_{KS}(r)$  within each sphere and constant  $V_{KS}(\mathbf{r}) \rightarrow V_0$  in the interstitial region.

In the region around each atom the potential is similar to the potential of the atom, in the interstitial region between atoms the potential is smooth.

The wavefunctions can be represented in terms of spherical harmonics around each atom and plane waves in between. The entire problem is recast into a matching or boundary condition problem.

APW treat highly localized atomic-like states using atomic-like spherical functions and delocalized states using delocalized planewaves, solving separate problems in different regions and trying to match them.

The disadvantage is the difficulty of this matching problem and solving non-linear equations in this basis.



# Characteristics of a pseudopotential

We make an arbitrary distinction between *valence* and *core* states:

- The core states are assumed to change very little due to change in the environment: their effect is replaced by a model potential derived in the atomic configuration (**transferability**).
- The valence states are seen to oscillate rapidly close to the core regions so we try to make them as smooth as possible (**smoothness**).

Accuracy and transferability generally lead to the choice of a small cutoff radius  $R_C$  and **hard** potentials, since one wants to describe the wavefunction as well as possible in the region near the atom.

Smoothness of the resulting pseudofunctions generally leads to the choice of a large cut-off radius  $R_C$  and **soft** potentials, since one wants to describe the wavefunctions with as few basis functions as possible.

# Pseudopotential method: Norm-conserving pseudopotentials

Conditions:

- 1 All-electron (AE) and pseudo-valence (PS) eigenvalues agree for the chosen atomic reference configuration

$$\hat{H}^{AE} |\Psi_i^{AE}\rangle = \epsilon_i |\Psi_i^{AE}\rangle$$

$$\hat{H}^{PS} |\Psi_i^{PS}\rangle = \epsilon_i |\Psi_i^{PS}\rangle$$

- 2 All-electron and pseudo-valence wavefunctions agree beyond a chosen core radius  $R_C$

$$|\Psi_i^{AE}\rangle = |\Psi_i^{PS}\rangle, \text{ for } r \geq R_C$$

- 3 Gauss's theorem: the total charge in the core region is correct

$$\int_0^{R_C} r^2 |\Psi_i^{AE}(\mathbf{r})|^2 dr = \int_0^{R_C} r^2 |\Psi_i^{PS}(\mathbf{r})|^2 dr$$

- 4 Phase shifts of the real potential, as those from scattering theory, must be reproduced;

# Pseudopotential method: Norm-conserving vs Ultrasoft pseudopotentials

Norm conservation and reproducibility of scattering properties  $\Rightarrow$  TRANSFERABILITY.  
Norm-conserving pseudopotentials achieve the goal of accuracy, usually at some sacrifice of smoothness thus requiring more basis functions. For some elements it is easy to obtain "soft" norm-conserving PP, but in general they are hard ( $E_{cut} > 70$  Ry).

The ultrasoft (Vanderbilt) pseudopotentials reach the goal of accurate calculations by a unitary projection that re-expresses the problem in terms of a smooth function and an auxiliary function around each ion core that represents the rapidly varying part of the density. Relaxation of the norm conserving condition allows smoother wavefunctions, and hence lower cut-off energies. This is advantageous in reducing the size of the plane wave basis set used. However this requires more operation per computational cycle, but the gain in terms of the number of basis functions needed dominates for large systems ( $E_{cut} \sim 25 \div 35$  Ry).