

Density Functional Theory for Beginners

Basic Principles and Practical Approaches

Fabio Finocchi

*Institut des NanoSciences de Paris (INSP)
CNRS and University Pierre et Marie Curie*

October 24, 2011

¹To the memory of Elena, who made me feel any small progress along the completion of these notes extremely important.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 2 |
| 2 | Basic DFT theory | 5 |
| 2.1 | Why is the problem difficult | 5 |
| 2.1.1 | Functionals | 6 |
| 2.2 | Electron density as the basic variable | 7 |
| 2.3 | Towards the definition of the energy functional | 9 |
| 2.4 | Finding the ground state: total energy and electron density | 11 |
| 2.4.1 | Minimizing $E[n]$ | 11 |
| 2.4.2 | The Kohn-Sham equations | 13 |
| 2.4.3 | Computing the total energy | 15 |
| 2.5 | The Exchange-Correlation energy functional | 16 |
| 3 | Practical Implementation | 20 |
| 3.1 | Solving the Kohn-Sham equations: basis sets | 20 |
| 3.1.1 | Pseudo-potentials | 22 |
| 3.1.2 | Plane waves | 23 |
| 3.2 | Solving the Kohn-Sham equations: the self-consistent cycle | 30 |
| 4 | Density Functional Perturbation Theory | 34 |
| 4.1 | The basic formalism. | 34 |
| 4.2 | First-order energy derivatives and atomic forces. | 35 |
| 4.3 | Second order | 38 |

Chapter 1

Introduction

Numerical Simulations and Materials Science Physicists have a talent for producing equations that they are quite unable to solve. This is likely the basic reason why numerical simulations have become ubiquitous in most fields of physics, including materials science, condensed matter physics and nano-sciences. Experiments can often be difficult to carry out, expensive and sometimes provide only indirect information. For instance, X-ray diffraction must be complemented by appropriate computer programs in order to find out atomic positions in a complex crystal. Moreover, X-ray diffraction can identify where atoms are, not why the actual structure is preferred to another one. The interpretation of the experimental data often requires hypothesis or models that should be validated independently. Numerical simulations can provide a valuable help to explain the experimental data on the basis of a microscopic model and give insights into the basic mechanisms of interaction between atoms. Simulations can also be useful to obtain the numerical solution of equations that cannot be solved by analytical means, and provide benchmarks for approximated theories or simplified models.

The standard use of numerical simulations is ultimately due to the fast improvements of the computer power and the development of new and more and more effective numerical methods. As far as materials science is concerned, some important dates must be recalled.

1. The invention of the first device that resembles modern computers in the 40's;
2. N. Metropolis proposed a new algorithm, based on the Monte Carlo method, (1) to sample a statistical distribution (2). It is worth noting, in this respect, that these methods benefited from algorithm for generating pseudo-random numbers and the fast development of computing power;
3. In 1959, B.J. Alder and T.E. Wainwright proposed a way to integrate the equations of motion for N particles numerically (3). This is the beginning of Molecular Dynamics simulations.
4. In 1985, R. Car and M. Parrinello designed a unified approach for Molecular Dynamics and first-principle calculations within the Density Functional Theory (4). Their work opened a way to simulate the dynamics of condensed matter systems for which semi-empirical interatomic potentials are inappropriate.

Ab initio methods In order to investigate systems at the nanoscale, it is necessary to model the interactions between atoms as accurately as possible, while ensuring that the calculations remain computationally feasible. Often, empirical interatomic potentials, which are usually fitted in order to reproduce some measured physical properties, are adopted. A well known example is provided by Lennard-Jones potentials, of the form: $V(r) = 4\epsilon \left[\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6} \right]$, where r is the interatomic distance and ϵ and σ two adjustable parameters. The Lennard-Jones potential is well suited to the simulation of van der Waals interactions. However, the chemical bond in strongly covalent systems, such as Carbon or Silicon, cannot be simulated by LJ potentials. Interatomic potentials are limited by the accuracy of the parametrisation, and correspondingly, their transferability to different environments or their ability to reproduce physical properties, other than those which they were designed for, can be poor. Indeed, the interaction between atoms does depend on their neighbors and, more generally, on their environment. For instance, surface atoms experience a different electrostatic field (generally weaker) and have a reduced number of neighbors

than atoms in the bulk. For example, bulk-derived potentials may fail in describing molecules, surfaces or systems at the nanoscale, where the atomic environment strongly differ from that in the perfect crystal.

Furthermore, the type of bonding ultimately depends on the electronic structure of the material, which is affected by its short and long range order. The interplay between the electronic and structural degrees of freedom (atomic positions and coordination number, space group if any, medium range order in disordered or polycrystalline materials, etc.) is well documented in materials science. Usually, interatomic potentials are fitted to a particular system or to a class of materials, with implicit assumptions regarding the type of bonding. Systems showing a chemical complexity (i.e. crystals with several chemical species, or disordered materials with many inequivalent impurities) are rather difficult to be treated by interatomic potentials. A large number of distinct bonds should be accounted for, which could be done by introducing many adjustable parameters in the potential, at the expense of reducing its simplicity and raising the problem of testing its reliability in a large number of cases. Even for elemental systems, interatomic potentials may be inappropriate, as documented by the case of a very important material such as Si. Most available interatomic potentials for Si are indeed fitted to its diamond structure, which is semiconducting, with Si atoms in the characteristic tetrahedral coordination. All of them provide a bad description of the liquid state, which is metallic with Si atoms having a coordination number close to 6. Another example is bond breaking or bond formation during chemical reactions, which are usually accompanied by the redistribution of valence electrons.

Last but not least, all the electronic properties of materials are out of reach of interatomic potentials, which do not account explicitly for the electronic structure. Nowadays, several techniques probe the electronic rather than the atomic structure. This is the case of Scanning Tunneling Microscopy, photon and electron spectroscopies and other experimental methods. For all these reasons, theoretical approaches that give access to the electronic structure are highly desirable in the field of nano- and materials sciences.

Modelling interactions with no a priori knowledge of the bonding requires the use of sophisticated methodologies, in which one attempts to solve the Schrödinger equation for the electrons. All the methods that are based upon the determination of the bonding starting from the basic theorems of quantum mechanics are called *ab initio* (or first-principles) methods. They are not necessarily exact, since some approximations are generally made at some stage of the theory or of its practical implementation, in order to make calculations feasible. However, first-principles approaches make no use of experimental input and, although quite computationally intensive, they have the advantage of being completely transferable, requiring only the atomic constituents of the system under consideration to be specified. *Ab initio* methods are thus more intellectually appealing; moreover, they may be used to calculate the properties of systems about which no a priori knowledge exists, that is, they are predictive. As a consequence, they can be used as effective methods to study the properties of materials, interpret the experiments, check the accuracy of simpler simulation schemes or even parametrise classical potentials allowing the computational restrictions associated with *ab initio* methods to be circumvented. It is for these reasons that *ab initio* approaches are widely and often routinely used nowadays in the simulation of materials. An example are infrared spectra, which yield the characteristic vibrational frequencies of a material, but give no insights on the nature of the vibrational modes unless reference spectra are available. Infrared spectra can now be interpreted through *ab initio* simulations (see chapter 4).

However, the straightforward resolution of the Schrödinger equation $H\Psi = E\Psi$ for many-electron systems is actually impossible (see 2.1). Therefore, approximated methods, such as Hartree-Fock theory and variational Monte Carlo approaches, have been implemented. Among *ab initio* methods, the density functional theory (DFT) has become the most popular one. The number of works about the implementation of DFT or using the DFT as a computational scheme is continuously increasing. DFT has become quite a standard approach, which accompanies and complements other laboratory techniques for studying materials. It seems therefore highly desirable for any materials scientist to know the basics of the theory, its practical implementation in the available codes, as well as its capabilities and shortcomings.

About this book Although many very good books on the DFT are available, most of them address a particular audience, which, very schematically, is made of graduate students who are willing to acquire a solid background in DFT in order to use it as a tool in material science and active researchers in the field of computational or theoretical physics and chemistry. All of these people must devote time to go deeply through the theory or to use DFT-based computational schemes for the calculation of materials properties. This is not the case of students at the Master level: in a limited amount of time, they have to

learn the fundamentals of DFT and go through short hands-on-computer sessions; they cannot generally dispose of a scientific environment that makes them aware of the main shortcomings and difficulties in the use (and abuse) of the DFT as a computational tool.

Having taught for some years a short course at the Master level on the DFT and its applications in Material Science, I am aware of the students' insatisfaction with the existing books or reviews. Previously, students were also provided with short notes, in the form of slides with extended text that I personally wrote. The former ones were thought of as too difficult, dense or long; the latter ones were generally felt as too schematic and not providing complementary matter, such as exercises, questions, extensive bibliography. Therefore, this book is specially conceived for them and aim at filling the lack of a simple and ready-to-use reference on the DFT and its use to compute materials properties. I do not go deeply into the theory and treat only the basic theorems that are at the foundations of DFT. More than formal proofs, I provide some simple exercises or illustrative examples, often taken from other physical problems. Such complementary material is an important part of the book, although formally distinguished from the main text. I mention some of the most recent advances in the field of DFT, without aiming at discussing them thoroughly. Unavoidably, I made use of my personal opinion and experience in order to select topics, give an advice about the shortcomings of a particular approach and suggest possible further reflexions to the reader. I apologise for any lack of citations or for having skipped relevant subjects; my scope is not to cover the entire DFT but to introduce the reader to the theory as quickly and simply as possible and, possibly, encourage her/him to read review articles or books on the subject. Sometimes, I summarize the main messages in each chapter, thus helping the reader to review the related concepts. The reader is highly advised to check her/his knowledge by reviewing the exercises that are proposed in the text.

Chapter 2

Basic DFT theory

2.1 Why is the problem difficult

Let's consider as a simple example benzene (C_6H_6), which is a rather common molecule, experimentally well known. C_6H_6 has $N = 42$ electrons in total. The many-electron wavefunction $\Psi(\vec{r}_1, \sigma_1; \dots \vec{r}_N, \sigma_N)$ ¹ is a complex scalar field. Its spatial part, which depends on the $3N$ electron coordinates, is defined in a cartesian space of dimension $3N = 126$. Let's suppose we choose a discrete mesh to represent the wave function in space. The diameter of the benzene molecule is about 4.5 Å; we take a cubic box 10 Å wide, in which the occupied electronic states of the molecule are well localized. We also take $M = 50$ points to interpolate the wave function on each spatial coordinate, resulting in a mesh with points that are 0.2 Å distant from each other. Therefore, the determination of any matrix element in the real space requires $M^{3N} = 50^{126}$ operations. In order to know the ground state wave function, by using the variational principle, one has to minimize the matrix element $\langle \Psi H \Psi \rangle / \langle \Psi \Psi \rangle$, where H is the Hamiltonian. This requires a repetition of 50^{126} operations, several times. For the time being, the most powerful computers can carry out about 10^{18} operations per second. Therefore, it is clear that it is impossible to find even the ground-state wavefunction of benzene by such a straightforward method (not to even mention the excited states), since this would require more than a human lifetime.

A solution to such a problem was proposed by Hartree and Fock (see, e.g., the book by Ashcroft and Mermin), who made a simplified assumption about the wavefunction; Ψ is written as a $N \times N$ determinant of N one-particle orbitals: $\Psi = \text{Det}(\phi_1, \dots, \phi_N)$. Then, the energy of the N -electron system is minimized giving the determinantal form of the wavefunction and a system of N equations for the one-particle orbitals $\langle \phi_i \rangle$ is derived. The potential term V_i for the i -th orbital ϕ_i depends on the other ϕ_j ($j \neq i$). V_i is usually called the *self-consistent field*, which means that it has to be found consistently with the wavefunction $\text{Det}(\phi_1, \dots, \phi_N)$. From mathematical point of view, the Hartree-Fock equations are integro-differential equations, thus more involved to be solved than a Schrodinger equations with a fixed external potential of the type $\left[-\frac{\hbar^2 \nabla^2}{2m} + V(\vec{r}) \right] \phi(\vec{r}) = \varepsilon \phi(\vec{r})$. Nevertheless, the many-electron problem can be treated in the framework of the Hartree-Fock method, which is implemented in several scientific packages. This is mainly due to the fact that, instead of dealing with the Schrodinger equation of the true N -electron wavefunction, the Hartree-Fock theory allows to solve N coupled one-electron equations, which is a much easier task from the computational point of view.

A basic idea to simplify the N -electron problem consists in finding some physical quantity that defines the system uniquely without growing in complexity as a function of the number of electrons N . Obviously, this cannot be the N -electron wavefunction, as shown before. A step further was done in the sixties with the introduction of the DFT (5; 6), which focuses on the electron density rather than on the wavefunction. We will see that the current implementations of the DFT are also based on a set of self-consistent equations, namely the Kohn-Sham equations, which are analogous in spirit to the Hartree-Fock equations.

¹ \vec{r}_j and σ_j are the spatial and spin coordinate relative to the j th electron, respectively. The wavefunction must obey the Pauli principle, that is, it must be anti-symmetric when two electrons are exchanged.

2.1.1 Functionals

Before going further, let's briefly recall what a functional is and give some basic rules to compute functional derivatives. A functional is traditionally a map from a vector space to the field underlying the vector space, which is usually made by real or complex numbers. In other words, a functional takes a function as its argument or input and returns a scalar. Its use goes back to the calculus of variations, where one searches for a function which minimizes a certain functional. A particularly important application in physics is to search for a state of a system which minimizes the energy functional. Formally, a functional is defined on a space of functions, with values in the real or complex domain:

$$F : f \longrightarrow \mathfrak{R} \quad \text{or} \quad F : f \longrightarrow \mathcal{C} \quad (2.1)$$

A functional is usually written using square brackets, in order to distinguish functionals from functions. Examples of some (simple) functionals are given below:

$$F[f] = \int dx f(x) \quad (2.2)$$

$$F[f] = \int dx \int dy f(x)f(y)g(x, y) \quad (2.3)$$

$$F[f] = \int dx \int dy \frac{\partial f(x, y)}{\partial x} \frac{\partial f(x, y)}{\partial y} g(x, y) \quad (2.4)$$

$$F[f] = \int dx g(f(x)) \quad (2.5)$$

where g is a well defined function. Note that the argument of f is completely irrelevant: the last functional $F[n]$ can be written as $F[f] = \int dx g(f(x))$ as well as $F[f] = \int dz g(f(z))$.

Given a function of one variable, let's say $z = f(x)$, the variation of z may be due to variations of x ($x \rightarrow x + dx$) at fixed f or to variations of f , that is, of the type of functional dependence. The first case is found in ordinary calculus, for which we know the basic derivation rules. The second case is encountered in variational calculus, for which, at a given point x , the functional variation δz measures how much z varies as a function of the changes $\delta f(x)$. In order to distinguish functional derivatives from the ordinary derivatives, we use the symbol $\delta f(x)$ instead of dx . There is a simple rule, although mathematically sloppy, to find out the functional derivative of a given functional, which is based on the analogy between functionals and scalar functions of M variables, for $M \rightarrow \infty$. For instance, in the case of $F[f] = \int dx f^2(x)$, the corresponding function of M variables (which we simply call the functional's analogous and indicate with a tilde) is $\tilde{F}(f_1, \dots, f_M) = \sum_{i=1}^M f_i^2$. The derivative $\partial \tilde{F} / \partial f_k$ (with $1 \leq k \leq M$) can be computed by the usual rules of ordinary calculus: $\partial \tilde{F} / \partial f_k = \sum_{i=1}^M 2f_i \delta_{ik} = 2f_k$ (here δ_{ik} is the Kronecker delta function). Correspondingly, $\delta F[f] / \delta f(y) = \int dx 2f(x) \delta(x - y) = 2f(y)$, making use of the Dirac delta function. Therefore, the correspondence is based on the integral discretization in a finite sum made of M terms. The discretized sum can be derived by using the rules of ordinary calculus, and the result is recasted in the form of a functional.

Examples/Exercises

1. Example: $\frac{\delta}{\delta f(x)} \int dy [f(y)]^\beta = \int dy \frac{\delta}{\delta f(x)} [f(y)]^\beta = \int dy \frac{\delta [f(y)]^\beta}{\delta f(y)} \frac{\delta f(y)}{\delta f(x)} = \int dy \beta [f(y)]^{\beta-1} \delta(x - y) = \beta [f(x)]^{\beta-1}$

2. Exercise: use the correspondence with functions of many variables to derive the functionals given in equations 2.2 to 2.5

2.2 Electron density as the basic variable

As schematically shown in the previous chapter, the many-electron wavefunction is an extremely complex mathematical object that contains more information than we need. Electron density $n(\vec{r})$, instead, is a function of the three spatial coordinates only, independently of the system size ²

From now on, let us consider a N -electron system that is subject to a fixed external potential $V_{\text{ext}}(\vec{r})$. The latter one can represent the Coulomb potential by the nuclei on the electrons, as well as any other electromagnetic external field. Here, we consider the non relativistic limit of a spin-compensated system, in absence of external magnetic fields. The corresponding many-electron Schrodinger equation writes:

$$H\Psi(\vec{r}_1, \dots, \vec{r}_N) = \varepsilon\Psi(\vec{r}_1, \dots, \vec{r}_N) \quad (2.6)$$

where the Hamiltonian is given by the sum of one-body and two-body terms:

$$H = \sum_{i=1}^N \left[-\frac{\hbar^2 \nabla_i^2}{2m} + V_{\text{ext}}(\vec{r}_i) \right] + \sum_{i>j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \quad (2.7)$$

When a system is in a state $|\Psi\rangle$, its energy can be computed as $E[\Psi] = \frac{\langle\Psi H\Psi\rangle}{\langle\Psi\Psi\rangle}$. The variational principle states that the minimization of the functional $E[\Psi]$ with respect to all N -electrons wavefunctions give the ground state $|\Psi_0\rangle$ and energy $E_0 = E[\Psi_0]$. For a system of N electrons in the external potential V_{ext} , the variational principle defines a procedure to determine the ground-state wave function $|\Psi_0\rangle$, the ground-state energy E_0 and other ground-state properties, among which the electron density. By varying V_{ext} at fixed N , $|\Psi_0\rangle$ and E_0 change; therefore, the *ground state* energy of the N -electron system is a functional of the external potential V_{ext} : $E_0 = E[V_{\text{ext}}]$.

The electron density can be obtained from the N -electron wavefunction $\Psi(\vec{r}_1, \dots, \vec{r}_N)$ by integration:

$$\begin{aligned} n(\vec{r}) &= \langle\Psi| \sum_{i=1}^N \delta(\vec{r} - \vec{r}_i) |\Psi\rangle = \int d^3r_2 \dots \int d^3r_N |\Psi(\vec{r}, \vec{r}_2, \dots, \vec{r}_N)|^2 + \\ &\dots + \int d^3r_1 \dots \int d^3r_{N-1} |\Psi(\vec{r}_1, \dots, \vec{r}_{N-1}, \vec{r})|^2 \\ &= N \int d^3r_2 \dots \int d^3r_N |\Psi(\vec{r}, \vec{r}_2, \dots, \vec{r}_N)|^2 \end{aligned} \quad (2.8)$$

where $\delta(\vec{r})$ is the Dirac delta function. We have used the fact that electrons are indistinguishable particles, so that a permutation of their coordinates can only change the wave function by a ± 1 factor.

The conventional way to determine the electron density is to solve the Schrodinger equation firstly, obtain the wavefunction and finally, by using equation 2.8, find $n(\vec{r})$. Therefore, once the Hamiltonian H is written, one can in principle determine the electron density, which we write as $H \Rightarrow n(\vec{r})$ (*from the Hamiltonian to the density*). It may be surprising to think that the Hamiltonian can be determined starting from the electron density, that is, $n(\vec{r}) \Rightarrow H$ (*from the density to the Hamiltonian*), but this can be done, actually. As an example, let's consider the ground state of an isolated atom, for which the electron density $n_0(\vec{r})$ is supposed to be exactly known, and try to derive the corresponding Hamiltonian. For such a case, the Hamiltonian can be written as:

$$H = \sum_{i=1}^N \left[-\frac{\hbar^2 \nabla_i^2}{2m} - \frac{Ze^2}{|\vec{r}_i - \vec{R}|} \right] + \sum_{i>j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \quad (2.9)$$

where Z is the atomic number (i.e. the positive charge of the nucleus) and \vec{R} is the position of the nucleus in space. Such Hamiltonian is determined once the three parameters Z , N and \vec{R} are obtained. The total

²For a spin-polarized system, one has to consider the two components of the electron density with spin projected along the direction \hat{z} , that is, $n_{\uparrow}(\vec{r})$ and $n_{\downarrow}(\vec{r})$. For simplicity sake, we restrict the following presentation to the case of spin-compensated systems, for which $n_{\uparrow}(\vec{r}) = n_{\downarrow}(\vec{r}) = \frac{1}{2}n(\vec{r})$.

number of electrons can be obtained by direct integration in space of the ground-state electron density as ³:

$$N = \int d^3r n_0(\vec{r}) \quad (2.10)$$

The position of the nucleus \vec{R} can be identified as the coordinate where the density has the cusp singularity. Taking \vec{R} as the origin of the coordinates, the electron density shows a spherical symmetry: $n_0(\vec{r}) = f(|\vec{r} - \vec{R}|)$. Moreover, Z can be determined by the density derivative close to the cusp as:

$$Z = - \left[\frac{a_0}{2n_0(\vec{r})} \frac{\partial n_0}{\partial r} \right]_{r \rightarrow |\vec{R}|} \quad (2.11)$$

where a_0 is the Bohr radius of the Hydrogen atom, $a_0 = \hbar^2/me^2$. Since the wavefunction can in principle be obtained from the Hamiltonian, in this particular case the detailed knowledge of the electron density is sufficient to determine all the physical properties of the system. This result is known as the Kato theorem (7).

The generalization of the previous result, that is, $n(\vec{r}) \Rightarrow H$, to a system with a fixed number of electrons N and for arbitrary external potentials was formally given by Hohenberg and Kohn in 1964, through two theorems:

- For non-degenerate ground states, two different Hamiltonians cannot have the same ground-state electron density. Therefore, it is possible to define the ground-state energy as a functional of $n(\vec{r})$: $E = E[n]$
- $E[n]$ is minimal when $n(\vec{r})$ is the actual ground-state density, among all possible electron densities.

The two previous theorems allow the ground state energy to be found by minimizing $E[n]$ instead of passing through the determination of the many-electron wavefunction. This is exactly what we were looking for! However, the HK theorems, although exact (a proof is given below) do not provide any hint on how the energy depends on the density via the $E[n]$ functional. We anticipate that, apart some few special cases, the exact $E[n]$ is unknown and only approximate functionals are used in practice. Moreover, the original formulation of the HK theorems is restricted to the ground state of a closed system of N electrons; as a consequence, it applies neither to excited states nor to compare systems with a varying number of electrons.

Proof of Hohenberg-Kohn theorems The *first HK theorem* is proved by contradiction. Assume that there exist two different external potentials V_{ext} and V'_{ext} which both give the same electron density $n(\vec{r})$. Then we have two Hamiltonians, H and H' with the same ground state density but different ground-state wavefunctions Ψ and Ψ' , respectively. Now we use the variational principle, taking Ψ' as a trial function for the H Hamiltonian, to obtain:

$$E_0 < \langle \Psi' | H | \Psi' \rangle = \langle \Psi' | H' | \Psi' \rangle + \langle \Psi' | (H - H') | \Psi' \rangle = E'_0 + \langle \Psi' | (V_{\text{ext}} - V'_{\text{ext}}) | \Psi' \rangle \quad (2.12)$$

In addition, we can take Ψ as a trial function for the H' Hamiltonian, to obtain:

$$E'_0 < \langle \Psi | H' | \Psi \rangle = \langle \Psi | H | \Psi \rangle + \langle \Psi | (H' - H) | \Psi \rangle = E_0 + \langle \Psi | (V'_{\text{ext}} - V_{\text{ext}}) | \Psi \rangle \quad (2.13)$$

Now we recognize that the expectation value of the difference in the external potentials differ only in sign, because we assumed that electron density is the same. When we add the two equations, we get the contradiction:

$$E_0 + E'_0 < E'_0 + E_0 \quad (2.14)$$

Thus, we conclude that there is a unique map between the external potential V_{ext} and the ground state density. This implies that the total energy is a functional of the density. We write this as $E = E[n]$. The density determines the Hamiltonian, and thereby, the wavefunction ($n \Rightarrow H \Rightarrow \Psi$).

Note that all of the previous Hamiltonians differ by the external potential term V_{ext} . The HK theorem establishes a one-to-one correspondence between a subset of external potentials and a subset of viable

³Note that we do not know N , since the atom may be partially ionized.

densities (non negative functions that integrate to N). In this respect, it can be viewed as a Legendre transformation, which makes it possible to express the energy as a function of density rather than of external potential.

The *second Hohenberg-Kohn theorem* states that the $E[n]$ functional reaches a minimum in correspondence to the actual ground state electron density n_0 . The proof is quite simple. Let's fix the external potential, which determines the (non degenerate) ground state Ψ_0 . Among all viable electron densities n such that $\int d^3r n(\vec{r}) = N$, let's consider anyone of them, which we note \tilde{n} . The first HK theorem ensures that two different wave functions cannot yield the same density; therefore, be $\tilde{\Psi}$ the many-electron wave function that corresponds to \tilde{n} . Since $\tilde{\Psi} \neq \Psi_0$, $E[\tilde{n}] = \langle \tilde{\Psi} | H | \tilde{\Psi} \rangle > \langle \Psi_0 | H | \Psi_0 \rangle = E[n_0]$. The second HK theorem generalises the variational principle from wave functions to electron densities.

Exercises

1. Consider a two-electron system with Hartree-Fock wave function $\text{Det}(\phi_1, \phi_2)$, with $\phi_1(\vec{r}) = A_1 e^{-\alpha_1 r}$ and $\phi_2(\vec{r}) = A_2 e^{-\alpha_2 r}$ (A_1 and A_2 are normalization constants). Write the 2-electron wave function explicitly as a function of \vec{r}_1 and \vec{r}_2 , and determine the corresponding electron density $n(\vec{r})$.
2. Application of the Kato theorem. Determine N , Z and \vec{R} for the following density: $n_0(\vec{r}) = \frac{4}{a_0^3} e^{-\frac{t}{a_0}}$, with $t = |\vec{r} - 3a_0\hat{x}|$. Which system is this?
3. Why is the ground-state energy a functional and not simply a function of the density?

2.3 Towards the definition of the energy functional

From the HK theorems, once $E[n]$ is given as a functional of the electron density, its minimization with respect to density variations that keep the total number of electrons gives the ground state energy of the system. As it will be explained in chapter 4, not only the ground state energy, but also its derivatives with respect to various parameters are in principle accessible. Among the many examples, the derivative of the total energy of a crystal with respect to the volume of the unit cell gives the pressure; the derivative of the total energy with respect to the atomic displacements yields the forces acting on the atoms, giving access to geometry optimization or to first-principle molecular dynamics. Therefore, many ground state properties and response functions can be computed from $E[n]$. However, as we have seen, the HK theorem merely states that the ground state energy is a functional of the electron density $n(\vec{r})$, but does not provide any definition of $E[n]$; its proof goes along a *reductio ad absurdum* and does not make any reference to the actual form of the functional itself. However, the construction of the energy functional is an unavoidable step to build up a theory that is able to compute the properties of real materials. A great part of the DFT focuses on the construction of energy functionals (especially their exchange-correlation part, see chapter 2.5). All of them are necessarily approximated and may perform differently in actual systems.

Here, we present the basic steps along the construction of one of the simplest functionals, that is, the Thomas-Fermi-Hartree functional $E^{(TFH)}[n]$. It is seldom used in current calculations, since it has severe drawbacks and the properties of materials that are computed by relying on $E^{(TFH)}[n]$ very much differ from the measured ones. However, its construction is quite straightforward and useful from a pedagogical point of view. Let us start from the non relativistic N -electron Hamiltonian (equation 2.7):

$$H = \sum_{i=1}^N \left[-\frac{\hbar^2 \nabla_i^2}{2m} + V_{\text{ext}}(\vec{r}_i) \right] + \sum_{i>j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \quad (2.15)$$

It consists of three distinct operators: the kinetic one (T), the external potential V_{ext} and the Coulomb

potential V_{Coul} ⁴:

$$T = - \sum_{i=1}^N \frac{\hbar^2 \nabla_i^2}{2m} \quad (2.16)$$

$$V_{\text{ext}} = \sum_{i=1}^N V_{\text{ext}}(\vec{r}_i) \quad (2.17)$$

$$V_{\text{Coul}} = \sum_{i>j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \quad (2.18)$$

For each of those terms, we search for the corresponding energy term that is expressed as a functional of the electron density.

Firstly, we begin with the external term. It can be easily proved (see exercise) that it writes *exactly* as:

$$E_{\text{ext}}[n] = \int d^3r V_{\text{ext}}(\vec{r}) n(\vec{r}) \quad (2.19)$$

Secondly, we adopt the following approximation for the Coulomb term:

$$U[n] \simeq \frac{e^2}{2} \int d^3r \int d^3r' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|}. \quad (2.20)$$

In doing this, we make at least two errors: (1) We consider the electron distribution as given by classical charges, instead of particles that obey the laws of quantum mechanics, as in the Hartree approximation (see the exercise below). (2) Even in the hypothetical case of classical charges, we have to discard the term $i = j$ in the double sum in equation 2.18, which involves $N(N - 1)/2$ terms. At odds, in equation 2.20 there are $N^2/2$ terms and the electron self-interaction ($i = j$ term) is erroneously included (two related exercises are given below). We will see in chapter 2.5 that in the current energy functionals another term, namely the exchange-correlation energy, should account for the expected correction (the quantum nature of the electron and the compensation for electron self-interaction effects). However, in most cases, such compensation is not exact.

Finally, we must write the kinetic energy functional in terms of the electron density explicitly, which in general cases is not at all trivial. Let us consider the non interacting homogeneous electron gas (HEG), which is a reference system in theory⁵. We obtain the following expression for the kinetic energy, as a function of the constant electron density $\bar{n} = N/V$:

$$T^{\text{HEG}}(\bar{n}) = N C \bar{n}^{2/3} \quad (2.21)$$

where C is a constant that can be easily determined (see the exercise below). This expression is very simple and could have been derived on the basis of a dimensional analysis ($[\bar{n}] = [L]^{-3}$ and $[T * 2m/\hbar^2] = [L]^{-2}$). The same result can be written as a real-space integral:

$$T^{\text{HEG}}(\bar{n}) = N C \bar{n}^{2/3} = \int \frac{d^3r}{V} N C \bar{n}^{2/3} = C \int d^3r \bar{n}^{5/3} \quad (2.22)$$

Now we generalize the previous result for electron densities that are not uniform in space; we assume that the functional dependence of T on $n(\vec{r})$ is the same as in the HEG:

$$T[n] \simeq C \int d^3r [n(\vec{r})]^{5/3} \quad (2.23)$$

Previous expression can be justified by hand-waving arguments: imagine that the real inhomogeneous system (with $n(\vec{r})$ varying from a point to another) could be decomposed in small cells, in which the electron density is almost constant; for each of the cells the expression 2.22 is adopted; finally, the number

⁴Note that V_{ext} denotes an operator in the sense of quantum mechanics, while $V_{\text{ext}}(\vec{r})$ is the function that such an operator takes in the spatial representation.

⁵As far as the usefulness of the HEG is concerned, it is a quite good model to look at the properties of simple metals, such as Na, K, Ag, Au, etc. Please refer to the classical textbooks for more details

of the cells tends to infinity while their volume goes to zero, so that one comes to equation 2.23. Note that this is a very crude approximation for the kinetic energy density in real space; indeed, whenever the density is not constant in space, it is conceivable to think that T depends on gradients at any order of the density, as well as more involved expressions of the density itself. In other words, for a virtual system of non interacting electrons, the exact expression for $T[n]$ reduces to $T^{\text{HEG}}(\bar{n})$ whenever $n(\vec{r}) = N/V \forall \vec{r}$, but not the reverse! (For instance, the function $f(x, y) = y^2$ is constant for $y = y_0$ and $\forall x$; however, this does not imply that the function is constant everywhere in space, whenever y varies). As an example, one can easily evaluate the kinetic energy for an Hydrogen atom, as an example of inhomogeneous electron gas, and compare with the expression provided by the Thomas-Fermi approximation (see exercise).

At the end, collecting all the previous results, we obtain the Thomas-Fermi-Hartree energy functional:

$$E^{(TFH)}[n] = C \int d^3r [n(\vec{r})]^{5/3} + \int d^3r V_{\text{ext}}(\vec{r}) n(\vec{r}) + \frac{e^2}{2} \int d^3r \int d^3r' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (2.24)$$

Such functional has several known drawbacks (see, e.g. (8)). For instance, it does not yield the shell structure in atoms; the binding energy of most homogeneous dimers is severely underestimated, so that molecules do not bind (the energy of the molecule is higher than the sum of the atoms that constitute it). Therefore, it has little practical utility. However, it is the first available explicit expression of $E[n]$, which makes it possible to do some simple calculations, derive a few equations. Last but not least, it provides the reader with a real, although very crude, expression for the energy functional; and, I hope, makes her/him familiar with some physical arguments that are currently used to derive even the more sophisticated functionals.

Exercises

1. Proof that $E_{\text{ext}}[n] = \int d^3r V_{\text{ext}}(\vec{r}) n(\vec{r})$. Start from the definition:
 $E_{\text{ext}}[n] = \langle \Psi | \sum_{i=1}^N V_{\text{ext}}(\vec{r}_i) | \Psi \rangle = \sum_{i=1}^N \int d^3r_1 \dots d^3r_N \Psi^*(r_1, \dots, r_N) V_{\text{ext}}(\vec{r}_i) \Psi(r_1, \dots, r_N)$;
then, use the fact that any permutation among the spatial coordinates r_1, \dots, r_N leaves the integral unchanged and make use of the definition of the electron density (equation 2.8).
2. Find out the electron self-interaction terms for a trial density of classic point charges. Evaluate $U[n]$ (eq.2.20) with $n(\vec{r}) = e \sum_{i=1}^N \delta(\vec{r} - \vec{r}_i)$ and compare with the exact definition $\sum_{i>j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|}$.
3. Compare $U[n]$ with the approximation that was given by Hartree in his theory (see the classical textbooks, for instance the one by Ashcroft and Mermin). Consider also the Hartree-Fock theory and deduce that $U[n]$ is equal to the Coulomb term, once the determinantal form of the wave function is assumed. Note that, in the Hartree-Fock theory, the double sum includes the case $i = j$ and that the exchange term exactly compensates the spurious self-interaction $i = j$ terms.
4. Kinetic energy for the homogeneous electron gas. Consider N electrons in the volume V , at constant density $\bar{n} = N/V$. Write the number of electrons and the kinetic energy as integrals over the wavevector space up to the Fermi wavevector k_F . Then rewrite the kinetic energy as a function of \bar{n} and find out the constant C in equation 2.21.
5. Evaluate, for an isolated hydrogen atom in its ground state, the kinetic energy term and compare with the value that equation 2.23 provides. The electron density of the Hydrogen atom can be easily derived in space by considering the square modulus of the $1s$ orbital.

2.4 Finding the ground state: total energy and electron density

2.4.1 Minimizing $E[n]$

Once that the energy is determined as a functional of $n(\vec{r})$, we can apply the second HK theorem, which states that $E[n]$ has an extremum in correspondence to the ground state electron density. Intuitively, one should search for the density that zeroes the functional derivative of the energy with respect to $n(\vec{r})$.

However, one has to be careful, since only variations of the density that keep the number of electrons constant are allowed: we must restrict our search within the subset of densities that fulfill $\int d^3r n(\vec{r}) = N$. To this purpose, we use the method of Lagrange multipliers⁶:

- First, we define a new functional that includes the constraint on the density: $E[n] - \mu \int d^3r n(\vec{r})$, where μ is the Lagrange multiplier associated to the conservation of the number of electrons.
- Second, we derive the new functional with respect to $n(\vec{r})$ (at this point we need to know the actual $E[n]$!), searching for its zeroes and leaving μ as a free parameter:

$$0 = \frac{\delta}{\delta n(\vec{r})} \left[E[n] - \mu \int d^3r n(\vec{r}) \right] = \frac{\delta E[n]}{\delta n(\vec{r})} - \mu \quad (2.25)$$

This way, we have a family of solutions for $n(\vec{r}; \mu)$ that depends on the Lagrange multipliers.

- Third, we determine μ such that the constraint $\int d^3r n(\vec{r}; \mu) = N$ be fulfilled.

Up to this point, the Lagrange multiplier μ is no more than a mathematical tool for taking the constraint on the number of electrons into account. However, it has a relevant physical meaning, since, as it is shown below, μ corresponds to the chemical potential of the electron system. At $T = 0K$, μ can be identified with the Fermi energy.

Proof The chemical potential of a system of particles is defined as the derivative of the energy with respect to the number of particles:

$$\frac{\partial E}{\partial N} = \int d^3r \frac{\delta E}{\delta n(\vec{r})} \frac{\partial n(\vec{r})}{\partial N} = \int d^3r \mu \frac{\partial n(\vec{r})}{\partial N} = \mu \frac{\partial}{\partial N} \int d^3r n(\vec{r}) = \mu \quad (2.26)$$

where we have used equation 2.25 and the normalization constraint $\int d^3r n(\vec{r}) = N$.

Apart from its theoretical interest, we would use the DFT as a tool for computing the properties of materials. For example, we may be interested to determine the total energy of a crystal as a function of its volume, in order to find its equilibrium volume, compressibility, etc. Next, we can compare the calculated values to the experimental measurements, or use the numerical data in situations in which measurements are technically impossible or not very precise (e.g. the behaviour of materials at very high pressure, like the one inside the Earth's core) or as a help for designing new materials, by calculating the electronic and mechanical properties of hypothetical compounds. In all the previous cases, it is essential to dispose of good functionals, which can provide the correct trends for the total energy and its dependence on external parameters, such as the interatomic distance and angles for a molecule, the volume and shape of the unit cell in a crystal, etc. In such context, the Thomas-Fermi-Hartree functional is not satisfactorily at all, and thus it is not useful in practice. The TFH functional contains two very crude approximations: the first one is on the kinetic part $T[n]$, the second one on the Coulomb energy. However, the Hartree theory can describe the electronic structure of the atoms more satisfactorily than the TFH approximation. Therefore, we guess that the kinetic energy $T[n]$ is badly accounted for by the TFH functional.

⁶In order to remind the method of Lagrange multipliers, consider a simple problem of finding the extrema of $f(x, y) = x^3 + y^3$ under the constraint $g(x, y) = x^2 + y^2 = 1$:

- First, consider the function $F(x, y) = f(x, y) - \mu g(x, y)$
- Second, solve $0 = \partial F / \partial x = 3x^2 - 2\mu x \Rightarrow x = 0; x = 2\mu/3$ and $0 = \partial F / \partial y = 3y^2 - 2\mu y \Rightarrow y = 0; y = 2\mu/3$. Apart from $x = y = 0$, there are three acceptable solutions as a function of the Lagrange multiplier μ : $(0, 2\mu/3)$, $(2\mu/3, 0)$ and $(2\mu/3, 2\mu/3)$.
- For each solution, impose the constraint and find the appropriate value for μ . For instance, for the third solution, we have $1 = 4\mu^2/9 + 4\mu^2/9$, from which $\mu = \pm 3/\sqrt{8}$. The corresponding solutions are: $x = \pm 1/\sqrt{2}; y = \pm 1/\sqrt{2}$

2.4.2 The Kohn-Sham equations

In 1965, Kohn and Sham suggested an alternative way to implement the DFT. In the fifties and sixties, one-electron Schrodinger equations with effective potentials were often used to compute the electronic structure of atoms, molecules and solids, rather satisfactorily. Those approaches were close in spirit to the Hartree method, but included a modified potential term that partially describes the exchange and correlation effects between electrons. Although relatively successful (especially considering that the electronic structure calculations were at their infancy), those approaches lacked a rigorous theoretical foundation. Kohn and Sham's intuition was to find out a one-particle equation within the framework of the DFT, which could be at least as successful as the empirical one-electron Schrodinger equations with effective potentials while being justified from a theoretical point of view. Those equations are now called the Kohn-Sham equations. A derivation is given in the following paragraphs.

In order to go beyond the TFH approximation, we introduce $T_s[n]$, the kinetic energy of a virtual non interacting N -electron system,⁷ which has exactly the same density than the interacting one (we will come back on that later). Therefore, the exact kinetic energy can be written as $T[n] = T_s[n] + T_c[n]$, where $T_c[n]$ is the reminder. From advanced many-body calculations that well account for the electron-electron interaction, $T_c[n]$ is known to be rather small with respect to $T[n]$, for several atoms and solids. Therefore, $T_s[n]$ represents a not too bad approximation to the exact kinetic energy functional. Similarly, the exact Coulomb functional $U[n]$ may be rewritten as a sum of the Hartree term $E_H[n]$, plus another contribution $\Delta U[n]$ that accounts for the quantum nature of the interacting electrons: $U[n] = E_H[n] + \Delta U[n]$. Next, we define the exchange-correlation energy functional as $E_{xc}[n] = T_c[n] + \Delta U[n]$. $E_{xc}[n]$ should account for: (1) the exchange effects (that is, the Pauli repulsion between electrons with the same spin) as the Fock term does in the Hartree-Fock theory; (2) the correction that compensates the spurious self-interaction term appearing in $E_H[n]$; (3) the correlation effects, that is the propensity for two electrons of unequal spin to choose different orbitals and to mutually avoid along their motion in space. For the moment, we assume that $E_{xc}[n]$ is available in some approximated way. A more detailed description of $E_{xc}[n]$ and the current approximations will be given in section 2.5.

Therefore, we can rewrite the exact energy functional as:

$$E[n] = T_s[n] + \int d^3r V_{\text{ext}}(\vec{r}) n(\vec{r}) + E_H[n] + E_{xc}[n] \quad (2.27)$$

The variational equation for the exact functional is thus:

$$\frac{\delta E[n]}{\delta n(\vec{r})} = \frac{\delta T_s[n]}{\delta n(\vec{r})} + V_{\text{ext}}(\vec{r}) + e^2 \int d^3r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + \frac{\delta E_{xc}[n]}{\delta n(\vec{r})} = \mu \quad (2.28)$$

where we used the fact that $\delta E_H[n]/\delta n(\vec{r}) = e^2 \int d^3r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|}$ (see exercise).

Kohn and Sham hypothesized that, for any system \mathbf{S} of N interacting electrons in a given external potential $V_{\text{ext}}(\vec{r})$, there is a virtual system \mathbf{S}_s of N non interacting electrons with exactly the same density as \mathbf{S} . Obviously, if such a correspondance exists, the non interacting electrons in \mathbf{S}_s must be subjected to another external potential, which must compensate the contributions from electron-electron interactions. We denote such an effective potential with $V_s(\vec{r})$, where s stands for "single particle". For \mathbf{S}_s , the variational equation writes:

$$\frac{\delta E[n]}{\delta n(\vec{r})} = \frac{\delta T_s[n]}{\delta n(\vec{r})} + V_s(\vec{r}) = \mu \quad (2.29)$$

The two equations 2.28 and 2.29 have the same mathematical form. They coincide whenever:

$$V_s(\vec{r}) = V_{\text{ext}}(\vec{r}) + e^2 \int d^3r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + V_{xc}(\vec{r}; [n]) \quad (2.30)$$

where $V_{xc}(\vec{r}; [n]) = \delta E_{xc}[n]/\delta n(\vec{r})$ is the functional derivative of the exchange-correlation energy, which is usually referred to as the "exchange-correlation potential" and is a functional of the electron density,

⁷ s stands for "single-particle", since, in the virtual system, the electrons are supposed not to interact between each other and to behave as independent single particles.

too. Since two identical equations have the same solutions, the electron densities of the interacting and the non interacting systems are the same if the effective potential $V_s(\vec{r})$ satisfies the equation 2.30. This equation is a necessary condition for the existence of the virtual system S_s of non interacting electrons at the same density as the physical system S .

Yet, we do not know an explicit expression of the $T_s[n]$ functional in terms of the density that is valid for any inhomogeneous system. However, we know how to solve the problem of N non interacting electrons in the potential $V_s(\vec{r})$, via N Schrodinger equations for the one-particle orbitals ψ_j ($j = 1, N$):

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + V_s(\vec{r}) \right] \psi_j(\vec{r}) = \varepsilon_j \psi_j(\vec{r}) \quad (2.31)$$

For such a system, we know also how to construct the corresponding electron density $n_s(\vec{r})$, by summing up the square moduli of the occupied orbitals, with f_j their occupation factor ($0 \leq f_j \leq 2$), where the factor 2 accounts for the double orbital occupancy with spin up and spin down electrons, since the virtual system is supposed to be spin-compensated as the physical one):

$$n_s(\vec{r}) = \sum_j f_j |\psi_j(\vec{r})|^2 \quad (2.32)$$

If $V_s(\vec{r})$ is given by equation 2.30, the density in equation 2.32 corresponds to that of the interacting system of N electrons, in its ground state, namely $n_s(\vec{r}) = n(\vec{r})$. Equations 2.30 and 2.31 thus represent an alternative and effective way to find the ground state density and energy of the system of interacting electrons, instead of minimizing $E[n]$ directly. They are the celebrated Kohn-Sham equations, which we rewrite below for a system of spin-compensated electrons:

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + V_{\text{ext}}(\vec{r}) + e^2 \int d^3 r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + V_{\text{xc}}(\vec{r}; [n]) \right] \psi_j(\vec{r}) = \varepsilon_j \psi_j(\vec{r}) \quad (2.33)$$

$$n(\vec{r}) = \sum_j f_j |\psi_j(\vec{r})|^2 \quad (2.34)$$

Some relevant observations on the Kohn-Sham (KS) equations follow:

- The first KS equation 2.33 looks like a simple Schrodinger equation. However, both $e^2 \int d^3 r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|}$ and $V_{\text{xc}}(\vec{r}; [n])$ depend on $n(\vec{r})$, which depends on the KS orbitals ψ_j , which in turn depend on the effective potential V_s . The problem is thus non linear; it is usually solved by starting from a trial potential or density and iterate to self-consistency. (i.e. further iterations would change neither the orbitals nor the potential below a pre-determined numerical precision). Some effective methods to achieve the numerical solution of the KS equations will be treated in section 3.2. Once the self-consistent density is found, the corresponding total energy can be computed via equation 2.27 (see also section 2.4.3). The KS equations are in this respect similar to Hartree-Fock equations, although they are mathematically distinct objects.
- In principle, the KS eigenvalues ε_j do not represent the eigen-energies of the physical electron system (apart from an exception, see below). Indeed, it could be shown that they are equivalent to Lagrange multipliers that ensure the orthogonality between any two KS orbitals of the virtual non interacting system: $\langle \psi_i | \psi_k \rangle = \delta_{ik}$. The virtual system was indeed introduced in order to reproduce the electron density of the physical system, not to yield the true wave functions or eigenvalues.
- For a finite system of N electrons, the only KS eigenvalue that has a physical meaning is the highest occupied one, which in principle represents the electron chemical potential. However, the common approximations to E_{xc} often spoil the agreement with the corresponding experimental data.
- Although in principle the KS eigenvalues ε_j do not represent the eigen-energies of the physical electron system, a quite common practice is to identify them with the physical electron eigenenergies. Sometime such an approximation is not too bad and provide the right orbital ordering as a function of their energies. However, it often fails and much care must be used when such an identification is made. This issue has been widely discussed in the literature (9).

Spin-polarized cases. Up to this point, we focused on spin-compensated systems, with N paired electrons: having chosen a quantization direction for S_z , $N/2$ electrons have spin-up (\uparrow) and $N/2$ electrons have spin-down (\downarrow). For systems containing unpaired electrons, we have to specify the number of spin-up and spin-down electrons, N_\uparrow and N_\downarrow , respectively. For each of the two spin components, a set of spin-polarized Kohn-Sham equations are solved, which leads to two set of KS orbitals, depending on their spin χ :

$$\left[-\frac{\hbar^2 \nabla^2}{2m} + V_{\text{ext}}(\vec{r}) + e^2 \int d^3 r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + V_{\text{xc}}(\vec{r}; [n, \sigma]) \right] \psi_{j,\chi}(\vec{r}) = \varepsilon_{j,\chi} \psi_{j,\chi}(\vec{r}) \quad (2.35)$$

$$n_\chi(\vec{r}) = \sum_j f_{j,\chi} |\psi_{j,\chi}(\vec{r})|^2 \quad (2.36)$$

where $\chi = \uparrow$ or $\chi = \downarrow$. The total electron density and the spin density are:

$$n(\vec{r}) = n_\uparrow(\vec{r}) + n_\downarrow(\vec{r}) \quad (2.37)$$

$$\sigma(\vec{r}) = n_\uparrow(\vec{r}) - n_\downarrow(\vec{r}) \quad (2.38)$$

In the spin-polarized case, the exchange-correlation energy and potentials obviously depend on both $n(\vec{r})$ and $\sigma(\vec{r})$. The number of spin-up and spin-down electrons (or equivalently, the total number N and the difference $N_\uparrow - N_\downarrow$) are independent variables and must be fixed before starting the calculation. Practically, if one does not know the spin polarization of the system in advance, distinct calculations must be carried out as a function of $(N_\uparrow - N_\downarrow)$ and the relative energies compared. Within the actual approximation for $E_{\text{xc}}[n, \sigma]$, the most stable spin configuration is that which minimizes the total energy $E[n, \sigma]$.

Exercises

1. For the three-dimensional, non interacting homogeneous electron gas at zero absolute temperature, show that $\mu = \epsilon_F = \partial E / \partial N$ (ϵ_F is the Fermi energy). Hint: find ϵ_F as a function of N and deduce that $\partial \epsilon_F / \partial N = 2\epsilon_F / 3N$. Then, compute the total energy E of the 3D non interacting HEG as a function of both N and ϵ_F and derive with respect to N , taking into account that ϵ_F depends on N (therefore, you will have to make a chain derivative, using the previous results).
2. Show that $\delta E_{\text{H}}[n] / \delta n(\vec{r}) = e^2 \int d^3 r_1 \frac{n(\vec{r}_1)}{|\vec{r} - \vec{r}_1|}$, where $E_{\text{H}}[n] = \frac{e^2}{2} \int d^3 r_1 \int d^3 r_2 \frac{n(\vec{r}_1)n(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|}$. Hint: use the "easy rules" for deriving functionals that have been presented in section 2.1.1.
3. For a system of N non interacting electrons, show that the density can be written like in equation 2.32. Hint: Write the wave function $\Psi(\vec{r}_1, \dots, \vec{r}_N)$ as a product of the one-particle orbitals ψ_j , like in the Hartree theory, and use equation 2.8.
4. Prove that the total energy of the virtual non interacting system is not equal to the sum of the eigenvalues ε_j in equation 2.33. Discuss why this is the case.

2.4.3 Computing the total energy

The total energy, previously defined in equation 2.27, is

$$E[n] = T_s[n] + \int d^3 r V_{\text{ext}}(\vec{r}) n(\vec{r}) + E_{\text{H}}[n] + E_{\text{xc}}[n] \quad (2.39)$$

Solving the Kohn-Sham equations is equivalent to find the density $n(\vec{r})$ that minimizes $E[n]$, for the actual exchange-correlation functional $E_{\text{xc}}[n]$, which is necessarily approximated for real systems. The corresponding single-particle orbitals $\psi_i(\vec{r})$ are thus self-consistent, like the density and the effective potential, as they generate themselves through the Kohn-Sham equations. Once self-consistent orbitals $\psi_i(\vec{r})$ and density $n(\vec{r})$ have been obtained, the total energy of the ground state, for the approximated

exchange-correlation functional, can be computed. In equation 2.39, the kinetic energy $T_s[n]$ of the virtual, non-interacting system can be expressed as:

$$T_s[n] = -\frac{\hbar^2}{2m} \sum_j f_j \psi_j^*(\vec{r}) \nabla^2 \psi_j(\vec{r}) \quad (2.40)$$

while the other terms usually depend solely on the density ⁸

An equivalent expression can be obtained by combining equations 2.45 and 2.33:

$$\sum_j f_j \int d^3r \psi_j^*(\vec{r}) \left[-\frac{\hbar^2 \nabla^2}{2m} + V_{\text{ext}}(\vec{r}) + e^2 \int d^3r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + V_{\text{xc}}(\vec{r}; [n]) \right] \psi_j(\vec{r}) = \sum_j f_j \int d^3r \varepsilon_j \psi_j^*(\vec{r}) \psi_j(\vec{r})$$

that is,

$$T_s[n] + \int d^3r V_{\text{ext}}(\vec{r}) n(\vec{r}) + e^2 \int d^3r \int d^3r' \frac{n(\vec{r}) n(\vec{r}')}{|\vec{r} - \vec{r}'|} + \int d^3r n(\vec{r}) V_{\text{xc}}(\vec{r}; [n]) = \sum_j f_j \varepsilon_j \quad (2.41)$$

From this equation, the kinetic energy can be expressed in terms of eigenvalues ε_j and density $n(\vec{r})$. By replacing the kinetic energy in equation 2.45 with this expression, one obtains for the total energy:

$$E[n] = \sum_j f_j \varepsilon_j - \frac{e^2}{2} \int d^3r \int d^3r' \frac{n(\vec{r}) n(\vec{r}')}{|\vec{r} - \vec{r}'|} + E_{\text{xc}}[n] - \int d^3r n(\vec{r}) V_{\text{xc}}(\vec{r}; [n]) \quad (2.42)$$

2.5 The Exchange-Correlation energy functional

The exchange-correlation energy $E_{\text{xc}}[n]$ was introduced as a reminder between the unknown exact energy functional $E[n]$ and the sum given by the kinetic energy of the non interacting electron gas at the same density, plus the Hartree term and the contribution of the external potential (see eq. 2.27). $E_{\text{xc}}[n]$ can be *defined* as:

$$E_{\text{xc}}[n] = E[n] - E_{\text{known}}[n] \quad \text{where} \quad (2.43)$$

$$E_{\text{known}}[n] = T_s[n] + \int d^3r V_{\text{ext}}(\vec{r}) n(\vec{r}) + E_{\text{H}}[n] \quad (2.44)$$

Therefore, $E_{\text{xc}}[n]$ should account for all the purely quantum effects, namely exchange and correlation, and eliminate the spurious electron self-interaction term that is present in $E_{\text{H}}[n]$ as well. Note that $E_{\text{xc}}[n]$ is simply a functional of the electron density and is independent of the external potential, so that it should work *for all materials*.

Often, $E_{\text{xc}}[n]$ is written as a sum of a term which is due to pure exchange plus a contribution from electron correlation as $E_{\text{xc}}[n] = E_{\text{x}}[n] + E_{\text{c}}[n]$. We know the exchange term explicitly in terms of the one-particle orbitals that forms the Fock determinant $\text{Det}(\phi_1, \dots, \phi_N)$ (see section 2.1), which can be written as:

$$E_{\text{x}}[\{\phi_i\}] = \frac{e^2}{2} \sum_{i,j} \int d^3r \int d^3r' \frac{\phi_i^*(\vec{r}) \phi_j^*(\vec{r}') \phi_i(\vec{r}') \phi_j(\vec{r})}{|\vec{r} - \vec{r}'|} \quad (2.45)$$

However, the expression of the exchange energy as a functional of the electron density is not exactly known, apart from few cases, such as the homogeneous electron gas (HEG), for which the density is equal to its mean value $\bar{n} = N/V$ everywhere ⁹:

$$E_{\text{x}}^{\text{HEG}}(\bar{n}) = -\frac{3e^2}{4} \left(\frac{3}{\pi}\right)^{1/3} V \bar{n}^{4/3} = -C_{\text{x}} V \frac{N}{V} \bar{n}^{1/3} \quad (2.46)$$

⁸An important exception is provided by hybrid exchange-correlation functionals, for which the corresponding XC energy should be expressed in terms of the single-particle orbitals $\psi_i(\vec{r})$ as the kinetic energy in 2.45.

⁹The following expression is valid in three dimensions; distinct formulae can be easily derived for $D = 1$ and $D = 2$.

Note that in the HEG the exchange contribution is always negative and therefore contributes to increase the cohesion by lowering the electron-electron repulsion.

In addition to the exchange term, there is an energy contribution from the dynamic correlations between electrons. Indeed, even for electrons with different spin, the probability of finding an electron in \vec{r}' when another is in \vec{r} , gets much smaller than 1 (which would be the limit for non interacting electrons, without correlation effects) when $\vec{r}' \rightarrow \vec{r}$. Therefore, along its trajectory in space, the electron sees around itself a depression of the density coming from the other electrons, whose average is usually referred to as the exchange-correlation hole $n_{xc}(\vec{r}, \vec{r}')$. Its exchange part is only effective between electrons with the same spin. The reminder is purely due to correlation effects and is sometime named the Coulomb hole.

Exchange and correlation effects can be properly treated within the many-body theory, which is largely based on Green functions and diagrammatic techniques. They are well beyond the scope of this book, so that I will just give some physical examples of correlation effects. The first example is the He atom, with just two electrons, which is given as exercise.

Obviously, the previous definition of $E_{xc}[n]$ that is given in equation 2.43 does not help in finding an expression for the exchange and correlation energy functional, since the complexity of the many-electron problem is simply displaced from $E[n]$ to $E_{xc}[n]$. On one hand, the main contributions to the total energy of atoms, molecules and solids come to the known part of the functional $E_{\text{known}}[n]$; as a consequence, the relative error on the total energy is small, independently of the approximations for $E_{xc}[n]$. On the other hand, usually one would like to compare two systems, which may differ by geometry, volume, atomic composition, number of electrons, presence of external fields, etc. In this case, what matters is the *difference* between the total energies of the two systems. The contribution of exchange and correlation effects to such energy differences may be quite relevant in many cases. Therefore, the problem of finding effective approximations for $E_{xc}[n]$ is a central one in DFT, as proved by the number of publications where the issue is discussed. It is still open, since the exact $E_{xc}[n]$ is not known. In the following, we give a very short account of some of the approximations that are currently used, more to inform the reader than to go deep into their derivation or to provide a thorough account on the performances of the various functionals. We refer the interested reader to the very good review articles and original papers on the subject.

The Local Density Approximation The Local Density Approximation (LDA) was introduced by Kohn and Sham in 1965 (6). It is the simplest approximation to $E_{xc}[n]$ and it was constructed exactly in the same spirit as the local approximation $E^{\text{HEG}}[n]$ to the kinetic energy functional that we discussed in section 2.3. Let us start again from the non interacting homogeneous electron gas (HEG). The exchange energy is exactly known in terms of the mean density \bar{n} (see equation 2.46). Another explicit expression in terms of \bar{n} can be derived for the correlation part in various ways. On one hand, from the high-density limit of the HEG, by extrapolation to the usual physical densities. Indeed, at high electron densities, where the kinetic contribution dominates, an expression for the correlation energy can be obtained through perturbation theory. On the other hand, from accurate quantum Monte-Carlo simulations of the homogeneous electron gas at selected densities. Numerical values obtained are then interpolated for any \bar{n} . All the expressions for the correlation energy of the HEG can be written as:

$$E_c^{\text{HEG}}(\bar{n}) = -C_c V \frac{N}{V} e_c(\bar{n}) \quad (2.47)$$

where the actual expression for the correlation energy per electron $e_c(\bar{n})$ is different in the various approximations. It is worth noting that, for the mean densities that are typical of simple metals like Na, K, Al, Ag or Au, the exchange contribution is more negative than its correlation counterpart.

Both $E_x^{\text{HEG}}(\bar{n})$ and $E_c^{\text{HEG}}(\bar{n})$ are usually rewritten as:

$$E_{xc}^{\text{HEG}}(\bar{n}) = V \frac{N}{V} e_{xc}^{\text{HEG}}(\bar{n}) \quad (2.48)$$

where $e_{xc}^{\text{HEG}}(\bar{n})$ is the exchange-correlation energy per electron in the HEG.

Now we generalize the previous result for electron densities that are not uniform in space. Let's assume that the functional dependence of E_{xc} on $n(\vec{r})$ is the same as in the HEG, in order to obtain the

LDA approximation:

$$E_{xc}^{\text{LDA}}[n] = \int d^3r n(\vec{r}) e_{xc}^{\text{HEG}}(n(\vec{r})) \quad (2.49)$$

where we have replaced \bar{n} with the local density $n(\vec{r})$ at the point \vec{r} and the volume V with the sum of small cells in which the system is supposed to be homogeneous. As a consequence, the LDA is expected to be a good approximation for systems in which the electron density does vary little and not too rapidly.

Despite its simplicity, the LDA works quite well in many systems, which are far from being quasi-homogeneous, as far as the electron density is concerned, and for many purposes. This is due to the fact that the LDA fulfills the sum rule on the exchange-correlation hole $\int d^3r' n_{xc}(\vec{r}, \vec{r}') = -1 \forall \vec{r}$. This gives rise to error compensation when computing the xc energy (typically, the LDA overestimates E_x and underestimates E_c). Moreover, the LDA approaches the correct limit when the electron density tends to very high values (as seen before, the kinetic energy dominates over the xc one for $N/V \rightarrow \infty$ in the HEG). Now, those regions where the electron density is high give the largest contributions to bonding in molecules and solids. Furthermore, although the LDA has many drawbacks, they are well known, so that the inexperienced user benefits from a large experience gained over past years when the LDA had been extensively used. A non exhaustive list of such drawbacks follows: the LDA tends to overbind (i.e. the computed cohesive energies are too large). The electrons are not localised enough in space; indeed, the LDA is a generalization of the HEG, which is appropriate for some s and p electrons, but not for d and f orbitals. The long-range effects (image effects, van der Waals bonds) are completely missing, due to the extremely local nature of the LDA. As a consequence, the potential that an electron feels when approaching an atom or a surface is badly described by the LDA. The hydrogen bond, which is so important in biochemistry and in many chemical reactions, is also poorly accounted for. As an example, the structure of liquid water in the LDA greatly differs from the experimental one.

The Local Spin Density Approximation. The generalization of the LDA to spin-polarized systems is called the Local Spin Density Approximation (LSDA). It is built relying on the fully polarized HEG, for which the exchange-correlation energy per electron $e_{xc}^{\text{HEG}}(\bar{n}_\chi)$ ($\chi = \uparrow, \downarrow$) can be constructed, as a function of the mean spin-polarized electron density \bar{n}_χ . For a virtual HEG that is partially polarized, the exchange-correlation energy per electron depends on both the total mean electron density $\bar{n} = \bar{n}_\uparrow + \bar{n}_\downarrow$ and the spin polarization $\bar{\sigma} = \bar{n}_\uparrow - \bar{n}_\downarrow$. It is obtained by interpolation between the unpolarized case ($\bar{\sigma} = 0$) and the fully polarized case ($\bar{\sigma} = \bar{n}$). The resulting LSDA approximation for the exchange-correlation energy writes as:

$$E_{xc}^{\text{LSDA}}[n, \sigma] = \int d^3r n(\vec{r}) e_{xc}^{\text{HEG}}(n(\vec{r}), \sigma(\vec{r})) \quad (2.50)$$

The Generalized Gradient Approximation As expected, any real electron system is non-homogeneous, that is, the electron density varies in space. A first attempt to describe the density variations was provided by the expansion of the density in terms of its gradients ($|\nabla n(\vec{r})|$, $\nabla^2 n(\vec{r})$, etc.), then adding the truncated expansion to the LDA. However, the approximations that contain the lowest-order gradients were unsuccessful, and in most cases they worsen the much simpler LDA. A relevant adimensional parameter in the gradient expansion is $\xi(\vec{r}) = \frac{1}{k_F(n(\vec{r}))} \left| \frac{\nabla n(\vec{r})}{n(\vec{r})} \right|$, where $k_F(n(\vec{r}))$ is the local Fermi wavelength. It can be seen that in most systems, there are regions of the space in which $\xi(\vec{r}) \geq 1$; this is the case of low-density regions in solids, surfaces, and atoms, for which the gradient expansion diverges. Furthermore, the gradient expansions that are truncated abruptly do not fulfill the sum rule on the exchange-correlation hole $\int d^3r' n_{xc}(\vec{r}, \vec{r}') = -1 \forall \vec{r}$ unlike the LDA, and no error compensation generally occurs.

A major improvement took place in the eighties, when it was realized that there was no need to proceed in the gradient expansion order by order. Instead, functionals of both the density and its gradient were constructed without relying on a gradient expansion, which are currently known as "generalized-gradient approximations" (GGA):

$$E_{xc}^{\text{GGA}}[n, \sigma] = E_{xc}^{\text{LDA}}[n, \sigma] + \int d^3r e_{xc}^{\text{GGA}}(n(\vec{r}), \nabla n(\vec{r})) \quad (2.51)$$

At variance with the LDA, where an exact expression for $e_{xc}^{\text{LDA}}(\bar{n})$ exists (although explicit LDA functionals differ for its practical implementation), there are many different prescriptions for choosing the function $e_{xc}^{\text{GGA}}(n(\vec{r}), \nabla n(\vec{r}))$, each of one leading to distinct GGA's. Very schematically, there are two big classes of

GGA's: on the one hand, e_{xc}^{GGA} is written in terms of the adimensional quantity $\xi(\vec{r}) = \frac{1}{k_F(n(\vec{r}))} \left| \frac{\nabla n(\vec{r})}{n(\vec{r})} \right|$ and few parameters; the latter ones are fitted to a large set of molecules for which accurate many-body calculations of the exchange-correlation energy are available. The Becke GGA for correlation belongs to this class. On the other hand, one can work at analytical forms of e_{xc}^{GGA} that: (1) fulfill the sum rule like the LDA; (2) reduce to the first terms of the naive gradient expansion whenever $\xi(\vec{r}) \ll 1$; (3) cut all the contributes coming from regions where $\xi(\vec{r}) > 1$ off. The class of approximations that were proposed by Perdew and coworkers in the nineties is an example. Revised-GGA are also improvements along these lines. Usually, the detailed forms of the GGA functionals are quite cumbersome, so that we address the interested reader to the original papers and the many reviews.

Some schematic remarks on the performances of GGA's follow:

- GGA sometimes overcorrects LDA: bond lengths are 0-2% larger than experimental values, cohesive energy is 10-20% too small.
- GGA describes XC effects in small-density regions of space much better than LDA. GGA is therefore usually employed for atoms and molecules.
- GGA cannot describe long-range effects, such as Van der Waals. However, hydrogen bonds are usually well accounted for.
- Strongly correlated electron systems are generally out of reach of the GGA.

Exercises

- Variational method applied to the ground state of the He atom. Consider the non relativistic Hamiltonian for the He atom (atomic units $\hbar = m = e^2 = 1$ are used throughout): $H = -\frac{\nabla_1^2}{2} - \frac{\nabla_2^2}{2} - \frac{2}{r_1} - \frac{2}{r_2} + \frac{1}{|\vec{r}_1 - \vec{r}_2|}$. In the ground state, the total angular momentum L and total spin S are both null. The two-electron wave function is a singlet, of the form $\Psi = f(r_1, r_2) \frac{\chi_{\uparrow}(1)\chi_{\downarrow}(2) - \chi_{\uparrow}(2)\chi_{\downarrow}(1)}{\sqrt{2}}$. $f(r_1, r_2)$ is a symmetric function of the radial coordinates that we have to determine in each of the following cases, by minimizing $E_0 = \langle \Psi H \Psi \rangle$. A small computer program can be written in order to make the calculations faster and easier.
 - First try: $f(r_1, r_2) = \phi(r_1; a)\phi(r_2; a)$, where $\phi(r; a) = \frac{a^{3/2}}{\sqrt{\pi}} e^{-ar}$ is a normalized atomic function of 1s type which depends on the parameter a . In this case, the two electrons are forced on the same orbital. Answer: optimal $a_0 = 2 - 5/16$, with $E_0 = -5.695 Ry$.
 - Second try: $f(r_1, r_2) = \frac{1}{\sqrt{2}} [\phi(r_1; b)\phi(r_2; a) + \phi(r_2; a)\phi(r_1; b)]$, where $\phi(r; a) = \frac{a^{3/2}}{\sqrt{\pi}} e^{-ar}$ as above. In this second case, the two electrons could in principle have different orbitals. The minimization of the ground state energy is achieved for: $a_0 \simeq 2.18$; $b_0 \simeq 1.19$, with $E_0 \simeq -5.751 Ry$.
 - Find in the literature the computed correlation energy for the He atom in its ground state. Why didn't we find the correct value? Which other orbital degree of freedom is lacking in the previous form of the spatial orbitals?
- Derive the form of the exchange-correlation potential in the LDA ($V_{xc}^{LDA}(\vec{r}) = \delta E_{xc}^{LDA} / \delta n(\vec{r})$) starting from the definition of E_{xc}^{LDA} given in equation 2.49 and applying the rules for functional derivatives (section 2.1.1). Do the same for the GGA, with the energy functional given by equation 2.51.

Chapter 3

Practical Implementation

The popularity of the DFT is due to the fact that it is a very effective way to compute material properties and that many implementations of DFT exist. The standard implementations go through the resolution of the Kohn-Sham (KS) equations, which we rewrite for a system of spin-compensated electrons:

$$h^{\text{KS}}[n] \psi_j(\vec{r}) = \varepsilon_j \psi_j(\vec{r}) \quad (3.1)$$

$$n(\vec{r}) = \sum_j f_j |\psi_j(\vec{r})|^2 \quad (3.2)$$

where the Kohn-Sham effective one-electron Hamiltonian, which depends on the electron density $n(\vec{r})$, is given by:

$$h^{\text{KS}}[n] = -\frac{\hbar^2 \nabla^2}{2m} + V_{\text{ext}}(\vec{r}) + e^2 \int d^3 r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} + V_{\text{xc}}(\vec{r}; [n]) \quad (3.3)$$

As explained in section 2.4, the KS equations must be solved self-consistently, which is usually achieved by iterations. Starting from a trial density $n^{(0)}(\vec{r})$ (which can be obtained from trial orbitals), $h^{\text{KS}}[n^{(0)}]$ is constructed according to equation 3.3. Then, the KS equation 3.1 is solved and new KS orbitals $\psi_j^{(1)}$ are obtained. A new density $n^{(1)}(\vec{r})$ is thus computed according to equation 3.2. From $n^{(1)}(\vec{r})$ a new KS effective one-electron hamiltonian is derived, and so on. The process goes on until the differences of the KS effective one-electron hamiltonians at two successive iterations (or the electron densities) are below a pre-determined small number. Therefore, two distinct steps can be identified in the resolution of the KS equations: the first one is to obtain the KS orbitals for a fixed one-electron KS hamiltonian (equation 3.1). The standard techniques for the resolution of the Schrodinger equation can be used, as explained in section 3.1, by adopting a basis set for the expansion of the ψ_j , $n(\vec{r})$ and h^{KS} . Further, in section 3.2, we will illustrate briefly the devised methods which make the KS cycle converge.

3.1 Solving the Kohn-Sham equations: basis sets

First of all, we focus on the expansion of the KS orbitals on a basis set that is composed by the functions $\{f_\alpha(\vec{r})\}$. Although here the basis functions are defined on the real space, the expansion is valid regardless of the actual representation of the $\{f_\alpha\}$:

$$\psi_j(\vec{r}) = \sum_\alpha c_{j,\alpha} f_\alpha(\vec{r}) \quad (3.4)$$

The $\{f_\alpha(\vec{r})\}$ are known functions, whereas $c_{j,\alpha}$ are the coefficient expansions that have to be determined in order to compute the Kohn-Sham orbitals. Now, we insert the expansion 3.4 into equation 3.3 and obtain:

$$h^{\text{KS}}[n] \sum_\alpha c_{j,\alpha} f_\alpha(\vec{r}) = \varepsilon_j \sum_\alpha c_{j,\alpha} f_\alpha(\vec{r}) \quad (3.5)$$

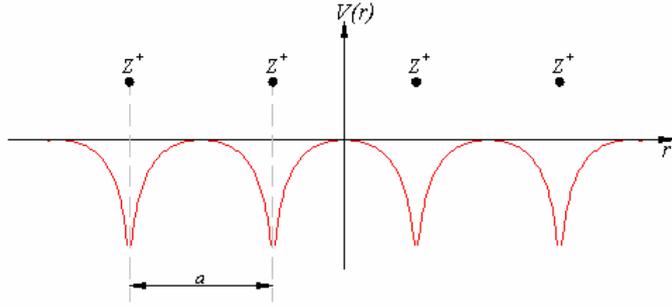


Figure 3.1: Potential for a periodic solid, in one dimension. The nuclei are separated by the distance a .

Then, we multiply both sides by a function belonging to the basis set, $f_\beta^*(\vec{r})$ and integrate over space:

$$\sum_{\alpha} c_{j,\alpha} \int d^3r f_\beta^*(\vec{r}) h^{\text{KS}}[n] f_\alpha(\vec{r}) = \sum_{\alpha} c_{j,\alpha} \varepsilon_j \int d^3r f_\beta^*(\vec{r}) f_\alpha(\vec{r}) \quad (3.6)$$

$$\sum_{\alpha} c_{j,\alpha} [h_{\alpha,\beta}^{\text{KS}} - \varepsilon_j S_{\alpha,\beta}] = 0 \quad (3.7)$$

where $h_{\alpha,\beta}^{\text{KS}}$ and $S_{\alpha,\beta}$ are the elements of the KS matrix \mathcal{H}^{KS} and the overlap matrix \mathcal{S} , respectively, which are expressed in the basis set $\{f_\alpha\}$. The homogeneous linear system 3.7 admits non zero solutions for the unknown coefficients $c_{j,\alpha}$ only if the determinant of the matrix in square bracket is null:

$$\text{Det} [\mathcal{H}^{\text{KS}} - \varepsilon_j \mathcal{S}] = 0 \quad (3.8)$$

Therefore, the problem of solving the N KS equations for each orbital $\psi_j(\vec{r})$ is transformed into a problem of linear algebra, which can be solved by standard diagonalization techniques.

In principle, the $\{f_\alpha(\vec{r})\}$ should form a complete functional space. In such a case, *any* arbitrary function could be expanded as in eq.3.4. However, for computational reasons, the $\{f_\alpha(\vec{r})\}$ set necessarily consists of a finite number of basis functions. Furthermore, the elements of the KS operator \mathcal{H}^{KS} and the overlap matrix \mathcal{S} should be as simple as possible to be calculated. Therefore, the choice of the basis set is always determined through a compromise between computational efficiency and numerical accuracy. This is the reason why there are many possible choices for the basis set $\{f_\alpha(\vec{r})\}$, which at first sight gives the beginner the feeling of a babel. This is not the place where an account of the different methods can be found (see, e.g., (10)). However, we briefly analyze in the following the nature of the physical problem and provide a short description of some of the numerical methods that are used nowadays.

For a collection of atoms, such as molecules and solids, the potential felt by an electron becomes very negative near the nuclei, where the Coulomb attraction is very strong, whereas it is relatively small in the interstitial regions (see Figure 3.1). Approaching a nucleus at \vec{R}_I , the potential on the electron behaves like $V(\vec{r}) \sim -\frac{Ze^2}{|\vec{r}-\vec{R}_I|}$ (Z is the atomic number), which is much stronger than the sum of the potentials that are given by all the remaining nuclei for $\vec{r} \rightarrow \vec{R}_I$. Moreover, $V(\vec{r})$ is approximately spherically symmetric around \vec{R}_I . At variance, in the interstitial regions between two or more nuclei, the strong Coulomb attraction by the nuclei is effectively screened by the core electrons, so that the potential varies more smoothly. As a consequence, the wave function of a valence electron has very different properties in those two regions: the wave function is smooth in the interstitial region between two atoms and oscillates rapidly in space close to the nuclei. The latter behaviour is due to the fact that it has to be orthogonal to the wave functions of the core electrons.

Therefore, the optimal basis set to expand the wave function of a valence electron should mimic such behaviour, being localized in space close to the nuclei, while smooth and rather delocalized in the interstitial regions at the same time. In practice, this set up is quite difficult to realize. The numerical

methods that are used nowadays may be divided into few big classes: First, those that adopt basis functions that are localized in space (atomic functions, gaussians, etc.). Second, methods that employ delocalized functions, such as plane waves (they will be treated more extensively). Third, those that make use of both localized and delocalized functions, which are either constructed by superposition of elementary functions (such as linear methods known as LMTO or FLAPW, or, more recently, PAW) or by adopting wavelets, which show a localized and delocalized character at the same time, as basis functions. It would be impossible to describe all of them, so that we limit ourselves to describe the methods that are based on plane waves in conjunction with pseudopotentials (section 3.1.2). This choice is not motivated by the fact that they are the most performing ones (indeed, they aren't!), but rather by convenience: in fact, a pseudopotential-plane wave code will be used to compute material properties in the hands-on-computer sessions.

3.1.1 Pseudo-potentials

Active electrons versus spectator electrons. In many problems in chemistry and physics, a distinction between *active* and *spectator* electronic orbitals can be made. The problem can thus be formulated in terms of the active electron wave-functions, while the spectator electrons can be treated within suitable approximations. This concept is very general and lies at the hearth of any simplified model to simulate electron systems. Among the criteria used to distinguish between active and spectator electrons, some are particularly general and relevant in most cases : (A) The spectator and active electron energy scales must differ considerably, by one order of magnitude or more; (B) They should react very unlikely to electronic perturbations; (C) Their respective density distributions are mostly localized in different regions of space.

If one is interested in the cohesive properties of condensed matter systems (solids, liquids, surfaces and interfaces, large molecules and clusters), a distinction between *active* electrons and *spectator* electrons is quite natural. The outer (valence) electrons are indeed responsible for the chemical behavior, while the inner electrons can be considered, in a first approximation, as inert. Core electron orbitals lie in energy well below the valence orbitals (of about -10^n Hartree, with $n \geq 1$), they are atomic-like, with a small spatial extent and large gradients ¹. All these features imply that the core orbitals are less polarizable than the valence ones (11). Moreover, the chemical behavior of real materials is essentially determined by bond breaking or formation, in which the inner orbitals do not play a crucial role.

In the pseudo-potential plane-wave approach, the core orbitals are frozen in their atomic state and the action of the core electrons on the valence electron wave-functions is represented by means of suitable operators, the pseudo-potentials; the valence electron orbitals are then expanded in plane waves.

Pseudopotentials: basic ideas. In order to understand some concepts that will be presented in the following (*non locality, non uniqueness, smoothness, etc.*), it is useful to remind the original formulation of the pseudo-potential in condensed matter systems (12). The all-electron valence orbital $|\psi_v\rangle$ is represented as a linear combination

$$|\psi_v\rangle = |\phi_v\rangle + \sum_c \alpha_{cv} |\psi_c\rangle \quad (3.9)$$

of a smooth wave-function $|\phi_v\rangle$ and core electron orbitals $|\psi_c\rangle$ with suitable coefficients to ensure the orthogonality among core and valence orbitals. By considering that $|\psi_v\rangle$ and $|\psi_c\rangle$ are solution of the Schrödinger equation with eigenvalues ϵ_v and ϵ_c , respectively, one can easily obtain the equation for $|\phi_v\rangle$:

$$\left[\hat{H} + \sum_c (\epsilon_v - \epsilon_c) |\psi_c\rangle \langle \psi_c| \right] |\phi_v\rangle = \epsilon_v |\phi_v\rangle \quad (3.10)$$

In equation 3.10, a smooth valence wave-function $|\phi_v\rangle$ is the lowest-energy solution of a new Hamiltonian, with the same eigenvalue as the all-electron valence wave-function $|\psi_v\rangle$. The new Hamiltonian contains the additional projector $\hat{\rho} = \sum_c (\epsilon_v - \epsilon_c) |\psi_c\rangle \langle \psi_c|$ which is non-local, that is, cannot be represented as a multiplicative term. Moreover, such a term is repulsive – which can be seen by considering that the matrix element $\langle \phi_v | \hat{\rho} | \phi_v \rangle$ is positive definite – and short-range, like the core orbitals. By replacing the all-electron problem with an effective Hamiltonian acting on smooth wave-functions that describe the valence electrons, the computational load is alleviated in two respects. First, we reduce the number of

¹As a consequence, their expansion in terms of delocalized functions such as plane waves is problematic.

the self-consistent KS orbitals to be determined. Second, we avoid representing the rapid oscillations of the all-electron valence orbitals close to the nucleus explicitly, while keeping the long tails that are mainly responsible for the formation of the chemical bonds.

The actual choice of pseudo-potentials that are employed to replace the core electrons is connected to the basis set used to represent the valence orbitals. In particular, when using a plane-wave basis set, the pseudo-potential should be *soft*, which means that its Fourier-transform should decrease as fast as possible with the magnitude of the wave-vector. Therefore, the expansion of a soft pseudopotential needs less plane-waves than a hard one, which is clearly advantageous from the computational point of view.

However, one must remark other important features of such an approach:

- (a) The pseudo wave function $|\phi_v\rangle$ as defined in equation 3.9 is not normalized.
- (b) The choice of $|\phi_v\rangle$ is not unique, which implies that the operator \wp is not unique, too (see the corresponding exercise).
- (c) The projector operator acts differently on the different angular momenta that are present in the core:

$$\wp = \sum_{n,l,m \in \text{core}} (\epsilon_v - \epsilon_{nlm}) |\psi_{nlm}\rangle \langle \psi_{nlm}| = \wp_s + \wp_p + \dots$$

Norm-conserving pseudopotentials. An important class of pseudo-potentials, which are still currently used, are the so-called norm-conserving pseudo-potentials. Along the hands-on-computer sessions, they will be given as an input for the numerical projects. A short description is given in the following. The interested reader is referred to (10) for more details on their generation and their use. Let us consider an isolated atom in the Coulomb potential. The solutions of the corresponding Kohn-Sham equation can be written as:

$$\psi_{nlm}(\vec{r}) = Y_{lm}(\theta, \varphi) \chi_{nl}(r) \quad (3.11)$$

where n , l and m are the principal, angular and azimuthal quantum numbers, respectively. To be more specific, let us consider a particular atom, for instance Al. Its atomic structure is $(1s)^2(2s)^2(2p)^6(3s)^2(3p)^1$. Firstly, we choose as core (spectator) electrons those belonging to the complete $(1s)^2(2s)^2(2p)^6$ shells, and the $3s$ and $3p$ electrons in the valence. The valence charge of the pseudo-atom is thus $Z_{\text{val}} = 3$. The all-electron $\chi_{3s}(r)$ wave function has two nodes (see figure); we replace it with a nodeless, smoother pseudo wave function, which coincides with the all-electron one for $r > r_c \simeq 1.5a_0$ (a_0 is the Bohr radius of the H atom). Therefore, the all-electron equation $H\chi_{3s} = \epsilon_{3s}\chi_{3s}$ is replaced with $H_{3s}^{(ps)}\chi_{3s}^{(ps)} = \epsilon_{3s}\chi_{3s}^{(ps)}$. Since $\chi_{3s}^{(ps)}$ is nodeless, the previous equation can be inverted and $V_{3s}^{(ps)}(r)$ computed. Obviously, the same procedure must be used to build a pseudopotential that corresponds to $3p$ states. $\chi_{3s}^{(ps)}$ is normalized exactly as χ_{3s} , which is the reason why the pseudopotentials are said to be norm-conserving. The resulting operator can thus be represented as:

$$\hat{V}^{(ps)} = \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^l V_l^{(ps)}(r) |Y_{lm}\rangle \langle Y_{lm}| \quad (3.12)$$

that is, as a sum of multiplicative terms in r and operators on the angular components. Basically, $V_l^{(ps)}(r)$ behaves like $-\frac{Z_{\text{val}}e^2}{r}$ for $r > r_c \forall l$, while it differs sensitively for different angular components in the core region, $r \leq r_c$. This is because the core states have a distinct behaviour for each l , and so do the pseudo-wave functions and the corresponding pseudo-potential.

3.1.2 Plane waves

The direct and reciprocal spaces. A perfect (ideal) crystal is a collection of atoms that are regularly placed on a three-dimensional lattice with long-range order. Because of the translational symmetry, the knowledge of the atomic structure within the elementary unit cell suffices to know the whole crystal structure. More precisely, any lattice point \vec{r}_l is given by linear combinations of the primitive lattice vectors $\vec{a}_1, \vec{a}_2, \vec{a}_3$, which form a linear space, the *direct* space:

$$\vec{r}_l = \sum_{i=1}^3 l_i \vec{a}_i \quad (3.13)$$

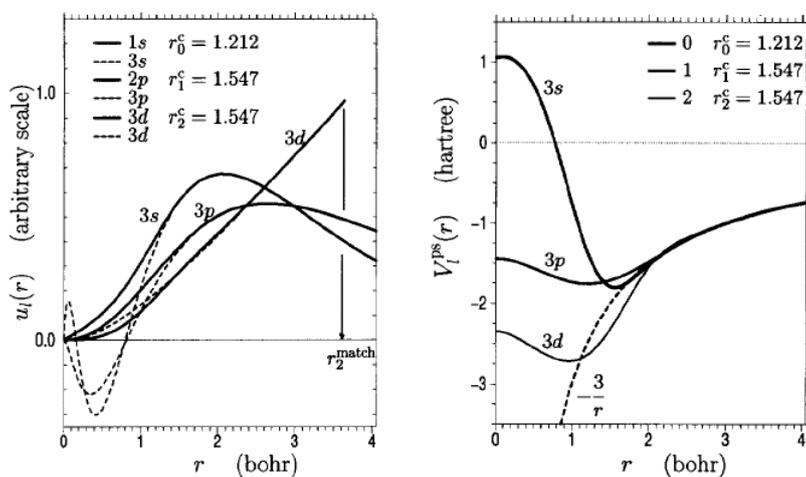


Figure 3.2: Left panel: the radial part of 3s, 3p and 3d all-electron orbitals (dashed line) and the corresponding nodeless pseudo-wave functions (full line) for Al. Right panel: ionic pseudo-potentials for the s, p and d components of the electronic wavefunctions. The radial 3s, 3p and 3d pseudo-wave functions are the ground state solutions for the corresponding pseudo-potentials. Note that all the pseudo-potentials behave like $-\frac{Z}{r}$ at larger distances than the cut off radii (“pseudization radii r_C ”) from the nucleus, where $Z = 3$ is the ionic charge of Al. The figure is taken from M. Fuchs, M. Scheffler, *Comp. Phys. Comm.* **119**, 67 (1999).

The l_i are arbitrary integers, which are in the range $0 < l_i < N_i$, for a crystal with $N = N_1 N_2 N_3$ unit cells. A virtual, infinite crystal is obtained in the limit $N_i \rightarrow \infty$. There are a finite number N_{at} of atoms within the unit cell. Their positions \vec{R}_I are defined by the basis vectors $\vec{\tau}_I$, where $1 \leq I \leq N_{\text{at}}$. The choice of the $\{\vec{a}_i\}$ and the basis is arbitrary; however, one usually employs the *primitive* lattice vectors. The corresponding unit cell (usually called Wigner-Seitz cell) has the smallest volume Ω_0 . Any point \vec{r} can be represented as a sum of a lattice vector and a vector \vec{v} belonging to the Wigner-Seitz cell: $\vec{r} = \sum_{i=1}^3 l_i \vec{a}_i + \vec{v}$. In addition to the translational symmetry, a perfect crystal is invariant under some point group operations. The set of the point group *and* the translational symmetry operations is called the *space group of the crystal*. There is only a limited number (230) of three-dimensional space groups, which are usually classified according to the Hermann-Mauguin notation (12; 13).

Now, define the (linearly independent) vectors $\vec{b}_1, \vec{b}_2, \vec{b}_3$ such that $\vec{a}_i \cdot \vec{b}_j = 2\pi\delta_{ij}$. Then consider the linear space that is formed by all vectors:

$$\vec{G}_h = \sum_{i=1}^3 h_i \vec{b}_i \quad (3.14)$$

where h_i are arbitrary integers. The set of $\{\vec{G}_h\}$ is usually called the *reciprocal space*, since it is the dual of the direct space. As in the case of the direct space, the three vectors $\vec{b}_1, \vec{b}_2, \vec{b}_3$ define an elementary volume in the reciprocal space, the Brillouin Zone (BZ). It is the reciprocal of the unit cell in the direct space and has volume $\Omega_{\text{BZ}} = \frac{(2\pi)^3}{\Omega_0}$. Any arbitrary wave-vector \vec{q} can thus be decomposed as $\vec{q} = \vec{G}_h + \vec{k}$, where $\vec{k} \in \text{BZ}$. Because of the duality between the direct and the reciprocal space, and the completeness of the space group, any symmetry operation in the direct space is also a symmetry in the reciprocal space. The portion of the BZ that can give rise to the entire one by means of the space group operations is usually called the *irreducible Brillouin Zone* (IBZ). A comprehensive account of the general properties and the symmetry-consistent representation of free-electron states in the most common crystals can be found in the book by Bassani and Pastori-Parravicini (13).

The Bloch’s theorem and the expansion in plane waves. In a perfect crystal that is invariant under direct lattice translations, the electron density and the external potential of the unperturbed system keep the crystal periodicity: $A(\vec{r}) = A(\vec{r} + \vec{\tau}_l)$ ($A = n, V_{\text{ext}}$). What are the consequences on the single-particle wave functions $\psi_m(\vec{r})$?² One can show (Bloch’s theorem) that the conservation of the crystal momentum \vec{k} follows from the discrete periodicity of the one-particle operator A (12). \vec{k} is a good quantum number for the wave function, which can thus be accordingly labelled. The Bloch theorem

²In the framework of effective one-particle theories, $\psi_m(\vec{r})$ may be the constituents of the Slater determinants (HF) or the Kohn-Sham orbitals (DFT).

states that the wave function has the following property:

$$\psi_{m,\vec{k}}(\vec{r} + \vec{r}_l) = e^{i\vec{k}\cdot\vec{r}_l} \psi_{m,\vec{k}}(\vec{r}) \quad (3.15)$$

As a consequence, the wave function can be decomposed in a product of a phase factor by a function $u_{m,\vec{k}}$ having the crystal periodicity as:

$$\psi_{m,\vec{k}}(\vec{r}) = \frac{1}{\sqrt{N\Omega_0}} e^{i\vec{k}\cdot\vec{r}} u_{m,\vec{k}}(\vec{r}) \quad \text{with} \quad u_{m,\vec{k}}(\vec{r} + \vec{r}_l) = u_{m,\vec{k}}(\vec{r}) \quad (3.16)$$

The periodic function $u_{m,\vec{k}}(\vec{r})$ is expanded in a discrete sum over the reciprocal lattice vectors as:

$$u_{m,\vec{k}}(\vec{r}) = \sum_{\{\vec{G}\}} e^{i\vec{G}\cdot\vec{r}} \tilde{u}_m(\vec{k} + \vec{G}) \quad (3.17)$$

The $\tilde{u}_m(\vec{k} + \vec{G})$ are the Fourier coefficients in the wavevector space. Since \vec{k} is a good quantum number, there are as many distinct expansions 3.17 as the number of \vec{k} points within the IBZ. ³ From the definition of the reciprocal lattice vectors \vec{G} (equation 3.14) and the construction of the basis vectors of the two linear spaces ($\vec{a}_i \cdot \vec{b}_j = 2\pi\delta_{ij}$), it is straightforward to verify that the expansion 3.17 guarantees that $u_{m,\vec{k}}(\vec{r} + \vec{r}_l) = u_{m,\vec{k}}(\vec{r}) \forall \vec{r}_l$.

In principle, any function having the lattice periodicity can be expanded according to equation 3.17, since the plane waves form a complete basis set. However, the expansion 3.17 contains an infinite number of terms, which is not feasible for numerical applications on a computer. From the basic principles of Fourier analysis, the need of including wave vectors of increasing magnitude is connected to the smoothness of the function to be expanded: the more smooth the function, the faster the expansion converges. Functions that vary rapidly in space need a very large number of wave vectors when expanded in plane waves. Therefore, the sum in 3.17 is generally rearranged summing up over wave vectors of increasing magnitude. Then, it is cut off, by excluding wave vectors such that $\frac{\hbar^2|\vec{k}+\vec{G}|^2}{2m} > E_{\text{cut}}$:

$$u_{m,\vec{k}}(\vec{r}) = \sum_{\{\vec{G}\}: \frac{\hbar^2|\vec{k}+\vec{G}|^2}{2m} \leq E_{\text{cut}}} e^{i\vec{G}\cdot\vec{r}} \tilde{u}_m(\vec{k} + \vec{G}) \quad (3.18)$$

E_{cut} is the maximum allowed kinetic energy of the plane waves involved in the truncated sum 3.18 and is a crucial parameter. On one hand, the larger E_{cut} , the better is the quality of the basis set, which tends to a complete one. ⁴ On the other hand, the smaller E_{cut} , the smaller is the number of Fourier coefficients to be calculated and stocked in the computer memory. As a consequence, E_{cut} is determined through a compromise between numerical accuracy and computational burden. In practice, E_{cut} is the only parameter that adjusts the quality of the PW basis set. The choice is usually done by looking to the behavior of several quantities (total energy, structural parameters, electronic structure, etc.) as a function of E_{cut} . As soon as they converge reasonably, the optimal E_{cut} is determined.

Born-Von Karman boundary conditions and BZ sampling. Consider a finite system with an integer number $N = N_1 N_2 N_3$ of unit cells. In order to simulate an infinite periodic system, the Born - von Karman (BvK) cyclic boundary conditions are often used:

$$\psi_{m,\vec{k}}(\vec{r} + N_i \vec{a}_i) = \psi_{m,\vec{k}}(\vec{r}) \quad (3.19)$$

m and \vec{k} label the band and the wave-vector, respectively. Because of the BvK conditions, the number of \vec{k} wave-vectors in the BZ is equal to $N = N_1 N_2 N_3$ and their density to $\Omega_0/(2\pi)^3$. Therefore, as the size of the system increases, the \vec{k} points become dense in the BZ and tend to be infinite for a truly unbound crystal. Clearly, there are computational restrictions that prevent the use of a too large set of \vec{k} points. The choice of an optimal finite set is often referred to as BZ sampling.

³Since \vec{k} and \vec{G} are homogeneous, the Fourier coefficient $\tilde{u}_m(\vec{k} + \vec{G})$ is a function of the generic wavevector $\vec{q} = \vec{k} + \vec{G}$.

⁴Roughly speaking, E_{cut} fixes the wave vector with the largest allowed magnitude, let's say G_{max} . The resolution in the direct space, that is, the distance between two points in the real space grid, is $\frac{2\pi}{G_{\text{max}}}$, as it can be easily proved (see exercise).

In practice - apart from the case in which the dependence on a particular wave-vector \vec{k} must be explicitly known - one has to compute *averages* of \vec{k} -dependent functions over the BZ:

$$\bar{f} = \frac{\Omega}{(2\pi)^3} \int_{\text{BZ}} d^3k f(\vec{k}). \quad (3.20)$$

In this case, the average \bar{f} is assumed to be a *c*-number, such as the expectation value of the kinetic energy operator in the DFT. In other cases, the function f depends on the other coordinates - for instance, the electron density in the DFT is computed through a sum over the BZ of a non-negative function defined in the real space. For such cases, several schemes have been proposed to consider as few as possible \vec{k} points in the BZ and thus reduce the computational cost. Among them, the special point technique (9) and the tetrahedron method are the most popular ones.

In the special point technique, the average of f over the BZ (Equation 3.20) can be approximated by a weighted sum over a finite number $N_{\vec{k}}$ of \vec{k} points:

$$\bar{f} = \sum_{i=1}^{N_{\vec{k}}} w_{\vec{k}_i} f(\vec{k}_i) + e(\{\vec{k}_i\}) \quad (3.21)$$

with normalized weights: $\sum w_{\vec{k}_i} = 1$. For a given $N_{\vec{k}}$, the set of special points $\{\vec{k}_i\}$ is such that the error $e(\{\vec{k}_i\})$ is minimized. A demonstration, which is based on the use of the lattice point group of the crystal and the symmetry of $f(\vec{k})$ for translations by reciprocal lattice vectors, can be found in the paper by Chadi and Cohen (14). In the same paper, a method for generating the special point sets is outlined and the determination of various special point sets explicitly carried out for cubic and hexagonal crystal lattices. A more general method for generating equispaced grids of \vec{k} points for crystals of any symmetry, which is equivalent to that introduced by Chadi and Cohen in special cases, was then provided by Monkhorst and Pack (15).

The special point technique is well suited for integrating smoothly varying functions in the BZ, particularly for insulators and semiconductors, where the number of occupied bands is independent of the wavevector \vec{k} , which simplifies its practical implementation. In metals, the number of occupied bands can vary abruptly with \vec{k} , as a consequence of band crossing, and the shape of the Fermi surface may be complicated and generally admits singular points. A practical way to deal with such singularities is to smoothen the Fermi surface by replacing the Heaviside function in Equation 2.9 with a smeared occupation function $\tilde{s}(\varepsilon_F - \varepsilon_{m,\vec{k}}; \{\alpha\})$ where the set of parameters $\{\alpha\}$ defines the actual choice of smearing (9). A simple example is provided by the Fermi-Dirac distribution, the width of which around ε_F is proportional to the temperature T . However, all physical quantities must generally be computed in the limit of no smearing (e.g. $T \rightarrow 0$ in the Fermi-Dirac distribution) in order to avoid the arbitrariness linked to the choice of the function \tilde{s} .

An alternative approach to sample the BZ is provided by the tetrahedron method (9), which is basically a simplex method for interpolating $f(\vec{k})$. The irreducible part of the BZ is partitioned into tetrahedra, within which $f(\vec{k})$ is linearized in \vec{k} . The linear approximation allows the integration (Eq.3.20) to be computed analytically, taking into account the complicated shape of the Fermi surface. While there are few doubts that at zero temperature the tetrahedron method is superior to the special point technique in metals, for insulators the latter one is often computationally faster and more robust.

The computation of electronic spectra in crystals, in contrast, represents a different picture. In this case, it is preferable to sample the BZ by using \vec{k} points that do not have any special symmetry. Indeed, the electronic bands may be flat or degenerate in highly symmetric \vec{k} points, which would artificially reinforce the computed weight of such electronic transitions. A practical alternative in those cases is to generate a special point grid that is successively shifted from the BZ center by a linear combination of the reciprocal lattice vectors with small coefficients, of the order of few tenths.

How to treat aperiodic systems. Obviously, not all the interesting physical systems are perfect crystals! On the contrary, since the advent of nano-sciences, there is a growing interest for systems that have one or more dimensions in the nanometer range, such as interfaces, graphene, supported clusters, nanowires and nanodots. Those systems are truly aperiodic and their properties are very much linked to their reduced dimensions. In principle, plane waves are not well suited for treating such cases, since their

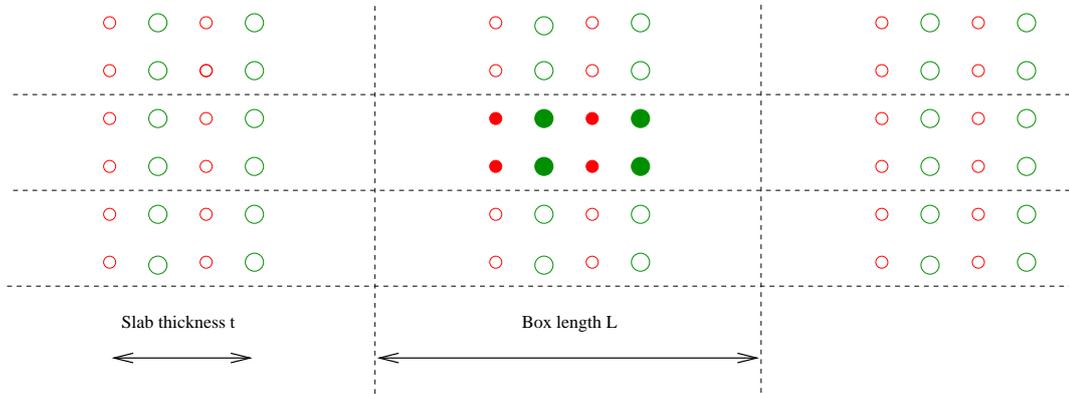


Figure 3.3: A simple representation of a slab with periodic boundary conditions. The atoms within the unit cell are represented as full circles, while the image atoms are drawn as empty circles.

use is restricted to expand periodic functions that satisfy the Bloch's theorem. It is a common practice, however, to represent aperiodic systems by adopting the so-called *supercell*.

Firstly, we focus on the simple case of a truly finite system, such as an atom or a molecule. We also assume that the system is globally neutral and consists of N_a atoms. We consider a cell containing some void space that surrounds the atom or the molecule, which we call the *supercell*. The supercell is then artificially repeated in space through Born-Von Karman (BvK) boundary conditions (BC). The model system is thus formally equivalent to a perfect crystal having N_a atoms per unit cell. Therefore, the mathematical framework that was developed for the perfect crystal model also applies to such a virtual periodic crystal. Therefore, the use of periodic basis sets, such as plane waves (see next section), is made possible even for truly aperiodic systems. The basic approximation here is the choice of supercell size, which is strictly related to N_a . The shape and size of the supercell must be consistent with the actual characteristics of the real system (i.e. isotropy, point symmetry if any, etc.) as much as possible. Moreover, the interaction between periodic images should be monitored and possibly evaluated. It is in principle possible to estimate finite-size effects, by adopting different cell sizes and checking the convergence of the relevant physical quantities with respect to the cell size. Let us assume, for instance, a uniform variation of the supercell lattice parameters such that the supercell volume scales as $\lambda\Omega$. As $\lambda \rightarrow \infty$, the supercell model becomes exact⁵. In that limit, the Brillouin Zone reduces to a single point (the Γ point $\vec{k} = 0$ – see exercise).

It is worth reminding that applying periodic boundary conditions to any intrinsically aperiodic system is not unobjectionable. A system that is characterized by a given charge distribution can be intrinsically different from its periodic analogous that is built by imposing a spurious periodicity, even in the limit of a very large lattice constant. The reason lies essentially in the long-range behavior of the Coulomb potential. Handling the Coulomb interaction sensitively depends on the system and on the kind of periodic boundary conditions. There are many practical recipes, each of them presenting advantages and shortcomings. For charged molecules or clusters, a cut-off Coulomb interaction can be introduced:

$$\begin{aligned}
 v(r) &= \frac{1}{r} & r < r_{\text{cut}} \\
 v(r) &= 0 & r \geq r_{\text{cut}}
 \end{aligned}
 \tag{3.22}$$

A careful evaluation of the bias introduced by such a sharp cut off in direct space or by the use of BvK boundary conditions on the computed quantities is in any case a necessary previous step when setting up the model system.

For charged defects or impurities in crystals, a *compensating uniform background* is often superimposed to ensure charge neutrality in the supercell. The defect formation energy is computed via a total-energy difference of the perfect and the defective crystals. It can be shown that energy differences between infinitely replicated periodic systems are in principle ill-defined, although the choice of a suitable reference system may remedy the situation. The treatment of the long-range Coulomb potential in reciprocal space is intimately related to the discretization of the Fourier transform into a Fourier series, which is limited

⁵In absence of long-range interactions and provided that the boundary conditions are consistent with those for the real system $\forall \lambda$

to reciprocal lattice vectors. By isolating the long-range contribution, which can be treated analytically, from the short-range one, which depends on the local details of the charge distribution and can be computed numerically, some ways of dealing with the Coulomb interaction in systems with periodic boundary conditions have been devised.

Another important case is provided by systems that are periodic in one or two dimensions and aperiodic in the remaining ones. A classical example is provided by surfaces. They are simulated by a *slab* consisting of some atomic layers, plus a surrounding vacuum region, with periodic boundary conditions along the surface plane (see Figure 3.1.2). The slab has to be thick enough to display the bulk characteristics (atomic positions, electronic structure, etc.) in the innermost layers. Despite the finite thickness, it is possible to extract convergent surface energies from this kind of calculations. The safest way, although not the most economic one as far as the computing time is concerned, is to increase the slab thickness t systematically, starting from a minimum t that allows bulk properties to be properly defined⁶. The slab total energy is then plotted as a function of the number of layers N_l , so that the averaged surface energy over the two slab terminations can be extrapolated at $N_l = 0$. However, especially when using a plane wave basis set (see Section 4.4), periodic boundary conditions are imposed along the normal to the surface plane, which can be source of numerical errors. This is the case, for example, of slabs with a net dipole moment, which give rise to intrinsically non-periodic electrostatic potentials. A way to overcome such difficulties still using 3-D PBC has been recently discussed.

The KS hamiltonian in plane waves. Here we provide a simple example on how the total energy and electron density are computed in plane waves. We will not be exhaustive, but give the simplest formulae to serve as a guideline. For those who wish to deeper their knowledge, a thorough account can be found in many books and articles (9).

First, we recall the expression of the Kohn-Sham orbitals in plane waves for a spin-compensated system of N electrons:

$$\psi_{m,\vec{k}}(\vec{r}) = \sum_{\{\vec{G}\}} e^{i(\vec{k}+\vec{G})\cdot\vec{r}} \tilde{u}_m(\vec{k} + \vec{G}) \quad (3.23)$$

where m is the band index and \vec{k} the crystal momentum. The electron density is given in terms of the Kohn-Sham orbitals as:

$$n(\vec{r}) = 2 \sum_m \sum_{\vec{k}} w_{\vec{k}} f_{m,\vec{k}} |\psi_{m,\vec{k}}(\vec{r})|^2 \quad (3.24)$$

where $w_{\vec{k}}$ is the weight of each \vec{k} point that sample the Brillouin zone, and $f_{m,\vec{k}}$ is the occupation factor (according to the Fermi-Dirac statistic or whatever), which ranges between 0 and 1. The factor 2 accounts for spin degeneracy.

If we replace in equation 3.24 the Kohn-Sham orbitals $\psi_{m,\vec{k}}(\vec{r})$ by their plane wave expansion 3.23, the electron density would be computed through a double sum over the reciprocal space. Generally, the number of PW's M_{PW} is large, making the straightforward calculation of $n(\vec{r})$ scaling as $N M_{\text{PW}}^2$ and becoming rather slow. However, we can adopt another faster and smarter method, which relies on the properties of Fast Fourier Transforms (FFT). For each point \vec{k} and each band m , starting from the knowledge of the Fourier coefficients $\tilde{u}_m(\vec{k} + \vec{G})$, we compute through the inverse FFT the corresponding periodic part of the KS orbital $u_{m,\vec{k}}(\vec{r})$. The computational load of such an operations is $M_{\text{PW}} \log M_{\text{PW}}$. Next, we compute its square modulus and sum up over the N_k points in the BZ and the occupied orbitals, which are $\simeq N$, obtaining the electron density in direct space via equation 3.24. Once $n(\vec{r})$ is computed, its Fourier coefficients in the reciprocal space $\tilde{n}(\vec{G})$ can be determined via direct FFT. Global scaling of such procedure is $\sim N N_k M_{\text{PW}} \log M_{\text{PW}}$, thus much better than the double sum in reciprocal space.

Regarding the calculation of the total energy, some terms can be evaluated in the reciprocal space, while some others in the direct space. First, we rewrite equation 2.27:

$$E[n] = T_s[n] + \int d^3r V_{\text{ext}}(\vec{r}) n(\vec{r}) + E_{\text{H}}[n] + E_{\text{xc}}[n] \quad (3.25)$$

⁶Please note that the space group of the slab is generally different from that of the infinite crystal. Therefore, one has to proceed carefully when considering the reference bulk state.

where $T_s[n]$ is the kinetic contribution of a virtual non interacting system with density $n(\vec{r})$, $E_H[n]$ is the classic Hartree contribution to the electron-electron interaction, and $E_{xc}[n]$ is written in the LDA, for simplicity sake:

$$T_s[n] = 2 \sum_m \sum_{\vec{k}} w_{\vec{k}} f_{m,\vec{k}} \int d^3r \psi_{m,\vec{k}}^*(\vec{r}) \left[-\frac{\hbar^2 \nabla^2}{2m} \psi_{m,\vec{k}}(\vec{r}) \right] \quad (3.26)$$

$$E_H[n] = \frac{e^2}{2} \int d^3r \int d^3r' \frac{n(\vec{r})n(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (3.27)$$

$$E_{xc}[n] \simeq E_{xc}^{\text{LDA}}[n] = \int d^3r n(\vec{r}) e_{xc}^{\text{HEG}}(n(\vec{r})) \quad (3.28)$$

Once the electron density $n(\vec{r})$ is determined as explained above, $E_{xc}^{\text{LDA}}[n]$ is easily computed in the direct space. On the contrary, the Hartree contribution involves a double integral in direct space, which also contains the factor $|\vec{r} - \vec{r}'|^{-1}$, which may diverge for arbitrarily close points. Making use of some analytical skill (see exercise), $E_H[n]$ can be conveniently rewritten in the reciprocal space as:

$$E_H[n] = 2\pi\Omega_0 \sum_{\vec{G} \neq 0} \frac{|\tilde{n}(\vec{G})|^2}{G^2} \quad (3.29)$$

The sum over the reciprocal lattice vectors (\vec{G}) excludes the diverging term $\vec{G} = 0$. It can be shown that such a long-wavelength divergence does not take place in globally neutral system, for which the positive charges, given by the ions, are equal to the number of electrons. Equation 3.29 provides a very simple and fast way to compute $E_H[n]$, which scales as M_{PW} .

- We end this section by summarizing the main characteristics of plane waves, pros (+) and cons (-):
- + Plane waves expansion is equivalent to interpolate functions (orbitals, density and potential) in terms of sinus and cosinus that have the lattice periodicity. The precision is the same everywhere, independently of the atomic positions. The improvement of the quality of the basis set is particularly simple to be done by increasing just one parameter, the cutoff on the wave vector energy.
 - + Different quantities can be computed either in the direct space or in the reciprocal space, according to the computational convenience. Plane waves based codes are thus quite easy to be written and implemented. The use of fast-Fourier transforms makes very simple and effective to go back and forth between the two dual spaces.
 - + The quality of the basis set is independent of the atomic positions. Therefore, the computation of energy derivatives with respect to the atomic positions (i.e. the forces on the atoms) is very simple and safe. Moreover, PW's can describe the electron density in regions of space where there are no nuclei, at odds with localized basis sets.
 - Plane waves are generally used in conjunction with pseudo-potentials, so that close to the nuclei the orbitals are different from their all-electron counterparts.
 - Many plane waves are commonly employed in the expansion, typically around 10^2 per electron pair (obviously, the precise number depends on the pseudopotentials, the cut off energy, etc.). The interpretation of the expansion coefficients is not physically intuitive and files handling is cumbersome.
 - PW's are used in conjunction with periodic boundary conditions, which poses problems for charged, non periodic systems. These systems are simulated by adopting supercells with void space, which increases the computational burden.

Exercises

1. BvK conditions in one dimension. The case of very large cells and Gamma point.
2. A single electron in a potential well of varying deepness.
3. Uniqueness of pseudo-potentials.

3.2 Solving the Kohn-Sham equations: the self-consistent cycle

As mentioned at the beginning of this chapter, the Kohn-Sham equations must be solved self-consistently, which is usually achieved by iterations. Starting from a trial density $n^{(0)}(\vec{r})$ (which can be obtained from trial orbitals), $h^{\text{KS}}[n^{(0)}]$ is constructed according to equation 3.3. Then, the KS equation 3.1 is solved and new KS orbitals $\psi_j^{(1)}$ are obtained. A new density $n^{(1)}(\vec{r})$ is thus computed according to equation 3.2. From $n^{(1)}(\vec{r})$ a new KS effective one-electron hamiltonian is derived, and so on. The process goes on until the differences of the KS effective one-electron hamiltonians at two successive iterations (or the electron densities) are below a pre-determined small number. Therefore, the solution of the KS equations is found when the input and output effective potential (or the electron density) coincide with each other. In other words, if we apply the KS machinery to a given function v , we want that the result to be v again. Mathematically, if K represents the ensemble of the operations that allow v to be recalculated according to the KS equations, we require that

$$v = K(v) \tag{3.30}$$

that is, we must find the fixed point v for the application K . Finding the fixed points of a general application, defined on a linear space, is a common mathematical problem. Many numerical approaches are available, where the solution is attained through iterations. However, it is not the scope of the present section to give a comprehensive account of these methods. Rather, we focus on selected examples showing why the convergence to the self-consistent solution may be delicate and rather time-consuming. The reader should consult "The Computational Physicist's Toolkit" by Philippe Depondt (chapter 3.2), where a few other examples are presented and discussed.

First of all, we imagine to have already found the fixed point of the application, such that $v_0 = K(v_0)$. What happens if we move a little away from the solution? Let's try with $v = v_0 + \delta v$, with small δv .⁷

$$K(v_0 + \delta v) = K(v_0) + \delta v \left[\frac{\partial K}{\partial v} \right]_0 + \dots \tag{3.31}$$

or, in other words:

$$\delta K = K(v_0 + \delta v) - K(v_0) \simeq \left[\frac{\partial K}{\partial v} \right]_0 \delta v \tag{3.32}$$

There are two different cases: if $\left| \left[\frac{\partial K}{\partial v} \right]_0 \right| < 1$, $\delta K < \delta v$. The new value $v_1 = K(v_0 + \delta v)$ thus satisfies: $|v_1 - v_0| < |\delta v|$. The sequence v_1, v_2, \dots, v_n would thus converge to v_0 as n grows and the fixed point

⁷Note that δv may also represent the numerical error when we approach the self-consistent solution.

v_0 is said to be stable. Eventually, the convergence may be very slow if $|\left[\frac{\partial K}{\partial v}\right]_0| \simeq 0$. In any case, straightforward iterations of the application will lead us to the nearest fixed point; slowly, but surely! On the other hand, if $|\left[\frac{\partial K}{\partial v}\right]_0| > 1$, simple iterations of the application will lead to divergence. In such a case, the fixed point is unstable: an arbitrary small perturbation to the solution would drive the system far away from the fixed point. This is actually the case of the KS cycle, for most physical problems. Before discussing the reasons that underlie such a divergence, I would like to anticipate that, even for diverging applications K (that is, those for which $|\left[\frac{\partial K}{\partial v}\right]_0| > 1$), it is always possible to define a new application, let's say K' , which has a stable fixed point. That's very simple! Let choose a positive number p such that $0 < p < \frac{1}{\max|\left[\frac{\partial K}{\partial v}\right]_0|}$, where the max is taken among all the directions of the derivative (v is supposed to be an element in a vector space of dimension >1 , or even a functional space, with infinite dimension). Then define the new application:

$$K'(v) = pK(v) + (1-p)v \quad (3.33)$$

which is a simple mixing of the input v and the output $K(v)$. It is straightforward to verify that K' has the same fixed point v_0 as K and that it is now stable, that is, $|\left[\frac{\partial K'}{\partial v}\right]_0| < 1$. The numerical scheme in equation 3.33 is usually referred to as simple mixing. This is a very robust algorithm to attain convergence! However, it can be shown that the convergence becomes very slow as $p \rightarrow 0$; an example can be found in the aforementioned book by Ph. Depoendt.

Why is the KS fixed point unstable? Clearly, one should linearize the KS application and find its derivative around the fixed point. However, each particular system has a distinct ground state density $n(\vec{r})$ and effective potential $V_s(\vec{r})$, as guaranteed by the Hohenberg-Kohn theorem. Therefore, it is a difficult task to derive general and rigorous criteria that are valid for any system of N electrons subject to an external potential. Rather, I will take an example from electrostatics, which is well suited to show the root of the instability. Consider a classical set of negative charges (they mimic the electrons) that are subject to a stabilizing external potential. This is a common electrostatic problem, which is equivalent to minimize:

$$E[n] = -e \int d^3r \left[\frac{1}{2}V(\vec{r}) + V_{\text{ext}}(\vec{r}) \right] n(\vec{r}) \quad (3.34)$$

where

$$V(\vec{r}) = -e \int d^3r' \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (3.35)$$

and the external potential may be thought as generated by a given (fixed) positive charge distribution $n_+(\vec{r})$:

$$V_{\text{ext}}(\vec{r}) = e \int d^3r' \frac{n_+(\vec{r}')}{|\vec{r} - \vec{r}'|} \quad (3.36)$$

The above relation between the negative charge density n and the electrostatic potential V is the solution of the Poisson equation ⁸:

$$\nabla^2 [V(\vec{r}) + V_{\text{ext}}(\vec{r})] = 4\pi e [n_+(\vec{r}) - n(\vec{r})] \quad (3.37)$$

which can be more easily analysed by passing to the wave vector space $\{\vec{q}\}$ through Fourier transforms:

$$q^2 [\tilde{V}(\vec{q}) + \tilde{V}_{\text{ext}}(\vec{q})] = 4\pi e [\tilde{n}_+(\vec{q}) - \tilde{n}(\vec{q})] \quad (3.38)$$

which has the solution:

$$\tilde{V}_0(\vec{q}) = -\tilde{V}_{\text{ext}}(\vec{q}) + 4\pi e \frac{\tilde{n}_+(\vec{q}) - \tilde{n}_0(\vec{q})}{q^2} \quad (3.39)$$

⁸Here we use the so called Gaussian units. In these units, Maxwell's first equation equation for the electric displacement $\vec{D}(\vec{r})$ in terms of the total charge distribution $\rho(\vec{r})$ reads $\nabla \cdot \vec{D} = 4\pi\rho$; the relation between $\vec{D}(\vec{r})$, electric field $\vec{E}(\vec{r})$ and polarization $c(\vec{r})$ is $\vec{D} = \vec{E} + 4\pi\vec{P}$. In Gaussian units the void electric permittivity is simply $\epsilon_0 = 1$ and the void magnetic constant $\mu_0 = 1$; as $\epsilon_0\mu_0 = c^{-2}$, with c the speed of light, the Lorentz force per unit charge is $\vec{F} = \vec{E} + \frac{\vec{v}}{c} \times \vec{B}$. This could be compared to the same equations in the MKSA unit system: $\nabla \cdot \vec{D} = \rho$; $\vec{D} = \epsilon_0\vec{E} + \vec{P}$ and $\vec{F} = \vec{E} + \vec{v} \times \vec{B}$; the void electric permittivity is $\epsilon_0 = \frac{10^7}{4\pi c^2}$.

The above equation represents the self-consistent relation between the charge distribution n and the electrostatic potential V it generates by mutual repulsion between the negatively charged classical particles. It can be easily linearized around the fixed point giving:

$$\delta\tilde{V}(\vec{q}) = -4\pi e \frac{\delta\tilde{n}_0(\vec{q})}{q^2} \quad (3.40)$$

It is important to focus on the previous equation by distinguishing two cases:

- The case of short wave lengths (or large wave vector q). In such a case, any fluctuation around the exact solution produces smaller fluctuations on the potential. The cycle will then converge to the stable fixed point.
- The case of long wave lengths (or small wave vector q). Any density fluctuation, such as an error that may be due to a lack of numerical precision, or an initial estimate that does not coincide with the analytical solution, yields large fluctuations on the electrostatic potential. The simple iteration of equations 3.35 and 3.37 does not attain convergence. The fixed point is unstable.

Physically speaking, if a system has bigger dimensions than a characteristic length $L_c \simeq [4\pi\bar{n}]^{-1/3}$, the long-range Coulomb interaction can produce oscillations in the electrostatic potential. Accordingly, the charge density, which is initially localized, let's say in a region A, at the following iteration may be localized in another portion of space, let's say B, which is quite far from A. Such a phenomenon is usually called *charge sloshing*. There is a quite easy way to avoid charge sloshing, simply by introducing a threshold wavelength $q_{\text{cut}} \simeq \frac{2\pi}{L_c}$. The equation 3.40 that links the variation of the potential at the next iteration $\delta\tilde{V}(\vec{q})$ as a function of the charge density fluctuations $\delta\tilde{n}(\vec{q})$ between the input and output $\tilde{n}(\vec{q})$ is modified as follows:

$$\delta\tilde{V}(\vec{q}) = -4\pi e \frac{\delta\tilde{n}(\vec{q})}{q_{\text{cut}}^2 + q^2} \quad (3.41)$$

The cycle will then converge to the fixed point. The readers who know electron screening, would likely recognize a function that they have already seen, the Thomas-Fermi static dielectric function in reciprocal space! Indeed, $\delta\tilde{V}(\vec{q})$ as given by 3.41 may be thought as a perturbation; the input potential $\tilde{V}_{in}^{(n+1)}(\vec{q})$ for the next iteration $n + 1$ would only uptake a part of it:

$$\tilde{V}_{in}^{(n+1)}(\vec{q}) = \tilde{V}_{in}^{(n)}(\vec{q}) + \frac{q^2}{q_{\text{cut}}^2 + q^2} \left[\tilde{V}_{out}^{(n)}(\vec{q}) - \tilde{V}_{in}^{(n)}(\vec{q}) \right] \quad (3.42)$$

which is equivalent to a simple mixing in the reciprocal space, with a q -dependent coefficient! Other popular, and usually more effective, mixing schemes are the Anderson and the Broyden methods. Their implementation for the Kohn-Sham problem is detailed in (17).

Another fundamental reason why the KS cycle may diverge is given to the presence of degenerated or almost degenerated energy levels around the Fermi energy. For simplicity sake, imagine to have two perfectly degenerated Kohn-Sham levels at the Fermi energy, that should be filled with two orbitals. The obvious solution is that they must be both half-filled. However, they are degenerated only when the self-consistent solution is attained! Therefore, at the starting point, one of them (let's say ε_A) would be lower than the other (ε_B). If the filling rules at zero electronic temperature are assumed, level A will be filled with two states, while B would be empty. At the following iteration, if the electron-electron repulsion is strong enough, $\varepsilon_A > \varepsilon_B$ and the reverse filling would result. The system would then go back and forth between these two states, without necessarily attaining convergence. A solution to such a problem is to adopt a finite-temperature function for electron filling. Such a procedure is usually known as *Fermi-surface smearing*. There are many available smearing schemes. A comprehensive account may be found in (16).

Now, let's come back to the Kohn-Sham problem and its numerical resolution. The KS cycle has two characteristics: first, the potential and electron density are constructed through the single-particle KS orbitals ψ_m . Each of them (there are about N , the number of electrons in the system) is expanded in a basis set, with M components. In the case of plane waves, the total number of degree of freedom may be $NM \sim 10^6$ and even more! Therefore, any good algorithm should not make use of many vectors, since they are quite cumbersome. Second, the diagonalization of the KS hamiltonian is computationally very demanding. A good algorithm should converge as fast as possible, in order to save computer time.

Conjugate gradient methods are widely used for the minimization of the DFT total electronic energy E given by equation 3.25. The principle of these kinds of algorithms is to find the optimal KS orbitals by relying on the generalized forces $\mathcal{F}_j = -\frac{\delta E}{\delta \psi_j}$. Iterations go on until $\mathcal{F}_j \simeq 0$. Information on the evolution of \mathcal{F}_j as iterations proceed enables the construction of an approximate metric that should tend towards the exact Hessian $\mathcal{H}_{ij} = -\frac{\delta^2 E}{\delta \psi_i \delta \psi_j}$. Since the KS orbitals are usually expanded on a basis set, the Hessian has the same dimensions as the KS hamiltonian h^{KS} . The convergence gets faster as the gap between the biggest and smallest generalized forces shrinks. This is usually achieved by *pre-conditioning*: the Hamiltonian matrix is multiplied by a diagonal matrix M_p so that the smallest and largest eigenvalues become closer in magnitude. When using PW's, it is usual to consider the inverse of the kinetic energy as the pre-conditioning matrix M_p . A comprehensive account of pre-conditioned conjugate gradient methods can be found in (17).

Another popular and very effective scheme for minimizing $\varsigma_j = \langle \psi_j | (h^{\text{KS}} - \varepsilon_j) | \psi_j \rangle$ ⁹ is the so-called *Direct Inversion of Iterative Subspace* method (DIIS). The self-consistent solution $\{\psi_j^{(\infty)}\}$ is written as a linear combination of the KS orbitals at the previous iterations $\{\psi_j^{(n)}\}$. The optimal linear combination is found by minimizing the residual ς_j . A thorough account of the use of DIIS for the KS problem is given in (18). The DIIS method is very effective whenever the KS orbitals $\{\psi_j^{(n)}\}$ are relatively close to the exact solution. Therefore, it is often used to refine the first guess, which has been found by other algorithms, such as conjugate gradients.

⁹Note that if $\varsigma_j = \langle \psi_j | (h^{\text{KS}} - \varepsilon_j) | \psi_j \rangle = 0 \forall j$ self-consistency is attained. Therefore, the minimization of ς_j is often used as a criterion for convergence towards the self-consistent solution of the KS equations.

Chapter 4

Density Functional Perturbation Theory

Up to this point, we have explained how to find the ground state electron density, the self-consistent effective potential and the energy for a system of N electrons. The computation of the previous quantities is made possible by solving the Kohn-Sham equations. However, what is generally measured by experimental techniques, is the response of a system (the target) to an external perturbation (the probe). For instance, we can apply an external pressure P to the system and measure the change in volume ΔV . Their product $P\Delta V$ is the energy lost by the system due to mechanical work. By using the language of thermodynamics, the pressure is defined as the first derivative of the internal energy with respect to the volume, that is $P = -\frac{\partial E}{\partial V}$, with all the remaining thermodynamical variables (temperature T , number of particles N , etc.) being kept fixed. Then, for each applied pressure, we can measure the corresponding equilibrium volume and obtain the empirical law $V = V(P)$. The (isothermal) bulk modulus $B = -\frac{1}{V} \frac{\partial P}{\partial V} |_{T,N} = \frac{1}{V} \frac{\partial^2 E}{\partial V^2} |_{T,N}$ can thus be obtained. The bulk modulus is a second-order derivative of the internal energy and provide information on the elastic compliance of a material that is subject to an hydrostatic pressure.

All the previous quantities can be computed in the framework of the DFT. An elegant and effective way is provided by the so-called Density Functional Perturbation Theory (DFPT). DFPT is related to DFT exactly as the static perturbation theory of quantum mechanics is related to the Schrodinger equation. Through the formalism of DFPT, energy derivatives at any order can be in principle computed and the response of the system to an external macroscopic static perturbation (mechanical, electrical, magnetic, etc.) determined. The perturbations can also be of microscopic nature (for instance, atomic displacements) or having no experimental counterparts, although well defined from the theoretical point of view. An example is provided by the so-called alchemical perturbations: an atom of Pb may be changed into Au, through a sequence of steps in which the Pb pseudo-potential is progressively changed into that of Au. This kind of calculation is useful to determine some thermodynamic properties of impurities. The old dream of transforming metals into the most precious one, namely gold, may be achieved through DFPT, although only in simulations!

4.1 The basic formalism.

Perturbative expansions. Within the DFT, the total energy of a N electron system is a functional of the electron density, as proven by the HK theorem. However, there are many *external parameters* that are implicitly assumed in order to define the system precisely. A non exhaustive list includes:

- The volume and shape of the unit cell.
- The number, the chemical nature and the positions of the atoms within the unit cell.
- The intensity, the polarization and characteristic wave vector of an external electro-magnetic field.

These parameters are considered as external ones since they are independent of the electronic structure of the system. In other words, for a given set of external parameters, the ground state of the system is

found by varying the electron density until the energy reaches a minimum. Let's note with $\{\lambda\}$ the set of all those external parameters and with $E_\lambda[n]$ the total energy of the system, which has a *functional* dependence on the electron density n and a *parametrical* dependence on $\{\lambda\}$. For any distinct set of $\{\lambda\}$, the electron density which minimizes $E_\lambda[n]$ can be found in the framework of the basic theory described in chapter 2 and the numerical methods in chapter 3. Formally, we consider the external potential V that depends on the parameter λ and define the variations of V as due to the variations of λ around the value $\lambda = 0$, which is assumed to denote the unperturbed system. We assume that dependence of V_λ with λ is known through all orders:

$$V_\lambda = V^{(0)} + \delta\lambda V^{(1)} + (\delta\lambda)^2 V^{(2)} + \dots \quad (4.1)$$

where $V^{(n)} = \frac{\partial^n V}{\delta\lambda^n}$. The previous equation differs from a Taylor expansion of V only for the absence of the $\frac{1}{n!}$ factors. The corresponding ground-state electron density and energy can be expanded in a similar way:

$$n_\lambda = n^{(0)} + \delta\lambda n^{(1)} + (\delta\lambda)^2 n^{(2)} + \dots \quad (4.2)$$

$$E_\lambda = E^{(0)} + \delta\lambda E^{(1)} + (\delta\lambda)^2 E^{(2)} + \dots \quad (4.3)$$

Adopting the terminology of thermodynamics, first-order derivatives of the total energy are usually called generalized forces; second-order derivatives are connected to response functions, in the linear response regime (i.e. generalized forces are proportional to the perturbation). Third and higher order terms are needed to go beyond the case of linear response. Obviously, there is matter for many written pages: one can go through derivatives at high orders, consider extensions to complex perturbations or discuss the subtleties of the theory. Instead, I prefer to focus on few selected cases, which are fundamental for the hands-on-computer sessions. The first one is the computation of the atomic forces (section 4.2), which give access to first-principle molecular dynamics (FPMD); in FPMD, the atom dynamics is determined via the electron distribution, through the basic law of quantum mechanics, without referring to any semi-empirical interatomic potential. The second case is the calculation of lattice dynamics (section 4.3), which allows the comparison with infrared or Raman spectra, as well as the determination of some temperature-dependent properties. Moreover, atomic vibrations in complex systems or in materials that are hardly accessible to experiments can be predicted; for instance, elastic and vibrational properties of materials in extreme conditions (high temperature and/or intense pressure) have been extensively studied through DFPT. Interested readers can find more thorough and rigorous accounts on DFPT in several reviews and books, such as (19; 10; 9).

4.2 First-order energy derivatives and atomic forces.

In the previous sections, we focused on the behaviour of electrons, while the nuclei are considered to be fixed. Obviously, in many interesting physical cases the exact atomic positions are not known, so that the ions must be relaxed in order to find the equilibrium configuration. Moreover, we are also interested to perform ion dynamics at finite temperature or pressure, as in classical molecular dynamics. In all these cases, the forces acting on the ions must be computed. First, I recall the Born-Oppenheimer approximation, and then I discuss how the atomic forces can be computed.

The Born-Oppenheimer approximation. In many cases, the electrons and the nuclei can be treated on two distinct grounds. Here, I come back to the basic formalism that allows the electronic and nuclear degrees of freedom to be decoupled. This is a basic theorem of quantum mechanics, the so called Born-Oppenheimer (BO) approximation. I give a very brief account of some of the physical situations where the BO approximation is no more valid.

Within quantum mechanics, the properties of a system of nuclei (described by the set of coordinates $\{\vec{R}_I\}$) and electrons (with $\{\vec{r}_j\}$) are derived from the solution of the Schrodinger equation:

$$H_{\text{tot}}\Phi(\{\vec{R}_I\}, \{\vec{r}_j\}) = \mathcal{E}\Phi(\{\vec{R}_I\}, \{\vec{r}_j\}) \quad (4.4)$$

The total (nuclei + electrons) Hamiltonian is given by:

$$H_{\text{tot}} = H_{(el)} + T_N + V_{N-N} \quad (4.5)$$

with $H_{(el)}$ the many-electron Hamiltonian that we have introduced at the very beginning (see section 2.1), T_N the nuclear kinetic energy and V_{N-N} the Coulomb interaction between the nuclei of atomic charge Z_I and mass M_I :

$$H_{(el)} = \sum_{i=1}^N \left[-\frac{\hbar^2 \nabla_i^2}{2m} + V_{\text{ext}}(\vec{r}_i; \{\vec{R}_I\}) \right] + \sum_{i>j} \frac{e^2}{|\vec{r}_i - \vec{r}_j|} \quad (4.6)$$

$$T_N = \sum_{i=1}^N \left[-\frac{\hbar^2 \nabla_I^2}{2M_I} \right] \quad (4.7)$$

$$V_{N-N} = \sum_{I>J} \frac{Z_I Z_J e^2}{|\vec{R}_I - \vec{R}_J|} \quad (4.8)$$

Note that the external potential felt by the electrons depends on the nuclear coordinates, which is made explicit by writing $V_{\text{ext}}(\vec{r}_i; \{\vec{R}_I\})$.¹ The operator ∇_I^2 acts on the nuclear coordinate \vec{R}_I , while ∇_j^2 acts on the electronic coordinate \vec{r}_j .

The solution of equation 4.4 for a realistic crystal is out of the scope of any analytical or numerical treatment. However, the problem can be much simplified due to the very different dynamics of electrons and nuclei. Classically, the dynamic response of nuclei is very much slower than that of electrons, because of the mass ratio: $m/M_I \ll 1$. Let's see how this is possible in quantum mechanics, that is, working on equation 4.4.

First, Born and Oppenheimer did the following ansatz on the whole wave function:

$$\Phi(\{\vec{R}_I\}, \{\vec{r}_j\}) \simeq \chi(\{\vec{R}_I\}) \Psi(\{\vec{R}_I\}, \{\vec{r}_j\}) \quad (4.9)$$

where $\Psi(\{\vec{R}_I\}, \{\vec{r}_j\})$ is the solution of the electron Hamiltonian for a given nuclear configuration, as if the nuclei were fixed. Their positions $\{\vec{R}_I\}$ are therefore considered as parameters:

$$H_{(el)} \Phi(\{\vec{R}_I\}, \{\vec{r}_j\}) = E(\{\vec{R}_I\}) \Phi(\{\vec{R}_I\}, \{\vec{r}_j\}) \quad (4.10)$$

Obviously, the electronic eigenvalues depend on the nuclear coordinates parametrically, which we note $E(\{\vec{R}_I\})$.² Inserting the ansatz 4.9 in equation 4.4, and making use of equation 4.10, one obtains:

$$\begin{aligned} & \Psi(\{\vec{R}_I\}, \{\vec{r}_j\}) \left[T_N + V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\}) \right] \chi(\{\vec{R}_I\}) + \\ & - \sum_{i=1}^N \frac{\hbar^2}{M_I} [\vec{\nabla}_I \chi(\{\vec{R}_I\})] \cdot [\vec{\nabla}_I \Psi(\{\vec{R}_I\}, \{\vec{r}_j\})] - \chi(\{\vec{R}_I\}) \sum_{i=1}^N \frac{\hbar^2}{2M_I} [\nabla_I^2 \Psi(\{\vec{R}_I\}, \{\vec{r}_j\})] = \\ & = \mathcal{E} \chi(\{\vec{R}_I\}) \Psi(\{\vec{R}_I\}, \{\vec{r}_j\}) \end{aligned} \quad (4.11)$$

where all operators act on the functions on their right. Multiplying the previous equation by $\Psi^*(\{\vec{R}_I\}, \{\vec{r}_j\})$ and integrating over the electronic coordinates $\int d^3 r_1 \int d^3 r_2 \dots$, we obtain:

$$\left[T_N + V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\}) \right] \chi(\{\vec{R}_I\}) + \mathcal{T} = \mathcal{E} \chi(\{\vec{R}_I\}) \quad (4.12)$$

where

$$\mathcal{T} = - \sum_I \left[\frac{\hbar^2}{2M_I} \langle \Psi | \nabla_I^2 | \Psi \rangle + \frac{\hbar^2}{M_I} \langle \Psi | \vec{\nabla}_I | \Psi \rangle \cdot \vec{\nabla}_I \right] \chi(\{\vec{R}_I\}) \quad (4.13)$$

¹When all electrons (core and valence) are explicitly treated, $V_{\text{ext}}(\vec{r}_i; \{\vec{R}_I\}) = - \sum_I \frac{Z_I e^2}{|\vec{R}_I - \vec{r}_i|}$; otherwise, $V_{\text{ext}}(\vec{r}_i; \{\vec{R}_I\})$ represents the ionic pseudopotential. In the latter case, Z_I are the charges of the N_{at} ions (nuclei + core electrons) at $\vec{R}_1, \dots, \vec{R}_{N_{\text{at}}}$.

²As usual, at fixed nuclear positions, one obtains from equation 4.10 a series of eigenstates $\Phi_m(\{\vec{R}_I\}, \{\vec{r}_j\})$ and eigenvalues $E_m(\{\vec{R}_I\})$, with $m = 0, 1, \dots$. $m = 0$ correspond to the electronic ground state, $m = 1$ to the first electronic excited state, etc. For simplicity sake, I omit the index m in equation 4.10; nevertheless, the reader should keep in mind that there are several distinct functions $E_m(\{\vec{R}_I\})$, which correspond to the electronic eigenvalues for given nuclear positions $\{\vec{R}_I\}$.

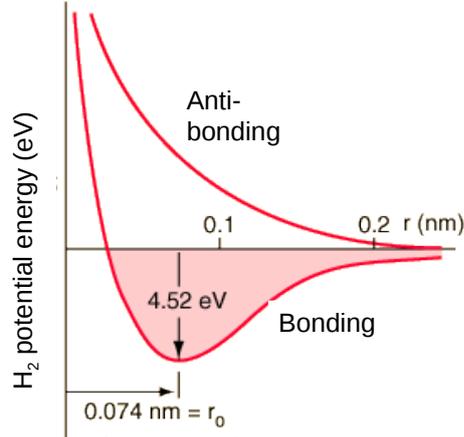


Figure 4.1: Potential energy as a function of the internuclear separation r for the H_2 molecule, with respect to the isolated H atoms. The equilibrium bond length is the internuclear distance corresponding to the depth of the potential minimum (4.52eV), with bonding molecular orbital. The upper curve describes the potential energy when the molecular orbital is of the anti-bonding type.

The second main approximation is to neglect \mathcal{T} in equation 4.13. Then, it can be rewritten as:

$$\left[T_N + V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\}) \right] \chi(\{\vec{R}_I\}) = \mathcal{E} \chi(\{\vec{R}_I\}) \quad (4.14)$$

This is a Schrodinger equation for the nuclear part of the wave function, with potential energy $V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\})$. The first term is simply the repulsive Coulomb contribution between the positively charged nuclei; the second one, instead, comes from the solution of the equation 4.10 for the electrons at fixed nuclear coordinates and is called the *potential energy surface*. $E(\{\vec{R}_I\})$ is a manifold in the nuclear space, in which the contribution from electrons is included. If the electronic ground state is computed, then $E(\{\vec{R}_I\})$ represents the cohesive contribution from the chemical bonds between the nuclei that are fixed in $\{\vec{R}_I\}$. As an example, one may refer to the energy of the H_2 molecule as a function of the distance r between the two protons. The lowest curve $E(r)$ is obtained for the electronic ground state, which is of the bonding type. Since in such a state the electronic distribution is mainly localized between the two protons, the internuclear Coulomb repulsion is substantially screened and the potential energy reaches a minimum at the equilibrium bond length r_0 .

As a consequence, the Born-Oppenheimer approximation allows the solution of the Schrodinger equation 4.4 to be obtained in two steps: First, the electronic problem is solved at fixed nuclear coordinates. The corresponding eigenvalues, which depend on the nuclear coordinates, provide the electronic contribution to the effective interaction between the nuclei $E(\{\vec{R}_I\})$. Next, the nuclear hamiltonian is solved with the potential energy $V_{N-N} + E$, where the second term depends on the detailed electronic structure. As the nuclear positions vary, $E(\{\vec{R}_I\})$ varies, too. Searching the minimum of $V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\})$ therefore yields the equilibrium atomic positions of the system. The basic approximation is that the electrons follow the nuclear motion adiabatically, thus to rearrange instantaneously to the ground state for the given atomic coordinates $\{\vec{R}_I\}$.

Why could \mathcal{T} be neglected? If one assumes that the momentum p_j of electrons and P_I of the nuclei³ is of the same order of magnitude, the expectation value of the ratio between the nuclear and electronic kinetic energy is $\simeq \frac{m}{M_I}$. Quantum mechanically, it is not unreasonable to assume that the momenta of the electrons and nuclei in a molecule are comparable in magnitude. If one could think the molecule as a box containing the electrons and nuclei, this is indeed the case (the momenta are independent of either the electron or the nuclear mass but is inversely proportional to the box dimensions). However, there are many situations in which the Born-Oppenheimer approximation breaks down. Let's consider, for instance, the collision between two atoms: there is a finite probability, which increases with the kinetic energy as measured in the center of mass reference system, that they will be ionized after the collision. Therefore, part of the nuclear kinetic energy can be transferred to the electrons and the corresponding dynamics cannot be decoupled. More advanced methods must be employed in these cases.

³Note that the corresponding quantum mechanical operators do not contain the mass.

The Hellmann-Feynman theorem in DFT and the atomic forces. The atomic forces are the derivatives of the potential energy $V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\})$ with respect to the atomic positions $\{\vec{R}_I\}$: $\vec{F}_I = -\nabla_I [V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\})]$. However, such a definition would imply to solve equation 4.11 and obtain the eigenvalue E as well as its explicit dependence on the nuclear coordinates $\{\vec{R}_I\}$. Even if we use finite-difference approximations for $E(\{\vec{R}_I\})$, the calculations would be unfeasible for systems consisting of more than a few atoms, since many Kohn-Sham equations with varying $\{\vec{R}_I\}$ should be solved.

Instead, we proceed differently and use the variational properties of the DFT. Consider the total energy of the system of nuclei and electrons, with all of the nuclei fixed:

$$V_{N-N}(\{\vec{R}_I\}) + E_{(\{\vec{R}_I\})}[n] \quad (4.15)$$

Here $E_{(\{\vec{R}_I\})}[n]$ is the electronic energy that is a functional of the electron density which, according to the DFT, reaches its minimum for the correct ground state density n . It also depends on the nuclear coordinates in a parametric way, as stated by the Born-Oppenheimer approximation. Therefore, the atomic force can be also defined as the total derivative of the previous expression with respect to the coordinate \vec{R}_J :

$$\vec{F}_J = -\frac{dV_{N-N}}{d\vec{R}_J} - \frac{dE}{d\vec{R}_J} \quad (4.16)$$

The total variation of E consists of two contribution, through the explicit dependence on the parameter \vec{R}_J , and the implicit dependence on the electron density, which also varies with \vec{R}_J :

$$\begin{aligned} \frac{dE}{d\vec{R}_J} &= \frac{\partial E}{\partial \vec{R}_J} + \int d^3r \frac{\delta E}{\delta n(\vec{r})} \frac{\partial n(\vec{r})}{\partial \vec{R}_J} = \frac{\partial E}{\partial \vec{R}_J} + \int d^3r \mu \frac{\partial n(\vec{r})}{\partial \vec{R}_J} \\ &= \frac{\partial E}{\partial \vec{R}_J} + \mu \frac{\partial}{\partial \vec{R}_J} \int d^3r n(\vec{r}) = \frac{\partial E}{\partial \vec{R}_J} \end{aligned} \quad (4.17)$$

where we have used equation 2.25 and the fact that $\int d^3r n(\vec{r}) = N$ independently of \vec{R}_J . Therefore, the *total* derivative of the density energy functional with respect to a given nuclear coordinates simply coincides with the *partial* derivative. This is a very important result: computing the partial derivative is quite simple, since it implies to derive only the terms that have an explicit dependence on \vec{R}_J . There is no need to determine the change of the electron density with respect to variations of the external parameters! This is a quite general result, which is valid for any external parameter λ (such as the volume, the intensity of a magnetic field, etc.) in any theory that is based on a variational principle, such as the DFT and Hartree-Fock method.

From a practical point of view, the partial derivatives $\frac{\partial E}{\partial \lambda}$ are usually computed at the same time as the energy, for all the most common parameters λ such as the atomic positions and the volume. Considering the breakdown of the total electronic energy (equation 3.25, only the external potential depends on the atomic coordinates explicitly. In all-electron methods, $V_{\text{ext}}(\vec{r}) = -\sum_I \frac{Z_I e^2}{|\vec{r} - \vec{R}_I|}$. The atomic force is therefore:

$$\begin{aligned} \vec{F}_I &= -\frac{\partial V_{\text{ext}} + \partial V_{N-N}}{\partial \vec{R}_I} = \\ &= -Z_I e^2 \int d^3r n(\vec{r}) \frac{(\vec{r} - \vec{R}_I)}{|\vec{r} - \vec{R}_I|^3} + \sum_{J \neq I} Z_J Z_I e^2 \frac{(\vec{R}_J - \vec{R}_I)}{|\vec{R}_J - \vec{R}_I|^3} \end{aligned} \quad (4.18)$$

Note that the correct force must be computed with the electron density at its ground state value. When using a plane waves basis set in conjunction with pseudo-potentials, the expression is more involved but still easy to be computed numerically. Calculating forces takes as much computer time as determining the total energy for a given density $n(\vec{r})$. The actual implementations obviously depend on the explicit representation used, and will not be detailed here.

4.3 Second order

We have seen that the first order variation of the energy ($E^{(n)}$ in the expansion 4.3) does not depend on $n^{(1)}$ but only on the ground state electron distribution $n^{(0)}$. This result, which is the DFT analogous of

the Hellmann-Feynman theorem in quantum mechanics, is based on the variational character of the total electronic energy. Similarly, we can consider the expansion of the energy at any order in the perturbation. An extension of Hellmann-Feynman theorem, which is called the "2n + 1 theorem", holds, which states that $E^{(2n+1)}$ depends on the variations of the electron density at all orders up to n (20). In particular, the second order variation of the energy $E^{(2)}$ depends on both $n^{(0)}$ and $n^{(1)}$. The ground state density is computed through the Kohn-Sham equations 2.33 and 2.34. From the latter one, the first-order variation of the electron density is:

$$n^{(1)} = \sum_j f_j \psi_j^{(1)} \psi_j^{(0)*} + C.C. \quad (4.19)$$

$\psi_j^{(1)}$ may be computed through the so called *Sternheimer equation*, which we derive in the following.

Consider the Schrodinger equation

$$\left[H^{(0)} + H^{(1)} + H^{(2)} + \dots - (\varepsilon_i^{(0)} - \varepsilon_i^{(1)} - \varepsilon_i^{(2)} + \dots) \right] (\psi_i^{(0)} - \psi_i^{(1)} - \psi_i^{(2)} + \dots) = 0 \quad (4.20)$$

where $H^{(n)}$ is known through all orders. We assume that it has been solved at zero order, so that we know $\varepsilon_i^{(0)}$ and $\psi_i^{(0)}$, while $\varepsilon_i^{(1)}$ and $\psi_i^{(1)}$ must be determined. According to standard perturbation theory, $\varepsilon_i^{(1)} = \langle \psi_i^{(0)} | H^{(1)} | \psi_i^{(0)} \rangle$. In order to find $\psi_i^{(1)}$, let's consider the first-order terms in the Schrodinger equation 4.20 (that is, we ignore all n -order terms with $n > 1$ as well products between first-order terms):

$$\left[H^{(0)} - \varepsilon_i^{(0)} \right] \psi_i^{(1)} = \left[H^{(1)} - \varepsilon_i^{(1)} \right] \psi_i^{(0)} \quad (4.21)$$

This is the Sternheimer equation. When the hamiltonian and the Kohn-Sham orbitals is discretized on a basis set, the equation 4.21 is of the form $A\vec{x} = \vec{b}$, with $A = H^{(0)} - \varepsilon_i^{(0)} I$, $\vec{x} = \psi_i^{(1)}$ and $\vec{b} = \left[H^{(1)} - \varepsilon_i^{(1)} \right] \psi_i^{(0)}$. Its solution is: $\vec{x} = A^{-1}\vec{b}$, whenever A^{-1} exists. When a plane wave basis set is used, A is a large matrix (consisting of up to $\sim 10^5$ elements), so that its inversion is cumbersome. However, it can be obtained by iterative processes that imply only matrix - vector multiplication without the need of computing A^{-1} by direct inversion.

Atomic vibrations. We assume that the Born-Oppenheimer separation holds and that the electron and nuclear dynamics can be decoupled. In equation 4.14, the potential energy term for the nuclear motion is $E_{\text{pot}}(\{\vec{R}_I\}) = V_{N-N}(\{\vec{R}_I\}) + E(\{\vec{R}_I\})$, where the last term represents the effective contribution by electrons relative to the fixed nuclei. We also make the assumption that the nuclei can be considered as classical particles, which is reasonable apart from light elements such as H or He. In order to find the small oscillations of the system, we expand the potential energy in terms of the nuclear displacements $\vec{u}_I = \vec{R}_I - \vec{R}_I^{(0)}$ from the equilibrium positions $\vec{R}_I^{(0)}$:

$$E_{\text{pot}}(\{\vec{R}_I\}) = E_{\text{pot}}(\{\vec{R}_I^{(0)}\}) + \sum_{I,\alpha} \left[\frac{\partial E_{\text{pot}}}{\partial R_{I,\alpha}} \right]_0 u_{I,\alpha} + \sum_{I,\alpha; J,\beta} \frac{1}{2} \left[\frac{\partial^2 E_{\text{pot}}}{\partial R_{I,\alpha} \partial R_{J,\beta}} \right]_0 u_{I,\alpha} u_{J,\beta} + \dots \quad (4.22)$$

where I and J runs on the atoms and $\alpha, \beta = 1, 3$ label the cartesian components. The first order term is null, since at equilibrium $\left[\frac{\partial E_{\text{pot}}}{\partial R_{I,\alpha}} \right]_0 = 0$. Within the harmonic approximation, we neglect in the expansion 4.22 all terms beyond the second order in the atomic displacements, that we suppose to be small. Therefore, the atomic force on the atom I that is due to the set of displacements is proportional to the atomic displacements $\{u_{J,\beta}\}$ ⁴:

$$F_{I,\alpha} = - \sum_{J,\beta} C_{I,\alpha; J,\beta} u_{J,\beta} \quad (4.23)$$

where we have introduced the $(3N_{\text{at}} \times 3N_{\text{at}})$ interatomic force constant matrix:

$$C_{I,\alpha; J,\beta} = \left[\frac{\partial^2 E_{\text{pot}}}{\partial R_{I,\alpha} \partial R_{J,\beta}} \right]_0 \quad (4.24)$$

⁴This is quite usual: we speak of linear response theory whenever we neglect terms beyond second order in the expansion of the energy as a function of the perturbations. This apparent semantic contradiction is due to the fact that the generalized forces are indeed linear in the perturbation!

It can be shown that the force constant matrix is symmetric ($C_{I,\alpha;J,\beta} = C_{J,\beta;I,\alpha}$) and that the sum of its elements along any row is zero ($\sum_{J,\beta} C_{I,\alpha;J,\beta} = 0$).

By considering the nuclei as classical particles, their motion is governed by the Newton equation:

$$M_I \frac{d^2 u_{I,\alpha}}{dt^2} = F_{I,\alpha} = - \sum_{J,\beta} C_{I,\alpha;J,\beta} u_{J,\beta} \quad (4.25)$$

We look an oscillatory solution with frequency $\omega/2\pi$ of the form:

$$u_{I,\alpha}(t) = \frac{u_{I,\alpha}^{(0)}}{\sqrt{M_I}} e^{i\omega t} \quad (4.26)$$

By using this expression in the equation of motion 4.25, we obtain the linear homogeneous system in the unknown amplitudes $u_{J,\beta}^{(0)}$:

$$\sum_{J,\beta} \left[\frac{C_{I,\alpha;J,\beta}}{\sqrt{M_I M_J}} - \omega^2 \delta_{IJ} \delta_{\alpha\beta} \right] u_{J,\beta}^{(0)} e^{i\omega t} = 0 \quad (4.27)$$

which has non trivial solutions ($u_{J,\beta}^{(0)} \neq 0$) only if the determinant of the coefficients is null:

$$\text{Det} \left[\frac{C_{I,\alpha;J,\beta}}{\sqrt{M_I M_J}} - \omega^2 I \right] = 0 \quad (4.28)$$

From the secular equation above, the frequencies ω^2 are computed. Since the matrix $C_{I,\alpha;J,\beta}/\sqrt{M_I M_J}$ is symmetric, the $3N_{\text{at}}$ solutions $\omega_k^2 \geq 0$ with $k = 1, 3N_{\text{at}}$. From the corresponding normalized eigenvectors $x_{i,k}$, the amplitudes for the mode k on each atom I and cartesian component α can be determined as $u_i^{(0)} \sim \frac{x_{i,k}}{\sqrt{M_I}}$ ⁵.

Once the matrix of the interatomic force constants is known, the vibrational frequencies (the eigenvalues) and the amplitudes for each atom (the eigenvectors) are thus obtained by standard diagonalization techniques. Therefore, the main problem is to compute $C_{I,\alpha;J,\beta}/\sqrt{M_I M_J}$. In practice, the perturbations consist of small atomic displacements. Once the equilibrium geometry is found, the atoms can be displaced "by hands" one by one, along the cartesian directions, in order to find the elements of the interatomic force constant matrix by finite differences (for instance, by computing the variations of the atomic forces that are due to the displacements). This method is known as the *frozen phonon* calculation. It usually gives very good results in finite systems, such as molecules and clusters, provided that force derivatives are well approximated by finite differences. The atomic displacements should be small enough not to include anharmonic contributions (i.e. forces must be linear in $u_{I,\alpha}^{(0)}$) and large enough in order to avoid numerical errors (the finite difference $\frac{F(x+dx) - F(x-dx)}{2dx} \rightarrow \frac{0}{0}$ if $dx \rightarrow 0$).

Periodic crystals: phonons. The case of crystal implies some additional complications. In a periodic solid, the matrix of interatomic force constant depends only on the mutual distances between the atoms: displacing the origin of coordinates from a point to another one that is equivalent by lattice vectors translations produces no physical changes. Such property implies that the atomic displacements within different unit cells of the crystal are related by:

$$u_{n,I,\alpha}^{(0)} = u_{m,I,\alpha}^{(0)} e^{i\vec{k} \cdot \vec{L}_{n-m}} \quad (4.29)$$

where \vec{L}_{n-m} is a direct lattice vector going from cell m to cell n and \vec{k} a wave vector within the Brillouin zone. The previous expression is similar to that of one-electron orbitals in a periodic potential. Indeed, the Bloch theorem holds for the atomic vibrations in a periodic crystal, too, from which Equation 4.29 can be derived. Accordingly, the number of independent vibrational degrees of freedom in a solid is given by $3N_{\text{at}}^{(0)}$, which is the number of atoms within the unit cell by the three spatial dimensions. By using equation 4.29, all the vibrational properties can be determined by looking to a single unit cell, in principle. As in the case of the single-particle electronic orbitals, it is much better to work in the reciprocal space.

⁵Here, for the sake of conciseness, the indexes run on both cartesian and atomic labels ($i = I, \alpha$)

Therefore, we introduce the dynamical matrix, which is the Fourier transform of the interatomic force constant matrix:

$$D_{I,\alpha;J,\beta}(\vec{q}) = \sum_{\vec{L}} C_{I,0,\alpha;J,\vec{L},\beta} e^{i\vec{L}\cdot\vec{q}} \quad (4.30)$$

where the cell at the origin (0) is related to another one by the lattice vector \vec{L} . The eigenvalues of $D_{I,\alpha;J,\beta}(\vec{q})$ at any wave vector within the Brillouin zone yield $3N_{\text{at}}^{(0)}$ values of the squared frequencies $\omega^2(\vec{q})$. The ω versus \vec{q} plot is the *phonon dispersion relation*, which should be known for the whole Brillouin zone, in principle.

If one uses periodic boundary conditions, each atomic displacement is periodically repeated in space, which poses practical restrictions on the phonon wave length (see exercise 1). Therefore, any calculation that is based on the frozen phonon technique can only give phonons for a few \vec{q} vectors in the Brillouin zone. For the case of phonon calculation in a solid, the perturbation is instead the variation of the external potential when the nuclei in \vec{R}_I are subject to the displacements \vec{u}_I :

$$\Delta V_{\text{ext}} = \sum_I V_{\text{ext}}(\vec{r} - (\vec{R}_I + \vec{u}_I)) - V_{\text{ext}}(\vec{r} - \vec{R}_I) \quad (4.31)$$

$$\text{with } \vec{u}_I = \lambda \vec{e}_I \cos(\vec{q} \cdot \vec{R}_I) \quad (4.32)$$

where λ is a small parameter, \vec{e}_I the phonon polarisation and \vec{q} the phonon wave vector.

In practice, the calculation of the vibrational properties of a molecule or a solid is performed in the following way:

1. Find the equilibrium geometry and the corresponding electronic ground state, that is, the Kohn-Sham orbitals $\{\psi_i^{(0)}\}$ and the electron density $n^{(0)}$.
2. For each perturbation of the type illustrated in equations 4.31 and 4.32, compute the first-order variations $\{\psi_i^{(1)}\}$ and $n^{(1)}$. The Sternheimer equation can be used, or alternative approaches that are not detailed here.
3. From $\{\psi_i^{(0)}\}$, $\{\psi_i^{(1)}\}$, $n^{(0)}$ and $n^{(1)}$, compute the energy variations $E^{(2)}$ at the second order in the phonon perturbation. Then, rely $E^{(2)}$ with the parameter of the perturbation (equations 4.31 and 4.32) and construct the dynamical matrix. Its diagonalization yields the phonon energy $\omega(\vec{q})$ and the corresponding eigenvectors.
4. For a molecule, only the calculation at $\vec{q} = 0$ is meaningful. For a periodic solid, at variance, repeat steps 2 and 3 for selected wave vectors \vec{q} within the Brillouin zone, in order to determine the phonon dispersion relation.

Exercises

1. Consider a periodic chain of atoms with two distinct atoms per unit cell. Be L the length of the unit cell and M their number ($M \gg 1$). Write the atomic displacements along the chain according to equation 4.28. Suppose now to change by hands the coordinate of an atom within the unit cell and apply the periodic boundary conditions. Which wave vector q does correspond to such perturbation? Imagine now to choose as unit cell the double of the primitive ones. Such new unit cell has length $2L$ and contains four atoms. Firstly, establish a correspondence between the reciprocal lattice vectors of the primitive and the new unit cells. Then, change by hands the coordinate of an atom within the unit cell and apply the periodic boundary conditions. Which wave vector q corresponds to such a perturbation, in the coordinates of the new unit cell of size $2L$? Which wave vector q corresponds to such a perturbation, in the coordinates of the (old) primitive cell of size L ? Discuss a method to obtain the vibrational frequencies at $q = \frac{\pi}{mL}$ by displacing atoms by hands and choosing unit cells of increasing lengths mL .

Bibliography

- [1] N. Metropolis and S. Ulam, "The Monte Carlo method", *J. Amer. Stat. Ass.* **44**, 335 (1949).
- [2] N. Metropolis et al., "Equation of state by fast computing machines", *J. Chem. Phys.* **21**, 1087 (1953).
- [3] B.J. Alder and T.E. Wainwright, "Studies in Molecular dynamics: general method", *J. Chem. Phys.* **31**, 459 (1959)
- [4] R. Car and M. Parrinello, "Unified approach for molecular dynamics and density functional theory", *Phys. Rev. Lett.* **55**, 2471 (1985).
- [5] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", *Phys. Rev.* **136**, B864 (1964).
- [6] W. Kohn and L.J. Sham, "Self-consistent equations including exchange and correlation effects", *Phys. Rev.* **A140**, 1133 (1965).
- [7] T. Kato, *Commun. Pure Appl. Math.* **10**, 151 (1957).
- [8] R.G. Parr and W. Yang, "Density Functional Theory of Atoms and Molecule", Oxford Univ. Press (1989); R.M. Dreizler and E.K.U. Gross, "Density Functional Theory", Springer-Verlag (1990).
- [9] R.M. Martin, "Electronic Structure: Basic Theory and Practical Methods", Cambridge Univ. Press (2004).
- [10] F. Finocchi, J. Goniakowski, X. Gonze, and C. Pisani, "An introduction to First-Principles simulations of extended systems", in *Handbook of Numerical Analysis*, vol.X (special volume on Computational Chemistry), edited by C. Le Bris, pages 377-451 (Elsevier, Amsterdam, 2003).
- [11] G.D. Mahan and K.R. Subbaswamy, "Local Density Theory of Polarizability", Plenum Press (1990).
- [12] N.W. Ashcroft and N.D. Mermin, "Solid State Physics", Holt-Saunders (1981).
- [13] F. Bassani and G. Pastori Parravicini, "Electronic states and optical transitions in solids", Pergamon Press (1975).
- [14] D.J. Chadi and M.L. Cohen, "Special Points in Brillouin Zone", *Phys. Rev. B* **8**, 5747 (1973)
- [15] H.J. Monkhorst and J.D. Pack, "Special points for Brillouin-zone integrations", *Phys. Rev. B* **13**, 5188 (1976).
- [16] M. Verstraete and X. Gonze, *Comp. Mater. Sci.* **30**, 27 (2004).
- [17] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias and J.D. Joannopoulos, "Iterative Minimization techniques for ab initio total energy calculations - Molecular Dynamics and Conjugate Gradients", *Rev. Mod. Phys.* **64**, 1045 (1992).
- [18] P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980).
- [19] S. Baroni, S. De Gironcoli, S. Dal Corso et al, "Phonons and related crystal properties from density functional perturbation theory", *Rev. Mod. Phys.* **73**, 515 (2001)
- [20] X. Gonze, "Perturbation expansion of variational principles at arbitrary order", *Phys. Rev. A* **52**, 1086 (1995).